UK National Screening Committee

Newborn screening for congenital adrenal hyperplasia

External review against programme appraisal criteria for the UK National Screening Committee

Version: Final

Author: School of Health and Related Research (ScHARR), University of Sheffield

Date: February 2022

The UK National Screening Committee secretariat is hosted by the Department of Health and Social Care

About the UK National Screening Committee (UK NSC)

The UK NSC advises ministers and the NHS in the 4 UK countries about all aspects of population screening and supports implementation of screening programmes.

Conditions are reviewed against evidence review criteria according to the UK NSC's evidence review process.

Read a complete list of UK NSC recommendations.

UK National Screening Committee, Southside, 39 Victoria Street, London, SW1H 0EU

www.gov.uk/uknsc

Blog: https://nationalscreening.blog.gov.uk/

For queries relating to this document, please contact: https://view-health-screening-recommendations.service.gov.uk/helpdesk/

Published: February 2022

Contents

Plain English summary

This review looked at screening in newborn babies for a rare genetic disorder called congenital adrenal hyperplasia (CAH).

Children with CAH have faulty genes that cause a lack of the enzymes that control the production of the hormone cortisol. Cortisol is important as it controls the amount of water and salt in the body. The genes responsible for CAH may also cause the body to produce too much of a hormone called androgen.

There are 3 main types of CAH. The most severe type of CAH is classic salt-wasting (SW). Newborns with SW-CAH cannot balance the salt levels in the body. This leads to dehydration, low blood pressure and can result in life-threatening crises. In addition, the production of too much androgen may cause female genitalia to appear more like male genitalia.

Classic simple virilising CAH (SV-CAH) tends to be less severe than SW-CAH and may not be noticed until children are older. Whilst the balance of salt in the body is better in this type of CAH, female genitalia may still appear more like male genitalia. Older children who have not received treatment may have other signs such as early puberty, and growth which starts and finishes early leading to a shorter final adult height. People with SV-CAH can suffer from infertility if it is not treated. Non-classic CAH (NC-CAH) is not usually noticed at birth, and symptoms do not tend to appear until later in life. People with NC-CAH may have early puberty and problems with infertility.

Newborn screening might find babies with CAH before symptoms appear. The UK National Screening Committee (UK NSC) last looked at the evidence for newborn screening for CAH in 2015. The review found that there was not enough evidence to recommend a screening programme.

This current evidence summary updates the previous UK NSC review. It looks at all new evidence published since 2015, and also at evidence since 2008 to try to understand the age at which CAH symptoms appear in newborns. The focus of this review is to:

- find out the how many babies are born each year with CAH in the UK
- understand the age at which CAH symptoms appear
- see whether current tests can accurately find babies with CAH

The review found that in Great Britain, the incidence is about 1 in 18,000 children. However, this evidence summary does not recommend screening for CAH in newborn babies. This is because there is still not enough evidence on:

- whether screening would find newborns with CAH before they suffer from a life-threatening crisis
- whether the screening test can accurately find all cases of CAH

Executive summary

Purpose of the review

This review aims to assess whether there have been significant developments in the evidence base since the previous review for newborn screening for CAH was conducted in 2015, and if sufficient support exists for a screening programme for CAH in newborn infants. The purpose of this evidence synthesis was to assess whether to reconsider the current UK NSC's position, which is not to recommend screening for CAH.

Background

Congenital adrenal hyperplasia (CAH) is a group of autosomal recessive disorders. These conditions are caused by enzymatic defects which lead to impairments in cortisol synthesis in the adrenal gland. If left undiagnosed and untreated, CAH can cause high morbidity and mortality. CAH manifests with a range of clinical and biochemical phenotypes. These are commonly categorised into severe (classic) or mild (non-classic) forms. The most common form of CAH is 21-hydroxylase deficiency (21-OHD) which accounts for over 90% of cases There are 3 main subtypes of 21-OHD. Classic salt wasting (SW) CAH accounts for 70-80% of all cases of 21-OHD and causes a life-threatening salt-wasting crisis in newborns and is also characterised by virilisation of female external genitalia. The classic simple virilising (SV) form is less severe than SW-CAH and it is characterised by virilised external genitalia in females but with no clinically apparent salt loss. Excess androgen leads to rapid skeletal growth and premature puberty in both girls and boys. Finally, non-classic (NC) CAH is a much rarer and milder form, which may not be detected until later in life. There may be no symptoms. However, female patients may exhibit hyperandrogenism, and both females and males may exhibit symptoms such as short adult stature, premature pubarche or accelerated bone maturation.

Newborn screening for CAH focuses on the detection of 21-OHD by measuring the enzyme's substrate, 17-hydroxyprogesterone (17-OHP). First-tier tests typically measure elevated 17-OHP concentrations in dried blood spots as a biomarker for 21-hydroxylase deficiency. The cut-off levels for 17-OHP are commonly optimised to achieve maximum sensitivity. However, this strategy results in a large number of false positive results. Premature, stressed or ill babies in particular can display elevated 17-OHP concentrations. Second-tier testing typically involves steroid profiling performed in real time by liquid chromatography tandem mass spectrometry and may improve specificity without a loss of sensitivity. This technology is increasingly used in newborn screening laboratories for other elements of the newborn screening programme.

Focus of the review

This review aimed to evaluate whether the evidence base has developed substantially and a screening programme for CAH has become viable since the previous UK NSC review was conducted in 2015. Specifically, new evidence was collected to answer the following 3 questions:

1. What is the incidence of congenital adrenal hyperplasia in the UK population, including discrete subgroups of the population? (criterion 1)
2. What is the median age of presentation of congenital adrenal hyperplasia? (criterion 1)
3. What is the accuracy of available screening tests using dried blood spots (DBS) to detect congenital adrenal hyperplasia? (criteria 4 and 5)

Recommendation under review

The UK National Screening Committee (NSC) previously considered evidence for newborn screening for CAH in 2015. Based on the evidence presented in the 2015 review, screening for CAH in newborns in the UK was not recommended. The review acknowledged that second-tier testing by mass spectrometry could improve the positive predictive value (PPV) and consequently the accuracy of the test, but the evidence on this was very limited. There was also insufficient evidence to show whether screening might take place too late to reduce the severity of patient outcomes, and it was noted that establishing the median age at presentation may be useful. In addition, the previous review did not completely clarify the uncertainties surrounding the UK incidence of CAH including discrete population subgroups.

Findings and gaps in the evidence of this review

Twenty-four publications were extracted and included in the evidence synthesis. A summary of question level results is presented below.

Criterion 1 – *'The condition should be an important health problem as judged by its frequency and/or severity. The epidemiology, incidence, prevalence and natural history of the condition should be understood, including development from latent to declared disease and/or there should be robust evidence about the association between the disease marker and serious or treatable disease'.*

Question 1 – What is the incidence of congenital adrenal hyperplasia in the UK population, including discrete subgroups of the population?

Fifteen studies were included in the review. No new studies of UK incidence of CAH were identified since the last review was published in 2015, however the 2 studies from Great Britain (GB) that featured in the previous review were also included in this

evidence summary. All the included studies reported data collected from newborn screening programmes, apart from those from GB which collected data through an active surveillance programme. Duration of data collection ranged from one year to 26 years, with most spanning 7 years and over. The duration of the GB data collection was 2 years. The average incidence rate across studies was calculated as 1:16,869. The incidence rate in Great Britain was around this international average, being reported as 1:18,248.

Question 2 – What is the median age of presentation of congenital adrenal hyperplasia?

Eight studies were included in the review. No study set out explicitly to determine the time to presentation of symptoms in newborns. Instead, age at diagnosis was most commonly reported, hence the relevance of these studies to appropriately address the question is limited. One international and one GB study reported data on presentation of symptoms in children in the first month of life as part of wider studies. The remaining studies reported data from selected cohorts of children with CAH recruited as part of studies with other aims. The GB study reported a median age of diagnosis of 15 days in an unscreened cohort, with the international study reporting a median age of presentation of 10 days. These data suggest that a number of children present with symptoms before the results of their screening tests would be currently available in the UK, and therefore they may not benefit from screening. However, the current evidence on median age of presentation in newborns is limited in terms of volume and quality and it is therefore difficult to draw definitive conclusions relating to the impact screening would have on clinical outcomes.

Overall, the incidence of CAH in Great Britain has been found to be approximately 1 in 18,000 children. However, the evidence on median age of presentation in newborns is limited and therefore, overall, criterion 1 is not met.

Criteria 4 and 5 – '*There should be a simple, safe, precise and validated screening test' and 'The distribution of test values in the target population should be known and a suitable cut-off level defined and agreed'.*

Seven studies published since the previous review (January 2015) were included in the review. One study had good generalisability to the UK setting, in terms of population and timing of test, but reported on multiple forms of CAH, including but not limited to the most common form of CAH due to 21-hydroxylase deficiency. All other studies had comparable populations, but took DBSs earlier than in the UK, and most also included various forms of CAH. Four studies reported fluoroimmunoassay of 17-OHP as a first-tier or single-tier test and one study reported a novel liquid chromatography tandem mass spectrometry (LC-MS/MS) first-tier test. Three studies reported second-tier testing using LC-MS/MS to measure various markers. The most accurate were 17-OHP, 21-deoxycortisol or a combination of 17-OHP and cortisol. One study reported the

combined results of a first-tier fluoroimmunoassay and second-tier LC-MS/MS. Cut-offs used to define a positive test were often defined differentially according to parameters such as gestational age, weight and/or day of sample to allow for differences in biomarker levels according to these parameters. Whilst there was not a consensus about which parameters to vary cut-offs by, or which set of cut-offs were most appropriate, the included studies showed that it is possible to describe cut-off values that, within their own specific context, can deliver adequate test performance. Within the context of screening in the UK, the actual cut-off selected could be population and method dependent and further study could clarify a number of variables including the optimum cut-off values to be used by fluoroimmunoassay and LC-MS/MS with respect to weight, gestational age, and age when sampled, including whether cut-offs defined with reference to population biomarker percentiles could be an appropriate method in the UK. Day of sample also varied across studies and was often not reported. Generally, the evidence base was at high risk of bias, since patients did not all receive the same reference standard, and what they received was often dictated by the index test. Fluoroimmunoassays had variable but generally good sensitivity and specificity, though positive predictive value was generally low. The evidence relating to LC-MS/MS had limitations in terms of study designs and sample sizes, though reported reasonably good sensitivity, specificity and PPV. Overall, based on the findings of this evidence summary and due to the limitations of the evidence base, criteria 4 and 5 are not met.

Recommendations on screening

There was some data relevant to UK incidence but limited evidence on the median age of presentation in newborns. The evidence is too limited to draw definitive conclusions relating to the age at which CAH presents and is diagnosed, and what impact screening would therefore have on clinical outcomes. The evidence relating to LC-MS/MS has limitations in terms of study design, sample size and quality, though it shows promise for reducing the number of false positives (but not false negatives) identified by fluoroimmunoassay. Due to the limitations of the evidence, the recommendation is that newborn screening for CAH is still not recommended and that further study is needed to address the evidence gaps.

Limitations

Overall, the review includes limited evidence from the UK, and it is characterised by a lack of studies designed specifically to answer each of the 3 review questions. The searches were limited to English language papers. However, it is unlikely that any pertinent studies would be lost due to the necessary focus on the applicability of screening to the UK and countries that are comparable to the UK. The review employed rapid reviewing methods and therefore sifting was conducted by one of 2 reviewers. However, 20% of all retrieved citations were independently sifted by a second reviewer and discrepancies were discussed and resolved ensuring consistency in the sifting

process. This accepted pragmatic approach should have minimised any risk of errors. Finally, only the author of the GB studies was contacted for further information regarding their methods. It was beyond the scope of this rapid review to contact authors of all included international studies in order to request clarifications of their methods and reported results. Further information regarding aspects of all studies may have provided a greater degree of confidence about the study designs in some cases.

Evidence uncertainties

For the GB studies, it was unclear whether any cases of CAH had been missed using the active surveillance method. Cut-offs used to define a positive test were often defined differentially according to gestational age, weight and/or day of sample to allow for differences in biomarker levels according to these parameters. There was not a consensus about which parameters to vary cut-offs by, or which set of cut-offs were most appropriate. Actual cut-offs could be population and context dependent; they could be defined with reference to UK population biomarker percentiles as has been done elsewhere. It is currently unclear which LC-MS/MS target biomarker may have optimal diagnostic accuracy. The reference standard was problematic in most studies. Whilst improvements may be challenging in large-scale studies, studies relating to second-tier tests with increased methodological consistency could be feasible and they could improve by testing all patients for CAH, since a smaller volume of patients are subjected to second-tier testing. It is unclear which combination of first and second-tier tests might be optimal, and whether diagnostic accuracy of second-tier tests varies according to the patient spectrum identified by the first-tier test. It is also unclear whether the day the DBS is taken impacts on the diagnostic accuracy of the tests, and under what circumstances a DBS might be retested or a second sample taken.

Introduction and approach

Background

Congenital adrenal hyperplasia (CAH) is a group of autosomal recessive disorders. These conditions are caused by enzymatic defects which lead to impairments in cortisol synthesis in the adrenal gland(1-3). If left undiagnosed and untreated, CAH can cause high morbidity and mortality. CAH manifests with a range of clinical and biochemical phenotypes(4). These are commonly categorised into severe (classic) or mild (non-classic) forms. The most common form of CAH is 21-hydroxylase deficiency (21-OHD) which accounts for over 90% of cases(3). 21-OHD can be further subdivided into 3 variants: 1) classic salt wasting (SW); 2) classic simple virilising (SV); and 3) non-classic (NC). Classic salt-wasting accounts for between 70-80% of all cases of 21-OHD and is the most severe. This form causes a life-threatening salt-wasting crisis commonly within the second week after birth and is also characterised by virilisation of female external genitalia. Severe outcomes associated with this form of CAH may include death, hospitalisation, and other complications such as shock, dehydration and vomiting. In the simple virilising form, female patients present with virilised external genitalia but with no clinically apparent salt loss. After birth, the excess androgen presence leads to rapid skeletal growth and premature puberty in all children(5, 6). The non-classic variant may not be detected until later in life and there may be no symptoms. Nonetheless, female patients may exhibit hyperandrogenism, as well as milder symptoms such as short adult stature, premature pubarche or accelerated bone maturation(6). However, it has been suggested that 21-hydroxylase deficiency may be a continuum of phenotypes rather than a number of distinct phenotypical entities(7).

Newborn screening for CAH focuses on the detection of 21-OHD by measuring the enzyme's substrate, 17-hydroxyprogesterone (17-OHP). First-tier tests typically measure elevated 17-OHP concentrations in dried blood spots as a biomarker for 21-hydroxylase deficiency. The cut-off levels for 17-OHP are commonly optimised to achieve maximum sensitivity. However, this strategy results in a large number of false positive results and consequently a low positive predictive value (PPV) of between 1.5-2.5%(8). Premature, stressed or ill babies in particular can display elevated 17-OHP concentrations and the specificity of diagnosis may be improved to some extent by gestational age stratification(9).

More recently, second-tier testing typically involving steroid profiling performed in real time by liquid chromatography tandem mass spectrometry (LC-MS/MS) has been advocated to improve specificity without a loss of sensitivity(10) and this technology is increasingly used in newborn screening laboratories for other elements of the newborn screening programme. Following screening, diagnostic confirmation requires stimulation

of adrenal function and the measurement of cortisol production(4). 21-OH deficiency can be diagnosed by a baseline panel of LC-MS/MS steroids, but often requires cosyntropin testing. This is based on a characteristic rise in adrenal hormones preceding the enzymatic blockage(4).

Many patients with non-classic CAH receive treatment on the basis of their specific symptoms. For severe forms, recommended treatment is with glucocorticoids, for example hydrocortisone, and mineralocorticoids to control electrolytes and plasma renin activity. Treatments aim to prevent any further virilisation and allow normal growth and development. Once growth is complete, more potent glucocorticoids, such as prednisolone and dexamethasone may be used. Babies with salt-wasting CAH require treatment with mineralcorticoids such as fludrocortisone to achieve a plasma renin activity in the healthy range. Sodium chloride and glucocorticoid therapy are also needed(4). Additional doses of glucocorticoids will be needed by all patients with classic CAH during periods of stress (for example, surgery, fever, shock(4, 11)). For patients with non-classic CAH, treatment is given on an ad hoc basis, depending on symptoms(4). Patients with CAH may also receive surgical intervention. For females with severe virilisation, recent guidelines recommend that early genital surgery be performed. For females with minimal virilisation, to the recommendation is to inform parents about surgical options, which include delaying surgery and/or observation until the child is older(12).

Current global landscape of newborn screening

Newborn screening for CAH due to 21-OHD is currently performed routinely in over 40 countries worldwide including 15 within the EU. Whilst the overall incidence is estimated at between 1:14,000 and 1:18,000, this rate varies between different populations. It is more frequent in small, genetically isolated groups, particularly in remote geographic regions(6).

Current policy context and previous reviews

The previous UK National Screening Committee (NSC) review of newborn screening for CAH was published in 2015(13). It concluded that a) around 40 babies are born each year with CAH in the UK, with a suggestion that it is more common in people of an Asian background; b) there are a high number of false positive results for the current screening test (using 17-OHP immunoassay), and that the test performance is particularly poor in premature babies and those with low birth weight; and c) screening might take place too late to reduce the severity of patient outcomes. Therefore, the current recommendation from the UK NSC is that newborn screening for CAH should not be offered.

It was acknowledged in the report that second-tier testing by mass spectrometry had been shown to improve the PPV and consequently the accuracy of the test, but only

limited evidence was available at the time to suggest that this was a viable option. In addition, the previous review did not completely clarify the uncertainties surrounding the UK incidence of CAH including discrete population subgroups. It has also been noted by experts that establishing the median age at presentation may be useful. Therefore, alongside updating the previous searches from 2015, the UK NSC has requested a focus on these issues relating to incidence in the UK and median age of presentation by assessing the available evidence since 2008.

Objectives

The review aims to assess whether there have been significant developments in the evidence base since the previous review was conducted in 2015, and if sufficient support exists for a screening programme for CAH in newborns. The review appraised evidence on the questions in Table 1, which each relate to the criteria set out by the UK NSC for assessing the suitability of a screening programme.

Table 1. Key questions for the evidence summary, and relationship to UK NSC screening criteria

	Criterion	Key questions	Studies Included
	THE CONDITION		
1	The condition should be an important health problem as judged by its frequency and/or severity. The epidemiology, incidence, prevalence and natural history of the condition should be understood, including development from latent to declared disease and/or there should be robust evidence about the association between the risk or disease marker and serious or treatable disease.	KQ1: What is the incidence of congenital adrenal hyperplasia in the UK population, including discrete subgroups of the population? KQ2: What is the median age of presentation of congenital adrenal hyperplasia?	Incidence n=15 studies Median age at presentation n=8 studies
	THE TEST		
4	There should be a simple, safe, precise and validated screening test.	KQ3: What is the accuracy of available screening tests using dried blood spots (DBS) to detect congenital adrenal hyperplasia?	Test performance n=7 studies
5	The distribution of test values in the target population should be known and a suitable cut-off level defined and agreed.	As above	As above

Abbreviations: KQ, key question

Methods

The current review was conducted by the School of Health and Related Research (ScHARR), University of Sheffield, in keeping with the UK National Screening Committee evidence review process(14). Database searches were conducted in July 2020 to identify studies relevant to the questions detailed in Table 1.

Databases/sources searched

All searches were conducted in July 2020 and covered the following electronic databases:

- MEDLINE (including Epub Ahead of Print & In-Process)
- EMBASE
- The Cochrane library

All searches were limited to English Language. For question 1, all searches were limited to publications between 2008 and July 2020 for UK studies, and between 2015 and July 2020 for international studies (update of previous UK NSC evidence summary). For question 2, all searches were limited to publications between 2008 and July 2020, and for question 3 between 2015 and July 2020. Editorials, comments and letters were excluded from the searches and all searches were limited to human studies. Searches were supplemented by hand-searching the reference list of included studies and searching for surveillance reports, for example the British Paediatric Surveillance Unit (BPSU), and the laboratory surveillance scheme comprising diagnostic and genetic laboratories performing testing for CAH. Full details of the searches, including the search strategy for each database, are presented in Appendix 1.

Eligibility for inclusion in the review

The following review process was followed:

1. Each title and abstract were reviewed against the inclusion/exclusion criteria by one of 2 reviewers. Where the applicability of the inclusion criteria was unclear, the article was initially included at this stage in order to ensure that all potentially relevant studies were captured. A third reviewer provided input in cases of uncertainty. Twenty percent of the first reviewer's screening decisions were validated by the second reviewer. Any disagreements were resolved by discussion until a consensus was met.
2. Full-text articles required for the full-text review stage were acquired.
3. Each full text article was scrutinised against the inclusion/exclusion criteria by one of the 2 reviewers. Articles that were determined to be relevant to one or more of the review questions were retained for inclusion in the review. Any disagreements were resolved by discussion until a consensus was met.

The inclusion and exclusion criteria for each review question are outlined in Table 2 below.

Table 2. Inclusion and exclusion criteria for the key questions

Key question	Inclusion criteria							Exclusion criteria
	Population	**Target condition**	**Intervention / Index test**	**Reference standard**	**Outcome**	**Study type**	**Study setting**	
1. What is the incidence of congenital adrenal hyperplasia in the UK population, including discrete subgroups of the population?	Newborn infants	Congenital adrenal hyperplasia due to 21-hydroxylase deficiency	NA	NA	Incidence (birth prevalence) of CAH due to 21-OH deficiency by subgroup: • Classic salt-wasting • Classic simple virilising • Non-classic	Cross-sectional studies; cohort studies; systematic reviews; surveillance reports	Studies from the UK to be prioritised but, in the absence of such studies, those from comparable countries may be reported, for example studies conducted in EEA or OECD countries (excluding Turkey, Israel and Japan)	UK studies before January 2008 or international studies before 2015, non-English language, non-human studies, case reports, editorials, comments and letters

2. What is the median age of presentation of congenital adrenal hyperplasia?	Newborn infants and or children ≤18 years	Congenital adrenal hyperplasia due to 21-hydroxylase deficiency	NA	NA	Age of presentation (median and/or average) of CAH due to 21-OH deficiency by subgroup: • Classic salt-wasting • Classic simple virilising • Non-classic CAH	Cross-sectional studies; cohort studies; systematic reviews; surveillance reports	As above	UK and international studies before January 2008, non-English language, non-classic CAH, non-human studies, case reports, editorials, comments and letters
3. What is the accuracy of the available screening tests using dried blood spots (DBS) to detect congenital adrenal hyperplasia?	Newborn infants	Congenital adrenal hyperplasia due to 21-hydroxylase deficiency	• 17-OHP immunoassay via automated time-resolved dissociated-enhanced lanthanide fluoroimmunoassay or other types of assays as a first-tier or standalone test • Liquid chromatography tandem mass spectrometry as a second-tier test following 17-OHP immunoassay or as a standalone test	• Cosyntropin stimulation test • Urine steroid profile • Genetic analysis • Diagnosis at birth of a female due to the apparent genital ambiguity • Any other specific "gold standard", as determined	Measures of screening accuracy: • Sensitivity • Specificity • Positive predictive value • Negative predictive value • Likelihood ratios • AUC	Studies in randomly assigned or consecutively enrolled populations Case-control studies where randomly assigned or consecutively enrolled studies not available	As above	UK and international studies before January 2015, non-English language; infants (over 28 days of age), children, adults, pregnant women; tests not using dried blood spots, non-human studies, case

	• Any other tests used for screening of CAH	by the study itself		reports, editorials, comments and letters

Abbreviations: AUC, area under the curve; CAH, congenital adrenal hyperplasia; NA, not applicable, 17-OHP, 17-hydroxyprogesterone

Appraisal for quality/risk of bias tool

The following tools were used to assess the quality and risk of bias of each study included in the review:

- Incidence studies: JBI Critical Appraisal Checklist for Studies Reporting Prevalence Data(15)
- Median age of presentation studies: a quality assessment tool developed by Murad *et al.(16)* and based on adaption of the Newcastle Ottawa scale for cohort and case-control studies (with removal of items related to comparability and adjustment), Bradford Hills and Pierson criteria
- Diagnostic accuracy studies: Quality Assessment of Diagnostic Accuracy Studies (QUADAS II) tool(17). The QUADAS II tool was adapted to the specifics of this review, and the scoring schedule with explanatory notes can be found in Appendix 4.

Quality assessment of all included studies was undertaken independently by one of 2 reviewers. Disagreements were resolved by consensus or through discussion with a third reviewer. Results of the quality assessments and appraisal of individual studies are presented in Appendix 3. It was beyond the scope of this rapid review to contact all study authors to ask for clarification on aspects of methodology or reporting of results. Due to the higher relevance of UK data, however, missing or unclear information was sought from authors of UK studies.

Prioritisation approach

The following approach was taken to prioritise studies for extraction:

1. Systematic reviews and meta-analyses would be considered the highest quality of evidence if any were found. Following this, study designs would be prioritised for each question as described in the inclusion/exclusion criteria (Table 2).

2. Studies would be prioritised if they considered a UK population, followed by studies from Western populations analogous to the UK.

In addition, the following criteria were applied after assessing the overall volume of evidence identified in the review:

3. Studies that were published prior to the earliest search dates in the previous UK NSC review (that is, studies that were completed in 2015 or earlier), were not extracted, with the exception of: Q1 UK studies and Q2 studies from January 2008 onwards, which were included.

4. Studies for Q3 (screening test performance) were deprioritised if they did not report all data necessary to calculate test accuracy.

Question level synthesis

Criterion 1 — Incidence and median age of presentation of CAH

1: 'The condition should be an important health problem as judged by its frequency and/or severity. The epidemiology, incidence, prevalence and natural history of the condition should be understood, including development from latent to declared disease and/or there should be robust evidence about the association between the disease marker and serious or treatable disease'.

Question 1 – What is the incidence of congenital adrenal hyperplasia in the UK population, including discrete subgroups of the population?

The previous (2015) UK NSC review(14) evaluated evidence relating to the incidence of congenital adrenal hyperplasia in the UK population. A limited volume of evidence was identified relating to the UK, with 2 publications in unscreened populations reporting CAH incidence in Great Britain(18, 19) (Khalid 2012, Knowles 2013). Incidence data was gathered through the British Paediatric Surveillance Unit (BPSU) over a period of 2 years and an incidence in Great Britain (GB) of approximately 1:18,000 was reported. Whilst the UK rate was noted to be within the range consistent with international incidence rates reported in 11 studies of screened populations, the study found 25% of CAH cases were children of Asian ethnicity, a finding not reflected in the international studies. The review concluded that evidence was limited by the short-term duration of the study and by the limited evidence regarding incidence within discrete subgroups of the UK population.

The aim of this question was to identify and synthesise evidence published since 2015 on incidence of CAH, including any studies since 2008 if they were conducted in the UK.

Eligibility for inclusion in the review

This review searched for cross-sectional studies, cohort studies, and relevant surveillance reports, along with systematic reviews/meta-analyses. Studies conducted in the UK were prioritised, however limited UK evidence was identified, therefore studies conducted in comparable countries were included. Studies were eligible if they reported the incidence of CAH due to 21-hydroxylase deficiency in its various forms: classic salt-wasting, classic simple virilising and non-classic. Studies that reported incidence in newborns were included, however due to a lack of UK newborn incidence data, 2 UK surveillance studies of children with CAH were included. Full details of eligibility criteria are presented in table 2.

Description of the evidence

Database searches yielded a total of 2,326 results, of which 11 were judged to be relevant to this question. An additional 4 relevant articles were identified through hand-searching the reference lists of included studies, so 15 articles were ultimately included in this review.

Appendix 2 contains a full PRISMA flow diagram (Figure 1), along with a table of the included publications and details of which questions these publications were identified as being relevant to the review questions. Appendix 3 contains individual summary characteristics for each included study. Table 4 below in the 'discussion of findings' presents a summary of the incidence data for included studies.

No additional UK studies since the 2015 UK NSC evidence summary were identified, however the 2 UK studies identified in the previous review were included(18, 19). A further 13 international studies were found since 2015. Of these, 8 were outside of Europe: 5 were in the USA(20-24), one in Canada(3), one in Australia(25) and one in New Zealand(26). Of the remaining 5 European studies: 2 were set in the Czech Republic(27, 28), one was from Spain(29), one from the Netherlands(30) and one from Sweden(31).

All studies apart from the 2 UK studies reported data from screened populations, more specifically from regional or national newborn screening programmes, with age at first sampling ranging between 24 and 72 hours in all studies. All first-tier screens were by measurement of 17-OHP from dried blood spots (DBS). The UK studies(18, 19) reported data from unscreened populations identified through a prospective study in the form of a national surveillance programme. This active national surveillance programme was undertaken in England, Scotland and Wales through the BPSU over a 2-year period from August 2007 to August 2009.

The BPSU is a national paediatric rare disease surveillance system. Around 3,000 consultant paediatricians in the UK complete a monthly orange card recording the numbers of children diagnosed during the previous month with those conditions under surveillance (approximately 10 at the time the study was published). Additional cases of CAH between August 2007 and January 2009 were identified through a laboratory surveillance scheme comprising 12 UK laboratories. CAH cases were defined by specified clinical features in association with elevated plasma 17-OHP. All notified cases were subsequently reviewed by an expert panel of clinicians who determined whether a child had CAH or not on the basis of the clinical presentation and investigations reported by local clinicians. The studies included children up to the age of 16, with Khalid 2012(18) reporting incidence for all data (with separate breakdown for age-specific incidence), whilst Knowles 2014(19) reported only data for children with late-presenting CAH (diagnosis at aged one and over).

Six publications shared study data with one other study. Khalid 2012(18) and Knowles 2014(19) both reported data from the BPSU 2-year study of CAH in UK children aged 16 and under, with Knowles reporting on a subset of the identified cases (late-presenting children over one year of age). Pearce 2016(23) and Pearce 2017(23) both reported data from the New York State screening programme between 2007 and 2014, with Pearce 2017 only reporting on Summer/Winter season. David 2018(27) and David 2019(28) both reported data from the Czech Republic newborn screening programme but over different overlapping time periods.

The shortest period of data collection was in Speiser 2020(20), where one year of newborn screening (NBS) data was collected from 17 states in the USA. The longest was Iniguez 2019(29) where data spanning 26 years collected from 6 autonomous communities in Spain was studied. Most of the other studies reported data over at least 7 years of screening (Pearce 2016, Pearce 2017 and David 2019 reported 7 years of data; Fox 2020 and Zetterstrom 8 years; David 2018 10 years; Van der Linde 11 years; Eshragh 14 years and Heather 19 years). The only exceptions were the 2 UK publications(18, 19) and the Australian study(25) which collected data over a comparatively short time period of 2 years. The average duration of incidence data collection across all included studies was 9 years. Sample sizes for screened populations ranged from 202,960 newborns(25), to 4,370,213(21). The average sample size for studies of screened populations was 1,633,659 newborns.

Discussion of findings

A study-level summary of data extracted from each included publication and stratified by question is presented in ‘Summary and appraisal of individual studies’ in Appendix 3.

Quality assessment

The quality of the included studies was appraised using the JBI Critical Appraisal Checklist for Studies Reporting Prevalence Data.

A summary of the risk of bias findings is presented below in Table 3. The full appraisal is presented in Table 22 Appendix 3.

Table 3. Summary of Joanna Briggs risk of bias appraisal for studies on incidence of CAH

	Target population	Appropriate recruitment	Adequate sample size	Setting/subjects described	Data analysis coverage	Measurement criteria	Measurement reliability	Appropriate statistical analysis	Response rate
David 2019	Yes	Yes	Yes	Unclear	Yes	Unclear	Unclear	Unclear	Yes
David 2018	Yes	Yes	Yes	Unclear	Yes	Unclear	Unclear	Unclear	Yes
Eshragh 2020	Yes	Unclear	Yes	Unclear	Yes	Unclear	Unclear	Unclear	Yes
Fox 2020	Yes	Unclear	Yes	Unclear	Yes	Unclear	Unclear	Unclear	Yes
Heather 2015	Yes	Yes	Yes	Yes	Yes	Unclear	Yes	Unclear	Yes
Held 2015	Yes	Unclear	Yes	Unclear	Yes	Unclear	Unclear	Unclear	Yes
Iniguez 2018	Yes	Unclear	Yes	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear
Khalid 2012	Yes	Yes	Yes	Yes	Yes	Unclear	Unclear	Partially	Unclear
Knowles 2014	Yes	Yes	Yes	Yes	Yes	Unclear	Unclear	Partially	Unclear
Lai 2020	Yes	Unclear	Yes	Unclear	Yes	Yes	Unclear	Unclear	Yes
Pearce 2017	No	No	No	No	No	Unclear	No	No	Unclear
Pearce 2016	Yes	Unclear	Yes	Unclear	Yes	Unclear	Unclear	Unclear	Yes
Speiser 2020	No	No	Yes	No	Yes	Unclear	No	Unclear	Unclear
van der Linde 2019	Yes	Yes	Yes	Yes	Yes	Partial	Partial	Yes	Yes
Zetterstrom 2020	Yes	Yes	Yes	Unclear	Yes	Unclear	Unclear	Unclear	Yes

Population – risks of bias
The majority of studies met the inclusion criteria for the target population, with national or regional newborn screening programmes all capturing data from DBSs. However, Pearce 2017(24) only studied seasonal data and Speiser 2020(20) excluded low birthweight newborns from some aspects of their analysis, although it is not clear which aspects. Most studies were judged to have conducted appropriate recruitment to capture the whole newborn population, with the exception of Pearce 2017(24) and Speiser 2020(20) for the reasons given above. For other studies it was unclear whether all live newborns had been included – in some cases this is due to lack of detailed reporting. In Khalid 2012(18) and Knowles 2014(19), it is unclear whether all CAH cases were captured as this relied upon reporting through the BPSU. All studies were judged to have an adequate sample size apart from Pearce 2017(24) which only studied seasonal data.

Measurement – risks of bias
For most studies there was limited description of the case identification methods. Screening protocols, for example cut-off thresholds, varied by region/country. The reliability of measurement of number of CAH cases was also unclear in many studies. This is most likely due to possible issues in the identification of false negatives. Cases missed by screening would be expected to be picked up by clinical presentation, however mild or late-presenting cases may be missed and not recorded. It is therefore difficult to ascertain whether false negative cases were being adequately recorded.

Results
The results for incidence of CAH are presented in Table 4 below. Full study details are provided in Appendix 3.

Table 4. Summary of incidence data

Study	Country	Sampling period (method)	Population	Age at sampling	Total N Sampled	Cases detected (Male:Female)	Incidence			
							Classic CAH (all)	Classic salt-wasting CAH	Classic simple virilising CAH	Non-classic CAH
David 2018 **NB data overlaps with David 2019**	Czech Republic	2006-16 (measurement of 17-OHP from DBS)	All newborns	2002– 2006: 5-7 days; 2006-2009: 3-4 days; 2009-2016: 2-3 days	1,196,387	100 M:F NR	1:11,964	NR	NR	NR
David 2019	Czech Republic	2010-2017 (measurement of 17-OHP from DBS)	All newborns	48-72 hours	888,891	71 M:F NR	1:12,520	NR	NR	NR
Eshragh 2020	USA	2003-2017 (measurement of 17-OHP from DBS)	Newborn screening in 7 states	First screening performed between 24-48h; second between 10 and 14 days of life	2,212,550	164 87M:77F	1:13,491	1:24049	1:30729	NR
Fox 2020	Canada	2010-2018 (NR)	Newborn screening Yukon region	24-48 hours	331,671	Screened cohort: 17 8M:9F Unscreened cohort: 40 21M:19F	1:19,510	1:20,729	1:331,671	NR

Study	Country	Sampling period (method)	Population	Age at sampling	Total N Sampled	Cases detected (Male:Female)	Incidence			
							Classic CAH (all)	Classic salt-wasting CAH	Classic simple virilising CAH	Non-classic CAH
Heather 2015	New Zealand	1994-2013 (measurement of 17-OHP from DBS, but changed to Delta immunoassay in 1998)	All newborns	>48hours	1,175, 973	44 16M:28F	1:26,727 New Zealand European 32/44 Maori 3/44 Pacific Islander 7/44 Other 2/44	NR	NR	NR
Held 2015	USA	2003-2011 (measurement of 17-OHP from DBS	Newborn screening in 7 states	First screen, NR; second screen between 8 and 14 days	4,370,213	374 M:F NR	1:11,685	1:20421[a] (LBW, <2500g: 1:25707[a]	1:74071[a]	1:49103[a]
Iniguez 2019	Spain	1990-2016 (measurement of 17-OHP from DBS)	Newborn screening in 6 Autonomous Communities	48 hours	3,086,015	142 M:F NR	1:21,732	NR	NR	NR
Khalid 2012 **NB data overlaps with Knowles 2014**	UK	2007-2009	Data from national case registry	Unscreened – children presenting with clinical features of CAH and elevated 17-OHP NR	Mid-year estimates from ONS of children <16 years (2007 and 2008): GB: 22,633,700 England: 19,990,900	GB: 135[d] All age groups: 58M:77F <1 year 34M:48F 1<5 year 8M:8F 5<16 16M:21F	Incidence of new diagnoses in children ≤16 years was 0.60 (95% CI 0.50 to 0.71) per 100 000 (1:166,666)[a] Incidence of diagnosis at <1 year of age 5.48 (95% CI 4.42 to 6.81) per 100 000 (1:18,248)[a]			

Study	Country	Sampling period (method)	Population	Age at sampling	Total N Sampled	Cases detected (Male:Female)	Incidence			
							Classic CAH (all)	Classic salt-wasting CAH	Classic simple virilising CAH	Non-classic CAH
					Wales: 1,145,000 Scotland: 1,497,800					
Knowles 2014 **As Khalid 2012 but excludes <1 year olds**	UK	2007-2009	Data from national case registry	Unscreened – children presenting late with clinical features of CAH and elevated 17-OHP	NR	58 26M:32F	1:434,782[c]			
Lai 2020	Australia	2018-2020 (measurement of 17-OHP from DBS)	Newborn screening in New South Wales	<8 days (99.8%)	202,960	10 5M:4F (1 undetermined)	1:20,296	1:22,551	1/202,960	NR
Pearce 2017 **NB data overlaps with Pearce 2016**	USA	2007-20014 (measurement of 17-OHP from DBS)	Newborn screening in New York State	24-48 hours	Summer/ Winter Season: 979,383	Summer/ Winter Season: 52 M:F NR	Summer/ Winter Season: 1:18,834	NR	NR	NR
Pearce 2016	USA	2007-20014 (measurement of 17-OHP from DBS)	Newborn screening in New York State	24-48 hours	1,962,433	108 55M:53F	1:18,170 [b]see footnote for ethnicity	1:21805[a]	1:178,403[a]	1:392,487[a]

Study	Country	Sampling period (method)	Population	Age at sampling	Total N Sampled	Cases detected (Male:Female)	Incidence			
							Classic CAH (all)	Classic salt-wasting CAH	Classic simple virilising CAH	Non-classic CAH
							breakdown by type			
Speiser 2020	USA	2017 (NR)	Newborn screening in 17 states	NR	1,564,756	93 M:F NR	1:16,825 (data also available by each region reported in individual data extraction sheet see appendix 3)	NR	NR	NR
Van der Linde 2019	The Netherlands	2002-2013 (NR)	All newborns	NR	2,235,931	Total:133 SW CAH (114) 73M:41F SV CAH (14) 10M:4F NCCAH (5) 4M:1F	1:16939	NR	1:19613	NR
Zetterstrom 2020	Sweden	2011-2019 (measurement of 17-OHP from DBS)	All newborns (also offered to older children moving to Sweden from other countries without NBS	48-72 hours	1,030,409	87 49M:38F	All CAH 92/1030409 (1:11,200) All classic CAH 1:12,300			1:79,300 (reported)

[a]calculated by review team

Study	Country	Sampling period (method)	Population	Age at sampling	Total N Sampled	Cases detected (Male:Female)	Incidence			
							Classic CAH (all)	Classic salt-wasting CAH	Classic simple virilising CAH	Non-classic CAH
[b]Ethnic breakdown by CAH type ●All CAH White 1:15,610 (56/ 874,066) Hispanic 1:17,450 (19/ 331,589) Black 1:24,840 (12/ 298,057) Asian 1:15,250 (9/ 137,269) Native American, NA (0/3009) Other 1:13,150 (12/157,777) ●Salt wasting CAH White 1:19001 (46/ 874,066) Hispanic 1:23684 (14/ 331,589) Black 1:24838 (12/ 298,057) Asian 1:19609 (7/ 137,269) Native American, NA (0/3009) Other 11/157,777 ●Simple virilising CAH White 1:109258 (8/ 874,066) Hispanic 1:165794 (2/ 331,589) Black 0/ 298,057 Asian 1/ 137,269 Native American, NA (0/3009) Other 0/157,777										
[c]Annual age-specific incidence (risk) of CAH diagnosis between 1 and 15 years of age (based on 52 children notified between 1 September 2007 and 31August 2009) was 0.23 per 100,000 children (Fisher's exact 95% CIs 0.19 to 0.33) 1:434,782										
[d]Excluded 9 cases from August 2007 as might have been prevalent not incident cases										

Study	Country	Sampling period (method)	Population	Age at sampling	Total N Sampled	Cases detected (Male:Female)	Incidence			
							Classic CAH (all)	Classic salt-wasting CAH	Classic simple virilising CAH	Non-classic CAH

Abbreviations: CAH, congenital adrenal hyperplasia; M:F, Male:Female ratio; SW, salt wasting; SV, simple-virilising; NC, non-classic; CI, confidence intervals; DBS, dried blood spots; GB, Great Britain; LBW, low birth weight; NA, not applicable; NBS, newborn screening; NR, not reported, ONS, Office for National Statistics; 17-OHP, 17-hydroxyprogesterone

The GB incidence rate was determined from data collected in an unscreened population spanning 2 years of data collection. All other identified incidence studies were of screened populations, and spanned longer periods, generally over 7 years and up to 26 years.

The highest incidence rates were reported in the Czech Republic over a 10-year period(27) and in the USA in 7 States over an 8-year period(21), both reporting overall CAH newborn incidence of approximately 1:11,000. The lowest overall incidence rates were reported by Heather 2015(26) in New Zealand, with an incidence rate of approximately 1:26,000. The average incidence rate across studies was calculated by the authors of this review as 1:16,869. Seven studies provided separate incidence data for the various types of CAH(3, 21-23, 25, 29, 31). Van der Linde(30) provided incidence data for classic SV type.

The BPSU surveillance methodology had a high reported response rate of 94% of the 3000 clinicians who return monthly cards. One hundred and forty-four children with CAH were identified. Eight of these were identified through the laboratory surveillance scheme described previously, where additional cases of CAH were identified between August 2007 and January 2009, through 12 UK laboratories. One hundred and thirty-two of the children were diagnosed with CAH with 21-OH deficiency. Nine cases from August 2007 were excluded as reporting errors are more likely in the initial month of BPSU reporting (prevalent not incident cases). Mid-year population estimates were obtained by the study authors from the Office of National Statistics (ONS) for 2007 and 2008. The incidence rate of CAH was calculated using the number of children with confirmed diagnosis between September 2007 and August 2009 (135 cases) and denominator data obtained from the ONS mid-year population estimates for 2007 and 2008. Annual incidence of new diagnoses of CAH for the Great Britain was 0.60 per 100,000 (95%CI 0.5 to 0.71), approximately 1 in 18,000 children. Incidence was highest in the first year of life. Annual age-specific incidence of diagnosis was:

- below one year of age: 5.48 per 100,000 (95%CI 4.42 to 6.81)
- between one and 5 years of age: 0.29 per 100,000 (95%CI 0.18 to 0.47)
- between 5 and 16 years of age: 0.26 per 100,000 (95%CI 0.19 to 0.35)

The authors acknowledged that the BPSU study may have under-ascertained asymptomatic or mild cases due to only including children that had been brought to clinical attention. In addition, BPSU surveillance may miss children who die before diagnosis. However, the GB incidence rate (approximately 1:18,248) was consistent with the average incidence rate found across all included studies.

Nine studies (3, 18, 19, 22, 23, 25, 26, 30, 31) reported the sex ratio of the identified CAH cases (see table 4). CAH is recessively inherited and therefore a sex ratio of 1:1 would be expected. An under representation of boys in a study of an unscreened

population may indicate under ascertainment due to a higher death rate in boys from undiagnosed CAH. The significance of the sex ratio in the included studies was rarely reported. Most studies were of screened populations, and very few reported the reason for diagnosis (for example, through clinical diagnosis or via screening). Heather (2015) report a sex ratio of 16M:28F, with females more commonly identified through clinical diagnosis (1M:22F), whilst males were more commonly identified through screening (15M:6F). This is probably due to the fact that girls are more likely to be identified soon after birth due to virilised genitalia, whereas boys might present later with salt-wasting crises, if they have not already been identified through screening.

In addition to the previously reported ethnicity incidence data reported in Khalid 2012, only 2 further studies reported the ethnic breakdown of CAH cases(23, 26). Heather 2015(26) reported an overall newborn CAH incidence rate from the New Zealand newborn screening programme of 1:26,727, with 32/44 cases New Zealand European; 3/44 cases Maori; 7/44 cases Pacific Islanders, and 2/44 cases from other ethnicities. Pearce 2016(23) reported a detailed breakdown of the ethnicities of the 108 detected CAH cases over 7 years of the New York State screening programme, with incidence rates for all CAH, salt-wasting and simple virilising. For all CAH cases, both White and Asian ethnicities showed an incidence rate of approximately 1:15,000, with a rate of approximately 1:17,000 in Hispanics and approximately 1:24,000 in those of Black ethnicity. A similar pattern was seen by type, with both White and Asian ethnicities showing an incidence rate of approximately 1:19,000, Hispanics 1:23,000 and those of Black ethnicity 1:24,000. For the simple-virilising form, incidence rates were low, with rates of 1:109,000 for those of White ethnicity, 1:137,000 for Asians, 1:165,000 for Hispanics and 0 cases in those of Black ethnicity.

For the GB results reported by Khalid 2012(18), 25% of identified CAH cases were of Asian ethnicity, despite only 9% of births in England and Wales being to parents of Asian ethnicity. This suggests that incidence is higher in Asian populations than in other populations. It is unclear why the UK data suggests a higher incidence in Asian populations, whilst the New York State data does not. The UK study did not appear to be designed *a priori* to determine differences in incidence across ethnicities, and no statistical tests were performed to compare these groups. Given the small absolute numbers of cases, the apparent difference may not be statistically significant. Other possible explanations include that there is a real difference, but that the composition of Asian ethnicity differed. In the UK, those of Asian ethnicity are predominantly from India, Pakistan and Bangladesh(32), whilst in the US Asian ethnicity is predominantly East Asian, particularly Chinese(33). Further, the response rate from the BPSU active surveillance was reported to be 94%, however the paper does not report the geographic areas that were represented in this data or which laboratories provided the additional data. These areas may differ considerably in their ethnic compositions compared to those who did not report data. However, the statistical significance of ethnic differences may not be a key factor in influencing decisions regarding the introduction of screening.

Whilst the incidence rate found in GB is broadly comparable with the included international studies, it is noted that incidence of genetic disorders including CAH varies across populations. Khalid (2012) cited differences in estimates of incidence of CAH in older (pre-2000) regional UK studies(34-37), therefore the active surveillance study reported by Knowles and Khalid involving the whole of Great Britain provides the most relevant data available in relation to incidence estimates of CAH in the UK.

Question 2 – What is the median age of presentation of congenital adrenal hyperplasia?

The previous (2015) UK NSC review(13) did not evaluate evidence relating to the median age of presentation of congenital adrenal hyperplasia. However, it was noted that one of the UK studies(18) included age at diagnosis in the context of current UK standards for the reporting of newborn screening results, which should be available on the child health information service system (CHISS) by 17 days of age or under (birth is day 0).(38) Children with CAH who remain undiagnosed at 17 days may therefore benefit from screening. The 2015 review also noted that in the study by Khalid 2012(18), 18 of the 27 newborns who presented with salt-wasting crisis (a severe and life-threatening clinical feature of CAH) presented on or after day 14, when screening results would have been available at the time the review was written.

The aim of this question was to identify and synthesise evidence published since 2008 on the median age of presentation of CAH in newborns.

Eligibility for inclusion in the review

This review searched for cross-sectional studies, cohort studies, and relevant surveillance reports, along with systematic reviews/meta-analyses. Studies conducted in the UK were prioritised, however limited UK evidence was identified, therefore studies conducted in comparable countries were included. Studies were eligible if they reported the median age of presentation of CAH due to 21-hydroxylase deficiency in its various forms: classic salt-wasting; classic simple virilising; and non-classic. Studies that reported median age of presentation in newborns were included, however due to a lack of studies with relevant data relating to newborns, studies of populations of children with CAH were included, provided they included clear data on the median age of presentation or diagnosis of CAH. Full details of eligibility criteria are presented in table 2.

Description of the evidence

Database searches yielded 2,326 results, of which 9 were judged to be relevant to this question. Two publications related to the same study using the same dataset, therefore Bomberg 2015(39) and Sarafoglou 2014(40) are listed here as the same study, resulting in 8 studies in total. Half of the included studies were set in the USA(39, 41-43) 2 were set in the UK(18, 19), one in Canada(3) and one in the Netherlands(44). Sample sizes ranged from 35(42) to 180(43) cases.

Appendix 2 contains a full PRISMA flow diagram (Figure 1), along with a table of the included publications and details of which questions these publications were identified as being relevant to (Table 13). Appendix 3 contains individual summary characteristics

for each included study. Table 6 below in the ‘discussion of findings’ presents a summary of the median age of presentation data for included studies.

Age at first presentation of symptoms was rarely reported. Instead, age at diagnosis was most commonly reported. Most of the studies that reported a median age of diagnosis were in unscreened children who were recruited to a study for a range of differing aims, for example to determine final height(39, 44), to examine the effect of anastrozole on bone mineral density(41), or to estimate the incidence of hypertension(43). Method of diagnosis was mostly by clinical assessment.

No study was identified that aimed explicitly to determine the time to presentation of CAH in newborns. However, one study in Canada(3) and a UK study(18) did provide evidence of median age at presentation in newborns (in the first month of life). A further UK study(19) presented evidence of age of presentation in children with CAH, however this was focused on late presenting CAH in children aged one and over. Both UK studies aimed to record incidence of CAH in children identified via active national surveillance through the BPSU. The aim of the study by Fox 2020(3) was to evaluate the clinical and cost impacts of a newborn screening programme, and therefore collected data both retrospectively from unscreened children with CAH and prospectively (screened children). Data spanning 20 years pre-screening was included, with 8 years of screened data.

From the UK, Khalid 2012(18) prospectively investigated all new diagnoses of CAH in children under 16 years of age in England, Scotland and Wales over a 25-month period between 2007-2009. In particular, the study aimed to assess the potential benefits of pre-symptomatic detection of CAH and explored sex and age-specific incidence of diagnosis, clinical features and time to presentation of children who presented within the first month of life. In order to maximise ascertainment, 12 UK laboratories taking part in a laboratory surveillance scheme were asked to identify CAH cases in order to match cases already identified through the BPSU. Unmatched cases were followed up with their paediatricians. The duration of this data collection was slightly shorter, taking place over 18 months of the whole study period. Notified cases were reviewed by an expert panel who diagnosed CAH based on karyotype, presence or absence of salt-wasting, virilisation, and biochemical and molecular genetic test results. CAH subtype was determined by evaluation of this clinical information. Khalid 2012 noted that the newborn screening programme would be expected to report results by age 14 days, therefore the study authors examined the proportion of newborns who remained undiagnosed at 14 days.

Discussion of findings

A study-level summary of data extracted from each included publication and stratified by question is presented in ‘Summary and appraisal of individual studies’ in Appendix 3.

Quality Assessment

The quality of the included studies was appraised using an adapted checklist developed by Murad *et al(16).* and based on an adaption of the Newcastle Ottawa scale for cohort and case-control studies (with removal of items related to comparability and adjustment), Bradford Hills and Pierson criteria.

A summary of the risk of bias for question 2 is presented in Table 5 below, and the full appraisals are presented in Table 23 Appendix 3.

Table 5. Summary of risk of bias for studies on median age of presentation of CAH

	Selection	Ascertainment		Reporting	Overall
		Exposure	Outcomes		
Bomberg 2015/Sarafoglou 2014	High	Low	Unclear	Unclear	Unclear
Fox 2020	Unclear	Low	Low	Low	Unclear
Halper 2019	Unclear	Unclear	Unclear	High	Unclear
Hsieh & White 2011	Low	Unclear	Low	High	Unclear
Khalid 2012	Unclear	Low	Low	Low	Unclear
Knowles 2014	Unclear	Low	Low	Low	Unclear
Maccabee-Ryaboy 2016	High	Low	Low	Unclear	Unclear
Pijnenburg-Kleizen 2019	Low	Low	Unclear	Unclear	Unclear

Risk of bias - Selection

Two studies were at high risk of selection bias(39, 43), whilst half had an unclear risk of selection bias(3, 18, 19, 41). Only 2 studies had a low risk of selection bias(42, 44). None of the studies' main aims was to determine the median age of presentation of CAH in newborns, although this was reported in Fox (2020) and Khalid (2012). Due to this, many of these studies stipulated specific selection criteria, such as a certain number of height/blood pressure measurements meaning these populations were highly selective. Due to the majority of studies being either at high risk or unclear risk the overall judgement is that there is a high risk of selection bias.

Risk of bias – Ascertainment

A key source of risk of ascertainment bias across studies was the possibility of under ascertainment of cases with asymptomatic or mild symptoms, due to the study populations being unscreened cohorts where presentation with symptoms was the method of identification.

- Exposure

Exposure was diagnosis of CAH. For most studies, description of the method of diagnosis was given despite most studies being focused on their main aims, which were not determining age at diagnosis. In 2 of these studies, however, cohorts of children with CAH were included by virtue of being registered patients at, for example, a particular

institution, and the details of their initial diagnosis was not described. Description of clinical diagnosis for the UK studies was described. Diagnosis was based on specified clinical features in association with elevated 17-OHP. Initial diagnosis was verified by an expert panel.

- Outcome

Outcome was age at presentation or diagnosis. Many studies did not provide a breakdown of age at presentation by sub-type of CAH. Most studies did not provide any detailed description comparing the age of presentation of symptoms with age at diagnosis, nor did they describe the nature of the presenting symptom. Overall, the studies were generally regarded as being at low or unclear risk of ascertainment bias.

Risk of bias – Reporting

The studies were not designed to ascertain the median age of presentation of CAH. As a result, the reporting of the details of CAH cases was generally insufficient for this purpose. Rather than focusing on factors pertinent to diagnostic criteria, reporting tended to focus on other factors, for example. hypertension measurements. Only Fox 2020(3), Khalid 2012(18) and Knowles 2014(19) reported sufficient detail of factors associated with diagnosis, a reflection of their different study focus.

Overall, the quality of all studies was judged to be at unclear risk of bias. It is worth noting that studies were judged on risk of bias for answering the question on median age of presentation. None of these studies was explicitly designed to answer this question, and overall judgements are therefore based on this and are not a judgement on the risks of bias for the studies' actual stated objectives.

Results

The results for the median age of presentation of CAH are presented in Table 6 below. Full study details are provided in Appendix 3.

Table 6. Summary of median age of presentation data for CAH

Study	Population	Method of diagnosis	Median age of presentation			
			All CAH	Classic SW	Classic SV	Non-classic
Bomberg 2015/Sarafoglou 2014 ***USA***	104 unscreened children with classic CAH from 3 medical institutions in Minnesota, identified by retrospective chart review, between 1955-2012	Diagnosis based on based on clinical, hormonal, biochemical, and in some cases, molecular testing.	NR	Mean age at diagnosis 1.6 +/- 3.1 months Female, n=38 (mean ±SD): 0.6 ±0.8 months Male, n=28 (mean ±SD): 3.1 ±4.4 months	Mean age at diagnosis was 37.4 +/-31.2 months Female, n=25 (mean ±SD): 35.2 ±34.1 months Male, n=13 (mean ±SD): 43.8 ±27.5 months	NR
Fox 2020 ***Canada***	Prospective analysis of regional screening programme of newborns in Yukon Territory between Nov 2010- March 2018. N=57 positive for CAH	Two-tier testing: immunoassay and LC-MS/MS	Median days to positive screen was 6 and age at diagnosis was 5 days (range, 0-30 days) and 6 days (range, 0-13 days) in unscreened and screened	NR	NR	NR

Study	Population	Method of diagnosis	Median age of presentation All CAH	Classic SW	Classic SV	Non-classic
			populations, respectively.			
Halper 2019 ***USA***	56 retrospectively identified children with CAH from Minnesota University Hospital	NR	Mean age at diagnosis given by treatment group. Using anastrozole group 3.27 years (SD 3.51), not using anastrozole group 1.76 (SD 3.21)	NR	NR	NR
Hsieh & White 2011 ***USA***	35 children with CAH from a cohort of 77 children with primary adrenal insufficiency, between 1999-2010	NR	22/35 were diagnosed as inpatients at CMC during infancy, with median (interquartile range, IQR) age of 9 d (2.5–12.7) at presentation	NR	NR	NR
Khalid 2012 ***UK***	144 unscreened children diagnosed with CAH (132 with 17-hydroxylase deficiency) of which 77 diagnosed at <1 month, identified via active national surveillance through British Paediatric Surveillance Unit	One or more of the criteria are met: - Virilisation of female genitalia - Adrenal crisis (and/or salt wasting crisis) - Adrenal insufficiency (non-life threatening)	Median age at presentation (days, IQR) 1 (0–14) Median age at presentation (days, IQR)	Median age at presentation of salt-wasting crisis (days, IQR) Males (n=24) 15 (11–20)	NR	NR

Study	Population	Method of diagnosis	Median age of presentation All CAH	Classic SW	Classic SV	Non-classic
		- Incomplete masculinisation of male genitalia - Precocious puberty - Accelerated skeletal age - Short stature - Hypertension - Family history AND one or more of the following: - Elevated 17-OHP in blood test - Positive synacthen stimulation test - Test diagnostic of rarer form of CAH (3β HSD deficiency)	Males (n=33) 14 (9–18) Females (n=44) 0 (0–1)	Females (n=3) 15 (13–16) Total (n=27) 15 (11-19) range 9 to 30 days		
Knowles 2013 ***UK***	As Khalid but 58 children with late-presenting CAH between 1-15 years of age	As Khalid	Median age at presentation (years, IQR) 21-hydroxylase deficiency (n=50) 5.6 (4.2-7.8) 11β-hydroxylase deficiency (n=6) 9.1 (6.4-13.9) Males (n=26) 5.4 (4.2-8.1)	NR	NR	NR

Study	Population	Method of diagnosis	Median age of presentation			
			All CAH	**Classic SW**	**Classic SV**	**Non-classic**
			Females (n=32) 6.4 (4.7-8.4) White (n=40) 5.5 (4.5-7.6) Asian/British Asian (n=12) 7.7 (3.1-13.3) Other (n=6) 7.4 (6.3-10.1)			
Maccabee-Ryaboy 2016 ***USA***	180 screened and unscreened patients with CAH at 3 paediatric centres in Minnesota, identified from retrospective chart review. Only patients who had BP measurements from at least three separate clinic visits were recruited, between 1970-2013.	Subjects were divided into two time periods, those born between 1970– 1994 (pre-newborn screening) and 1995–2013 (shortly after newborn screening for CAH was initiated in Minnesota)	NR	Males average age at diagnosis 3.7 months +/-8 and females 1.5 months +/-4	Males average age at diagnosis 47 months +/-36 and females 40 months +/-36	NR
Pijnenburg-Kleizen 2019 ***The Netherlands***	39 unscreened patients treated during childhood for classic CAH due to 21-hydroxylase deficiency between 1980-1997, identified for a retrospective evaluation of longitudinal data.	Diagnosis of SW-CAH or SV-CAH was based on clinical and biochemical data and confirmed by mutation analysis	NR	24/25 were diagnosed neonatally because of ambiguous genitalia and/or SW crises. 1/25 diagnosed at 2 years.	Female median age of diagnosis 2.3 years (range 0–4.0 years), male age at diagnosis 4.4 years (range	NR

Study	Population	Method of diagnosis	Median age of presentation			
			All CAH	Classic SW	Classic SV	Non-classic
					2.5–6.3 years)	

Abbreviations: CAH, congenital adrenal hyperplasia; IQR, Inter Quartile Range; LC-MS/MS, liquid chromatography tandem mass spectrometry; NR, not reported, SD, standard deviation; SV, simple virilising; SW, salt-wasting; 21OHD, 21-hydroxylase deficiency

Whilst none of the included studies were explicitly designed to evaluate time to presentation of symptoms of CAH in newborns, data was identified in a small number of studies. For many studies the method of data collection was by retrospective chart review of all children with CAH being treated at a particular institution. Age at first presentation of symptoms was rarely reported. Instead, age at diagnosis was most commonly reported. There was little evidence indicating time elapsed from presentation of symptoms to diagnosis in unscreened children.

A large-scale study by Fox 2020(3) reported time to presentation of symptoms in screened and unscreened newborns from a single institution. The study reported data from a sample of 17 screened and 40 unscreened children with CAH diagnosed in a tertiary hospital in Canada over a 20-year period. Males and females were similarly represented in both screened (8/17 were males) and unscreened (21/40 were males) cohorts. Age at diagnosis for the screened cohort was 6 days (on receipt of screening results). The median age of diagnosis for the unscreened cohort was earlier at 5 days and was due mainly to ambiguous genitalia (48%), family history (2%) or a salt wasting crisis (25%). No child in the screened cohort had a salt-wasting crisis. Ten of the 40 children in the unscreened cohort were diagnosed due to a salt wasting crisis, of whom the majority were boys (8 of the 21 males in the cohort). For male children in the screened cohort, the median age at diagnosis was 5.5 days, compared with the unscreened cohort where median age of diagnosis was 14 days. This data was not presented for females. Data for the unscreened cohort was collected retrospectively, with age at presentation defined as time of the first endocrine consultation or by telephone where advice was given. It is unclear whether later-presenting children with milder forms of CAH were captured by the data collection method, or whether only children presenting with early symptoms were included.

Results from the UK study focusing on late presenting CAH(19) reported a median age at presentation of 5.9 years. Specifically, diagnosis through presentation of secondary sexual characteristics occurred at a median age of 5.8 years. Khalid 2012(18) reported data from the same dataset but for all children with CAH, with a specific focus on evaluating time to presentation within the first month of life. The study was conducted in an unscreened population and therefore gives an indication of the potential benefits of newborn screening for CAH. During the study period, 144 children were diagnosed with CAH, of which 77 presented in the first month of life. Of these, 27 presented with a salt-wasting crisis, at a median age of 15 days. The authors reported a median age at presentation of a salt-wasting crisis of 15 days for both males and females. However, only 3 girls presented with a salt wasting crisis in the first 30 days compared to 24 boys. Three children (2 boys and 1 girl) presented with adrenal insufficiency, at a median age of 10 days. Forty-seven children presented at a median age of 0 days, of which 11 were diagnosed due to having an affected sibling, 2 presented with incomplete masculinisation, and 34 presented with virilisation of female genitalia. For children who presented in the first month of life, the median age at clinical presentation was the day of

birth for girls and 14 days for boys. The earliest salt-wasting crisis was 9 days, with 67% presenting with a salt-wasting crisis at 14 days or later. By day 14, when newborn screening results were presumed to be available, only 6% of girls but 50% of boys remained undiagnosed. Sixteen out of 33 boys were diagnosed by day 14, and 23 out of 33 boys were diagnosed by day 17 when current guidelines suggest results should be available. Forty-one out of 44 girls were diagnosed by day 14 whilst all 44 were diagnosed by day 17 (unpublished data from authors). These data and the observation that females presented most commonly at birth due to virilisation of genitalia suggests that girls may be less likely to benefit from screening. In contrast, boys are more likely to present with a severe and potentially life-threatening salt-wasting crisis around 14 days from birth or later and therefore could benefit more from screening.

Khalid 2012(18) reported the median age of presentation of a salt-wasting crisis as 15 days for both boys and girls. However, in other studies females with classic salt-wasting CAH were reportedly diagnosed at an earlier age than males. For example, in Maccabbee-Ryaboy 2016(43), males were diagnosed with a salt wasting crisis at 3.7 months compared to females at 1.5 months; in Bomberg 2015/Sarafoglou 2014(39, 40), more females than males were diagnosed with a salt wasting crisis (38 compared to 28), and at an earlier time (females presented with a salt wasting crisis at 0.6 months compared to males at 3.1 months). These latter studies are notable in that they were retrospective reviews of medical records of highly selected populations, with the studies conducted for other specific aims. Classic simple virilising CAH was consistently diagnosed later than classic salt-wasting No breakdown was given for ethnicity for time to presentation for children under one year old(18). However, Knowles 2014 reported median age of diagnosis of CAH for late-presenting cases (over one year of age to 16 years) and of these children, those of Asian ethnicity were diagnosed at a median age of 7.7 years, whilst children of White ethnicity were diagnosed at an earlier median age of 5.5 years. For late-presenting cases, boys were diagnosed slightly earlier at median 5.4 years compared to girls who were diagnosed at a median age of 6.4 years.

Overall, the included studies provide a limited amount of evidence to indicate patterns of development and presentation of symptoms in children with CAH in their first month of life. It is not possible to draw firm conclusions on what impact the introduction of screening would have. Only one study of newborns was set in the UK(18). The number of affected babies might be already clinically diagnosed by the time any screening result would be available, depending on when screening takes place, although the data presented in Khalid 2012(18) indicates that a small proportion of salt-wasting crises occurs earlier to this and therefore screening would be of no benefit to these children. This is further reflected in the study of newborns by Fox 2020(3) who report a median age of presentation of 10 days for a salt-wasting crisis in their unscreened cohort. Further studies designed to address this evidence gap and establish the age at presentation of CAH may be achievable and would allow for an informative evaluation of the impact that newborn screening could reasonably have.

Summary of Findings Relevant to Criterion 1: Criterion 'not met'[1]

Quantity: this review identified 20 studies that offered evidence towards criterion 1. Fifteen studies were identified that provided CAH incidence data, with no new studies (not already included in the previous review) from the UK, although there was a moderate number of studies from international countries (13). Incidence studies were large scale reports of regional and national newborn screening programmes. Whilst the GB incidence data is based on an unscreened population identified through BPSU, the indicated incidence of 1:18,000 is comparable with the incidence rates reported by the international screening programme data. No studies were identified that were designed to identify the median age at presentation of newborns with CAH. Only 2 studies reported median age of presentation in newborns with CAH. Most studies that reported age at presentation did so as a demographic characteristic of a population taking part in a study designed with a different aim.

Quality: for studies reporting incidence of CAH, most studies were at low risk of bias for their targeting of newborns with CAH, with adequate sample sizes, response rates and data analysis coverage. Much of the reporting for measurement criteria and reliability was unclear, and it was beyond the scope of this rapid review to contact authors of international studies to request clarifications regarding methods and reporting of results. Further information regarding aspects of all studies may have provided a greater degree of confidence about the study designs in some cases. The duration of study for the UK incidence study was at the lowest end of the range at only 2 years. Most studies reporting median age at presentation were not considered to be high quality due to the explicit aim of the studies not being to explore median age of presentation and diagnosis in newborns. Populations were highly selective, with inclusion criteria designed for the purpose of the study.

Applicability: the GB BPSU study provided data on incidence of CAH in an unscreened population, with the reported incidence rate comparable with the international data from screened populations. Few studies reported ethnic breakdown of CAH cases, and the differences in ethnic classifications between countries make any comparisons between studies difficult. The 2 studies that specifically addressed age at presentation in the first month of life reported males more commonly presenting

[1] **Met** -for example, this should be applied in circumstances in which there is a sufficient volume of evidence of sufficient quality to judge an outcome or effect which is unlikely to be changed by further research or systematic review.
Not Met - for example, this should be applied in circumstances where there is insufficient evidence to clearly judge an outcome or effect or where there is sufficient evidence of poor performance.
Uncertain -for example, this should be applied in circumstances in which the constraints of an evidence summary prevent a reliable answer to the question. An example of this may be when the need for a systematic review and meta-analysis is identified by the rapid review.

with salt wasting crises. Females were reported to present more commonly at birth due to virilised genitalia. Screening may therefore be of more benefit to boys, if the results are available before a severe or potentially life-threatening salt wasting crisis.

Consistency: there were a range of reported incidence rates, however some variation between populations would be expected. Median age of diagnosis of sub-types of CAH was consistent between studies.

Conclusions: despite the availability of incidence data relevant to the UK, the criterion is not met because the current evidence on median age of presentation in newborns is too limited in terms of volume and quality to draw definitive conclusions relating to the impact screening would have on clinical outcomes.

Criteria 4 and 5 — Test accuracy of available screening tests in dried blood spots to detect CAH

4: 'There should be a simple, safe, precise and validated screening test'.

5: 'The distribution of test values in the target population should be known and a suitable cut-off level defined and agreed'.

Question 3 – What is the accuracy of available screening tests using dried blood spots (DBS) to detect congenital adrenal hyperplasia?

The previous (2015) UK NSC review(13) evaluated evidence relating to the incidence of congenital adrenal hyperplasia in the UK population. It concluded that criterion 4 was not met since cases were missed and consequently clinical diagnoses delayed as a result of screening. The authors noted that a number of patients presenting with clinical signs and symptoms were detected before screening could be offered. The review also noted that false positives were more prevalent in pre-term neonates, and that liquid chromatography tandem mass spectrometry (LC-MS/MS) may be able to improve positive predictive values (PPVs). However, the 2015 evidence summary also noted that, at the time, data for strategies such as second-tier LC-MS/MS was very limited, with no large cohort studies undertaken in an unselected population that adequately demonstrate an improvement applicable to a UK screening population.

Eligibility for inclusion in the review

This review searched for studies in randomly assigned or consecutively enrolled populations, however due to a limited number of these studies, case-control studies were also included. Studies conducted in the UK or countries comparable to the UK were included. Studies were eligible if they reported the diagnostic accuracy of an index test used to diagnose CAH due to 21-hydroxylase deficiency in newborns using dried blood spots. Studies reporting first-tier tests (mass screening), or second-tier tests (screening those identified by a first-tier test) or the combination of both were included. A range of reference standards were eligible for inclusion. This is because a perfect reference standard could not be defined: for example, genetic testing may miss some patients where the mutation is novel; LC-MS/MS is not a perfect test in all situations, and long term follow-up may miss some patients with mild disease. Studies providing both sensitivity and specificity, or providing data allowing their calculation, were prioritised. Studies reporting incomplete data on diagnostic accuracy were identified and summarised narratively, but they were not formally included nor quality appraised. Full details of eligibility criteria are presented in table 2.

Description of the evidence

Database searches yielded 2,326 results, of which 7 were judged to be relevant to this question.

Appendix 2 contains a full PRISMA flow diagram (Figure 1), along with a table of the included publications and details of which questions these publications were identified as being relevant to (Table 13). A summary of test performance data is presented in Table 8 in the 'discussion of findings'.

In total, 7 studies were identified that were published since 2015 and that reported data on sensitivity and specificity of CAH screening tests or provided data that allowed these to be calculated. No studies were from the UK. There were 3 studies from the USA(23, 45, 46), 2 from the Netherlands(30, 47), one from Germany(48) and one from New Zealand(26). These countries were selected to ensure applicability of findings to the UK, since they have comparable population characteristics and healthcare systems. Five studies(23, 26, 30, 45, 47) were retrospective analyses of national or regional newborn screening programmes. Three of these recruited large samples(23, 30, 45) (largest reported n=2,235,931(30); lowest reported n=439,227(45)). One(26) study did not report the total sample size but recruited all newborns in New Zealand from 2011-2013. Another study(47) recruited a small sample of first-tier positive patients (n=100). The remaining 2 studies were case control studies which drew samples from national screening programmes (n=100(48) and n=45(46)).

A further 5 studies were identified that reported partial diagnostic accuracy data or assumed there were no false negatives. These 5 were deprioritised since more methodologically robust studies with more relevant information were available. These were not fully data extracted and were not quality appraised. However, they are described in brief in the 'deprioritised studies' section in the 'discussion of findings' below and in Appendix 5.

The mass screening strategies reported across the 7 included studies were either single-tier (one mass screening test, sometimes with repeat testing or repeat sampling), or two-tier (a first-tier mass screening test, followed by a second-tier test for those who had a positive first-tier test). The aim of second-tier tests was to reduce the number of newborns referred to specialist services. Of the 7 studies included in the review, 4 reported results for a fluoroimmunoassay as first-tier or stand-alone tests(23, 26, 30, 48) and one for a novel version of LC-MS/MS as a first-tier test(48). Three reported second-tier testing using LC-MS/MS. One study(45) did not report results for the tests separately, but for fluoroimmunoassay and LC-MS/MS as a whole.

All the fluoroimmunoassy studies assayed 17-OHP levels(23, 26, 30, 48). It was unclear what marker the study(48) that assessed a novel first-tier LC-MS/MS test was using, but

this may have been 17-OHP or a ratio of the sum of 17-OHP and androstenedione, to cortisol. The second-tier LC-MS/MS studies reported a range of different markers: 2 studies used 17-OHP(46, 48), one study used 21-deoxycortisol and several other candidate markers (11-deoxycortisol; 11-deoxycorticosterone; 17-hydroxyprogesteron; Δ4-androstenedion; corticosterone; cortisone; cortisol), and one study used the ratio of 17-OHP to cortisol, or a combination of 17-OHP and the ratio of 17-OHP to cortisol.

Three studies(46-48) did not report the time that the DBS was taken. One study took the DBS sample at 12-23 hours after birth(45), one at 24-48 hours after birth(23), one at more than 48 hours after birth(26), and one at 3-7 days after birth(30). This may limit the applicability of the evidence base to the UK where samples are taken 5-8 days after birth. It is unclear to what extent the timing of the sample will impact on estimates of diagnostic accuracy. One study considered the effect of timing of sample on diagnostic accuracy(45), though only considered the difference between 12-23 hours and 24-48 hours, both of which are before the sampling time frame in the UK. One other study(23) compared no re-sampling to re-sampling any for whom the DBS was taken before 24 hours after birth. One screening programme in the Netherlands mandated that a second sample is taken within 7 days of the first sample if the sample is inconclusive. Re-testing (using the original DBS) or re-sampling positive newborns (not necessarily according to the timing of the sample) was also a feature in most of the 14 US states included in one of the deprioritised studies(21), and in a study from New York(23). In New Zealand(26), positive samples were re-tested after diethyl extraction. Re-testing or re-sampling is usually undertaken to reduce the number of false positives.

There was high heterogeneity between studies regarding the exact cut-offs used to define a positive test. Five studies used variable cut-offs. These varied according to age and weight(23), gestational age (or birth weight if gestational age not known(30)), weight alone(26, 45), or gestational age alone(47).

Only one study(30) aimed to identify 21-hydroxylase deficiency. One(23) also reported positive screening test results for other forms of CAH. In the remaining 5 studies it was unclear which CAH diagnoses were included(26, 45-48). This should be noted when interpreting the evidence. The reference standards also varied across studies. Four of the studies that retrospectively analysed the performance of national or regional screening programmes(23, 26, 30, 45) had reference standards where patients with negative first-tier or one-tier tests received no further testing, but where clinicians could alert national or regional registries of false negative cases that were identified later. Those with positive tests were subjected to variable and often poorly described diagnostic work-up strategies, that usually included at least clinical signs and symptoms, blood tests and urine tests and sometimes genetic confirmation, abdominal ultrasound and LC-MS/MS. One retrospective analysis of a national screening programme(47), that sampled first-tier positive patients, had a reference standard of routine diagnostic work-up, with 21-hydroxylase deficiency confirmed through genetic testing. Both case-control

studies(46, 48) sampled from CAH cases, and first-tier negative controls diagnosed through routine screening. It was unclear for these latter 3 studies if any first/one-tier false negative patients were included, but the small sample size of the studies and low prevalence of CAH makes it unlikely that any were.

Discussion of findings

Quality Assessment

The quality of the included studies was appraised using an adapted QUADAS II checklist. A summary of the risk of bias and applicability to the UK setting is presented in table 7 below, and the full appraisal is presented in Table 24, Appendix 3.

Table 7. Summary of QUADAS II assessment for included studies

	Risk of Bias				Applicability		
Author, year	**Patient selection**	**Index test**	**Reference standard**	**Flow and timing**	**Patient selection**	**Index test**	**Reference standard**
Boelen 2016	Unclear	Unclear	High	Unclear	Unclear	Unclear	Low
Gaudl 2019	High	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear
Han 2019	High	High	Unclear	Unclear	Unclear	Unclear	Unclear
Heather 2015	Unclear	Unclear	High	High	Unclear	Unclear	Unclear
Pearce 2016	Low	Unclear	High	High	Low	Unclear	High
Tang 2016	High	Unclear	High	High	Low	Unclear	Unclear
van der Linde 2019	Low	Unclear	High	High	Low	Low	High

Overall, the quality and applicability of the studies was generally poor or unclear. This is largely because most studies were not designed to assess the diagnostic accuracy of an index test, but to assess the diagnostic performance of a screening programme already in operation.

Patient selection

Of the 5 studies(23, 26, 30, 45, 47) that recruited a national or regional screening cohort, 2 also avoided inappropriate exclusions and scored low risk of bias for patient selection. Of the remaining 3 studies: one(45) included patients from different years for false-negative and false-positive outcomes and was deemed at high risk of bias; one study(26) did not state how non-severe cases were included in the analysis and was judged at unclear risk of bias; and one study(47) was judged at unclear risk of bias as it did not state how patients were selected from the national cohort, except that they had a positive first-tier test. The 2 case-control studies(46, 48) were at high risk of bias because of the intrinsic bias of this type of study for assessing diagnostic accuracy, and

since it was unclear whether false negative patients were included in the CAH positive samples.

Index test

It was unclear whether the index test's conduct could have introduced bias in 6 studies(23, 26, 30, 45, 47, 48), and there was a high risk of bias in one study(46). All tests were unclear about whether the results of the index test were interpreted without knowledge of the reference standard, and were therefore judged to be at unclear risk of bias with respect to the conduct of the index test. This was because the reference standard either included clinical signs, which could be known when the index test was interpreted, or were unclear on what the reference standard was. In the case-control studies, it was unclear if the cases and controls were blinded(46, 48). Five studies(23, 26, 30, 45, 47) pre-specified the cut-off values used. One study(46) was judged to be at high risk of bias as the cut-off was not pre-specified [receiver operating characteristic (ROC) curves were used to determine the best cut-off], and one study(48) was unclear on this point.

Reference standard

The reference standards applied varied. Four large-scale studies(23, 26, 30, 45) used routine diagnostic work-up and included ascertainment of false negatives through clinical registries or clinical reporting, one study used genetic testing(47), and the 2 case-control studies(46, 48) selected DBS screen-positive and screen-negative patients from routine diagnostic practice. The conduct of the reference standard was judged to be at high or unclear risk of bias in all studies because all studies had a differential reference standard dictated by the index test or were unclear on this point. Additionally, all but one study(47) identified 21-hydroxylase deficiency CAH as well as other variants, or were unclear regarding this matter.

Flow and timing

Studies were at unclear or high risk of bias in this domain. The large-scale studies(23, 26, 30, 45) suffered from partial or differential verification bias, since it was not possible to give the same reference standard to all patients. Often the index test dictated which patients received a reference standard (for example, those with a positive DBS test result went on for further tests, whilst those with a negative DBS test results received only long-term follow-up(23, 26, 30, 45)). False negatives were assessed through voluntary reports to national registries, leaving studies at high risk of missing false negatives. There were also problems with the interval between index test and reference standard, since not enough time has elapsed since patient recruitment to pick up cases that may present clinically after several years. This would most likely lead to overestimates of sensitivity and negative predictive value (NPV), but the extent to which this has occurred is unknown. A more robust reference standard, applied to all patients, would have been challenging when assessing screening programmes of this size.

The small-scale case control studies(46, 48) probably also suffered from partial or differential verification bias, as participants were drawn from routine screening samples, though the details were not reported. The one other small-scale study(47) drew its sample from first-tier screen positive newborns and used routine diagnostic work-up with genetic confirmation as the reference standard, which is also at risk of differential verification bias.

Applicability
No studies were fully applicable to the UK setting. The closest match was a study from the Netherlands(30), which took its samples between 3 and 7 days after birth, but aimed to identify 21-hydroxylase deficiency and other forms of CAH.

The newborn population included in 3 of the studies recruiting national or regional samples(23, 30, 45) was judged to be similar to the UK population, whilst the remaining studies were deemed at high or unclear risk for applicability due to exclusions or a lack of clarity. Only one study was judged to be at low risk for applicability with respect to the index test as it took samples between 3 and 7 days, which is fairly comparable to the UK setting. Since it is unclear if timing of sample will impact on estimates of diagnostic accuracy, all other studies were at unclear risk of bias: 3 studies did not state at what time the sample was taken, one study(26) took the sample at an unspecified point after 48 hours, and 2 studies(23, 45) took samples before 48 hours after birth. Only one study(47) used a reference standard that only identified 21-hydroxylase deficiency, whilst the remaining studies either included other forms of 21-hyroxylase deficiency or did not clarify what the target condition was.

Results
The results of screening test accuracy for CAH are presented in Table 8 below. Full study details are provided in Appendix 3.

The results are organised as follows:

i. studies reporting single-tier or fist-tier tests
ii. studies assessing second-tier tests
iii. studies reporting the results of two-tier screening (that is, they did not report results for the 2 tests separately)
iv. data by CAH type
v. data by gestational age/birth weight

The studies that were deprioritised are briefly discussed in under the section labelled ‘Deprioritised studies – narrative summary only’

Table 8. Summary of test accuracy data

Study	Population	Index test	Repeat testing	Day of test	Cut-off	Target condition and reference standard	Accuracy data
Single-tier or first-tier tests[a]							
Gaudl 2019 ***Germany*** ***Case-control***	100	Fluoroimmunoassay 17-OHP	NR	NR	NR	TC: unclear which CAH types RS: Unclear, possibly routine screening methodology in Germany	TP: 3 FP: 61 FN: 0 TN: 36 PPV: 4.7%[b] NPV: 100%[b] Sensitivity: 100%[b] Specificity: 37.11%[b]
		Novel high-throughput LC-MS/MS methodology Unclear if 17-OHP or (17-OHP+androstenedione)/ Cortisol	NR	NR	NR		TP: 3 FP: 4 FN: 0 TN: 93 PPV: 86%[b] NPV:100%[b] Sensitivity: 100%[b] Specificity: 95.88%[b]
Pearce 2016 ***USA (New York State)*** ***Cohort study: Retrospective analysis of***	1,962,433 newborns 2007-2014	Fluoroimmunoassay of 17-OHP (AutoDELFIA) Immediate referral for very high values[c] after 2010	New sample for babies <24 hours old when DBS taken, after 2010	24–48 h after birth	Varies by age when sampled and birth weight	TC: 21-hydroxylase deficiency and other forms of CAH RS: Diagnostic testing - not	TP: 105 (90 SW CAH, 8 SV CAH, 5 NC CAH, 2 other enzyme deficiency) FP: 2,371 FN: 3 (NR which type) TN: 1,959,954 PPV: 4.24%

regional programme			Re-test sample twice if elevated 17-OHP; if mean of 3 test elevated, second sample taken			further described. False negatives reported, but authors assume this is underreported.	NPV: 100% Sensitivity: 97.22% Specificity: 99.88%
	1,001,820 newborns 2007-2010	As above, but no retest for babies <24 hours old when DBS taken and no immediate referral for very high values			As above		PPV: 2.90% NPV: 100% Sensitivity: 95.00% Specificity: 100%
	960,430 newborns 2010-2014	As above, with retest for babies <24 hours old when DBS taken, and immediate referral for very high values			As above		PPV: 5.80% NPV: 100% Sensitivity: 98.5% Specificity: 100%
Van der Linde 2019[d] ***The Netherlands*** ***Cohort study: Retrospective analysis of national programme***	2,235,931 newborns	Fluoroimmunoassay of 17-OHP (AutoDELFIA or GSP®-instruments)	Second sample within 7 days of the first sample if test inconclusive	3-7 days of birth	Adjusted according to gestational age, or birth weight if gestational age not known.	TC: 21-hydroylase deficiency CAH and other forms RS: Second and third tier testing, including physical examination, blood tests, LC-MS/MS, ultrasound, genetic analysis. Paediatricians report FNs to	TP: 133 (Salt-wasting CAH: 114 (73 M, 41 F); Simple virilising CAH: 14 (10 M, 4 F); Non-classic CAH: 5 (4 M, 1 F); Other enzyme deficiency: 2) [d] FP: 327[d] FN: 0[d] TN: 2,235,452[d] PPV: 24.7%[d] NPV: 100%[d] Sensitivity: 100%[d] Specificity: 99.98%[d]

						Dutch Paediatric Surveillance System.	
Heather 2015 ***New Zealand*** ***Cohort study: Retrospective analysis of national programme***	NR (all newborns 2011-2013)	Fluoroimmunoassay of 17-OHP	Re-test positive samples (after diethyl extraction)	After 48 hours of life	17-OHP concentrations above 23 nmol/L for babies with a birth weight above 1500g, and 32 nmol/L for those less than 1500g	TC: "severe CAH" (not defined further). RS: Routine diagnostic work-up plus babies diagnosed with CAH in the neonatal period, sent to the NMSP by paediatricians.	TP: 4 FP: 364 FN: 0 TN: NR PPV: 1.08% NPV: 100% Sensitivity: 100% Specificity: 99.8%[e]
Second-tier tests – DBS LC-MS/MS[a]							
Boelen 2016 ***The Netherlands*** ***Cohort study: Retrospective analysis of national programme***	92 newborns positive at first-tier test	LC–MS/MS of 21-deoxycortisol	NR	NR	According to gestational age	TC: 21-hydroxylase deficiency RS: Routine diagnostic work-up, with 21-Hydroxylase deficiency confirmed by mutation analysis	TP: 8[f] FP: 0 FN: 0 TN: 84 PPV: 100% NPV: 100% Sensitivity: 100% Specificity: 100%
		LC-MS/MS of other biomarkers[g]			As above	As above	None were 100% specific [f]

Gaudl 2019 ***Germany*** ***Case control***	100 sampled so at least 50% were false positive at first-tier screening	LC-MS/MS of 17-OHP	NR	NR	NR	TC: unclear which CAH types RS: Unclear, possibly routine screening methodology in Germany	TP: 3 FP: 0 FN: 0 TN: 97 PPV: 100% [b] NPV: 100% [b] Sensitivity: 100% [b] Specificity: 100% [b]
Han 2019 ***USA*** ***Case control***	45 newborns (17 cases confirmed CAH, 28 controls, first-tier negative)	LC–MS/MS 17-OHP	NR	NR	> 39.1 ng/mL	TC: unclear which CAH types RS: Diagnostic testing results and clinical symptoms	Sensitivity: 100% Specificity: 96.4%
		LC-MS/MS ratio of 17-OHP/cortisol			Ratio: 1		Sensitivity: 88.2% Specificity: 75%
		LC-MS/MS combination of 17-OHP and the 17-OHP/cortisol ratio			17-OHP > 39.1 ng/mL Ratio of 1 for 17-OHP/cortisol		Sensitivity: 94.1% Specificity: 100%
Two-tier testing (combined data from first and second-tier tests)[a]							
Tang 2016 ***USA (California)*** ***Cohort study: Retrospective analysis of regional programme***	Newborns, 108,429 False positive ascertainment, 2013 newborns; false negative ascertainment, 2006-2013 newborns	Immunofluorescence (17-OHP) and for intermediate results, MS/MS (17-OHP, androstenedione and cortisol), DBS collected at 12-23 hours	NR	12-23 hours after birth	According to birth weight	TC: Unclear which CAH types RS: Referral to endocrine centre for diagnostic evaluation (no	TP: 42 FP: 108 FN: 2 TN: 106,277 Sensitivity: 95.45% Specificity: 99.90%

				further details provided), genetic confirmation. False negatives identified via state registry.	
	Newborns, 332,798 False positive ascertainment, 2013 newborns; false negative ascertainment, 2006-2013 newborns	Immunofluorescence (17-OHP) and MS/MS (17-OHP, androstenedione and cortisol), DBS collected at 24-48 hours	24-48 hours after birth	As above	TP: 145 FP: 467 FN: 10 TN: 332,176 Sensitivity: 93.58% Specificity: 99.86%
	[a] single-tier tests are mass-screening stand-alone tests that lead to a specialist referral; first-tier tests are also mass-screening tests, but are followed by one or more tests before specialist referral; second-tier tests come after first-tier tests, are suitable for smaller volumes of test, and usually lead to a specialist referral; two-tier testing studies used a first and second-tier test, but did not report the results for these separately, but rather for the screening programme as a whole. [b] calculated by reviewer [c] emergency cut-off: ≥1751g birthweight: 110ng/ml; ≤1750g birthweight: 150ng/ml [d] excluding 17 missing diagnoses and counting 2 other diagnoses as TP [e] specificity was reported in the article as 99.8%, which may refer to the first-tier test. The reviewers have calculated specificity as follows: 8 referrals under this protocol, but only 4 diagnoses = 4 TP, 4 FP. Therefore, there were 364 TN (372 total minus 8 positive test (TP+FP), minus 0 FN = 364). Specificity = TN/TN+FP = 364/ (364+4) = 98.9% [f] all values read off graph and calculated by reviewer. It was not possible to extract more detailed data for markers other than 21-deosycortisol. [g] 11-deoxycortisol; 11-deoxycorticosterone; 17-hydroxyprogesteron; Δ4-androstenedion; Corticosterone; Cortisone; Cortisol.				

Abbreviations: 17-OHP, 17α-hydroxyprogesterone; DBS, dried blood spot; FP, false positive, FN, false negative; ICU, intensive care unit; LC–MS/MS, liquid chromatography with tandem mass spectrometry; MS/MS, tandem mass spectrometry; NC, non-classic; NMSP, Newborn Metabolic Screening Programme; NPV, negative predictive value; NR, not reported; PPV, positive predictive value; RS, reference standard; SW, salt wasting; SV, simple virilising; TC, target condition; TP, true positive; TN, true negative

i. Studies reporting single-tier or first-tier tests

Fluoroimmunoassay is widely used around the world as single-tier or first-tier test, though the homogeneity of the technical details relating to test methodology were beyond the scope of this review. Generally, this test tends to have very good sensitivity and specificity, but poor PPV, generating a large number of false positives.

Three national or regional cohort studies assessed the performance of fluoroimmunoassay. One study did not report sample size, whilst in the other 2 studies, sample size ranged between 1,962,433(23) and 2,235,931(30) cases. Sensitivity varied somewhat, ranging from 95%(23) to 100%(26, 30), whilst specificity was consistent, ranging from 99.8% to 100%. PPV varied quite widely, ranging from 1.08%(26) to 24.7%(30), and NPV was 100% in all studies. This is because the number of true negatives was large (sample size over 1,000,000, or unclear where not reported(26)) compared to the number of false negatives (ranging from 0(26) to 3(30)). One case-control study(48) (n=100) also assessed fluoroimmunassay but reported outlier results for specificity (37.11% compared with over 99% in the other 3 studies(23, 26, 30, 49) that reported this metric). The reasons for this are unclear.

It is worth noting that Pearce *et al.(23)* (n=1,962,433, USA) introduced a routine retest for samples taken less than 24 hours after birth, and an emergency cut-off for immediate referral in 2010, to reduce false positives. The manufacturers also changed the antibody in the testing kit around the same time. The sensitivity, specificity, PPV and NPV from 2007–2010 were 95%, 100%, 2.9% and 100% respectively. For the period 2010–2014, the values were 98.5%, 100%, 5.8% and 100% respectively, indicating a decrease in false positives.

One case-control study(48) (n=100) assessed a novel high-throughput LC-MS/MS as a first-tier test. Sensitivity, specificity, PPV and NPV were 100%, 96%, 43% and 100% respectively. This is higher than the specificity reported for fluoroimmunonassay in the same study, but lower than the specificities reported from the large scale cohort studies(23, 26, 30). However, the PPV was improved, at 42.9%(48) compared to the highest value of 24.7%(30) for fluoroimmunoassay.

ii. Studies assessing second-tier tests

Three studies(46-48) (n range: 45(46) to 100(48)) assessed second-tier tests. Two were case-control studies(46, 48) whilst one was a retrospective cohort study(47) which recruited a sample of first-tier positive DBSs. The target biomarkers varied. Two studies reported 100% accuracy, one for LC–MS/MS of 21-deoxycortisol(47), and one for LC-MS/MS of 17-OHP(48). However, another study reported lower specificity for LC-M/MS of 17-OHP (96.4%(46)). This study identified a combination of 17-OHP and the 17-

OHP/cortisol ratio as another promising target biomarker with 100% specificity, but this reported poorer sensitivity, at 94.1%, indicating some patients with CAH would be missed.

iii. Studies reporting the results of two-tier screening

One study reported the results of two-tier screening, but did not report the results of each tier individually. The first-tier measured 17-OHP by fluoroimmunoassay and the second-tier measured 17-OHP, androstenedione and cortisol by tandem mass spectrometry (MS/MS). This study aimed to assess whether age at sample (12–23 hours compared with 24–48 hours after birth) affected diagnostic accuracy. It was complicated by the fact that true positives (TPs) and false negatives (FN) were reported for newborns between 2006–2013, whilst false positives (FPs) and true negatives (TNs) were reported for newborns in 2013 only. As such, PPV and NPV could not be calculated. For DBSs taken between 12 and 23 hours after birth, sensitivity and specificity were calculated to be 95.5% and 99.9% respectively. For DBSs taken between 24 and 48 hours after birth, sensitivity and specificity were similar at 93.5% and 99.9% respectively.

iv. Data by CAH type

No study reported the diagnostic accuracy for each type of 21-hydroxylase deficiency. Two studies of first-tier fluoroimmunoassays reported some data relating to CAH type but this was largely incomplete. One study(23) reported the proportion of patients with different types of CAH but not which type the FNs were. One study(30) reported the proportion of cases with different types of CAH; there were no FNs in this study.

v. Data by gestational age/birth weight

Stratification of studies by gestational age was pre-planned in the scope and protocol. However, this was not possible due to underreporting in the evidence base. Most studies did not report mean gestational age, but many used different cut-off values by birth weight and/or gestational age. The details of these different cut-offs have not been extracted in this rapid review due to time constraints. Tang *et al*.(45) reported specificity and NPV for different weights of screened newborns (<1,000, 1,000–1,499, 1,500–2,499, >2,499 g), and for samples taken at 12–23 and 24–48 hours. The first-tier measured 17-OHP by immunofluorescence and the second-tier measured 17-OHP, androstenedione and cortisol by MS/MS. They concluded that false positive rates were higher in the lower birth weight groups (and this observation was true for both first-tier and second-tier tests), regardless of timing of sample collection. For those with low birth weight, the false positive rate was lower in the early collection group. Only for those with birthweight greater than 2,499g was false positive rate higher in the early collection group.

Deprioritised studies – narrative summary only

A table summarising these deprioritised studies is provided in Appendix 5. One large study(20) analysed the screening programme in 17 states in the USA (total n= 1,564,756). However, this was deprioritised as it assumed there were no false negatives. All states used a single-tier test, though some mandated a second sample, and most utilised a later or second sample for newborns in intensive care units to reduce false positives. The study only recruited normal birth weight infants. The authors noted that 17-OHP cut-off points varied widely. Because it was assumed there were no false negatives, sensitivity and NPV were 100% in all states. Specificity varied from 99.19% to 99.99% and PPV from 1.19 to 50%. Two of the 3 states mandating a second sample had the 2 highest PPVs, at 20% and 50%. The authors concluded that newborn screening protocols in the USA would benefit from standardisation to improve the PPV.

Four national or regional cohort studies studies (from New Zealand(50), Canada(3), USA(21) and Romania(51)) did not report sufficient diagnostic data to be included in this evidence summary. Three of these studies(3, 21, 51) assessed a first-tier or single-tier fluoroimmunoassay of 17-OHP. The PPVs were low at 0.61%(51) (PPV calculated by reviewer) and 0.10%(3) (PPV calculated by reviewer); one did not report PPV(21) and reported a sensitivity of 94%. Three(3, 50, 51) of the studies also reported results for a second-tier test. In one study(3), LC-MS/MS measured 21-deoxycortisol and 17-OHP/cortisol and PPV was calculated by the reviewer as 14.05%. In another study(51), LC-MS/MS measured several biomarkers and was able to identify one case of CAH out of 163 first-tier positive tests, though there was no reference standard in this study, so the accuracy is unknown. A third study(50) reported PPV for 2 second-tier tests, LC-MS/MS (11.11%) and a fluoroimmunoassay (1.71%), both of 17-OHP.

One study(21) included a number of USA states, but for a different time period (2003–2011) than Speiser *et al(20)*. (2017). This study sub-grouped states in to one-screen or two-screen states (both using fluoroimmunoassays), after an observation that the first screen tended to identify SW cases, whilst the second screen (on a new sample) identified SV and NC cases. One-screen states had a sensitivity of 93.9% (calculated by reviewer) and two-screen states had a sensitivity of 98.6% (calculated by reviewer). The authors concluded that SW was usually identified by the first screen, and that the second screen tended to identify SV and NC cases.

Summary of Findings Relevant to Criterion 4 & 5: Criteria not met[2]

Quantity: this review identified 7 studies reporting the diagnostic accuracy of screening tests using dried blood spots to detect congenital adrenal hyperplasia. Five studies(23, 26, 30, 45, 47) were retrospective analyses of national or regional newborn screening programmes; 3 of these recruited large unselected samples (largest reported n=2,235,931(30); lowest reported n=439,227(45)), one(26) did not report the total sample size, but recruited all newborns in New Zealand from 2011-2013, and one study(47) recruited a small sample of first-tier positive patients (n=100). The remaining 2 studies were case control studies which drew samples from national screening programmes (n=100(48) and n=45(46)).

Quality: the quality of the studies was generally poor. Studies were often judged at unclear or high risk of bias for patient selection due to inappropriate exclusions or a lack of clarity. No study reported blinding of index test or reference standard. Since the index test often dictated which reference standard a patient would receive, flow and timing were also at risk of bias. It was beyond the scope of this rapid review to contact authors of international studies for clarifications regarding unclear aspects of methods. Further information regarding aspects of all studies may have provided a greater degree of confidence about the study designs in some cases.

Applicability: no studies were from the UK. Only studies from countries with comparable population characteristics and healthcare systems were selected for inclusion. Where reported, the timing of the test was either earlier than in the UK (2 studies less than 48 hours after birth), vague (one study greater than 48 hours) or similar (one study between 3 and 7 days). Since DBSs are taken between 5 and 8 days after birth in the UK, this limits the applicability of the evidence base, though the extent to which diagnostic accuracy is altered by age is unclear. Not all studies aimed to identify only CAH due to 21-hydroxylase deficiency, and data could not be extracted relating to just this target condition. Where it was reported, it would appear that most missed cases were SV or NC CAH. It is unclear which combination of first and second-tier tests might be optimal, and whether diagnostic accuracy of

[2] **Met** -for example, this should be applied in circumstances in which there is a sufficient volume of evidence of sufficient quality to judge an outcome or effect which is unlikely to be changed by further research or systematic review.
Not Met - for example, this should be applied in circumstances where there is insufficient evidence to clearly judge an outcome or effect or where there is sufficient evidence of poor performance.
Uncertain -for example, this should be applied in circumstances in which the constraints of an evidence summary prevent a reliable answer to the question. An example of this may be when the need for a systematic review and meta-analysis is identified by the rapid review.

second-tier tests varies according to the patient spectrum identified by the first-tier test.

Consistency: generally, first-tier tests used fluoroimmunoassays to measure 17-OHP, and second-tier tests used LC-MS/MS to measure various target biomarkers, including 17-OHP. Cut-offs used to define a positive test were often defined differentially according to parameters such as gestational age, weight and/or day of sample to allow for differences in biomarker levels according to these parameters. However, there was not a consensus about which parameters to vary cut-offs by, or which set of cut-offs were most appropriate. Fluoroimmunoassays tended to have very good sensitivity and specificity, but poor PPV, generating a large number of false positives. LC-MS/MS second-tier tests reduced the number of false positives being referred to specialist services, with sometimes perfect diagnostic accuracy, though there was heterogeneity in estimates even when the same marker was measured. There was heterogeneity (and lack of clarity) in terms of who should be retested, when this should occur, whether a new sample should be drawn, and how often repeat tests should be done.

Conclusions: Seven studies were identified, and one had fairly high generalisability to the UK. Several countries use fluoroimmunoassay of 17-OHP to screen for CAH, and some use an additional second-tier LC-MS/MS test to reduce the number of false positives referred to specialist services. Whilst there was not a consensus about which parameters to vary cut-offs by, or which set of cut-offs were most appropriate, the included studies showed that it is possible to describe cut-off values that, within their own specific context, can deliver adequate test performance. Within the context of screening in the UK, the actual cut-off selected could be population and method dependent and further study could clarify a number of variables including the optimum cut-off values to be used by fluoroimmunoassay and LC-MS/MS with respect to weight, gestational age, and age when sampled, and whether cut-offs defined with reference to population biomarker percentiles could be an appropriate method in the UK. There is a fair amount of evidence relating to fluoroimmunoassay from large scale studies, however all studies had high or unclear risk of bias. The evidence relating to LC-MS/MS is limited in terms of study design, sample size and quality, but shows promise for reducing the number of false positives (but not false negatives) identified by fluoroimmunoassay if used as a second-tier test. Further research could also clarify the impact of testing DBSs taken during the 5-8 day window used in the UK; the best biomarker to be measured by LC-MS/MS; whether LC-MS/MS could be used as first tier test or whether it is best reserved as a second-tier test; and the impact of second samples or repeat testing. Further studies could have superior methodological quality

and reporting clarity, especially with respect to the reference standards used and blinding of index test and reference standards, though such studies may have a small sample size. Studies could also ascertain what combination of first and second-tier tests are optimal. Improvements to the evidence base would help to clarify whether screening can be recommended. However, at present based on the findings of this evidence summary, the test criteria are judged as “not met”.

Review summary

Conclusions and implications for policy

Based on the overall synthesis of the evidence published since the previous UK NSC review in 2015, further evidence is required before a change to the current recommendation can be made. Therefore, newborn screening for CAH is still not recommended.

Three key questions were considered in the review: whether there has been a significant development in the evidence base relating to 1) the incidence of CAH in the UK population, 2) the median age at presentation of CAH, and 3) the accuracy of available screening tests using dried blood spots to detect CAH.

Despite a number of published studies that met the inclusion criteria, there are limitations to the evidence. There were no newly identified studies of UK incidence, and the evidence base relies on the 2 papers reporting data obtained through an active surveillance programme already identified in the previous review. It is unclear whether all cases of CAH have been captured by this surveillance. In addition, the study duration was 2 years, which is the lowest end of the range for incidence studies compared with the included international studies. Nevertheless, the reported incidence rate for Great Britain of approximately 1 in 18,000 children is comparable with those reported in the newly identified international studies as it was in the previous review. It is also noted that for incidence studies of longer duration the instruments used to analyse dried blood spots (DBS), or clinical practice in diagnosing and managing CAH, are more likely to change over time, therefore cases identified in the earlier years may not be comparable to those identified later. The evidence base to determine the median age of presentation of CAH is very limited. No studies specifically designed to answer this question were identified in the review, and data therefore is drawn from studies that have described median age at diagnosis as one of a number of demographic characteristics in studies designed to meet other unrelated aims. The limited amount of evidence of time to presentation in newborns suggested that some children present with salt-wasting crises before 15 days, whilst the results from the UK newborn screening programme would potentially be available at around 17 days. These children may therefore not benefit from inclusion of CAH into the newborn screening programme. Conversely, screening might be of benefit to those babies at risk of presenting later with severe and life-threatening symptoms or of dying undiagnosed. The current evidence on median age of presentation in newborns is limited in terms of volume and quality and it is therefore difficult to offer definitive conclusions relating to the impact screening would have on clinical outcomes.

Similarly, the quality of the studies assessing the performance of screening tests was generally poor, largely due to most studies not being designed with the purpose of assessing diagnostic accuracy of an index test, but to assess the diagnostic performance of a screening programme already in operation. There were also problems with or a lack of clarity about the reference standard used by each study. The evidence included was in relation to fluoroimmunoassays and LC-MS/MS as first-tier and second-tier screening tests respectively. Both types of tests can be considered simple and safe since they require a DBS, and can be used for mass screening, as is the case in other countries. However, LC-MS/MS is somewhat more technical and, due to the constraints of processing large volumes of tests, it is more suited to a second-tier setting. The fluoroimmunoassays identified as first-tier tests generally showed a poor PPV, and the review found that they may generate a small number of false negatives. Across studies, identification of false negatives was problematic, with uncertainty around whether all false negatives were being recorded. False negatives may be clinically significant in particular for those with classic salt-wasting CAH. A false negative may also hinder subsequent clinical diagnosis when presenting with symptoms. A limited number of studies of moderate size outlined the use of LC-MS/MS as a second-tier test. Whilst some evidence suggest that LC-MS/MS can increase PPV, this can lead to an increase in false negatives in some instances. Moreover, the diagnostic accuracy of LC-MS/MS has not been exhaustively characterised, and it appears to vary according to methodology and target marker used.

Limitations

The review includes limited evidence from the UK, and it is characterised by a lack of studies designed specifically to answer each of the 3 review questions. The searches were limited to English language papers. However, this is unlikely to have led to the exclusion of any pertinent or pivotal studies due to the necessary focus on the applicability of screening to the UK and countries that are comparable to the UK. The review employed rapid reviewing methods and therefore sifting was conducted by one of 2 reviewers. However, 20% of all retrieved citations were independently sifted by a second reviewer and discrepancies were discussed and resolved ensuring consistency in the sifting process. This accepted pragmatic strategy should have minimised any risk of errors. Finally, only the author of the GB studies was contacted for further information regarding their methods. It was beyond the scope of this rapid review to contact authors of all included international studies in order to request clarifications of their methods and reported results. Further information regarding aspects of all studies may have provided a greater degree of confidence about the study designs in some cases.

Appendix 1 — Search strategy

Electronic databases

MEDLINE (including Epub Ahead of Print & In-Process), Embase and the Cochrane Library were searched in July 2020.

The database searches were supplemented by a search for surveillance reports in the British Paediatric Surveillance Unit (BPSU) and the laboratory surveillance scheme comprising diagnostic and genetic laboratories performing testing for CAH.

Table 9. Summary of electronic database searches and dates for Q1&2

Database	Platform	Searched on date	Date range of search
MEDLINE, MEDLINE In-Process, MEDLINE Daily, Epub Ahead of Print	Ovid SP	14/07/2020	2008 to 2020 July 10
Embase	Ovid SP	14/07/2020	2008 to 2020 July 10
The Cochrane Library, including: - Cochrane Database of Systematic Reviews (CDSR) - Cochrane Central Register of Controlled Trials (CENTRAL)	Wiley Online	14/07/2020	2008-2020

Table 10. Summary of electronic database searches and dates for Q3

Database	Platform	Searched on date	Date range of search
MEDLINE, MEDLINE In-Process, MEDLINE Daily, Epub Ahead of Print	Ovid SP	15/07/2020	2015 to 2020 July 14
Embase	Ovid SP	15/07/2020	2015 to 2020 July 13
The Cochrane Library, including: - Cochrane Database of Systematic Reviews (CDSR) - Cochrane Central Register of Controlled Trials (CENTRAL)	Wiley Online	15/07/2020	2015-2020

Search Terms

The search strategy consists of thesaurus and free text terms relating to congenital adrenal hyperplasia and epidemiology for questions 1 and 2 and for questions 3 terms for congenital adrenal hyperplasia, screening and relevant tests and test accuracy.

All searches were limited to English Language and for question 1: to publications between 2008 and July 2020 for UK studies, and 2015 and July 2020 for international studies; for question 2 to publications between 2008 and July 2020; for question 3 between 2015 and July 2020. Editorials,

comments and letters were excluded from the searches and all searches were limited to human studies.

Search terms for MEDLINE, MEDLINE In-Process, MEDLINE Daily, Epub Ahead of Print and Embase for Q1 and 2 and Q3 are shown in **Error! Reference source not found.**11-12.

Q1 and 2

Table 11. Search strategy for MEDLINE, MEDLINE In-Process, MEDLINE Daily, Epub Ahead of Print and Embase

Term Group	#	Search terms	Results
MEDLINE, MEDLINE In-Process, MEDLINE Daily, Epub Ahead of Print			
Disease Area	1	Adrenal Hyperplasia, Congenital/	6658
	2	congenital adrenal hyperplasia$.ab,ti.	4757
	3	STEROID 21-HYDROXYLASE/	1749
	4	17-alpha-Hydroxyprogesterone/	1996
	5	("17" and hydroxyprogesterone).ab,ti.	3741
	6	("21" and hydroxylase).ab,ti.	4752
	7	cah.ab,ti.	4232
	8	21-OH.af.	275
	9	classic salt-wasting.ab,ti.	34
	10	classic simple virilising.ab,ti.	1
	11	non-classic.ab,ti	577
	12	17-OHP.af.	771
	13	11B-hydroxylase.af.	15
	14	17a-hydroxylase.af.	22
	15	3B-hydroxysteroid dehydrogenase.af.	21
	16	Steroid Acute Regulatory protein.ab,ti.	48
	17	OR/1-16	16559
Epidemiology	18	exp Epidemiologic Methods/	6201280
	19	exp Epidemiologic Studies/	2506013
	20	exp Sentinel Surveillance/	6310
	21	exp Seroepidemiologic Studies/	23351

	22	exp Cohort Studies/	2009171
	23	exp Cross-Sectional Studies/	331674
	24	exp Longitudinal Studies/	135652
	25	exp Follow-Up Studies/	643482
	26	exp Prospective Studies/	542781
	27	or/18-26	6201280
	28	exp Incidence/	261692
	29	exp Prevalence/	290967
	30	Epidemiology/	12360
	31	incidence.ab,ti.	746373
	32	prevalence.ab,ti.	622140
	33	"Age Distribution"/	662211
	34	"Sex Distribution"/	55179
	35	((age or sex) adj3 distribution$).ab,ti.	19761
	36	or/28-35	1536860
	37	27 or 36	6766612
Disease and Epidemiology terms combined	38	17 and 37	3667
Limits	39	limit 38 to english language	3235
	40	limit 39 to yr="2008 -Current"	1620
	41	limit 40 to humans	1446
	42	(comment or editorial or letter).pt.	1863621
	43	41 NOT 42	1421
Embase			
Disease Area	1	congenital adrenal hyperplasia/	8389
	2	congenital adrenal hyperplasia$.ab,ti.	6641
	3	steroid 21 monooxygenase/	3086
	4	hydroxyprogesterone/	6673
	5	("17" and hydroxyprogesterone).ab,ti.	4067
	6	("21" and hydroxylase).ab,ti.	6384
	7	cah.ab,ti.	5540
	8	21-OH.af.	383
	9	classic salt-wasting.ab,ti.	53
	10	classic simple virilising.ab,ti.	1
	11	non-classic.ab,ti.	969
	12	17-OHP.af.	1248

	13	11B-hydroxylase.af.	40
	14	17a-hydroxylase.af.	63
	15	3B-hydroxysteroid dehydrogenase.af.	69
	16	Steroid Acute Regulatory protein.ab,ti.	60
	17	or/1-16	22484
Epidemiology terms	18	exp epidemiology/	3547726
	19	sentinel surveillance/	2465
	20	seroepidemiology/	4261
	21	cohort analysis/	637115
	22	cohort analysis/	637115
	23	exp longitudinal study/	147889
	24	follow up/	1610731
	25	prospective study/	642650
	26	or/18-25	5404977
	27	exp incidence/	504770
	28	exp prevalence/	776853
	29	epidemiology.fs.	1042989
	30	incidence.ab,ti.	1092848
	31	prevalence.ab,ti.	912633
	32	age distribution/	144360
	33	sex ratio/	71743
	34	((age or sex) adj3 distribution$).ab,ti.	29623
	35	or/27-34	2921270
	36	26 or 35	6467348
Disease and epidemiology terms combined	37	17 and 36	4405
Limits	38	limit 37 to english language	4144
	39	limit 38 to yr="2008 -Current"	3043
	40	limit 39 to human	2819
	41	(editorial or letter).pt.	1827375
	42	40 not 41	2769
	43	limit 42 to embase	1549

Q3

Table 12. Search strategy for MEDLINE, MEDLINE In-Process, MEDLINE Daily, Epub Ahead of Print and Embase

Term Group	#	Search terms	Results
MEDLINE, MEDLINE In-Process, MEDLINE Daily, Epub Ahead of Print			
Disease Area	1	Adrenal Hyperplasia, Congenital/	6658
	2	congenital adrenal hyperplasia$.ab,ti.	4757
	3	STEROID 21-HYDROXYLASE/	1749
	4	17-alpha-Hydroxyprogesterone/	1996
	5	("17" and hydroxyprogesterone).ab,ti.	3741
	6	("21" and hydroxylase).ab,ti.	4752
	7	cah.ab,ti.	4232
	8	21-OH.af.	275
	9	classic salt-wasting.ab,ti.	34
	10	classic simple virilising.ab,ti.	1
	11	non-classic.ab,ti	577
	12	17-OHP.af.	771
	13	11B-hydroxylase.af.	15
	14	17a-hydroxylase.af.	22
	15	3B-hydroxysteroid dehydrogenase.af.	21
	16	Steroid Acute Regulatory protein.ab,ti.	48
	17	OR/1-16	16559
Tests	18	exp Mass Screening/	127828
	19	(detect$ or test or tests or testing or screen$).ti,ab.	4733782
	20	exp Immunoassay/	487730
	21	immunoassay.ab,ti.	53909
	22	automated time- resolved dissociation-enhanced lanthanide fluoroimmunoassay.ab,ti.	0
	23	DELFIA.ab,ti.	394
	24	fluoroimmunoassay.ab,ti.	1048
	25	Fluoroimmunoassay/	1739
	26	Radioimmunoassay/	64579
	27	radioimmunoassay.ab,ti.	49319

	28	Enzyme-Linked Immunosorbent Assay/	147920
	29	enzyme-linked immunosorbent assay.ab,ti.	80419
	30	ELISA.ab,ti.	167468
	31	assay$.ab,ti.	1068776
	32	Tandem Mass Spectrometry/	47990
	33	liquid chromatography tandem mass spectrometry.ab,ti	22792
	34	LC-MSMS.ab,ti.	173
	35	Chromatography, High Pressure Liquid/	184861
	36	hplc.ab,ti.	127761
	37	high performance liquid chromatography.ab,ti.	91095
	38	or/18-37	5843098
Disease and test terms combined	39	17 and 38	5565
Test accuracy terms	40	(sensitiv$ or specific$).ti,ab.	4062842
	41	"predictive value".ti,ab.	91469
	42	"false positiv$".ti,ab.	59398
	43	"false negativ$".ti,ab.	33376
	44	accuracy.ti,ab.	410768
	45	or/40-44	4394374
	46	(sensitiv: or diagnos:).mp. or di.fs.	6229389
	47	45 or 46	8697900
Disease and test and test accuracy terms combined	48	39 and 47	3134
Limits	49	limit 48 to english language	2761
	50	limit 49 to yr="2015 -Current"	548
	51	limit 50 to humans	374
	52	(comment or editorial or letter).pt.	1865895
	53	51 not 52	373
Embase			
Disease Area	1	congenital adrenal hyperplasia/	8389
	2	congenital adrenal hyperplasia$.ab,ti.	6641
	3	steroid 21 monooxygenase/	3086

	4	hydroxyprogesterone/	6673
	5	("17" and hydroxyprogesterone).ab,ti.	4067
	6	("21" and hydroxylase).ab,ti.	6384
	7	cah.ab,ti.	5540
	8	21-OH.af.	383
	9	classic salt-wasting.ab,ti.	53
	10	classic simple virilising.ab,ti.	1
	11	non-classic.ab,ti.	969
	12	17-OHP.af.	1248
	13	11B-hydroxylase.af.	40
	14	17a-hydroxylase.af.	63
	15	3B-hydroxysteroid dehydrogenase.af.	69
	16	Steroid Acute Regulatory protein.ab,ti.	60
	17	or/1-16	22484
Test terms	18	exp mass screening/	249618
	19	(detect$ or test or tests or testing or screen$).ti,ab.	6461098
	20	immunoassay/	69109
	21	immunoassay.ab,ti	73145
	22	automated time-resolved dissociation-enhanced lanthanide fluoroimmunoassay.ti,ab.	0
	23	DELFIA.ab,ti.	646
	24	fluoroimmunoassay.ab,ti.	1214
	25	fluoroimmunoassay/	1717
	26	exp immunoassay/	580636
	27	radioimmunoassay/	78383
	28	radioimmunoassay.ab,ti.	55777
	29	exp enzyme linked immunosorbent assay/	377099
	30	enzyme-linked immunosorbent assay.ab,ti.	93771
	31	ELISA.ab,ti.	273064
	32	assay$.ab,ti.	1434238
	33	exp tandem mass spectrometry/	72539
	34	liquid chromatography tandem mass spectrometry.ab,ti.	28246
	35	LC-MSMS.ab,ti.	490

	36	exp high performance liquid chromatography/	328283
	37	hplc.ab,ti.	182035
	38	high performance liquid chromatography.ab,ti.	110557
	39	or/18-38	7818029
Disease and test terms combined	40	17 and 39	8900
Test accuracy terms	41	(sensitiv$ or specific$).ti,ab.	5188867
	42	"predictive value".ti,ab.	143559
	43	"false positiv$".ti,ab.	81809
	44	"false negativ$".ti,ab.	47946
	45	accuracy.ti,ab.	537015
	46	or/41-45	5625756
	47	di.fs. or predict:.tw. or specificity.tw.	5501003
	48	46 or 47	9650882
Disease, test and test accuracy terms combined	49	40 and 48	4194
Limits	50	limit 49 to english language	3827
	51	limit 50 to yr="2015 -Current"	1108
	52	limit 51 to human	990
	53	(editorial or letter).pt.	1827375
	54	52 not 53	978
	55	Limit 54 to embase	575

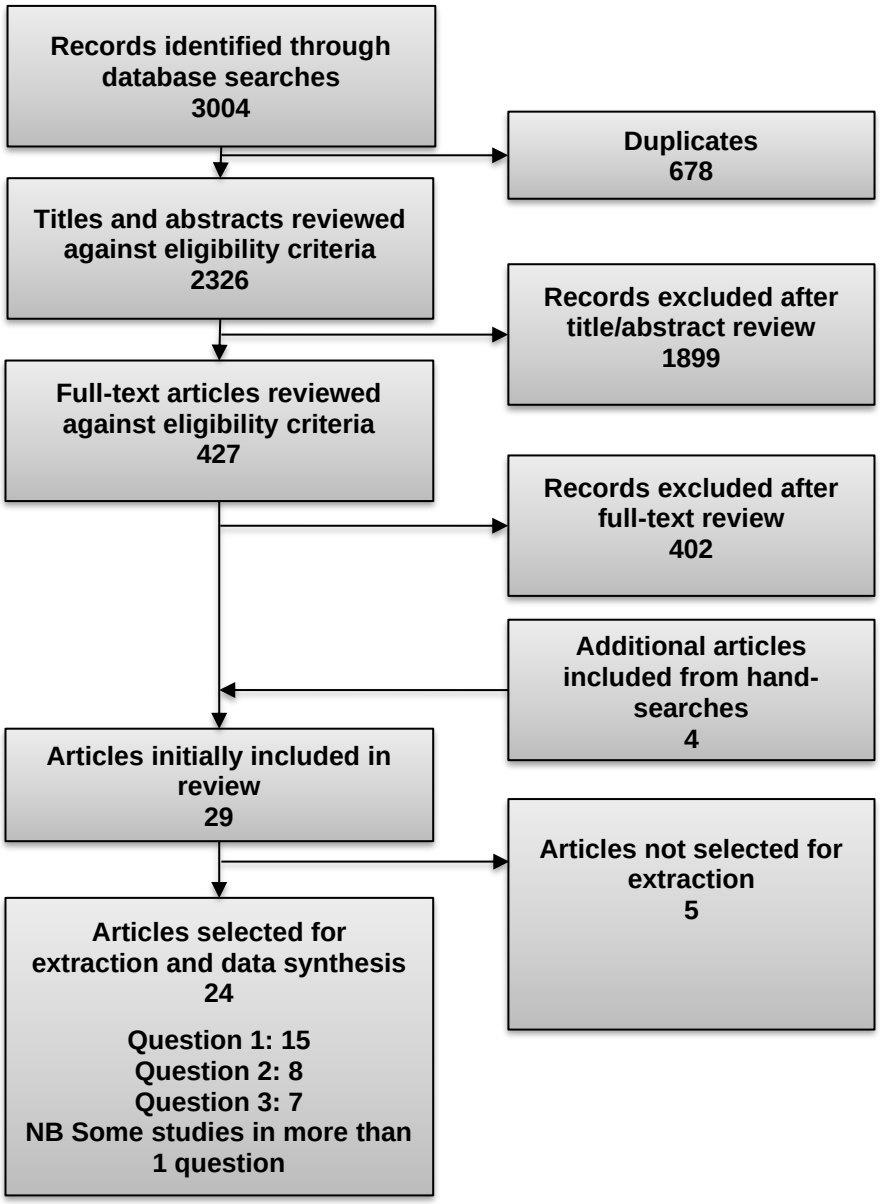
Results were imported into EndNote and de-duplicated.

Appendix 2 — Included and excluded studies

PRISMA flowchart

Figure 1 summarises the volume of publications included and excluded at each stage of the review. Twenty-four publications were ultimately judged to be relevant to one or more review questions and were considered for extraction. Publications that were included or excluded after the review of full-text articles are detailed below.

Figure 1. Summary of publications included and excluded at each stage of the review



Publications included after review of full-text articles

The 29 publications included after review of full-texts are summarised in Table **13. Summary o** 13 below. Twenty-four studies were prioritised for extraction and data synthesis. Five studies were ultimately deprioritised due to one of the reasons described in the Methods section.

Table 13. Summary of publications included after review of full-text articles, and the question(s) each publication was identified as being relevant to

Study	Question	CAH subtypes (Q1/2)/screening test (Q3)
David 2018	Q1	All classic
David 2019	Q1	All classic
Eshragh 2020	Q1	All classic, classic SW and SV
Fox 2020	Q1/Q2	SW, SV and non-classic/All classic
Heather 2015	Q1/Q3	All classic, classic SV/Immunoassay
Held 2015	Q1	All classic, classic SW and SV, non-classic
Iniguez 2019	Q1	All classic
Khalid 2012	Q1/Q2	All CAH/Not reported by type
Knowles 2013	Q1/Q2	All CAH/All types
Lai 2020	Q1	All classic, classic SW and SV
Pearce 2017	Q1	All CAH/Immunoassay
Pearce 2016	Q1/Q3	All classic, classic SW and SV, non-classic
Speiser 2020	Q1	All CAH
Van der Linde 2019	Q1/Q3	SW + SV, SV/Immunoassay
Zetterstrom 2020	Q1	All CAH, all classic CAH, non-classic
Bomberg 2015/Sarafoglou 2014	Q2	Classic SW and SV
Halper 2019	Q2	Not reported by type
Hsieh & White 2011	Q2	All classic
Maccabee-Ryaboy 2016	Q2	Classic SW and SV
Pijnenburg-Kleizen 2019	Q2	Classic SW and SV
Boelen 2016	Q3	UPLC-MS/MS
Gaudl 2019	Q3	Immunoassay + LC-MS/MS
Han 2019	Q3	LC-MS/MS
Tang 2016	Q3	Immunoassay + LC-MS/MS

Table 14. Summary of publications deprioritised after review of full-text articles, and the question(s) each publication was identified as being relevant to

Study	Question	Screening test (Q3)
De Hora 2020	Q3	Immunoassay + LC-MS/MS
Fox 2020	Q3	LC-MS/MS
Held 2015	Q3	Immunoassay
Monostori 2015	Q3	Immunoassay + LC-MS/MS
Speiser 2020	Q3	Immunoassay

Publications not selected for extraction and data synthesis are clearly detailed in Table **13. Summary o**15-18 below.

Publications excluded after review of full-text articles

Of the 427 publications included after the review of titles and abstracts, 402 were ultimately judged not to be relevant to this review. These publications, along with reasons for exclusion, are listed in tables 15-18.

Publications excluded after review of full-text articles

Table 15. Publications excluded from Q1

Reference	Reason for exclusion
	Studies excluded due to date of publication pre-2015
Gleeson, H. K., et al. (2008). "Two-year pilot study of newborn screening for congenital adrenal hyperplasia in New South Wales compared with nationwide case surveillance in Australia." Journal of Paediatrics & Child Health 44(10): 554-559.	Date of publication pre-2015
Cavarzere, P., et al. (2009). "Transient hyper-17-hydroxyprogesteronemia: a clinical subgroup of patients diagnosed at neonatal screening for congenital adrenal hyperplasia." European Journal of Endocrinology 161(2): 285-292.	Date of publication pre-2015
Kasper, D. C., et al. (2010). "The national Austrian newborn screening program - eight years experience with mass spectrometry. past, present, and future goals." Wiener Klinische Wochenschrift 122(21-22): 607-613.	Date of publication pre-2015
Liivak, K., et al. (2008). "Incidence of classical 21-hydroxylase deficiency and distribution of CYP21A2 mutations in Estonia." Hormone Research 69(4): 227-232.	Date of publication pre-2015
Nagasaki, K., et al. (2011). "The occurrence of neonatal acute respiratory disorders in 21-hydroxylase deficiency." Endocrine Journal 58(7): 603-606.	Date of publication pre-2015
Nermoen, I., et al. (2010). "Subjective health status in men and women with congenital adrenal hyperplasia: a population-based survey in Norway." European Journal of Endocrinology 163(3): 453-459.	Date of publication pre-2015
Schweizer, R., et al. (2010). "Prevalence and incidence of endocrine disorders in children: results of a survey in Baden-Wuerttemberg and Bavaria (EndoPrIn BB) 2000-2001." Klinische Padiatrie 222(2): 67-72.	Date of publication pre-2015
Shetty, V. B., et al. (2012). "Ethnic and gender differences in rates of congenital adrenal hyperplasia in Western Australia over a 21 year period." Journal of Paediatrics & Child Health 48(11): 1029-1032.	Date of publication pre-2015
	Studies excluded due to setting being a non-comparable country
Abdullah, M. A., et al. (2012). "Disorders of sex development among Sudanese children: 5-year experience of a pediatric endocrinology clinic." Journal of Pediatric Endocrinology & Metabolism 25(11-12): 1065-1072	Sudan

Abid, F., et al. (2008). "CYP21A2 gene mutation analysis in Moroccan patients with classic form of 21-hydroxylase deficiency: high regional prevalence of p.Q318X mutation and identification of a novel p.L353R mutation." Clinical Chemistry & Laboratory Medicine 46(12): 1707-1713.	Tunisia
Alfadhel, M., et al. (2017). "Expanded Newborn Screening Program in Saudi Arabia: Incidence of screened disorders." Journal of Paediatrics & Child Health 53(6): 585-591.	Saudi Arabia
Al Hosani, H., et al. (2013). "Expanding the comprehensive national neonatal screening programme in the United Arab Emirates from 1995 to 2011." Eastern Mediterranean Health Journal 20(1): 17-23.	United Arab Emirates
Alratrout, R., et al. (2017). "The frequency of inherited metabolic and endocrine disorders in the eastern and north-western Jawf provinces of Saudi Arabia: Four years data from the newborn screening department, ministry of health, Dammam." Current Pediatric Research 21(4): 665-673.	Saudi Arabia
Ameyaw, E., et al. (2019). "Incidence of disorders of sexual development in neonates in Ghana: prospective study." Archives of Disease in Childhood 104(7): 636-638.	Ghana
Barra, C. B., et al. (2012). "Neonatal screening for congenital adrenal hyperplasia." Revista Da Associacao Medica Brasileira 58(4): 459-464.	Brazil
Cantu-Reyna, C., et al. (2016). "Incidence of inborn errors of metabolism by expanded newborn screening in a Mexican hospital." Journal of Inborn Errors of Metabolism and Screening 4(no pagination).	Mexico
Castro, P. S., et al. (2019). "High frequency of non-classical congenital adrenal hyperplasia form among children with persistently elevated levels of 17-hydroxyprogesterone after newborn screening." Journal of Pediatric Endocrinology & Metabolism 32(5): 499-504.	Brazil
De Miranda, M. C., et al. (2020). "Adverse outcomes and economic burden of congenital adrenal hyperplasia late diagnosis in the newborn screening absence." Journal of the Endocrine Society 4(2).	Brazil
Gong, L. F., et al. (2019). "A pilot study on newborn screening for congenital adrenal hyperplasia in Beijing." Journal of Pediatric Endocrinology & Metabolism 32(3): 253-258.	China
Gruneiro-Papendieck, L., et al. (2008). "Neonatal screening for congenital adrenal hyperplasia: Experience and results in Argentina." Journal of Pediatric Endocrinology and Metabolism 21(1): 73-78.	Argentina

Hoehn, T., et al. (2013). "Establishment of the first newborn screening program in the People's Democratic Republic of Laos." Journal of Tropical Pediatrics 59(2): 95-99.	Laos
Jiang, X., et al. (2019). "The adjustment of 17-hydroxyprogesterone cut-off values for congenital adrenal hyperplasia neonatal screening by GSP according to gestational age and age at sampling." Journal of Pediatric Endocrinology & Metabolism 32(11): 1253-1258.	China
Kamath, S. S. (2015). "Newborn screening in India." Indian Pediatrics 52(5): 373-374.	India
Kashimada, K., Ishii, T., Nagasaki, K., Ono, M., Tajima, T., Yokota, I., & Hasegawa, Y. (2015). Clinical, biochemical, and genetic features of non-classical 21-hydroxylase deficiency in Japanese children. *Endocrine journal*, *62*(3), 277-282.	Japan
Kaur, G., et al. (2010). "Preliminary report on neonatal screening for congenital hypothyroidism, congenital adrenal hyperplasia and glucose-6-phosphate dehydrogenase deficiency: a Chandigarh experience." Indian Journal of Pediatrics 77(9): 969-973.	India
Kopacek, C., et al. (2017). "Neonatal screening for congenital adrenal hyperplasia in Southern Brazil: a population based study with 108,409 infants." BMC Pediatrics 17(1): 22.	Brazil
Kopacek, C., et al. (2019). "Clinical and molecular profile of newborns with confirmed or suspicious congenital adrenal hyperplasia detected after a public screening program implementation." Jornal de Pediatria 95(3): 282-290.	Brazil
Mazen, I., et al. (2010). "Screening of genital anomalies in newborns and infants in two Egyptian governorates." Hormone Research in Paediatrics 73(6): 438-442.	Egypt
Pode-Shakked, N., Blau, A., Pode-Shakked, B., Tiosano, D., Weintrob, N., Eyal, O., ... & Almashanu, S. (2019). Combined Gestational Age-and Birth Weight–Adjusted Cutoffs for Newborn Screening of Congenital Adrenal Hyperplasia. *The Journal of Clinical Endocrinology & Metabolism*, *104*(8), 3172-3180.	Israel
Rodrigues, L. P., et al. (2019). "Heel prick test: maternal-fetal conditions that may have an effect on the test results in newborns admitted to the intensive care unit." Revista Brasileira de Terapia Intensiva 31(2): 186-192.	Brazil

Sharma, P., et al. (2018). "Prevalence of inborn errors of metabolism in neonates." Journal of Clinical and Diagnostic Research 12(5): BC07-BC13.	India
Somboonnithiphol, K., et al. (2011). "Newborn screening for congenital adrenal hyperplasia in Srinagarind Hospital, Khon Kaen University, Thailand." Asian Biomedicine 5(6): 855-859.	Thailand
Tsuji, A., Konishi, K., Hasegawa, S., Anazawa, A., Onishi, T., Ono, M., ... & Kashimada, K. (2015). Newborn screening for congenital adrenal hyperplasia in Tokyo, Japan from 1989 to 2013: a retrospective population-based study. *BMC pediatrics*, *15*(1), 209.	Japan
Verma, J., et al. (2020). "Newborn Screening for Congenital Hypothyroidism, Congenital Adrenal Hyperplasia, and Glucose-6-Phosphate Dehydrogenase Deficiency for Improving Health Care in India." Journal of Pediatric Intensive Care 9(1): 40-44.	India
Yeung, M. C. W., et al. (2020). "Clinical utility of second-tier testing in newborn screening for congenital adrenal hyperplasia: The Hong Kong experience." Hong Kong Journal of Paediatrics 25(1): 3-7.	Hong Kong
	Studies excluded for other reasons
Odenwald, B., et al. (2015). "Classic Congenital Adrenal Hyperplasia due to 21-Hydroxylase-Deficiency: 13 Years of Neonatal Screening and Follow-up in Bavaria." Klinische Padiatrie 227(5): 278-283.	Non-English language (in German)
Lund, A., et al. (2020). "Danish expanded newborn screening is a successful preventive public health programme." Danish Medical Journal 67(1).	No incidence data
Xu, Z., et al. (2013). "Comprehensive mutation analysis of the CYP21A2 gene: an efficient multistep approach to the molecular diagnosis of congenital adrenal hyperplasia." Journal of Molecular Diagnostics 15(6): 745-753.	Not newborns
Zlotogora, J. (2019). "Autosomal recessive diseases among the Israeli Arabs." Human Genetics 138(10): 1117-1122.	No incidence data

Table 16. Publications excluded from Q2 due to country non-comparable to UK

Reference	Reason for exclusion
Abdullah, M. A., et al. (2012). "Disorders of sex development among Sudanese children: 5-year experience of a pediatric endocrinology clinic." Journal of Pediatric Endocrinology & Metabolism 25(11-12): 1065-1072.	Sudan
Abid, F., et al. (2008). "CYP21A2 gene mutation analysis in Moroccan patients with classic form of 21-hydroxylase deficiency: high regional prevalence of p.Q318X mutation and identification of a novel p.L353R mutation." Clinical Chemistry & Laboratory Medicine 46(12): 1707-1713.	Tunisia
Al Shaikh, A., et al. (2019). "Clinical patterns and linear growth in children with congenital adrenal hyperplasia, an 11-year experience." Indian Journal of Endocrinology and Metabolism 23(3): 298-306.	Saudi Arabia
Al-Jurayyan, N. A. (2011). "Ambiguous genitalia: two decades of experience." Annals of Saudi Medicine 31(3): 284-288.	Saudi Arabia
Al-Jurayyan, N. A. M. and A. A. Al-Hakami (2018). "Psychological impact of congenital adrenal hyperplasia on adolescent and young girls in Saudi Arabia." Biomedical Research (India) 29(20): 3742-3746.	Saudi Arabia
Al-Mulhim, A. N. and H. M. Kamal (2010). "Ambiguous genitalia in neonates: a 4-year prospective study in a localized area." Eastern Mediterranean Health Journal 16(2): 214-217.	Saudi Arabia
Al-Obaidi, R. G. Y., et al. (2016). "Molecular Analysis of CYP21A2 Gene Mutations among Iraqi Patients with Congenital Adrenal Hyperplasia." Enzyme Research 2016 (no pagination).	Iraq
Alzanbagi, M. A., et al. (2018). "Growth characteristics in children with congenital adrenal hyperplasia." Saudi Medical Journal 39(7): 674-678.	Saudi Arabia
Amr, N. H., et al. (2014). "Carotid intima media thickness and other cardiovascular risk factors in children with congenital adrenal hyperplasia." Journal of Endocrinological Investigation 37(10): 1001-1008.	Egypt
Amr, N. H., et al. (2019). "Cognitive functions in children with congenital adrenal hyperplasia." Archives of Endocrinology & Metabolism 63(2): 113-120.	Egypt
Ariyawatkul, K., et al. (2017). "Cardio-metabolic risk factors in youth with classical 21-hydroxylase deficiency." European Journal of Pediatrics 176(4): 537-545.	Thailand
Ballerini, M. G., et al. (2014). "Serum concentration of 17alpha-hydroxyprogesterone in children from birth to adolescence." Hormone Research in Paediatrics 81(2): 118-125.	Argentina
Bunraungsak, S., et al. (2013). "Growth pattern and pubertal development in patients with classic 21-hydroxylase deficiency." Asian Biomedicine 7(6): 787-794.	Thailand

Cakir, E. D., et al. (2012). "Testicular adrenal rest tumors in patients with congenital adrenal hyperplasia." Journal of clinical research in pediatric endocrinology 4(2): 94-100.	Turkey
Coeli-Lacchini, F. B., et al. (2020). "Clinical, Molecular, Functional, and Structural Characterization of CYP17A1 Mutations in Brazilian Patients with 17-Hydroxylase Deficiency." Hormone and Metabolic Research 52(3): 186-193.	Brazil
De Miranda, M. C., et al. (2020). "Adverse outcomes and economic burden of congenital adrenal hyperplasia late diagnosis in the newborn screening absence." Journal of the Endocrine Society 4(2).	Brazil
Demirel, F., et al. (2014). "Bone mineral density and vitamin D status in children and adolescents with congenital adrenal hyperplasia." Turkish Journal of Medical Sciences 44(1): 109-114.	Turkey
Elmougy, F., et al. (2020). "Genetic profiling of CAH Egyptian children: rapid guide to clinical interpretation of common mutations." Journal of Endocrinological Investigation.	Egypt
Elnecave, R. H., et al. (2008). "Bone mineral density in girls with classical congenital adrenal hyperplasia due to CYP21 deficiency." Journal of Pediatric Endocrinology & Metabolism 21(12): 1155-1162.	Brazil
Engels, M., et al. (2019). "Glucocorticoid Activity of Adrenal Steroid Precursors in Untreated Patients With Congenital Adrenal Hyperplasia." Journal of Clinical Endocrinology & Metabolism 104(11): 5065-5072.	Indonesia
Fan, L., et al. (2019). "Novel phenotypes and genotypes in Antley-Bixler syndrome caused by cytochrome P450 oxidoreductase deficiency: Based on the first cohort of Chinese children." Orphanet Journal Of Rare Diseases 14(1).	China
Fernandez, C. S., et al. (2020). "Genetic characterization of a large cohort of Argentine 21-hydroxylase Deficiency." Clinical Endocrinology 93(1): 19-27.	Argentina
Fontenele, R., et al. (2018). "17alpha-HYDROXYLASE DEFICIENCY IS AN UNDERDIAGNOSED DISEASE: HIGH FREQUENCY OF MISDIAGNOSES IN A LARGE COHORT OF BRAZILIAN PATIENTS." Endocrine Practice 24(2): 170-178.	Brazil
Ganesh, R., et al. (2018). "Bone Mineral Content and Density in Indian Children with Congenital Adrenal Hyperplasia." Indian Pediatrics 55(10): 880-882.	India
Ganesh, R., et al. (2016). "Correlation of Bone Mineral Parameters with Anthropometric Measurements and the Effect of Glucocorticoids on Bone Mineral Parameters in Congenital Adrenal Hyperplasia." Indian Journal of Pediatrics 83(2): 126-130.	India
Ganie, Y., et al. (2017). "Disorders of sex development in children in KwaZulu-Natal Durban South Africa: 20-year experience in a tertiary centre." Journal of Pediatric Endocrinology & Metabolism 30(1): 11-18.	South Africa

Ganie, Y., et al. (2018). "Congenital adrenal hyperplasia due to 21-hydroxylase deficiency in south africa." South African Medical Journal 108(2): 132-137.	South Africa
Gazzaneo, I. F. P., et al. (2016). "Profile of patients with genitourinary anomalies treated in a clinical genetics service in the Brazilian unified health system." Revista Paulista de Pediatria 34(1): 91-98.	Brazil
Gilban, D. L., et al. (2014). "Health related quality of life of children and adolescents with congenital adrenal hyperplasia in Brazil." Health & Quality of Life Outcomes 12: 107.	Brazil
Gomes, L. G., et al. (2013). "Mineralocorticoid replacement during infancy for salt wasting congenital adrenal hyperplasia due to 21-hydroxylase deficiency." Clinics (Sao Paulo, Brazil) 68(2): 147-152.	Brazil
Goncalves, E. M., et al. (2009). "Impairment in anthropometric parameters and body composition in females with classical 21-hydroxylase deficiency." Journal of Pediatric Endocrinology & Metabolism 22(6): 519-529.	Brazil
Goncalves, E. M., et al. (2014). "Performance of phalangeal quantitative ultrasound parameters in the evaluation of reduced bone mineral density assessed by DX in patients with 21 hydroxylase deficiency." Ultrasound in Medicine & Biology 40(7): 1414-1419.	Brazil
Goncalves, E. M., et al. (2013). "Estimation of percent body fat based on anthropometric measurements in children and adolescents with congenital adrenal hyperplasia due to 21-hydroxylase deficiency." Clinical Nutrition 32(1): 45-50.	Brazil
Goncalves, E. M., et al. (2012). "Accuracy of anthropometric measurements in estimating fat mass in individuals with 21-hydroxylase deficiency." Nutrition 28(10): 984-990.	Brazil
Guven, A., et al. (2015). "Gonadotropin releasing hormone analog treatment in children with congenital adrenal hyperplasia complicated by central precocious puberty." Hormones 14(2): 265-271.	Turkey
Hamed, S. A., et al. (2018). "Cognitive function in children with classic congenital adrenal hyperplasia." European Journal of Pediatrics 177(11): 1633-1640.	Eygpt
Hassan, M. M., et al. (2013). "Growth in infants with congenital adrenal hyperplasia due to 21-hydroxylase deficiency: An analysis of the factors affecting height." Egyptian Pediatric Association Gazette 61(2): 57-62.	Eygpt
Hou, L., et al. (2019). "Analysis of phenotypes and genotypes in 84 patients with 21-Hydroxylase deficiency in southern China." Steroids 151: 108474.	China
Houben, C. H., et al. (2014). "Reconstructive surgery for females with congenital adrenal hyperplasia due to 21-hydroxylase deficiency: a review from the Prince of Wales Hospital." Hong Kong Medical Journal 20(6): 481-485.	Hong Kong

Huang, Z., et al. (2016). "Identification of five novel STAR variants in ten Chinese patients with congenital lipoid adrenal hyperplasia." Steroids 108: 85-91.	China
Liu, S. Y., et al. (2018). "Clinical characteristics of Taiwanese children with congenital adrenal hyperplasia due to 21-hydroxylase deficiency detected by neonatal screening." Journal of the Formosan Medical Association 117(2): 126-131.	Taiwan
Idris, A. N., et al. (2014). "Behavioural outcome in children with congenital adrenal hyperplasia: Experience of a single centre." International Journal of Endocrinology 2014 (no pagination).	Malaysia
Inozemtseva, O., et al. (2008). "Learning disabilities spectrum and sexual dimorphic abilities in girls with congenital adrenal hyperplasia." Journal of Child Neurology 23(8): 862-869.	Mexico
Iqbal, S. and A. H. Khan (2013). "Raised 17-hydroxyprogesterone levels in congenital adrenal hyperplasia." Jcpsp, Journal of the College of Physicians & Surgeons - Pakistan 23(5): 373-374.	Pakistan
Jan, I. A., et al. (2011). "Management of children with disorders of sexual development (DSD): A retrospective analysis." Pakistan Journal of Medical Sciences 27(4): 729-733.	Pakistan
Jaruratanasirikul, S. and V. Engchaun (2014). "Management of children with disorders of sex development: 20-year experience in southern Thailand." World Journal of Pediatrics 10(2): 168-174.	Thailand
Jaruratanasirikul, S. and M. Thaiwong (2012). "Precocious pubarche in Thai children." Journal of the Medical Association of Thailand 95(11): 1404-1410.	Thailand
Jaruratanasirikul, S. and T. Thongseiratch (2013). "Diagnosis and management of congenital adrenal hyperplasia: 20-year experience in Songklanagarind Hospital." Journal of the Medical Association of Thailand 96(3): 288-293.	Thailand
Jiang, J. F., et al. (2016). "Surgical Therapy of 17alpha-hydroxylase Deficiency in 30 Patients." Chung-Kuo i Hsueh Ko Hsueh Yuan Hsueh Pao Acta Academiae Medicinae Sinicae 38(5): 559-562.	China
Jiang, S. and Y. Kuang (2019). "The Cycle Characteristics and Outcomes of Infertile Nonclassic 21-Hydroxylase Deficiency Patients Undergoing Ovarian Stimulation for In Vitro Fertilization." Hormone & Metabolic Research 51(5): 315-325.	China
Juan, L., et al. (2016). "Near-final height in 82 Chinese patients with congenital adrenal hyperplasia due to classic 21-hydroxylase deficiency: a single-center study from China." Journal of Pediatric Endocrinology & Metabolism 29(7): 841-848.	China
Kang, M. J., et al. (2011). "The prevalence of testicular adrenal rest tumors and associated factors in postpubertal patients with congenital adrenal hyperplasia caused by 21-hydroxylase deficiency." Endocrine Journal 58(6): 501-508.	Korea

Kaupert, L. C., et al. (2016). "A Single Nucleotide Variant in the Promoter Region of 17beta-HSD Type 5 Gene Influences External Genitalia Virilization in Females with 21-Hydroxylase Deficiency." Hormone Research in Paediatrics 85(5): 333-338.	Brazil
Kharrat, M., et al. (2011). "Detection of a frequent duplicated CYP21A2 gene carrying a Q318X mutation in a general population with quantitative PCR methods." Diagnostic Molecular Pathology 20(2): 123-127.	Tunisia
Kharrat, M., et al. (2010). "Only two mutations detected in 15 Tunisian patients with 11beta-hydroxylase deficiency: the p.Q356X and the novel p.G379V." Clinical Genetics 78(4): 398-401.	Tunisia
Khorashad, B. S., et al. (2018). "Childhood Sex-Typed Behavior and Gender Change in Individuals with 46,XY and 46,XX Disorders of Sex Development: An Iranian Multicenter Study." Archives of Sexual Behavior 47(8): 2287-2298.	Iran
Kim, J. H., et al. (2017). "Long-term Consequences of Congenital Adrenal Hyperplasia due to Classic 21-hydroxylase Deficiency in Adolescents and Adults." Experimental & Clinical Endocrinology & Diabetes 125(3): 196-201.	Korea
Koh, J. W., et al. (2013). "Clinical features of congenital adrenal insufficiency including growth patterns and significance of ACTH stimulation test." Journal of Korean Medical Science 28(11): 1650-1656.	Korea
Kollahi, N. A., et al. (2019). "Complex alleles of cyp21a2 are the most frequent causes of congenital adrenal hyperplasia in Iranian population." Iranian Journal of Pediatrics 29(6).	Iran
Kulshreshtha, B., et al. (2008). "Fertility among women with classical congenital adrenal hyperplasia: report of seven cases where treatment was started after 9 years of age." Gynecological Endocrinology 24(5): 267-272.	India
Lai, S., et al. (2015). "Clinical profile of congenital adrenal hyperplasia and short-term response to treatment." Rawal Medical Journal 40(1): 44-47.	Pakistan
Lee, H. H., et al. (2008). "Low frequency of the CYP21A2 deletion in ethnic Chinese (Taiwanese) patients with 21-hydroxylase deficiency." Molecular Genetics and Metabolism 93(4): 450-457.	Taiwan
Leong, K. S. W. and L. L. Wu (2019). "Case series of testicular adrenal rest tumours in boys with congenital adrenal hyperplasia: A single centre experience." Medical Journal of Malaysia 74(1): 92-93.	Malaysia
Liang, H. Y., et al. (2008). "Psychiatric manifestations in young females with congenital adrenal hyperplasia in Taiwan." Chang Gung Medical Journal 31(1): 66-73.	Taiwan
Longui, C. A., et al. (2011). "Near-final height in patients with congenital adrenal hyperplasia treated with combined therapy using GH and GnRHa." Arquivos Brasileiros de Endocrinologia e Metabologia 55(8): 661-664.	Brazil
Ma, L., et al. (2019). "Sonographic features of the testicular adrenal rests tumors in patients with congenital adrenal hyperplasia: a single-center experience and literature review." Orphanet Journal Of Rare Diseases 14(1): 242.	China

Maheshwari, A., et al. (2019). "Long-term Growth in Congenital Adrenal Hyperplasia." Indian Journal of Pediatrics 86(2): 154-158.	India
Maiti, A. and S. Chatterjee (2011). "Congenital adrenal hyperplasia: an Indian experience." Journal of Paediatrics & Child Health 47(12): 883-887.	India
Manzoor, J., et al. (2019). "Ambiguous genitalia: An overview of 7 years experience at the children's hospital & institute of child health, lahore, pakistan." Pakistan Journal of Medical Sciences 35(1): 151-155.	Pakistan
Marei, M. M., et al. (2016). "Anatomical measurements of the urogenital sinus in virilized female children due to congenital adrenal hyperplasia." Journal of Pediatric Urology 12(5): 282.e281-282.e288.	Eygpt
Marei, M. M., et al. (2016). "Timing and Outcome Concerns regarding Feminizing Genitoplasty from the Perspective of Egyptian Families of Girls with Virilized External Genitalia." Hormone Research in Paediatrics 85(1): 49-57.	Eygpt
Meena, H., et al. (2019). "Growth Pattern and Clinical Profile of Indian Children with Classical 21-Hydroxylase Deficiency Congenital Adrenal Hyperplasia on Treatment." Indian Journal of Pediatrics 86(6): 496-502.	India
Mendes-Dos-Santos, C. T., et al. (2011). "Normalization of height and excess body fat in children with salt-wasting 21-hydroxylase deficiency." Jornal de Pediatria 87(3): 263-268.	Brazil
Mendes-Dos-Santos, C. T., et al. (2018). "Prevalence of Testicular Adrenal Rest Tumor and Factors Associated with Its Development in Congenital Adrenal Hyperplasia." Hormone Research in Paediatrics 90(3): 161-168.	Brazil
Metwalley, K. A., et al. (2019). "Epicardial Fat Thickness in Children with Classic Congenital Adrenal Hyperplasia." Journal of clinical research in pediatric endocrinology 11(1): 61-69.	Eygpt
Metwalley, K. A., et al. (2016). "Left ventricular dysfunction and subclinical atherosclerosis in children with classic congenital adrenal hyperplasia: a single-center study from upper Egypt." European Journal of Pediatrics 175(3): 405-412.	Eygpt
Milyani, A. A., et al. (2018). "Initial presentations and associated clinical findings in patients with classical congenital adrenal hyperplasia." Journal of Pediatric Endocrinology & Metabolism 31(6): 671-673.	Saudi Arabia
Moreira, R. P., et al. (2013). "Obesity and familial predisposition are significant determining factors of an adverse metabolic profile in young patients with congenital adrenal hyperplasia." Hormone Research in Paediatrics 80(2): 111-118.	Brazil
Moreira, R. P. P., et al. (2012). "Impact of Glucocorticoid Receptor Gene Polymorphisms on the Metabolic Profile of Adult Patients with the Classical Form of 21-Hydroxylase Deficiency." PLoS ONE 7(9).	Brazil
Moura-Massari, V. O., et al. (2013). "CYP21A2 genotypes do not predict the severity of hyperandrogenic manifestations in the nonclassical form of congenital adrenal hyperplasia." Hormone & Metabolic Research 45(4): 301-307.	Brazil

Moura-Massari, V. O., et al. (2016). "The Presence of Clitoromegaly in the Nonclassical Form of 21-Hydroxylase Deficiency Could Be Partially Modulated by the CAG Polymorphic Tract of the Androgen Receptor Gene." PLoS ONE [Electronic Resource] 11(2): e0148548.	Brazil
Musa, N., et al. (2020). "Assessment of health-related quality of life in Egyptian children and adolescents with congenital adrenal hyperplasia." Journal of Pediatric Endocrinology and Metabolism 33(2): 295-304.	Eygpt
Nageshwari, R., et al. (2017). "Common CYP21A2 gene mutations in South Indian congenital adrenal hyperplasia patients." International Journal of Human Genetics 17(3): 103-108.	India
Neres, M. S., et al. (2010). "Distinctive profile of the 17-hydroxylase and 17,20-lyase activities revealed by urinary steroid metabolomes of patients with CYP17 deficiency." Arquivos Brasileiros de Endocrinologia e Metabologia 54(9): 826-832.	Brazil
Osifo, O. D. and T. I. Amusan (2009). "Female children with ambiguous genitalia in awareness-poor subregion." African Journal of Reproductive Health 13(4): 129-136.	Nigeria
Osuwannaratana, P., et al. (2008). "The etiologies of adrenal insufficiency in 73 Thai children: 20 years experience." Journal of the Medical Association of Thailand 91(10): 1544-1550.	Thailand
Park, S., et al. (2011). "Long-term follow-up after feminizing genital reconstruction in patients with ambiguous genitalia and high vaginal confluence." Journal of Korean Medical Science 26(3): 399-403.	Korea
Rabbani, B., et al. (2012). "Mutation analysis of the CYP21A2 gene in the Iranian population." Genetic Testing & Molecular Biomarkers 16(2): 82-90.	Iran
Rehman, U. L., et al. (2016). "Clinical Spectrum of Disorders of Sexual Differentiation." Jcpsp, Journal of the College of Physicians & Surgeons - Pakistan 26(3): 199-203.	Pakistan
Rodrigues, T. M., et al. (2015). "Cardiovascular risk factors and increased carotid intima-media thickness in young patients with congenital adrenal hyperplasia due to 21-hydroxylase deficiency." Archives of Endocrinology & Metabolism 59(6): 541-547.	Brazil
Sahakitrungruang, T., et al. (2008). "Bone mineral density and body composition in prepubertal and adolescent patients with the classical form of 21-hydroxylase deficiency." Journal of the Medical Association of Thailand 91(5): 705-710.	Thailand
Sap, S. N. U., et al. (2020). "Observational study of disorders of sex development in Yaounde, Cameroon." Journal of Pediatric Endocrinology and Metabolism 33(3): 417-423.	Cameroon
Schnaider-Rezek, G. S., et al. (2011). "Metabolic evaluation of young women with congenital adrenal hyperplasia." Arquivos Brasileiros de Endocrinologia e Metabologia 55(8): 646-652.	Brazil

Seraphim, C. E., et al. (2019). "Impact of Long-Term Dexamethasone Therapy on the Metabolic Profile of Patients with 21-Hydroxylase Deficiency." Journal of the Endocrine Society 3(8): 1574-1582.	Brazil
Seyam, R. M., et al. (2013). "Long-term outcome of genital reconstruction of Middle Eastern women with congenital adrenal hyperplasia." Urology Annals 5(4): 277-282.	Middle Eastern population
Sharaf, S., et al. (2015). "High frequency of splice site mutation in 21-hydroxylase deficiency children." Journal of Endocrinological Investigation 38(5): 505-511.	Egypt
Su, L., et al. (2018). "Clinical presentation and mutational spectrum in a series of 166 patients with classical 21-hydroxylase deficiency from South China." Clinica Chimica Acta 486: 142-150.	China
Tony Nengom, J., et al. (2017). "Assessment of cardiac function in children with congenital adrenal hyperplasia: a case control study in Cameroon." BMC Pediatrics 17(1): 109.	Cameroon
Vijayan, R., et al. (2019). "Metabolic profile, cardiovascular risk factors and health-related quality of life in children, adolescents and young adults with congenital adrenal hyperplasia." Journal of Pediatric Endocrinology & Metabolism 32(8): 871-877.	India
Wang, C. and Q. Tian (2015). "The investigation of quality of life in 87 Chinese patients with disorders of sex development." BioMed Research International 2015: 342420.	China
Werneck, G., et al. (2019). "Testicular adrenal rest tumors in patients with congenital adrenal hyperplasia: 6 years of follow-up." Journal of Pediatric Endocrinology & Metabolism 32(5): 519-526.	Brazil
Wijaya, M., et al. (2019). "Etiology of primary adrenal insufficiency in children: a 29-year single-center experience." Journal of Pediatric Endocrinology & Metabolism 32(6): 615-622.	China
Wu, C., et al. (2017). "17alpha-hydroxylase/17, 20-lyase deficiency: Clinical and molecular characterization of eight Chinese patients." Endocrine Practice 23(5): 576-582.	China
Xu, C., et al. (2019). "Genotype-phenotype correlation study and mutational and hormonal analysis in a Chinese cohort with 21-hydroxylase deficiency." Molecular Genetics & Genomic Medicine 7(6): e671.	China
Xu, J. and P. Li (2019). "Identification of novel and rare CYP21A2 variants in Chinese patients with congenital adrenal hyperplasia due to 21-hydroxylase deficiency." Clinical Biochemistry 68: 44-49.	China
Yu, M. K., et al. (2015). "Clinical manifestations of testicular adrenal rest tumor in males with congenital adrenal hyperplasia." Annals of Pediatric Endocrinology and Metabolism 20(3): 155-161.	Korea
Zainuddin, A. A., et al. (2020). "A Multicenter Cross-Sectional Study of Malaysian Females With Congenital Adrenal Hyperplasia: Their Body Image and Their Perspectives on Feminizing Surgery." Journal of Pediatric and Adolescent Gynecology.	Malaysia

Zainuddin, A. A., et al. (2019). "Malaysian females with congenital adrenal hyperplasia: Surgical outcomes and attitudes." Frontiers in Pediatrics 7(MAR).	Malaysia
Zhang, B., et al. (2017). "Molecular diagnosis of Chinese patients with 21-hydroxylase deficiency and analysis of genotype-phenotype correlations." Journal of International Medical Research 45(2): 481-492.	China
Zhang, H. J., et al. (2010). "Metabolic disorders in newly diagnosed young adult female patients with simple virilizing 21-hydroxylase deficiency." Endocrine 38(2): 260-265.	China
Zhang, M., et al. (2015). "New, recurrent, and prevalent mutations: Clinical and molecular characterization of 26 Chinese patients with 17alpha-hydroxylase/17,20-lyase deficiency." Journal of Steroid Biochemistry & Molecular Biology 150: 11-16.	China
Zhu, D., et al. (2012). "Quality of life evaluation in juveniles with disorders of sexual development." Pediatric Surgery International 28(11): 1119-1123.	China
Zou, C. C., et al. (2008). "Peripheral precocious puberty: a retrospective study for six years in Hangzhou, China." Journal of Paediatrics & Child Health 44(7-8): 415-418.	China

Table 17. Studies excluded from Q2 for other reasons

Reference	Reason for exclusion
Agladioglu, S. Y., et al. (2011). "Does pseudohypoaldosteronism mask the diagnosis of congenital adrenal hyperplasia?" JCRPE Journal of Clinical Research in Pediatric Endocrinology 3(4): 219-221.	Only two cases
Alwashih, M. A., et al. (2017). "Plasma metabolomic profile varies with glucocorticoid dose in patients with congenital adrenal hyperplasia." Scientific Reports 7(1): 17092.	No age at presentation data
Ambroziak, U., et al. (2016). "The diagnosis of nonclassic congenital adrenal hyperplasia due to 21-hydroxylase deficiency, based on serum basal or post-ACTH stimulation 17-hydroxyprogesterone, can lead to false-positive diagnosis." Clinical Endocrinology 84(1): 23-29.	Diagnosis was purpose of study
Ambroziak, U., et al. (2015). "LC-MS/MS improves screening towards 21-hydroxylase deficiency." Gynecological Endocrinology 31(4): 296-300.	Diagnosis was purpose of study
Arlt, W., et al. (2010). "Health status of adults with congenital adrenal hyperplasia: a cohort study of 203 patients." Journal of Clinical Endocrinology & Metabolism 95(11): 5110-5121.	No age at presentation data

Armengaud, J. B., et al. (2009). "Precocious pubarche: distinguishing late-onset congenital adrenal hyperplasia from premature adrenarche." Journal of Clinical Endocrinology & Metabolism 94(8): 2835-2840.	Diagnosis was purpose of study
Auchus, R. J., et al. (2014). "Abiraterone acetate to lower androgens in women with classic 21-hydroxylase deficiency." Journal of Clinical Endocrinology & Metabolism 99(8): 2763-2770.	No age at presentation data
Bachelot, A., et al. (2012). "Influence of hormonal control on LH pulsatility and secretion in women with classical congenital adrenal hyperplasia." European Journal of Endocrinology 167(4): 499-505.	No age at presentation data
Baskin, A., et al. (2020). "Post-operative complications following feminizing genitoplasty in moderate to severe genital atypia: Results from a multicenter, observational prospective cohort study." Journal of Pediatric Urology.	No age at presentation data
Binay, C., et al. (2014). "Prevalence of nonclassic congenital adrenal hyperplasia in turkish children presenting with premature pubarche, hirsutism, or oligomenorrhoea." International Journal of Endocrinology 2014 (no pagination).	No separate age at presentation data for CAH subgroup
Binet, A., et al. (2016). "Should we question early feminizing genitoplasty for patients with congenital adrenal hyperplasia and XX karyotype?" Journal of Pediatric Surgery 51(3): 465-468.	No age at presentation data
Bizzarri, C., et al. (2009). "Growth hormone response to physical exercise in growing patients with classic congenital adrenal hyperplasia." Journal of Endocrinological Investigation 32(11): 903-907.	No age at presentation data
Bleicken, B., et al. (2012). "Improvement of health-related quality of life in adult women with 21-hydroxylase deficiency over a seven-year period." Endocrine Journal 59(10): 931-939.	No age at presentation data
Bogdanska, M., et al. (2018). "Long-term urinary symptoms in adolescent and adult women with congenital adrenal hyperplasia." Journal of Pediatric Urology 14(3): 240.e241-240.e245.	No age at presentation data
Bonfig, W., et al. (2018). "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 89(1): 7-12.	76% diagnosed by newborn screening. Age of diagnosis for remainder of sample not given.
Bonfig, W., et al. (2011). "Growth patterns in the first three years of life in children with classical congenital adrenal hyperplasia diagnosed by newborn screening and treated with low doses of hydrocortisone." Hormone Research in Paediatrics 75(1): 32-37.	Whole sample diagnosed by newborn screening.
Bonfig, W. and H. P. Schwarz (2014). "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 81(6): 871-875.	Whole sample diagnosed by newborn screening.
Bougneres, P., et al. (2017). "Deferring surgical treatment of ambiguous genitalia into adolescence in girls with 21-hydroxylase deficiency: A feasibility study." International Journal of Pediatric Endocrinology 2017(1).	Only states pre- or post-natal diagnosis

Bouvattier, C., et al. (2015). "Clinical Outcome, Hormonal Status, Gonadotrope Axis, and Testicular Function in 219 Adult Men Born With Classic 21-Hydroxylase Deficiency. A French National Survey." Journal of Clinical Endocrinology & Metabolism 100(6): 2303-2313.	No age at presentation data
Brownfoot, F. C., et al. (2014). "Complex atypical hyperplasia of the endometrium: differences in outcome following conservative management of pre- and postmenopausal women." Reproductive Sciences 21(10): 1244-1248.	No age at presentation data
Cavarzere, P., et al. (2010). "Possible andrologic markers in elevated neonatal 17-hydroxyprogesterone." Fertility and Sterility 94(6): 2350-2352.	Newborns with increased 17 -OHP but not CAH
Cavarzere, P., et al. (2013). "Genotype in the diagnosis of 21-hydroxylase deficiency: who should undergo CYP21A2 analysis?" Journal of Endocrinological Investigation 36(11): 1083-1089.	Gives detail on method of diagnosis but not age if not newborn
Ceccato, F., et al. (2016). "Long-term glucocorticoid effect on bone mineral density in patients with congenital adrenal hyperplasia due to 21-hydroxylase deficiency." European Journal of Endocrinology 175(2): 101-106.	Only age of diagnosis year range per patients given in addition to number identified through newborn screening
Chakhtoura, Z., et al. (2008). "Impact of total cumulative glucocorticoid dose on bone mineral density in patients with 21-hydroxylase deficiency." European Journal of Endocrinology 158(6): 879-887.	No age at presentation data
Chesover, A. D., et al. (2020). "Screening for nonclassic congenital adrenal hyperplasia in the era of liquid chromatography-tandem mass spectrometry." Journal of the Endocrine Society 4(2).	No age at presentation and not newborn testing
Chrisp, G. L., et al. (2018). "Variations in the management of acute illness in children with congenital adrenal hyperplasia: An audit of three paediatric hospitals." Clinical Endocrinology 89(5): 577-585.	No age at presentation
Collaer, M. L., et al. (2016). "Reduced short term memory in congenital adrenal hyperplasia (CAH) and its relationship to spatial and quantitative performance." Psychoneuroendocrinology 64: 164-173.	No age at presentation data
Crouch, N. S., et al. (2008). "Sexual function and genital sensitivity following feminizing genitoplasty for congenital adrenal hyperplasia." Journal of Urology 179(2): 634-638.	No age at presentation data
Dangle, P. P., et al. (2017). "Surgical Complications Following Early Genitourinary Reconstructive Surgery for Congenital Adrenal Hyperplasia-Interim Analysis at 6 Years." Urology 101: 111-115.	All but 2 patients diagnosed at birth. Individual ages given for n=2.
de Groot, M. J., et al. (2015). "Salivary morning androstenedione and 17alpha-OH progesterone levels in childhood and puberty in patients with classic congenital adrenal hyperplasia." Clinical Chemistry & Laboratory Medicine 53(3): 461-468.	Only gives median age of therapeutic follow-up
Debono, M., et al. (2015). "Hormonal circadian rhythms in patients with congenital adrenal hyperplasia: identifying optimal monitoring times and novel disease biomarkers." European Journal of Endocrinology 173(6): 727-737.	No age of presentation data
Delfino, M., et al. (2012). "Testicular adrenal rest tumors in patients with congenital adrenal hyperplasia: prevalence and sonographic, hormonal, and seminal characteristics." Journal of Ultrasound in Medicine 31(3): 383-388.	No age of presentation data

Delvecchio, M., et al. (2015). "Evaluation of impact of steroid replacement treatment on bone health in children with 21-hydroxylase deficiency." Endocrine 48(3): 995-1000.	Only states ten patients diagnosed at birth and the remainder before aged 16
Dobrowolska-Glazar, B., et al. (2020). "Sexual function and health status in adult patients with Congenital Adrenal Hyperplasia." Journal of Pediatric Urology.	Age at surgery not presentation
Doherty, P. J., et al. (2012). "Absence of prostatic growth in large cohort of adult female patients with congenital adrenal hyperplasia." Journal of Urology 188(4 SUPPL.): 1588-1595.	No age of presentation data
Dorr, H. G., et al. (2018). "Miscarriages in families with an offspring that have classic congenital adrenal hyperplasia and 21-hydroxylase deficiency." BMC Pregnancy & Childbirth 18(1): 456.	No age at presentation data
Dorr, H. G., et al. (2019). "Birth Size in Neonates with Congenital Adrenal Hyperplasia due to 21-hydroxylase Deficiency." Journal of clinical research in pediatric endocrinology 11(1): 41-45.	All diagnosed at birth
Dudzinska, B., et al. (2014). "Sexual well-being in adult male patients with congenital adrenal hyperplasia." International Journal of Endocrinology 2014 (no pagination).	Only a generalised statement on age of presentation given
Dumic, K. K., et al. (2017). "Molecular genetic analysis in 93 patients and 193 family members with classical congenital adrenal hyperplasia due to 21-hydroxylase deficiency in Croatia." Journal of Steroid Biochemistry & Molecular Biology 165(Pt A): 51-56.	Only a generalised statement on age of presentation given
El-Maouche, D., et al. (2015). "Cortical bone mineral density in patients with congenital adrenal hyperplasia due to 21-hydroxylase deficiency." Clinical Endocrinology 82(3): 330-337.	No age at presentation data
El-Maouche, D., et al. (2019). "Adrenal morphology and associated comorbidities in congenital adrenal hyperplasia." Clinical Endocrinology 91(2): 247-255.	No age at presentation data
Engberg, H., et al. (2015). "Congenital adrenal hyperplasia and risk for psychiatric disorders in girls and women born between 1915 and 2010: A total population study." Psychoneuroendocrinology 60: 195-205.	Stratifies data pre and post newborn screening in 1986 but doesn't give median diagnosis age pre screening.
Engels, M., et al. (2018). "Gonadal function in adult male patients with congenital adrenal hyperplasia." European Journal of Endocrinology 178(3): 285-294.	No age at presentation
Faienza, M. F., et al. (2009). "Osteoclastogenesis in children with 21-hydroxylase deficiency on long-term glucocorticoid therapy: the role of receptor activator of nuclear factor-kappaB ligand/osteoprotegerin imbalance." Journal of Clinical Endocrinology & Metabolism 94(7): 2269-2276.	Only general statement of age at diagnosis given
Falhammar, H., et al. (2014). "Increased psychiatric morbidity in men with congenital adrenal hyperplasia due to 21-hydroxylase deficiency." Journal of Clinical Endocrinology & Metabolism 99(3): E554-560.	No age of presentation given
Falhammar, H., et al. (2013). "Bone mineral density, bone markers, and fractures in adult males with congenital adrenal hyperplasia." European Journal of Endocrinology 168(3): 331-341.	No age at presentation data

Falhammar, H., et al. (2019). "Increased risk of autoimmune disorders in 21-hydroxylase deficiency: A Swedish population-based national cohort study." Journal of the Endocrine Society 3(5): 1039-1052.	No age at presentation data
Falhammar, H., et al. (2015). "Increased Cardiovascular and Metabolic Morbidity in Patients With 21-Hydroxylase Deficiency: A Swedish Population-Based National Cohort Study." Journal of Clinical Endocrinology & Metabolism 100(9): 3520-3528.	No age at presentation data
Falhammar, H., et al. (2017). "Reduced Frequency of Biological and Increased Frequency of Adopted Children in Males With 21-Hydroxylase Deficiency: A Swedish Population-Based National Cohort Study." Journal of Clinical Endocrinology & Metabolism 102(11): 4191-4199.	No age at presentation data
Falhammar, H., et al. (2012). "Fertility, sexuality and testicular adrenal rest tumors in adult males with congenital adrenal hyperplasia." European Journal of Endocrinology 166(3): 441-449.	No age at presentation data
Falhammar, H., et al. (2014). "Quality of life, social situation, and sexual satisfaction, in adult males with congenital adrenal hyperplasia." Endocrine 47(1): 299-307.	Early versus late diagnosis but no specific details on age of presentation
Fleming, L., et al. (2017). "Parental management of adrenal crisis in children with congenital adrenal hyperplasia." Journal for Specialists in Pediatric Nursing: JSPN 22(4): 10.	No age at presentation data
Fleming, L. K., et al. (2011). "Caregiver knowledge and self-confidence of stress dosing of hydrocortisone in children with congenital adrenal hyperplasia." Journal of Pediatric Nursing 26(6): e55-60.	Population is caregivers
Frey, K. R., et al. (2018). "Prednisolone is associated with a worse bone mineral density in primary adrenal insufficiency." Endocrine Connections 7(6): 811-818.	No age at presentation data
Frisen, L., et al. (2009). "Gender role behavior, sexuality, and psychosocial adaptation in women with congenital adrenal hyperplasia due to CYP21A2 deficiency." Journal of Clinical Endocrinology and Metabolism 94(9): 3432-3439.	No age at presentation data
Gehrmann, K., et al. (2019). "Sexuality in Males with Congenital Adrenal Hyperplasia Resulting from 21-Hydroxylase Deficiency." Journal of the Endocrine Society 3(8): 1445-1456.	No age at presentation data
German, A., et al. (2008). "Control of childhood congenital adrenal hyperplasia and sleep activity and quality with morning or evening glucocorticoid therapy." Journal of Clinical Endocrinology & Metabolism 93(12): 4707-4710.	No age at presentation data
Ghanny, B. A., et al. (2016). "Should children with isolated premature adrenarche be routinely evaluated for non-classical congenital adrenal hyperplasia?" Journal of Pediatric Endocrinology & Metabolism 29(3): 351-356.	No age at presentation data
Ghizzoni, L., et al. (2011). "Relationship of CYP21A2 genotype and serum 17-hydroxyprogesterone and cortisol levels in a large cohort of Italian children with premature pubarche." European Journal of Endocrinology 165(2): 307-314.	Not newborns
Gialluisi, A., et al. (2018). "A genetic epidemiology study of congenital adrenal hyperplasia in Italy." Clinical Genetics 93(2): 223-227.	No age at presentation data

Giebels, V., et al. (2014). "Severe fatigue in patients with adrenal insufficiency: physical, psychosocial and endocrine determinants." Journal of Endocrinological Investigation 37(3): 293-301.	No age at presentation data
Gonc, E. N., et al. (2011). "Is basal serum 17-OH progesterone a reliable parameter to predict nonclassical congenital adrenal hyperplasia in premature adrenarche?" Turkish Journal of Pediatrics 53(3): 274-280.	Population not newborns so exclude for screening. Age of onset given but only for whole population which is not CAH
Guarnotta, V., et al. (2020). "Clinical and hormonal characteristics in heterozygote carriers of congenital adrenal hyperplasia." Journal of Steroid Biochemistry and Molecular Biology 198 (no pagination).	Exclude - no age at presentation data
Gunnarsson, C., et al. (2017). "Health care burden in patients with adrenal insufficiency." Journal of the Endocrine Society 1(5): 512-523.	Exclude - no age at presentation data
Guran, T., et al. (2016). "Rare Causes of Primary Adrenal Insufficiency: Genetic and Clinical Characterization of a Large Nationwide Cohort." Journal of Clinical Endocrinology & Metabolism 101(1): 284-292.	CAH specifically excluded
Halper, A., et al. (2017). "Health-related quality of life in children with congenital adrenal hyperplasia." Health & Quality of Life Outcomes 15(1): 194.	No age at presentation data
Halper, A., et al. (2018). "Bone mineral density and body composition in children with congenital adrenal hyperplasia." Clinical Endocrinology 88(6): 813-819.	Controls for age at diagnosis but doesn't state what this is
Han, T. S., et al. (2013). "Quality of life in adults with congenital adrenal hyperplasia relates to glucocorticoid treatment, adiposity and insulin resistance: United Kingdom Congenital adrenal Hyperplasia Adult Study Executive (CaHASE)." European Journal of Endocrinology 168(6): 887-893.	No age at presentation data
Han, T. S., et al. (2013). "Glucocorticoid treatment regimen and health outcomes in adults with congenital adrenal hyperplasia." Clinical Endocrinology 78(2): 197-203.	No age at presentation data
Jenkins-Jones, S., et al. (2018). "Poor compliance and increased mortality, depression and healthcare costs in patients with congenital adrenal hyperplasia." European Journal of Endocrinology 178(4): 309-320.	No age at presentation data
Hirschberg, A. L., et al. (2017). "Fertility outcome and information on fertility issues in individuals with different forms of disorders of sex development: findings from the dsd-LIFE study." Fertility and Sterility 108(5): 822-831.	No age at presentation data
Idkowiak, J., et al. (2011). "Pubertal presentation in seven patients with congenital adrenal hyperplasia due to P450 oxidoreductase deficiency." Journal of Clinical Endocrinology & Metabolism 96(3): E453-462.	CAH due to ORD - gives individual age at investigation for each of 7 patients
Ishii, T., et al. (2018). "Incidence and Characteristics of Adrenal Crisis in Children Younger than 7 Years with 21-Hydroxylase Deficiency: A Nationwide Survey in Japan." Hormone Research in Paediatrics 89(3): 166-171.	Only age at evaluation for study
Janin, C., et al. (2013). "Clinical audit concerning the quality of management in patients with classic form of congenital adrenal hyperplasia." Annales d Endocrinologie 74(1): 13-26.	No age at presentation data

Jayakrishnan, R., et al. (2019). "Revisiting the association of HLA alleles and haplotypes with CYP21A2 mutations in a large cohort of patients with congenital adrenal hyperplasia." Gene 687: 30-34.	Jayakrishnan, R., et al. (2019). "Revisiting the association of HLA alleles and haplotypes with CYP21A2 mutations in a large cohort of patients with congenital adrenal hyperplasia." Gene 687: 30-34.
Jurgensen, M., et al. (2014). "Health-related quality of life in children with disorders of sex development (DSD)." European Journal of Pediatrics 173(7): 893-903.	Exclude - no CAH
Kandemir, N., et al. (2017). "Novel and prevalent CYP11B1 gene mutations in Turkish patients with 11-beta hydroxylase deficiency." Journal of Steroid Biochemistry & Molecular Biology 165(Pt A): 57-63.	Only general statements of age of presentation given (e.g. '7 patients were diagnosed in the first year of life')
Karlsson, L., et al. (2019). "Epigenetic alterations associated with early prenatal dexamethasone treatment." Journal of the Endocrine Society 3(1): 250-263.	No age at presentation data
Karlsson, L., et al. (2017). "Cognitive impairment in adolescents and adults with congenital adrenal hyperplasia." Clinical Endocrinology 87(6): 651-659.	76% were diagnosed through the National Screening Programme
Karunasena, N., et al. (2017). "Androgens correlate with increased erythropoiesis in women with congenital adrenal hyperplasia." Clinical Endocrinology 86(1): 19-25.	No age at presentation data
Karunasena, N., et al. (2017). "Impact of food, alcohol and pH on modified-release hydrocortisone developed to treat congenital adrenal hyperplasia." European Journal of Endocrinology 176(4): 405-411.	Exclude not CAH patients
Kaur, J. and P. Dey (2010). "Mean nuclear volume in complex atypical hyperplasia and adenocarcinoma of the endometrium." Analytical & Quantitative Cytology & Histology 32(5): 291-294.	Not CAH
Khan, K. M., et al. (2015). "Use of Automated Bone Age for Critical Growth Assessment." Clinical Pediatrics 54(11): 1038-1043.	No age at presentation data
Kim, M. S., et al. (2017). "Testicular Adrenal Rest Tumors in Boys and Young Adults with Congenital Adrenal Hyperplasia." Journal of Urology 197(3 Pt 2): 931-936.	No age at presentation data
Kim, M. S., et al. (2014). "Decreased adrenomedullary function in infants with classical congenital adrenal hyperplasia." Journal of Clinical Endocrinology & Metabolism 99(8): E1597-1601.	All diagnosed through newborn screening
Kim, M. S., et al. (2015). "Increased Abdominal Adiposity in Adolescents and Young Adults With Classical Congenital Adrenal Hyperplasia due to 21-Hydroxylase Deficiency." Journal of Clinical Endocrinology & Metabolism 100(8): E1153-1159.	No age at presentation data
King, T. F., et al. (2016). "Experience in optimizing fertility outcomes in men with congenital adrenal hyperplasia due to 21 hydroxylase deficiency." Clinical Endocrinology 84(6): 830-836.	No age at presentation data

Kirac, D., et al. (2014). "The frequency and the effects of 21-hydroxylase gene defects in congenital adrenal hyperplasia patients." Annals of Human Genetics 78(6): 399-409.	No age at presentation data
Kirli, E. A., et al. (2013). "An unexpected diagnosis in children with male phenotype and bilateral nonpalpable gonad: congenital adrenal hyperplasia with female genotype." Pediatric Surgery International 29(7): 719-724.	Exclude - only individual age at presentation given
Kocova, M., et al. (2018). "Testicular adrenal rest tumors in boys with 21-hydroxylase deficiency, timely diagnosis and follow-up." Endocrine Connections 7(4): 544-552.	Exclude - individual age at presentation given for n=8
Kok, H. K., et al. (2015). "Imaging features of poorly controlled congenital adrenal hyperplasia in adults." British Journal of Radiology 88(1053).	Exclude - no age at presentation data
Korkmaz, H. A., et al. (2019). "The impact of 21-hydroxylase deficiency on cardiac repolarization changes in children with 21-hydroxylase-deficient congenital adrenal hyperplasia." Turkish Journal of Pediatrics 61(2): 228-235.	Exclude - no age at presentation data
Kreukels, B. P. C., et al. (2018). "Gender Dysphoria and Gender Change in Disorders of Sex Development/Intersex Conditions: Results From the dsd-LIFE Study." Journal of Sexual Medicine 15(5): 777-785.	DsD life CAH data reported elsewhere
Krone, N., et al. (2013). "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 and Metabolism 98(2): E346-E354.	No age at presentation data
Krone, N., et al. (2012). "Genotype-phenotype analysis in congenital adrenal hyperplasia due to P450 oxidoreductase deficiency." Journal of Clinical Endocrinology & Metabolism 97(2): E257-267.	No age at presentation data
Krysiak, R., et al. (2019). "The effect of atorvastatin on cardiometabolic risk factors in women with non-classic congenital adrenal hyperplasia: A pilot study." Pharmacological Reports: PR 71(3): 417-421.	No median age of presentation, diagnosis performed 3-11 months before start of study
Kung, K. T. F., et al. (2018). "Emotional and behavioral adjustment in 4 to 11-year-old boys and girls with classic congenital adrenal hyperplasia and unaffected siblings." Psychoneuroendocrinology 97: 104-110.	No age at presentation data
Lao, Q., et al. (2018). "Complement component 4 variations may influence psychopathology risk in patients with congenital adrenal hyperplasia due to 21-hydroxylase deficiency." Human Genetics.	No age at presentation data
Lindert, J., et al. (2016). "Perineal ultrasound offers useful information in girls with congenital adrenal hyperplasia." Journal of Pediatric Urology 12(6): 427.e421-427.e426.	No age at diagnosis/presentation
Lin-Su, K., et al. (2011). "Final adult height in children with congenital adrenal hyperplasia treated with growth hormone." Journal of Clinical Endocrinology & Metabolism 96(6): 1710-1717.	No age at presentation data
Louise Rushworth, R., et al. (2016). "Hospital Admission Patterns in Children with CAH: Admission Rates and Adrenal Crises Decline with Age." International Journal of Endocrinology 2016 (no pagination).	No age at presentation data

Mallappa, A., et al. (2015). "A phase 2 study of Chronocort, a modified-release formulation of hydrocortisone, in the treatment of adults with classic congenital adrenal hyperplasia." Journal of Clinical Endocrinology and Metabolism 100(3): 1137-1145.	No age at presentation data
Marra, A. M., et al. (2015). "Cardiovascular abnormalities and impaired exercise performance in adolescents with congenital adrenal hyperplasia." Journal of Clinical Endocrinology & Metabolism 100(2): 644-652.	No age at presentation data
Martin, D. D., et al. (2013). "Validation of automatic bone age determination in children with congenital adrenal hyperplasia." Pediatric Radiology 43(12): 1615-1621.	No age at presentation data
Mathews, G. A., et al. (2009). "Personality and congenital adrenal hyperplasia: possible effects of prenatal androgen exposure." Hormones & Behavior 55(2): 285-291.	No age at presentation data
Matsubara, Y., et al. (2013). "Longitudinal analysis of growth and body composition of Japanese 21-OHD patients in childhood." Endocrine Journal 60(2): 149-154.	Diagnosed through newborn screening
Mattila, A. K., et al. (2012). "Gender identity and gender role orientation in female assigned patients with disorders of sex development." Journal of Urology 188(5): 1930-1934.	No age at presentation data
Mazzilli, R., et al. (2019). "The high prevalence of testicular adrenal rest tumors in adult men with congenital adrenal hyperplasia is correlated with ACTH levels." Frontiers in Endocrinology 10(JUN).	No age at presentation data
Mazzone, L., et al. (2011). "Emotional memory in early steroid abnormalities: an FMRI study of adolescents with congenital adrenal hyperplasia." Developmental Neuropsychology 36(4): 473-492.	No age at presentation data
Menabo, S., et al. (2012). "A sequence variation in 3'UTR of CYP21A2 gene correlates with a mild form of congenital adrenal hyperplasia." Journal of Endocrinological Investigation 35(3): 298-305.	No age at presentation data
Merke, D. P., et al. (2013). "Tenascin-X haploinsufficiency associated with ehlers-danlos syndrome in patients with congenital adrenal hyperplasia." Journal of Clinical Endocrinology and Metabolism 98(2): E379-E387.	No age at presentation data
Messina, V., et al. (2020). "Good overall behavioural adjustment in children and adolescents with classic congenital adrenal hyperplasia." Endocrine 68(2): 427-437.	No age at presentation data
Meyer-Bahlburg, H. F., et al. (2008). "Sexual orientation in women with classical or non-classical congenital adrenal hyperplasia as a function of degree of prenatal androgen excess." Archives of Sexual Behavior 37(1): 85-99.	No age at presentation data
Meyer-Bahlburg, H. F., et al. (2012). "Cognitive outcome of offspring from dexamethasone-treated pregnancies at risk for congenital adrenal hyperplasia due to 21-hydroxylase deficiency." European Journal of Endocrinology 167(1): 103-110.	Prenatal diagnosis
Meyer-Bahlburg, H. F., et al. (2017). "Syndrome-Related Stigma in the General Social Environment as Reported by Women with Classical Congenital Adrenal Hyperplasia." Archives of Sexual Behavior 46(2): 341-351.	No age at presentation data

Meyer-Bahlburg, H. F. L., et al. (2018). "Stigma Associated with Classical Congenital Adrenal Hyperplasia in Women's Sexual Lives." Archives of Sexual Behavior 47(4): 943-951.	No age at presentation data
Meyer-Bahlburg, H. F. L., et al. (2017). "Stigma in Medical Settings As Reported Retrospectively by Women With Congenital Adrenal Hyperplasia (CAH) for Their Childhood and Adolescence." Journal of Pediatric Psychology 42(5): 496-503.	No age at presentation data
Minette, M. S., et al. (2013). "Cardiac function in congenital adrenal hyperplasia: a pattern of reversible cardiomyopathy." Journal of Pediatrics 162(6): 1193-1198, 1198.e1191.	All diagnosed at birth through newborn screening or ambiguous genitalia
Mooij, C. F., et al. (2010). "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 74(5): 328-332.	All diagnosed through newborn screening
Mooij, C. F., et al. (2011). "Adult patients with congenital adrenal hyperplasia have elevated blood pressure but otherwise a normal cardiovascular risk profile." PLoS ONE [Electronic Resource] 6(9): e24204.	No age at presentation data
Mooij, C. F., et al. (2018). "Cardiac function in paediatric patients with congenital adrenal hyperplasia due to 21 hydroxylase deficiency." Clinical Endocrinology 88(3): 364-371.	No age at presentation data
Mooij, C. F., et al. (2017). "Cardiovascular and metabolic risk in pediatric patients with congenital adrenal hyperplasia due to 21 hydroxylase deficiency." Journal of Pediatric Endocrinology & Metabolism 30(9): 957-966.	No age at presentation data
Morissette, R., et al. (2015). "Broadening the Spectrum of Ehlers Danlos Syndrome in Patients With Congenital Adrenal Hyperplasia." Journal of Clinical Endocrinology & Metabolism 100(8): E1143-1152.	No age at presentation data
Mueller, S. C., et al. (2010). "Psychiatric characterization of children with genetic causes of hyperandrogenism." European Journal of Endocrinology 163(5): 801-810.	No age at presentation data
Nandagopal, R., et al. (2011). "Phenotypic profiling of parents with cryptic nonclassic congenital adrenal hyperplasia: findings in 145 unrelated families." European Journal of Endocrinology 164(6): 977-984.	No age at presentation data
Napolitano, E., et al. (2011). "Correlation between genotype and hormonal levels in heterozygous mutation carriers and non-carriers of 21-hydroxylase deficiency." Journal of Endocrinological Investigation 34(7): 498-501.	No age at presentation data
Nella, A. A., et al. (2016). "A phase 2 study of continuous subcutaneous hydrocortisone infusion in adults with congenital adrenal hyperplasia." Journal of Clinical Endocrinology and Metabolism 101(12): 4690-4698.	No age at presentation data
Neocleous, V., et al. (2018). "Genotype is associated to the degree of virilization in patients with classic congenital adrenal hyperplasia." Frontiers in Endocrinology 9 (no pagination).	No age at presentation data
Neocleous, V., et al. (2014). "Phenotypic variability of hyperandrogenemia in females heterozygous for CYP21A2 mutations." Indian Journal of Endocrinology and Metabolism 18(Supplement 1): S72-S79.	No suitable age at presentation data

Neocleous, V., et al. (2012). "Genetic defects in the cyp21a2 gene in heterozygous girls with premature adrenarche and adolescent females with hyperandrogenemia." Georgian Medical News(210): 40-47.	Association of hyperandrogenism and carrier status
Nermoen, I., et al. (2012). "Genetic, anthropometric and metabolic features of adult Norwegian patients with 21-hydroxylase deficiency." European Journal of Endocrinology 167(4): 507-516.	No age at presentation data
Nermoen, I., et al. (2011). "High frequency of adrenal myelolipomas and testicular adrenal rest tumours in adult Norwegian patients with classical congenital adrenal hyperplasia because of 21-hydroxylase deficiency." Clinical Endocrinology 75(6): 753-759.	No age at presentation data
New, M. I., et al. (2013). "Genotype-phenotype correlation in 1,507 families with congenital adrenal hyperplasia owing to 21-hydroxylase deficiency." Proceedings of the National Academy of Sciences of the United States of America 110(7): 2611-2616.	No age at presentation data
Nguyen, L. S., et al. (2019). "Influence of hormones on the immunotolerogenic molecule HLA-G: a cross-sectional study in patients with congenital adrenal hyperplasia." European Journal of Endocrinology 181(5): 481-488.	No age at presentation data
Niceta, M., et al. (2011). "A large view of CYP21 locus among Sicilians and other populations: identification of a novel CYP21A2 variant in Sicily." Journal of Endocrinological Investigation 34(11): 847-854.	No age at presentation data
Ning, C., et al. (2008). "Body image in adolescents with disorders of steroidogenesis." Journal of Pediatric Endocrinology & Metabolism 21(8): 771-780.	No age at presentation data
Nokoff, N. J., et al. (2017). "Prospective assessment of cosmesis before and after genital surgery." Journal of Pediatric Urology 13(1): 28.e21-28.e26.	No separate age of diagnosis for CAH group
Nordenskjold, A., et al. (2008). "Type of mutation and surgical procedure affect long-term quality of life for women with congenital adrenal hyperplasia." Journal of Clinical Endocrinology & Metabolism 93(2): 380-386.	No age at presentation data
Nordenstrom, A., et al. (2010). "Sexual function and surgical outcome in women with congenital adrenal hyperplasia due to CYP21A2 deficiency: clinical perspective and the patients' perception." Journal of Clinical Endocrinology & Metabolism 95(8): 3633-3640.	No age at presentation data
Nordenstrom, A., et al. (2018). "Hormone therapy and patient satisfaction with treatment, in a large cohort of diverse disorders of sex development." Clinical Endocrinology 88(3): 397-408.	No age at presentation data
Nygren, U., et al. (2013). "Voice problems due to virilization in adult women with congenital adrenal hyperplasia due to 21-hydroxylase deficiency." Clinical Endocrinology 79(6): 859-866.	No age at presentation data
Nygren, U., et al. (2009). "Voice characteristics in women with congenital adrenal hyperplasia due to 21-hydroxylase deficiency." Clinical Endocrinology 70(1): 18-25.	Age at diagnosis data by voice group
Ocal, G., et al. (2015). "Clinical review of 95 patients with 46,xx disorders of sex development based on the new chicago classification." Journal of Pediatric and Adolescent Gynecology 28(1): 6-11.	No data for CAH

Odenwald, B., et al. (2016). "Children with classic congenital adrenal hyperplasia experience salt loss and hypoglycemia: evaluation of adrenal crises during the first 6 years of life." European Journal of Endocrinology 174(2): 177-186.	All diagnosed through newborn screening
Ohlsson Gotby, A., et al. (2015). "Congenital Adrenal Hyperplasia, Polycystic Ovary Syndrome and criminal behavior: A Swedish population based study." Psychiatry Research 229(3): 953-959.	No age at presentation data
Ortiz-Flores, A. E., et al. (2018). "Role of sampling times and serum cortisol cut-off concentrations on the routine assessment of adrenal function using the standard cosyntropin test in an academic hospital from Spain: a retrospective chart review." BMJ Open 8(5): e019273.	No age at presentation data
Oswiecimska, J. M., et al. (2012). "Androgens concentrations and second-to fourth-digit ratio (2D:4D) in girls with congenital adrenal hyperplasia (21-hydroxylase deficiency)." Neuroendocrinology Letters 33(8): 787-791.	No age at presentation data
Ozdemir, R., et al. (2017). "Assessment of early atherosclerosis and left ventricular dysfunction in children with 21-hydroxylase deficiency." Clinical Endocrinology 86(4): 473-479.	No age at presentation data
Paizoni, L., et al. (2020). "Effect of androgen excess and glucocorticoid exposure on metabolic risk profiles in patients with congenital adrenal hyperplasia due to 21-hydroxylase deficiency." Journal of Steroid Biochemistry & Molecular Biology 197: 105540.	No age at presentation data
Palmer, B. W., et al. (2012). "Total and partial urogenital mobilization: focus on urinary continence." Journal of Urology 187(4): 1422-1426.	No age at presentation data
Pasterski, V., et al. (2011). "Prenatal hormones and childhood sex segregation: playmate and play style preferences in girls with congenital adrenal hyperplasia." Hormones & Behavior 59(4): 549-555.	No age at presentation data
Pasterski, V., et al. (2015). "Increased Cross-Gender Identification Independent of Gender Role Behavior in Girls with Congenital Adrenal Hyperplasia: Results from a Standardized Assessment of 4- to 11-Year-Old Children." Archives of Sexual Behavior 44(5): 1363-1375.	No age at presentation data
Pierre, P., et al. (2012). "Adrenal rest tissue in gonads of patients with classical congenital adrenal hyperplasia: multicenter study of 45 French male patients." Annales d Endocrinologie 73(6): 515-522.	No age at presentation data
Quinkler, M., et al. (2015). "Modified-release hydrocortisone decreases BMI and HbA1c in patients with primary and secondary adrenal insufficiency." European Journal of Endocrinology 172(5): 619-626.	No age at presentation data
Rapp, M., et al. (2018). "Multicentre cross-sectional clinical evaluation study about quality of life in adults with disorders/differences of sex development (DSD) compared to country specific reference populations (dsd-LIFE)." Health and Quality of Life Outcomes 16(1).	No age at presentation data

Reifsnyder, J. E., et al. (2016). "Nerve Sparing Clitoroplasty is an Option for Adolescent and Adult Female Patients with Congenital Adrenal Hyperplasia and Clitoral Pain following Prior Clitoral Recession or Incomplete Reduction." Journal of Urology 195(4 Pt 2): 1270-1273.	No age at presentation data
Reisch, N., et al. (2009). "High prevalence of reduced fecundity in men with congenital adrenal hyperplasia." Journal of Clinical Endocrinology and Metabolism 94(5): 1665-1670.	No age at presentation data
Reisch, N., et al. (2011). "Quality of life is less impaired in adults with congenital adrenal hyperplasia because of 21-hydroxylase deficiency than in patients with primary adrenal insufficiency." Clinical Endocrinology 74(2): 166-173.	No age at presentation data
Reisch, N., et al. (2010). "Total adrenal volume but not testicular adrenal rest tumor volume is associated with hormonal control in patients with 21-hydroxylase deficiency." Journal of Clinical Endocrinology & Metabolism 95(5): 2065-2072.	No age at presentation data
Reisch, N., et al. (2012). "Frequency and causes of adrenal crises over lifetime in patients with 21-hydroxylase deficiency." European Journal of Endocrinology 167(1): 35-42.	No age at presentation data
Riehl, G., et al. (2020). "Bone mineral density and fractures in congenital adrenal hyperplasia: Findings from the dsd-LIFE study." Clinical Endocrinology 92(4): 284-294.	No age at presentation data
Sahmay, S., et al. (2014). "Diagnosis of late-onset congenital adrenal hyperplasia in clinical practice: current evaluation." Minerva Endocrinologica 39(3): 215-222.	Diagnosis was due to study
Salle, J. L., et al. (2012). "Surgical treatment of high urogenital sinuses using the anterior sagittal transrectal approach: a useful strategy to optimize exposure and outcomes." Journal of Urology 187(3): 1024-1031.	Not a study
Sarafoglou, K., et al. (2017). "Obesity in children with congenital adrenal hyperplasia in the Minnesota cohort: importance of adjusting body mass index for height-age." Clinical Endocrinology 86(5): 708-716.	No age at presentation data
Schoelwer, M. J., et al. (2017). "Infants with congenital adrenal hyperplasia are at risk for hypercalcemia, hypercalciuria, and nephrocalcinosis." Journal of the Endocrine Society 1(9): 1160-1167.	Diagnosis before 2 years was inclusion criteria
Schulz, J., et al. (2016). "Reduction in daily hydrocortisone dose improves bone health in primary adrenal insufficiency." European Journal of Endocrinology 174(4): 531-538.	Dissertation
Shammas, C., et al. (2016). "Genetic screening of non-classic CAH females with hyperandrogenemia identifies a novel CYP11B1 gene mutation." Hormones 15(2): 235-242.	No age at presentation data
Skordis, N., et al. (2011). "Endocrine profile and phenotype-genotype correlation in unrelated patients with non-classical congenital adrenal hyperplasia." Clinical Biochemistry 44(12): 959-963.	No age at presentation data
Skordis, N., et al. (2015). "Genetic defects of the CYP21A2 gene in girls with premature adrenarche." Journal of Endocrinological Investigation 38(5): 535-539.	Diagnosis was due to study

Spencer, D., et al. (2017). "Prenatal androgen exposure and children's aggressive behavior and activity level." Hormones & Behavior 96: 156-165.	No age at presentation data
Stewart, P. M., et al. (2016). "Exploring Inpatient Hospitalizations and Morbidity in Patients With Adrenal Insufficiency." Journal of Clinical Endocrinology & Metabolism 101(12): 4843-4850.	No age at presentation data
Stites, J., et al. (2017). "Urinary continence outcomes following vaginoplasty in patients with congenital adrenal hyperplasia." Journal of Pediatric Urology 13(1): 38.e31-38.e37.	No age at presentation data
Strandqvist, A., et al. (2014). "Suboptimal psychosocial outcomes in patients with congenital adrenal hyperplasia: epidemiological studies in a nonbiased national cohort in Sweden." Journal of Clinical Endocrinology & Metabolism 99(4): 1425-1432.	No age at presentation data
Sturm, R. M., et al. (2015). "Congenital adrenal hyperplasia: Current surgical management at academic medical centers in the United States." Journal of Urology Part S. 193(5): 1796-1801.	No age at presentation data
Swendiman, R. A., et al. (2020). "Histrelin implantation and growth outcomes in children with congenital adrenal hyperplasia: An institutional experience." Journal of the Endocrine Society 4(2).	No age at presentation data
Szymanski, K. M., et al. (2018). "What about my daughter's future? Parental concerns when considering female genital restoration surgery in girls with congenital adrenal hyperplasia." Journal of Pediatric Urology 14(5): 417.e411-417.e415.	No age at presentation data
Szymanski, K. M., et al. (2019). "Validation and Preliminary Results of the Parental Assessment of Children's External Genitalia Scale for Females (PACE-F) for Girls With Congenital Adrenal Hyperplasia." Urology 130: 132-137.	No age at presentation data
Takishima, S., et al. (2016). "Lower body weight and BMI at birth were associated with early adiposity rebound in 21-hydroxylase deficiency patients." Endocrine Journal 63(11): 983-990.	All patients identified through newborn screening
Thyen, U., et al. (2018). "Quality of health care in adolescents and adults with disorders/differences of sex development (DSD) in six European countries (dsd-LIFE)." BMC Health Services Research 18(1): 527.	No age at presentation data
Thyen, U., et al. (2014). "Utilization of health care services and satisfaction with care in adults affected by disorders of sex development (DSD)." Journal of General Internal Medicine 29 Suppl 3: S752-759.	No CAH data
Tica, S. S. and E. A. Eugster (2017). "How often are clinicians performing genital exams in children with disorders of sex development?" Journal of Pediatric Endocrinology & Metabolism 30(12): 1281-1284.	No age at presentation data
Toraman, B., et al. (2013). "Investigation of CYP21A2 mutations in Turkish patients with 21-hydroxylase deficiency and a novel founder mutation." Gene 513(1): 202-208.	No age at presentation data
Trakakis, E., et al. (2008). "Non classical congenital adrenal hyperplasia due to 21-hydroxylase deficiency in families from a Greek island with a closed society." Clinical & Experimental Obstetrics & Gynecology 35(4): 267-271.	One family only

Trakakis, E., et al. (2013). "Prevalence of non classical congenital adrenal hyperplasia due to 21-hydroxylase deficiency in Greek women with acne: a hospital-based cross-sectional study." Journal of the European Academy of Dermatology & Venereology 27(11): 1448-1451.	Diagnosis was due to study
Tresoldi, A. S., et al. (2020). "Increased Infection Risk in Addison's Disease and Congenital Adrenal Hyperplasia." Journal of Clinical Endocrinology and Metabolism 105(2).	No age at presentation data
Tugtepe, H., et al. (2014). "Does common channel length affect surgical choice in female congenital adrenal hyperplasia patients?" Journal of Pediatric Urology 10(5): 948-954.	No age at presentation data
Turan, I., et al. (2020). "21-Hydroxylase deficiency: Mutational spectrum and Genotype-Phenotype relations analyses by next-generation sequencing and multiplex ligation-dependent probe amplification." European Journal of Medical Genetics 63(4).	No age at presentation data
Turcu, A. F., et al. (2017). "11-Oxygenated Androgens Are Biomarkers of Adrenal Volume and Testicular Adrenal Rest Tumors in 21-Hydroxylase Deficiency." Journal of Clinical Endocrinology & Metabolism 102(8): 2701-2710.	No age at presentation data
Turcu, A. F., et al. (2015). "Profiles of 21-Carbon Steroids in 21-hydroxylase Deficiency." Journal of Clinical Endocrinology & Metabolism 100(6): 2283-2290.	No age at presentation data
Unluhizarci, K., et al. (2010). "The prevalence of non-classic adrenal hyperplasia among Turkish women with hyperandrogenism." Gynecological Endocrinology 26(2): 139-143.	No age at presentation data
Van De Grift, T. C. and B. P. C. Kreukels (2019). "Breast development and satisfaction in women with disorders/differences of sex development." Human Reproduction 34(12): 2410-2417.	No age at presentation data
van der Zwan, Y. G., et al. (2013). "Severity of virilization is associated with cosmetic appearance and sexual function in women with congenital adrenal hyperplasia: a cross-sectional study." Journal of Sexual Medicine 10(3): 866-875.	No age at presentation data
Vanderbrink, B. A., et al. (2010). "Does preoperative genitography in congenital adrenal hyperplasia cases affect surgical approach to feminizing genitoplasty?" Journal of Urology 184(4 Suppl): 1793-1798.	No age at presentation data
Veerapandiyan, A., et al. (2011). "Chromosome 22q11.2 deletion syndrome in African-American patients: a diagnostic challenge." American Journal of Medical Genetics. Part A 155A(9): 2186-2195.	Not CAH
Verma, S., et al. (2010). "A pharmacokinetic and pharmacodynamic study of delayed- and extended-release hydrocortisone (Chronocort) vs. conventional hydrocortisone (Cortef) in the treatment of congenital adrenal hyperplasia." Clinical Endocrinology 72(4): 441-447.	No age at presentation data
Vidmar, A. P., et al. (2018). "Improved medical-alert ID ownership and utilization in youth with congenital adrenal hyperplasia following a parent educational intervention." Journal of Pediatric Endocrinology & Metabolism 31(2): 213-219.	No age at presentation data

Volkl, T. M., et al. (2011). "Adrenarche and puberty in children with classic congenital adrenal hyperplasia due to 21-hydroxylase deficiency." Hormone Research in Paediatrics 76(6): 400-410.	No age at presentation data
Volkl, T. M., et al. (2011). "IGF-I-IGFBP-3-acid-labile subunit (ALS) complex in children and adolescents with classical congenital adrenal hyperplasia due to 21-hydroxylase deficiency (CAH)." Growth Hormone & Igf Research 21(4): 191-198.	No age at presentation data
Volkl, T. M., et al. (2009). "Adiponectin levels are high in children with classic congenital adrenal hyperplasia (CAH) due to 21-hydroxylase deficiency." Acta Paediatrica 98(5): 885-891.	No age at presentation data
Volkl, T. M., et al. (2009). "Does an altered leptin axis play a role in obesity among children and adolescents with classical congenital adrenal hyperplasia due to 21-hydroxylase deficiency?" European Journal of Endocrinology 160(2): 239-247.	No age at presentation data
Volkl, T. M. K., et al. (2011). "IGF-I-IGFBP-3-acid-labile subunit (ALS) complex in children and adolescents with classical congenital adrenal hyperplasia due to 21-hydroxylase deficiency (CAH)." Growth Hormone and IGF Research 21(4): 191-198.	No age at presentation data
Volkl, T. M. K., et al. (2009). "Adiponectin levels are high in children with classic congenital adrenal hyperplasia (CAH) due to 21-hydroxylase deficiency." Acta Paediatrica, International Journal of Paediatrics 98(5): 885-891.	No age at presentation data
Wallensteen, L., et al. (2018). "Evaluation of behavioral problems after prenatal dexamethasone treatment in Swedish children and adolescents at risk of congenital adrenal hyperplasia." Hormones & Behavior 98: 219-224.	No age at presentation data
Wang, L. C. and D. P. Poppas (2017). "Surgical outcomes and complications of reconstructive surgery in the female congenital adrenal hyperplasia patient: What every endocrinologist should know." Journal of Steroid Biochemistry & Molecular Biology 165(Pt A): 137-144.	Review of surgery
Wasniewska, M., et al. (2013). "Increased large artery intima media thickness in adolescents with either classical or non-classical congenital adrenal hyperplasia." Journal of Endocrinological Investigation 36(1): 12-15.	No age at presentation data
Webb, E. A., et al. (2018). "Quantitative Brain MRI in Congenital Adrenal Hyperplasia: In Vivo Assessment of the Cognitive and Structural Impact of Steroid Hormones." Journal of Clinical Endocrinology & Metabolism 103(4): 1330-1341.	No age at presentation data
Wild, S. H., et al. (2010). "Health status of adults with congenital adrenal hyperplasia: A cohort study of 203 patients." Journal of Clinical Endocrinology and Metabolism 95(11): 5110-5121.	CaHASE study but no age at presentation data
Wolffenbuttel, K. P., et al. (2017). "Clitoral hoodplasty in females with disorders of sex development." Journal of Pediatric Urology 13(1): 61.e61-61.e65.	No age at presentation data
Yang, M. and P. C. White (2017). "Risk factors for hospitalization of children with congenital adrenal hyperplasia." Clinical Endocrinology 86(5): 669-673.	No age at presentation data

Yilmaz, R., et al. (2017). "Sonography and magnetic resonance imaging characteristics of testicular adrenal rest tumors." Polish Journal of Radiology 82: 583-588.	No age at presentation data
Zopf, K., et al. (2017). "BclI polymorphism of the glucocorticoid receptor and adrenal crisis in primary adrenal insufficiency." Endocrine Connections 6(8): 685-691.	No age at presentation data
	Studies (with country) excluded due to non-child population
Han, T. S., et al. (2014). "Relationship between final height and health outcomes in adults with congenital adrenal hyperplasia: United Kingdom congenital adrenal hyperplasia adult study executive (CaHASE)." Journal of Clinical Endocrinology and Metabolism 99(8): E1547-E1555.	UK.
Abe, Y., et al. (2016). "Manifestations and characteristics of congenital adrenal hyperplasia-associated encephalopathy." Brain & Development 38(7): 638-647.	Japan
Anastasovska, V., et al. (2014). "Direct molecular diagnosis of CYP21A2 point mutations in Macedonian and Serbian patients with 21-hydroxylase deficiency." Journal of Medical Biochemistry 34(1): 52-57.	Serbia and Macedonia
Aycan, Z., et al. (2013). "Prevalence and long-term follow-up outcomes of testicular adrenal rest tumours in children and adolescent males with congenital adrenal hyperplasia." Clinical Endocrinology 78(5): 667-672.	Turkey
Bas, F., et al. (2018). "Prevalence, clinical characteristics and long-term outcomes of classical 11 beta-hydroxylase deficiency (11BOHD) in Turkish population and novel mutations in CYP11B1 gene." Journal of Steroid Biochemistry & Molecular Biology 181: 88-97.	Turkey
Bechtold, S., et al. (2014). "Sexual difference in bone geometry of adult patients with classical congenital adrenal hyperplasia: data using peripheral quantitative computed tomography." Hormone Research in Paediatrics 82(3): 171-178.	Germany
Bello, R., et al. (2017). "Basal 17-hydroxyprogesterone cannot accurately predict nonclassical congenital adrenal hyperplasia in children and adolescents." Acta Paediatrica 106(1): 155-160.	Israel
Berglund, A., et al. (2017). "Incidence, prevalence, diagnostic delay, morbidity, mortality and socioeconomic status in males with 46,XX disorders of sex development: A nationwide study." Human Reproduction 32(8): 1751-1760.	Denmark
Bidet, M., et al. (2010). "Fertility in women with nonclassical congenital adrenal hyperplasia due to 21-hydroxylase deficiency." Journal of Clinical Endocrinology & Metabolism 95(3): 1182-1190.	France
Bidet, M., et al. (2009). "Clinical and molecular characterization of a cohort of 161 unrelated women with nonclassical congenital adrenal hyperplasia due to 21-hydroxylase deficiency and 330 family members." Journal of Clinical Endocrinology & Metabolism 94(5): 1570-1578.	France

Bodian, D. L., et al. (2016). "Utility of whole-genome sequencing for detection of newborn screening disorders in a population cohort of 1,696 neonates." Genetics in Medicine 18(3): 221-230.	USA
Bonfig, W. and H. P. Schwarz (2011). "Growth pattern of untreated boys with simple virilizing congenital adrenal hyperplasia indicates relative androgen insensitivity during the first six months of life." Hormone Research in Paediatrics 75(4): 264-268.	Germany
Breil, T., et al. (2019). "Typical characteristics of children with congenital adrenal hyperplasia due to 11beta-hydroxylase deficiency: a single-centre experience and review of the literature." Journal of Pediatric Endocrinology & Metabolism 32(3): 259-267.	Germany
Chaudhari, M., et al. (2018). "Testicular adrenal rest tumor screening and fertility counseling among males with congenital adrenal hyperplasia." Journal of Pediatric Urology 14(2): 155.e151-155.e156.	USA
Dangle, P. P., et al. (2017). "Surgical Complications Following Early Genitourinary Reconstructive Surgery for Congenital Adrenal Hyperplasia-Interim Analysis at 6 Years." Urology 101: 111-115.	USA
de Vries, L., et al. (2019). "Obesity and Cardiometabolic Risk Factors in Children and Young Adults With Non-classical 21-Hydroxylase Deficiency." Frontiers in Endocrinology 10 (no pagination).	Israel
Dumic, M., et al. (2017). "Testicular adrenal rest tumors in congenital adrenal hyperplasia-cross-sectional study of 51 Croatian male patients." European Journal of Pediatrics 176(10): 1393-1404.	Croatia
Dundar, A., et al. (2019). "The molecular basis and genotype-phenotype correlations of congenital adrenal hyperplasia (CAH) in Anatolian population." Molecular Biology Reports 46(4): 3677-3690.	Anatolia
El-Maouche, D., et al. (2018). "Longitudinal Assessment of Illnesses, Stress Dosing, and Illness Sequelae in Patients With Congenital Adrenal Hyperplasia." Journal of Clinical Endocrinology & Metabolism 103(6): 2336-2345.	USA
Eyal, O., et al. (2013). "Adult height of subjects with nonclassical 21-hydroxylase deficiency." Acta Paediatrica 102(4): 419-423.	Israel
Falhammar, H., et al. (2018). "Health status in 1040 adults with disorders of sex development (DSD): A European multicenter study." Endocrine Connections 7(3): 466-478.	Europe
Falhammar, H., et al. (2011). "Cardiovascular risk, metabolic profile, and body composition in adult males with congenital adrenal hyperplasia due to 21-hydroxylase deficiency." European Journal of Endocrinology 164(2): 285-293.	Sweden
Halper, A., et al. (2019). "Use of an aromatase inhibitor in children with congenital adrenal hyperplasia: Impact of anastrozole on bone mineral density and visceral adipose tissue." Clinical Endocrinology 91(1): 124-130.	USA
Harrington, J., et al. (2012). "Adolescents with congenital adrenal hyperplasia because of 21-hydroxylase deficiency have vascular dysfunction." Clinical Endocrinology 76(6): 837-842.	Australia

Khattab, A., et al. (2017). "Clinical, genetic, and structural basis of congenital adrenal hyperplasia due to 11beta-hydroxylase deficiency." Proceedings of the National Academy of Sciences of the United States of America 114(10): E1933-E1940.	Multi national
Kirli, E. A., et al. (2013). "An unexpected diagnosis in children with male phenotype and bilateral nonpalpable gonad: congenital adrenal hyperplasia with female genotype." Pediatric Surgery International 29(7): 719-724.	Turkey
Kocova, M., et al. (2018). "Testicular adrenal rest tumors in boys with 21-hydroxylase deficiency, timely diagnosis and follow-up." Endocrine Connections 7(4): 544-552.	Macedonia
Neeman, B., et al. (2019). "Central Precocious Puberty as a Presenting Sign of Nonclassical Congenital Adrenal Hyperplasia: Clinical Characteristics." Journal of Clinical Endocrinology & Metabolism 104(7): 2695-2700.	Israel
Neocleous, V., et al. (2019). "The Spectrum of Genetic Defects in Congenital Adrenal Hyperplasia in the Population of Cyprus: A Retrospective Analysis." Hormone & Metabolic Research 51(9): 586-594.	Cyprus
Neocleous, V., et al. (2014). "Phenotypic variability of hyperandrogenemia in females heterozygous for CYP21A2 mutations." Indian Journal of Endocrinology and Metabolism 18(Supplement 1): S72-S79.	Cyprus
Nygren, U., et al. (2009). "Voice characteristics in women with congenital adrenal hyperplasia due to 21-hydroxylase deficiency." Clinical Endocrinology 70(1): 18-25.	Sweden
Reisch, N., et al. (2013). "Testicular adrenal rest tumors develop independently of long-term disease control: a longitudinal analysis of 50 adult men with congenital adrenal hyperplasia due to classic 21-hydroxylase deficiency." Journal of Clinical Endocrinology & Metabolism 98(11): E1820-1826.	Germany
Sanches, S. A., et al. (2012). "Physical, social and societal functioning of children with congenital adrenal hyperplasia (CAH) and their parents, in a Dutch population." International Journal of Pediatric Endocrinology 2012(1).	Netherlands
Sarafoglou, K., et al. (2014). "Impact of hydrocortisone on adult height in congenital adrenal hyperplasia-the Minnesota cohort." Journal of Pediatrics 164(5): 1141-1146.e1141.	USA
Schernthaner-Reiter, M. H., et al. (2019). "Influence of Genotype and Hyperandrogenism on Sexual Function in Women With Congenital Adrenal Hyperplasia." Journal of Sexual Medicine 16(10): 1529-1540.	Austria
van de Grift, T. C., et al. (2018). "Body image and self-esteem in disorders of sex development: A European multicenter study." Health Psychology 37(4): 334-343.	Europe - DsD Life study
Zimmermann, A., et al. (2009). "Bone mineral density and bone turnover in Romanian children and young adults with classical 21-hydroxylase deficiency are influenced by glucocorticoid replacement therapy." Clinical Endocrinology 71(4): 477-484.	Germany

Table 18. Publications excluded from Q3

Reference	Reason for exclusion
	Studies excluded as prior to 2015
Cavarzere, P., et al. (2009). "Transient hyper-17-hydroxyprogesteronemia: a clinical subgroup of patients diagnosed at neonatal screening for congenital adrenal hyperplasia." European Journal of Endocrinology 161(2): 285-292.	pre 2015
De Jesus, V. R., et al. (2010). "Pilot proficiency testing study for second tier congenital adrenal hyperplasia newborn screening." Clinica Chimica Acta 411(21-22): 1684-1687.	pre 2015
Gleeson, H. K., et al. (2008). "Two-year pilot study of newborn screening for congenital adrenal hyperplasia in New South Wales compared with nationwide case surveillance in Australia." Journal of Paediatrics & Child Health 44(10): 554-559.	pre 2015
Holtkamp, U., et al. (2008). "EDTA in dried blood spots leads to false results in neonatal endocrinologic screening." Clinical Chemistry 54(3): 602-605.	pre 2015
Huidobro Fernandez, B., et al. (2011). "Neonatal screening for congenital adrenal hyperplasia: transitory elevation of 17-hydroxyprogesterone." Journal of Pediatric Endocrinology & Metabolism 24(3-4): 155-162.	pre 2015
Kamrath, C., et al. (2014). "The balance of cortisol-cortisone interconversion is shifted towards cortisol in neonates with congenital adrenal hyperplasia due to 21-hydroxylase deficiency." Journal of Steroid Biochemistry & Molecular Biology 143: 386-391.	pre 2015
Koyama, Y., et al. (2012). "Two-step biochemical differential diagnosis of classic 21-hydroxylase deficiency and cytochrome P450 oxidoreductase deficiency in Japanese infants by GC-MS measurement of urinary pregnanetriolone/ tetrahydroxycortisone ratio and 11beta-hydroxyandrosterone." Clinical Chemistry 58(4): 741-747.	pre 2015
Magnisali, P., et al. (2011). "Simultaneous quantification of 17alpha-OH progesterone, 11-deoxycortisol, DELTA4-androstenedione, cortisol and cortisone in newborn blood spots using liquid chromatography-tandem mass spectrometry." Journal of Chromatography B: Analytical Technologies in the Biomedical & Life Sciences 879(19): 1565-1572.	pre 2015
Malikova, J., et al. (2012). "Genetic analysis of the CYP21A2 gene in neonatal dried blood spots from children with transiently elevated 17-hydroxyprogesterone." Clinical Endocrinology 77(2): 187-194.	pre 2015
Nemeth, S., et al. (2012). "Reverse-hybridization assay for rapid detection of common CYP21A2 mutations in dried blood spots from newborns with elevated 17-OH progesterone." Clinica Chimica Acta 414: 211-214.	pre 2015
Pauwels, G., et al. (2012). "Risk factors for elevated levels of 17-hydroxyprogesterone during neonatal intensive care unit admission." Acta Clinica Belgica 67(2): 88-93.	pre 2015
Rossi, C., et al. (2010). "Serum steroid profiling for congenital adrenal hyperplasia using liquid chromatography-tandem mass spectrometry." Clinica Chimica Acta 411(3-4): 222-228.	pre 2015

Ryckman, K. K., et al. (2012). "Replication of clinical associations with 17-hydroxyprogesterone in preterm newborns." Journal of Pediatric Endocrinology & Metabolism 25(3-4): 301-305.	pre 2015
Sarafoglou, K., et al. (2014). "Comparison of newborn screening protocols for congenital adrenal hyperplasia in preterm infants." Journal of Pediatrics 164(5): 1136-1140.	pre 2015
Sarafoglou, K., et al. (2011). "Comparison of multiple steroid concentrations in serum and dried blood spots throughout the day of patients with congenital adrenal hyperplasia." Hormone Research in Paediatrics 75(1): 19-25.	pre 2015
Sarafoglou, K., et al. (2012). "Molecular testing in congenital adrenal hyperplasia due to 21alpha-hydroxylase deficiency in the era of newborn screening." Clinical Genetics 82(1): 64-70.	pre 2015
Schwarz, E., et al. (2009). "Use of steroid profiling by UPLC-MS/MS as a second tier test in newborn screening for congenital adrenal hyperplasia: the Utah experience." Pediatric Research 66(2): 230-235.	pre 2015
Slaughter, J. L., et al. (2010). "The effects of gestational age and birth weight on false-positive newborn-screening rates." Pediatrics 126(5): 910-916.	pre 2015
Sorensen, K. M., et al. (2008). "Multiplex ligation-dependent probe amplification technique for copy number analysis on small amounts of DNA material." Analytical Chemistry 80(23): 9363-9368.	pre 2015
	Studies excluded for non-comparable country setting
Anandi, V. S. and B. Shaila (2017). "Evaluation of factors associated with elevated newborn 17-hydroxyprogesterone levels." Journal of Pediatric Endocrinology & Metabolism 30(6): 677-681.	India
Barra, C. B., et al. (2012). "Neonatal screening for congenital adrenal hyperplasia." Revista Da Associacao Medica Brasileira 58(4): 459-464.	Brasil
Choi, R., et al. (2019). "Dried Blood Spot Multiplexed Steroid Profiling Using Liquid Chromatography Tandem Mass Spectrometry in Korean Neonates." Annals of Laboratory Medicine 39(3): 263-270.	Korea
Gruneiro-Papendieck, L., et al. (2008). "Neonatal screening for congenital adrenal hyperplasia: Experience and results in Argentina." Journal of Pediatric Endocrinology and Metabolism 21(1): 73-78.	Argentina
Hayashi, G. Y., et al. (2017). "Neonatal 17-hydroxyprogesterone levels adjusted according to age at sample collection and birthweight improve the efficacy of congenital adrenal hyperplasia newborn screening." Clinical Endocrinology 86(4): 480-487.	Brazil
Jiang, X., et al. (2019). "The adjustment of 17-hydroxyprogesterone cut-off values for congenital adrenal hyperplasia neonatal screening by GSP according to gestational age and age at sampling." Journal of Pediatric Endocrinology & Metabolism 32(11): 1253-1258.	China

Kaur, G., et al. (2010). "Preliminary report on neonatal screening for congenital hypothyroidism, congenital adrenal hyperplasia and glucose-6-phosphate dehydrogenase deficiency: a Chandigarh experience." Indian Journal of Pediatrics 77(9): 969-973.	India
Marino, S., et al. (2020). "Molecular analysis of the CYP21A2 gene in dried blood spot samples." Medicina 80(3): 197-202.	Argentina
Silveira, E. L., et al. (2009). "Molecular analysis of CYP21A2 can optimize the follow-up of positive results in newborn screening for congenital adrenal hyperplasia." Clinical Genetics 76(6): 503-510.	Brasil
Yeung, M. C. W., et al. (2020). "Clinical utility of second-tier testing in newborn screening for congenital adrenal hyperplasia: The Hong Kong experience." Hong Kong Journal of Paediatrics 25(1): 3-7.	Hong Kong
Zhong, K., et al. (2015). "Neonatal screening external quality assessment in China, 2014." Journal of Medical Screening 22(4): 175-181.	China
Zhong, K., et al. (2016). "The status of neonatal screening in China, 2013." Journal of Medical Screening 23(2): 59-61.	China
	Studies excluded for other reasons
Bodian, D. L., et al. (2016). "Utility of whole-genome sequencing for detection of newborn screening disorders in a population cohort of 1,696 neonates." Genetics in Medicine 18(3): 221-230.	Only one case of CAH identified.
Cavarzere, P., et al. (2018). "Children with premature pubarche: is an alterated neonatal 17-Ohp screening test a predictive factor?" Italian Journal of Pediatrics 44(1): 10.	Insufficient data
Faurschou, S., et al. (2015). "Hormonal disturbances due to severe and mild forms of congenital adrenal hyperplasia are already detectable in neonatal life." Acta Paediatrica 104(2): e57-62.	Not DBS
Kamrath, C., et al. (2016). "Diagnosis of 21-hydroxylase deficiency by urinary metabolite ratios using gas chromatography-mass spectrometry analysis: Reference values for neonates and infants." Journal of Steroid Biochemistry & Molecular Biology 156: 10-16.	Urinary spots not blood spots

Lao, Q., et al. (2019). "High-Throughput Screening for CYP21A1P-TNXA/TNXB Chimeric Genes Responsible for Ehlers-Danlos Syndrome in Patients with Congenital Adrenal Hyperplasia." Journal of Molecular Diagnostics 21(5): 924-931.	Target condition is CAH-X
Lund, A., et al. (2020). "Danish expanded newborn screening is a successful preventive public health programme." Danish Medical Journal 67(1).	Doesn't report CAH separately
Odenwald, B., et al. (2015). "Classic Congenital Adrenal Hyperplasia due to 21-Hydroxylase-Deficiency: 13 Years of Neonatal Screening and Follow-up in Bavaria." Klinische Padiatrie 227(5): 278-283.	German language
Tieh, P. Y., et al. (2017). "Utility of a precursor-to-product ratio in the evaluation of presumptive positives in newborn screening of congenital adrenal hyperplasia." Journal of Perinatology 37(3): 283-287.	Insufficient data

Appendix 3 — Summary and appraisal of individual studies

Data Extraction

Table 19. Studies relevant to criterion 1: Q1 (incidence)

Study Reference	David 2018(27)
Study Design	Design Retrospective cohort study Objective To analyse the birth frequency of diseases detected by neonatal screening (NBS) in normal and low birth weight LBW neonates in the Czech Republic Dates 2006–2016 Country Czech Republic
Population characteristics	Screening method Collection- Dried blood spots were collected from the heel pricks of neonates and mailed to laboratories. January 2002–January 2006 blood spots were taken between the 120th and 168th hour of neonatal life, February 2006–September 2009 between the 72nd and 96th hour, October 2009–December 2016 between the 48th and 72th hour. Analysis - For detection of CAH (from February 2006 to 2016), 17-hydroxyprogesterone (17-OHP) was determined using fluorescence immunoassay (Delfia® a AutoDelfia® produced by Perkin-Elmer, Waltham, MA, USA) Diagnosis - Neonates with positive NBS findings were referred for follow-up to appropriate clinical departments. In the studied

	period, all screening positive patients underwent confirmatory testing. Basal level of 17-hydroxyprogesterone (17-OHP) above reference range and/or positive cosyntropin test and casual mutation in CYP21A2 gene. Cut-offs according to birthweight and sample timing. Duration 10 years Sample size 1,196,387 Age at sampling January 2002–January 2006: 120th and 168th hour of neonatal life; February 2006–September 2009: 72nd and 96th hour, and; October 2009–December 2016: 48th and 72th hour of neonatal life
Outcomes	Cases identified 100 Incidence all CAH 1:11,964 (calculated) Low birthweight neonates: 8/89,911

Study Reference	**David 2019(28) (NB crossover with David 2018)**
Study Design	Design Retrospective cohort study Objective To analyse the epidemiology of Rare Diseases screened by NBS in the Czech Republic

	Dates 2010 to 2017 Country Czech Republic
Population characteristics	Screening method Collection - DBS on filter paper collected from heel pricks and sent to specified laboratories by mail. DBS were taken between the 48th –72th hour of newborn's life. Analysis - For detection of CAH (from February 2006 to 2016), 17-hydroxyprogesterone (17-OHP) was determined using fluorescence immunoassay (Delfia® a AutoDelfia® produced by Perkin-Elmer, Waltham, MA, USA) Diagnosis - Newborns with positive NBS findings were referred for follow-up to appropriate clinical centres to confirm the diagnosis using generally accepted diagnostic standards. Decision limit (capillary blood): 17-OHP according to birthweight/gestational age, range 20.0–160 nmol/L, example: 20.0 nmol/L for ≥ 2700 g (≥ 37 gestational week). Confirmatory test criteria (venous blood): Basal level of 17-hydroxyprogesterone (17-OHP) above reference range and/or positive cosyntropin test and casual mutation in CYP21A2 gene. Duration 7 years Sample size 888,891
Outcomes	Cases identified 71 Incidence all CAH 1: 12,520 FPs: 3696 PPV 0.02

Study Reference	Eshragh 2020(22)
Study Design	Design Cohort study Objective To determine the accuracy of screening, severity of CAH, and biochemical and clinical outcomes of cases detected by the Northwest Regional Newborn Screening Program Dates 2003–2017 Country USA - Newborns screened in 5 states (Oregon, Idaho, Alaska, Nevada and New Mexico) and some military bases in Washington and the South Pacific)
Population characteristics	Screening method Measurement of 17-OHP in dried blood spot specimen. Two stage NBS - first screening performed between 24–48h; second between 10 and 14 days of life. Duration 14 years Sample size 2,212,550 Sex 87M:77F (calculated) Ethnicity N/R
Outcomes	Cases identified 164 Incidence all CAH

	1:13,491 (first screen cases detected, n=126; second screen, n=41 not possible to calculate incidence as no details of denominator for each screening level) Incidence classic salt-wasting 1:24,049 (92/2212550) (numerator calculated and denominator assumed) Incidence simple virilising 1:30,729 (72/2212550) (numerator calculated and denominator assumed) Out of 164 cases of CAH, 25% were detected on the second screen.

Study Reference	**Fox 2020(3)** (extraction table covers both Q1 and Q2 data)
Study Design	Design Ambispective cohort study. Data collected retrospectively (unscreened cohort) and prospectively (screened cohort). Objective To evaluate the clinical impact of a congenital adrenal hyperplasia (CAH) newborn screening program and incremental costs relative to benefits in screened vs unscreened infants Dates 1988-2008 (pre-screening); 2010 to 2018 (screened cohort) Country Canada Setting British Columbia Children's Hospital, a tertiary care hospital Screening protocol Blood spot card, ideally collected between 24 and 48 hours of age, unless discharge <24 hours then test before discharge and repeat sample before 2 weeks of age. All NBS samples are shipped to BC Children's Hospital for review within 24 hours of collection, for receipt by the laboratory within 72 hours of collection.

	Results before target 8 days of age. For CAH, this should allow for diagnosis before a salt wasting crisis develops. 2-tiered method for CAH screening - first tier screening test measures 17OHP using the AutoDELFIA Neo17OHP immunoassay (Perkin Elmer Canada Inc, Woodbridge, Ontario). Cut-offs for advancement to the second-tier test are based on age at sample collection (<72 hours or >72 hours) and birth weight (<1500 g, 1500-2500 g, >2500 g). Second-tier test using the same NBS card measures 17OHP, cortisol, androstenedione, 11- deoxycortisol, and 21-deoxycortisol via tandem mass spectrometry. Cut-off values for a positive screen over the majority of the study period were based on the algorithm suggested by Janzen et al in 2007, using absolute 17OHP levels, and the ratio of (17OHP + 21-deoxycortisol)/cortisol. In April 2017, the ratio was changed to 17OHP/11-deoxycortisol to reduce false positives from premature infants with nonspecific low cortisol levels.
Population characteristics	Inclusion criteria Infants diagnosed with CAH at British Columbia Children's Hospital, Vancouver Diagnosis Pre-screened cohort - age at diagnosis was defined as the age at the time of the first endocrine consult or telephone call where medical advice was provided. In the screened cohort, the time from birth to reporting of the positive screen was recorded. Duration 20 years Sample size N=57 Screened cohort n=17; unscreened cohort n=40 Sex Male sex: 8/17 (screened cohort) 21/40 (unscreened cohort) Ethnicity NR

Outcomes	<u>Median age at diagnosis</u> Screened cohort Unscreened cohort Reason for suspected CAH: Ambiguous genitalia 6 (35%) 19 (48%) Family history 1 (6%) 9 (2%) Salt wasting crisis 0 (0) 10 (25%) Positive CAH screen 10 (59%) Age at time of positive screen (days) 6 (0-13) Age at diagnosis (days) 6 (0-13) 5 (0-30) Age at diagnosis for males (days) 5.5 (0-13) 14 (1-30) There were no salt wasting crises in the screened cohort (as defined by a need for resuscitation), compared with 25% of subjects in the unscreened cohort

Study Reference	**Heather 2015(26)**
Study Design	<u>Design</u> Retrospective cohort study <u>Objective</u> To evaluate the efficacy of national newborn screening for severe congenital adrenal hyperplasia (CAH) in New Zealand over the past 20 years <u>Dates</u> 1994–2013 Diagnostic accuracy analysis: 2011–2013 <u>Country</u> New Zealand
Population characteristics	<u>Screening method</u>

	Initially based on estimation of 17-hydroxyprogesterone (17-OHP) by an in-house radioimmunoassay (RIA) then changed to the Delfia immunoassay in 1998. Assays carried out on whole-blood samples collected on Guthrie cards after 48 hours of life. Samples posted to the laboratory. After the measurement of 17-OHP, samples with values more than 2 standard deviation (SD) above the assay mean are re-assayed after diethyl ether extraction. Positive screen indicated by 17-OHP concentrations above 23 nmol/L for babies with a birth weight above 1500 g, and 32 nmol/L for those less than 1500 g. Second whole-blood sample requested if cut-off met. The intraassay and interassay coefficients of variation for a whole-blood 17-OHP value of 23 nmol/L are 10%. Sample size 1,175,973 Diagnostic accuracy analysis: 372 Ethnicity -New Zealand European: 32 -Maori: 3 -Pacific Islander: 7 -Other: 2
Outcomes	Cases identified 44 (28 females, 16 males) Incidence all CAH 1:26 727
Diagnostic test characteristics	Tier of test: First tier and second-tier tests Day of test: After 48 hours of life Test: Fluoroimmunoassay of 17-OHP as first tier test Cut-off: A positive screening result is indicated by 17-OHP concentrations above 23 nmol/L for babies with a birth weight above 1500 g, and 32 nmol/L for those less than 1500 g. These cut-offs are used for request of a second whole-blood sample. In addition, for values of greater than 50 and 100 nmol/L in babies with birth weights of greater than 1500 g and less than 1500 g, respectively, the laboratory directly informs a paediatrician or paediatric endocrinologist, who arranges for same-day review of the baby.

Reference standard	Data on newborn screening for CAH were obtained from Newborn Metabolic Screening Programme (NMSP) records for the time period 1994–2013. In addition to screening results, the NMSP collects clinical data on babies diagnosed with CAH in the neonatal period by means of standardized forms sent out to paediatric endocrinologists and paediatricians. Cases were defined as clinically detected if the diagnosis was suspected prior to screening results being available. CAH was classified as simple virilising if pre-treatment serum electrolytes from day 7 onward did not reveal hyponatremia. Type of CAH: defined as “severe” cases
Diagnostic test outcomes ***reviewer calculated**	

TP	**FP**	**FN**	**TN**
4	364	0	NR
Sensitivity	**Specificity**	**PPV**	**NPV**
100%	99.8%	1.08%	100%*

Study Reference	**Held 2015(21)**
Study Design	Design Retrospective cohort study Objective To examine the effectiveness of the routine second screen for CAH by evaluating laboratory practices along with biochemical and medical characteristics of CAH cases (1) detected in the one-screen compared to two-screen states, and (2) detected on the first versus the second screen in the two-screen states Dates 2003–2011 Country USA

Population characteristics	Screening method 17-OHP quantified as the analytic marker for CAH using a dissociation-enhanced lanthanide fluoroimmunoassay. A fixed cut-off based on birth weight was used to identify newborns at risk for CAH in both of the one-screen states and in 4 of the two-screen states. One of the two-screen states used a floating cut-off that was determined daily based on a percent from the mean 17-OHP value obtained on the normal population and on low BW newborns. Depending upon the 17-OHP concentration, states either recommended repeating the newborn screen (by collecting the second specimen) or performing confirmatory testing and a clinical assessment. Timing: First screen, NR; second screen between 8 and 14 days Sample size 4,370,213
Outcomes	Cases identified 374 (first screen: 253; second screen: 99; targeted second screen: 5; Unknown if detected on first screen due to unsatisfactory initial sample: 7; not detected by NBS: 10) Incidence all CAH 1:11,685 Incidence classic CAH 1: 20,421 (214/4370213) (assumed) Incidence classic salt-wasting 1: 74,071 (59/4370213) (assumed) Incidence simple virilising 1: 49,103 (89/4370213) (assumed)

Study Reference	**Iniguez 2019(29)**

Study Design	Design Cohort study Objective Dates 1990–2016 Country Spain
Population characteristics	Screening method Newborn screening programme in 6 Autonomous Communities covering 29.8% of all newborn infants (measurement of 17-OHP from dried blood spot [assumed]) Sample size 3,086,015
Outcomes	Cases identified 142 Incidence all CAH 1:21,732

Study Reference	**Khalid 2012(18)** (extraction table covers both Q1 and Q2 data)
Study Design	Design Active surveillance of newly diagnosed CAH through the British Paediatric Surveillance Unit Objective To estimate the incidence of clinically diagnosed CAH in the UK. To evaluate the clinical features and time to presentation of those presenting clinically within the first month of life in order to assess the potential benefit of pre-symptomatic detection of CAH Dates August 2007 to August 2009

	Country UK
Population characteristics	Recruitment Paediatricians taking part in BPSU notified any child under the age of 16 years who presented in the preceding month with clinical features of CAH and elevated 17-OHP. To maximise case ascertainment, children with CAH were also notified from August 2007 to January 2009 through a laboratory surveillance scheme comprising 12 UK laboratories which measure 17-OHP in children (Cambridge, Leeds, Liverpool, London, 5 Manchester, Scotland, Southampton, Wales) and matched to BPSU reports. Non-matches identified through laboratory surveillance were followed-up. Diagnosis All notified children were reviewed by an expert diagnostic review panel and categorised as 'definite' or 'not' CAH based on karyotype, presence or absence of salt-wasting, virilisation, and biochemical and molecular genetic test results Duration 25 months Sample size 144 unscreened children diagnosed with CAH (132 with 17-hydroxylase deficiency) of which 77 diagnosed at <1 month
Outcomes	Incidence of new diagnoses of CAH in children under the age of 16 years estimated to be 0.60 (95% CI 0.50 to 0.71) per 100 000 in Great Britain (1:166,666) Incidence of diagnosis at <1 year of age 5.48 (95% CI 4.42 to 6.81) per 100,000 (1:18,248) Median age at presentation (all CAH) (days, Inter Quartile Range (IQR): 1 (0–14) Median age at presentation (all CAH) (days, IQR): Males (n=33)14 (9–18); Females (n=44) 0 (0–1)

	Median age at presentation of salt-wasting crisis (days, IQR): Males (n=24) 15 (11–20); Females (n=3) 15 (13–16) Total (n=27) 15 (11-19) range 9 to 30 days Sex Boys n=62 (43%) Girls n=82 (57%) Ethnicity Children of Asian ethnicity 36/144 (25%) all CAH and 24/86 (28%) of children notified in the first year of life. 96/144 (66%) children were white, 2/144 (1%) were black, 7/144 (5%) were of mixed or 'other' ethnicity, 3/144 (2%) had no details recorded

Study Reference	**Knowles 2014(19)** (extraction table covers both Q1 and Q2 data)
Study Design	Design Active surveillance of newly diagnosed CAH through the British Paediatric Surveillance Unit Objective To estimate the incidence of late clinical presentation of CAH and its associated complications Dates August 2007 to August 2009 Country UK
Population characteristics	Recruitment Paediatricians taking part in BPSU notified any child under the age of 16 years who presented in the preceding month with clinical features of CAH and elevated 17-OHP. To maximise case ascertainment, children with CAH were also notified from August 2007 to January 2009 through a laboratory surveillance scheme comprising 12 UK laboratories which measure 17-OHP in children (Cambridge, Leeds, Liverpool, London, 5 Manchester, Scotland, Southampton, Wales) and matched to BPSU reports. Non-matches identified through laboratory surveillance were followed-up. Population was the children who presented clinically with

	CAH aged 1–15 years (children <1 year excluded from the dataset) Diagnosis All notified children were reviewed by an expert diagnostic review panel and categorised as 'definite' or 'not' CAH based on karyotype, presence or absence of salt-wasting, virilisation, and biochemical and molecular genetic test results Duration 25 months Sample size 58
Outcomes	Annual age-specific incidence (risk) of CAH diagnosis between 1 and 15 years of age (based on 52 children notified between 1 September 2007 and 31August 2009) was 0.23 per 100,000 children (Fisher's exact 95% CIs 0.19 to 0.33) (1:434,782) Median age at presentation 5.9 years (Inter Quartile Range 4.7 to 8.4) Diagnosis through presentation with secondary sexual characteristics occurred at a median age of 5.8 (IQR 4.8, 7.6) years. Sex Boys n=26 (45%) Girls n=32 (55%) Ethnicity White n=40 (69%) Asian n=12 (21%) Other n=6 (10%)

Study Reference	**Lai 2020(25)**

Study Design	Design Retrospective cohort study Objective To evaluate the first 2 years of implementation of screening for CAH in New South Wales. Dates 2018-2020 Country Australia
Population characteristics	Screening method Two-tier screening protocol determining 17α-hydroxyprogesterone (17OHP) concentration by immunoassay followed by steroid profile. A threshold level of 17OHP from first tier immunoassay over 22 nmol/L and/or top 2% of the daily assay was further tested using liquid chromatography tandem mass spectrometry (LC-MS/MS) steroid profiling for 17OHP (MS17OHP), androstenedione (A4) and cortisol. Samples with a ratio of (MS17OHP + A4)/cortisol > 2 and MS17OHP > 200 nmol/L were considered as presumptive positive. These newborns were referred for clinical review with a request for diagnostic testing and a confirmatory repeat dried blood spot (DBS). Timing of test <8 days (99.8%) Sample size 202,960
Outcomes	Cases identified 10 (5M:4F and 1 sex undetermined) Incidence classic CAH 10/202960 (1:20,296 calculated) Incidence classic salt-wasting 9/202,960 (1:22,551) Incidence simple virilising 1/202,960

Study Reference	Pearce 2017(24) NB data overlap with Pearce 2016
Study Design	Design Retrospective cohort study Objective To investigate the effect of seasonal changes and kit lot changes on 17-OHP values to determine whether either contributes to false positive rates in the screening program. Dates 2007–2014 Country USA
Population characteristics	Screening method A blood specimen is collected via a heel stick from all newborns on a Guthrie filter card 24–48 h after birth. Since 2010, for any specimen that was collected when the infant was <24 h old, even though it was tested for 17-OHP, a repeat specimen was nevertheless requested. There is a high rate of false positive and false negative results for specimens that are collected in the first 24 h of life. Three millimeter dried blot spots were punched into 96-well plates and the AutoDELFIA Neonatal 17α-hydroxyprogesterone (Perkin Elmer, Turku, Finland) kit was used to measure the 17-OHP level. Multi-tiered thresholds were established based on the age and weight of the infant when specimens were collected. An emergency cut-off was established if the 17-OHP was extremely high (from July 2010 to 2014: 110 ng/ml for infants ≥1751 g and 150 ng/ml for infants ≤1750 g). The effect of seasonality and the AutoDELFIA Neonatal 17-OHP kit changes on 17-OHP value were investigated. Sample size 979,383
Outcomes	Cases identified 52

	Incidence all CAH 1:18,834 (calculated)

Study Reference	**Pearce 2016(23) NB data overlap with Pearce 2017**
Study Design	Design Retrospective cohort study Objective To report the results of the screening programme from 2007 to 2014 and to include algorithm changes that were made to reduce the number of false positive referrals Dates 2007–2014 Country USA
Population characteristics	Screening method A blood specimen is collected via a heel stick from all newborns on a Guthrie filter card 24–48 h after birth. Since 2010, for any specimen that was collected when the infant was <24 h old, even though it was tested for 17-OHP, a repeat specimen was nevertheless requested. There is a high rate of false positive and false negative results for specimens that are collected in the first 24 h of life. Three millimeter dried blot spots were punched into 96-well plates and the AutoDELFIA Neonatal 17α-hydroxyprogesterone (Perkin Elmer, Turku, Finland) kit was used to measure the 17-OHP level. Multi-tiered thresholds were established based on the age and weight of the infant when specimens were collected. An emergency cut-off was established if the 17-OHP was extremely high (from July 2010 to 2014: 110 ng/ml for infants ≥1751 g and 150 ng/ml for infants ≤1750 g). Sample size 1,962,433
Outcomes	Cases identified

108 including 3 FN cases

Incidence all CAH
All CAH (108/1,962,433)
1:18,170

Incidence classic CAH
90/1962433 (assumed)

Incidence classic salt-wasting
11/1962433 (assumed)

Incidence simple virilising
5/1962433 (assumed)

Sex
-All CAH (as reported)
Male, 1:18,280 (55/1,005,444)
Female, 1:18,050 (53/956,856)

-Salt wasting CAH cases
Male: 49/1,005,444
Female: 41/956,856

-Simple virilising CAH cases
Male: 4/1,005,444
Female: 7/956,856

-Non-classic CAH cases
Male: 2/1,005,444
Female: 3/956,856

Ethnicity
-All CAH
White 1:15,610 (56/ 874,066)

	Hispanic 1:17,450 (19/ 331,589) Black 1:24,840 (12/ 298,057) Asian 1:15,250 (9/ 137,269) Native American, NA (0/3009) Other 1:13,150 (12/157,777) -Salt wasting CAH White 46/ 874,066 Hispanic 14/ 331,589 Black 12/ 298,057 Asian (7/ 137,269 Native American, NA (0/3009) Other 11/157,777 -Simple virilising CAH White 8/ 874,066 Hispanic 2/ 331,589 Black 0/ 298,057 Asian 1/ 137,269 Native American, NA (0/3009) Other 0/157,777
Diagnostic test characteristics	Tier of test: First tier test algorithm Day of test: 24-48 hours of life Test: 1. Fluoroimmunoassay of 17-OHP: Three millimeter dried blot spots were punched into 96-well plates and the AutoDELFIA Neonatal 17α-hydroxyprogesterone (Perkin Elmer, Turku, Finland) kit was used to measure the 17-OHP level. The test is an FDA-approved time-resolved fluoroimmunoassay. Unrelated to CAH screening, in July 2014, the program made a concerted effort to encourage hospitals to submit specimens from neonatal intensive care unit (NICU) babies prior to transfusion. This effort led to an increase in submission of day of birth specimens. Multitiered thresholds were established based on the age and weight of the infant when specimens were collected. An emergency cut-off was established if the 17-OHP was extremely high (from July

2010 to 2014: 110 ng/ml for infants ≥1751 g and 150 ng/ml for infants ≤1750 g). In this case the program made the referral by phone prior to performing repeat confirmation testing. Elevated 17-OHP values triggered repeat testing in duplicate and the baby was referred based on the average value of three tests if the value remained above the cut-off level. If the average value of the three blood spots was borderline, then a repeat specimen was requested. In July 2010 the manufacturer changed the antibody used in the kit. Subsequent to the kit change, the cut-off for retesting for the kit was changed to a floating cut-off of 3% for a year after which new cut-off values were established. In order to decrease the number of false positives in normal weight infants whose specimens were collected on day of birth, the referral cut-off for infants was raised in September 2014 to ≥110 ng/ml). To reduce the false positive referrals additional changes were made to the algorithm that would create a borderline result and thus require an additional specimen be sent to the program: in October 2011 a borderline cut-off was established for low birth weight (LBW) infants ≥14 days of age, and, in September 2014 the cut-off level for the borderline category for ≥1751 g infants whose specimens were collected on DOB was raised. Additionally, borderline and referral levels for ≤1750 g babies whose specimens were collected between 14 and 40 days were also raised. Referrals were made by the NBS program to the paediatrician and appropriate specialty care center who then contacted parents to arrange diagnostic testing. The NBS program closed the case when they received a diagnosis and independent results. In some cases, a final diagnosis was not received because the infant was lost to follow up, parents refused further testing or the infant expired before diagnostic testing could be completed.

2. As above, but no retest for babies <24 hours old when DBS taken and no immediate referral for very high values
3. As above, with retest for babies <24 hours old when DBS taken, and immediate referral for very high values

Cut-off: See test description

Reference standard

"diagnostic testing" - not further described. False negatives reported, but authors assume this is underreported.

Type of CAH: 21-hydroxylase deficiency and other CAH enzyme deficiencies (not defined)

Diagnostic test outcomes

***reviewer calculated**

1. Fluoroimmunassay – all patients

TP	**FP**	**FN**	**TN**
105	2371	3	1959954
Sensitivity	**Specificity**	**PPV**	**NPV**
97.2%*	99.9%*	4.24%*	100%*

NB: 14 cases of possible disease have been counted as false positives by the reviewer to calculate these statistics

2. As 1, but no retest for babies <24 hours old when DBS taken and no immediate referral for very high values

TP	FP	FN	TN
NR	NR	NR	NR
Sensitivity	**Specificity**	**PPV**	**NPV**
95%	100%	2.9%	100%

3. As 1, As above, with retest for babies <24 hours old when DBS taken, and immediate referral for very high values

TP	FP	FN	TN
NR	NR	NR	NR
Sensitivity	**Specificity**	**PPV**	**NPV**
98.5%	100%	5.8%	100%

Study Reference	**Speiser 2020(20)**
Study Design	Design Retrospective survey of regional screening units Objective To highlight and describe differences in protocols among the US state laboratories protocols for newborn screening for congenital adrenal hyperplasia (CAH) caused by 21-hydroxylase deficiency. Dates The calendar year of 2017 Country USA

Population characteristics

Screening method

Varied by State, in particular cut-off levels. Generally, all laboratories used fluoroimmunoassay for 17OHP measurement in filter paper blood samples (DELFIA time-resolved fluorescence assay, Perkin Elmer, Waltham, MA, USA) obtained after 24 h of life. Most states performed the assay in their own laboratories, while some sent the samples to Perkin Elmer's facilities using the same assay kits. All but one of the programs utilized birthweight cut-points, but cut-offs varied widely, from 17OHP values of 25 to 75 ng/mL (mean 41.2 ng/mL) for normal birthweights >2250–2500 g. Cut-off points for lower birth weights were generally higher. Most states utilized a later or second screen for infants in intensive care units, as early screening of sick, premature, or low birth weight infants gives many false positive results, contributing to lower positive predictive values.

Infants with borderline or mildly elevated results usually underwent repeat screening. Four states mandated a second screen regardless of initial screening results. Those with more markedly abnormal results were most often referred directly to the paediatric endocrinologist at the state-designated centre and underwent further confirmatory tests, such as a cosyntropin stimulation test if indicated.

Duration

1 year

Sample size

1,564,756

Outcomes

Prevalence total and by state:

Screened Total 1,564,756	Referred positive Total 3217	Confirmed CAH (Total 93)	Prevalence (Weighted Mean = 1:16,825)	17OHP Cut-Off Point (ng/mL) (Mean = 41.2)
230,431	146	11	1:20,948	35
171,964	506	6	1:28,661	55
138,226	608	9	1:15,358	70
135,590	66	6	1:22,598	30
109,740	94	9	1:12,193	65
104,000	387	7	1:14,857	25
84,000	247	5	1:16,800	30
81,117	112	6	1:13,520	50
79,948	144	5	1:15,990	37

	79,000	135	4	1:19,750	35	
	72,440	36	3	1:24,147	75	
	61,500	8	4	1:15,375	25	
	59,643	113	6	1:9941	NR	
	55,935	121	5	1:11,187	25	
	53,361	434	3	1:17,787	30	
	36,361	55	3	1:12,120	38.5	
	11,500	5	1	1:11,500	35	

Study Reference	**van der Linde 2019(30)** (extraction table covers both Q1 and Q3 data)
Study Design	**Design:** Retrospective assessment of national screening programme **Objective:** To evaluate the validity of the Dutch neonatal CAH screening programme and to assess how many newborns with SW CAH have already been clinically diagnosed before the screening result was known. **Dates:** Recruitment: 1st Jan 2002 to 31st Dec 20134 **Country:** The Netherlands
Inclusion	All neonates. The participation was 99.7%
Exclusion	None reported
Study sample size	2,235,931
Test	Automated fluorescence immunoassay (AutoDelfia or GSP-instruments; PerkinElmer, Turku, Finland)
Tier; day of test	First tier; day 3-7

Cut offs	Cut-off levels of 17-OHP related to gestational age (GA) are applied (not extracted here). When GA is not known, birth weight categories are used
Reference Standard and target condition	Identified 21-hydroxylase deficiency and other forms of CAH (see below) Second (and third?) tier test: if positive 1st test, referral to paediatric endocrinology centre for physical examination, sodium and potassium levels in serum and urine, glucose, renin levels or plasma renin activity, and blood gas analysis, adrenocorticotropic hormone, steroid profile on liquid chromatography with tandem mass spectrometry (LC-MS/MS) (17-OHP, androstenedione and 21-deoxycortisol level). An ultrasound of the abdomen for evaluation of the adrenal glands and kidneys is performed. If indicated, the uterus and presence and appearance of gonads are checked. In case of elevated 17-OHP level, genetic analysis for CYP21A2 mutations is performed. After careful physical examination and first results of diagnostic workup (sodium, potassium, renin) children are diagnosed with CAH or as false positive. Afterwards, when also results of genetic analysis is known, patients with CAH can be classified as SW, simple virilising (SV) or non-classic (NC) or having other enzymatic defects or as a false-positive screening result. Paediatricians are asked to report false-negative screening results to the Dutch Paediatric Surveillance System (www.nsck.nl)
Subgroup data of interest?	Thirty per cent were premature (n=106; GA ≤36+0 weeks). In four premature patients (3.8%), CAH was confirmed (two SW, two SV).
Outcomes	

TP	FP	FN	TN
133	327	0 (though follow-up limited to <20 years)	2235452
Sensitivity	Specificity	PPV	NPV
Total CAH: 100% Salt-wasting CAH: 100%	Total CAH: Salt-wasting CAH: 99.98%	Salt-wasting CAH: 24.7%	Salt-wasting CAH: 100% (calculated by reviewer)

Excluding 17 missing diagnoses and counting 2 other diagnoses as TP
Salt-wasting CAH: 114 (73 M, 41 F)
Simple virilising CAH: 14 (10 M, 4 F)
Non-classic CAH: 5 (4 M, 1 F)

	Other enzyme deficiency: 2: one child with suspected P450 oxidoreductase deficiency; no mutation analysis has been performed; one child with compound heterozygous 3β-hydroxysteroid dehydrogenase

Study Reference	**Zetterstrom 2020(31)**
Study Design	Design Cohort study Objective To report on CAH screening from the Swedish national neonatal screening programme Dates From January 2011 until December 2019 Country Sweden
Population characteristics	Screening method 17OHP measured using dried blood spot (DBS) samples collected as soon as possible after 48 h after birth (48–72 h) on Perkin Elmer 226 Ahlstrom paper. DBS screening was also offered to older children, below the age of 9 years, that moved to Sweden from countries lacking a national newborn screening program for CAH Duration 8 years Sample size 1,030,409 newborns and 34,713 older children Sex 56%M:44%F Ethnicity N/R
Outcomes	Cases identified 87 newborns, 12 older children

	Incidence: Incidence of 21OHD was 1:11,200. 92% of the detected cases had the classic form of CAH (SW and SV forms) The incidence of classic CAH was 1:12,300 and the NC form 1:79,300 Incidence classic CAH 70% of newborns had SW CAH, overall incidence of classic CAH: 1:12,300 Incidence non-classic CAH: 1:79,300

Table 20. Studies relevant to criterion 1: Q2 (median age of presentation)

Study Reference	**Bomberg 2015(39)/Sarafoglou 2014(40)**
Study Design	Design Retrospective chart review Objective To determine whether the shorter final adult height seen in glucocorticoid-treated patients with CAH was attributable to reduced pubertal height gain or changes in the peri-pubertal timeframe Dates 1955 and 2012 Country USA Setting 3 medical institutions in Minnesota
Population characteristics	Recruitment All patients with a diagnosis of CAH at 3 institutions were identified through medical record review.

	<u>Diagnosis</u> Subtype classification (salt-wasting vs simple-virilising) was assigned by a paediatric endocrinologist at each participating institution and was based on clinical, hormonal, biochemical, and in some cases, molecular testing <u>Duration</u> 57 years <u>Sample size</u> N=104 Male (n = 41): 28 salt-wasting, 13 simple-virilising; female (n = 63): 38 salt-wasting, 25 simple-virilising) <u>Sex</u> Male n=41 female n=63 <u>Ethnicity</u> NR
Outcomes **NB male/female data from Bomberg, all data from Sarafoglou.**	Males with CAH Salt Wasting / Simple Virilising Age at diagnosis (months): 3.1 +/- 4.4 / 43.8 +/- 27.5 Females with CAH 0.6 +/- 0.8 / 35.2 +/- 34.1 Age of diagnosis for boys with simple-virilising was 40.7 months later than boys with salt-wasting, and the age at diagnosis for girls with simple-virilising was 34.6 months later than girls with salt-wasting. Both sexes: Salt wasting mean age at diagnosis 1.6 +/- 3.1 months Simple virilising mean age at diagnosis 37.4 +/-31.2 months

Study Reference	**Halper 2019(41)**

Study Design	Design Cross-sectional study with retrospective identification Objective To examine anastrozole's effect on bone mineral density (BMD) or body composition in children with congenital adrenal hyperplasia (CAH) due to 21-hydroxylase deficiency. Dates N/R Country USA Setting University of Minnesota Masonic Children's Hospital Multidisciplinary CAH Clinic
Population characteristics	Recruitment A retrospective cohort of 56 children with CAH identified from the University of Minnesota Masonic Children's Hospital Multidisciplinary CAH clinic Diagnosis N/R Duration N/R Sample size 25 children with CAH treated with anastrozole vs 31 children with CAH not treated with anastrozole. Children receiving anastrozole: 28% diagnosed with salt-wasting CAH, 40% with simple-virilising CAH and 32% with non-classic CAH. Children not receiving anastrozole: 68% diagnosed with salt-wasting CAH, 19% with simple-virilising CAH and 13% with non-classic CAH.

	Age Anastrozole group mean age 11.3 [SD 3.0] years No anastrozole group mean age 13.5 [SD 4.6] years Sex Anastrozole group: 56% males No anastrozole group: 29% males Ethnicity All white non-Hispanic
Outcomes	Mean age at diagnosis given by treatment group. Children using anastrozole group mean age at diagnosis 3.27 years (SD 3.51) Children not using anastrozole group mean age at diagnosis 1.76 years (SD 3.21)

Study Reference	**Hsieh & White 2011(42)**
Study Design	Design Retrospective chart review Objective To determine aetiologies, signs, and symptoms of primary adrenal insufficiency presenting in childhood. Dates 1999-2010 Country USA Setting Tertiary care paediatric hospital
Population characteristics	Recruitment Patients with primary adrenal insufficiency were identified by reviewing medical records of Children's Medical Center Dallas (CMC) carrying primary, secondary, or tertiary ICD-9 codes for corticoadrenal insufficiency,

	glucocorticoid deficiency, and mineralocorticoid deficiency. Cases with CAH were excluded from further study however age of presentation data was recorded. Diagnosis N/R for CAH sample Duration 11 years Sample size 32 CAH patients were identified out of a total sample of 77 Sex Not reported for CAH sample Ethnicity Not reported for CAH sample
Outcomes	22 of the 32 CAH patients were diagnosed as inpatients at CMC during infancy, with median (interquartile range, IQR) age of 9 d (2.5–12.7) at presentation. Age at presentation not given for the remaining 10 CAH patients. Subtypes not reported.

Study Reference	**Maccabee-Ryaboy 2016(43)**
Study Design	Design Retrospective medical chart review Objective To estimate the incidence of hypertension by CAH subtype and sex, and to assess its association with body mass index, hydrocortisone and fludrocortisone Dates 1970–2013

	Country USA Setting 3 paediatric centres in Minnesota (University of Minnesota Masonic Children's Hospital, The Mayo Clinic and Children's Hospitals of Minnesota)
Population characteristics	Recruitment Diagnosis CAH diagnosis and subtype assigned by the treating paediatric endocrinologists based on hormonal data, clinical and biochemical presentation and, in some cases, molecular testing of the CYP21A2 gene using a common mutation panel or sequencing Duration 43 years Sample size Total CAH patients identified 247. Study focuses on N=180 Salt Wasting and Simple Virilising patients (73%) who had at least three Blood Pressure measurements documented. Patients were divided into two time periods, those born between 1970–1994 (pre-newborn screening) and those born between 1995–2013 (after newborn screening for CAH was initiated in Minnesota). Sex 93 female; 87 male Ethnicity N/R
Outcomes	Pre-screening: 124/180 children (69%) were followed from CAH diagnosis. 88% of these 124 had Salt Wasting type and were diagnosed in the first year, males on average at 3.7 months and females at 1.5 months. 26% of 124 children had Simple Virilising type and were diagnosed in the first year, males on average at 47 months and females at 40 months

Study Reference	Pijnenburg-Kleizen 2019(44)
Study Design	Design Retrospective chart review Objective To identify age-dependent factors that influence Final Height in CAH patients, resulting in age-specific treatment goals Dates 1980–1997 Country The Netherlands Setting Department of Pediatric Endocrinology of the Radboud University Medical Center, Nijmegen
Population characteristics	Recruitment Patients diagnosed with classic CAH due to 21-hydroxylase deficiency, who were born between 1980 and 1997 and were treated in Radboud University Medical Center from early childhood onwards were included. These patients were born prior to screening. Diagnosis Diagnosis of Salt Wasting CAH or Simple Virilising CAH was based on clinical and biochemical data and confirmed by mutation analysis. The CYP21A2 mutations were categorized into different mutation groups based on in vitro 21-hydroxylase activity. Non-classic CAH cases were excluded. Duration 17 years Sample size N=39 CAH patients after exclusions for co-morbidities (severe spasticity and anorexia nervosa), limited data, possible different treatment strategies in the first years of life, for those diagnosed and treated elsewhere up to 8 years of age or older, lack of salivary steroid measurements due to non-compliance

	25 classified as Salt Wasting (12 females, 13 males); 14 classified as Simple Virilising (six females, eight males) based on their genotype and phenotype <u>Sex</u> 18 females; 21 males <u>Ethnicity</u> N/R
Outcomes	24/25 SW-CAH patients diagnosed neonatally because of ambiguous genitalia and/or SW crises. One boy with SW-CAH was diagnosed at the age of 2 years. Female SV-CAH patients were diagnosed earlier than male SV-CAH patients, at a median age of 2.3 years (range 0–4.0 years) compared to a median of 4.4 years (range 2.5–6.3 years)

Table 21. Studies relevant to criteria 4 and 5: Q3 (test accuracy)

Study Reference	**Gaudl 2019(48)**
Study Design	**Design:** Case-control **Objective:** To investigate method optimization strategies to increase specificity for serum cortisol and dried blood 17-OHP. Furthermore, we provide a decision tree for the user to navigate the different routes of problem solving that are presented here **Dates:** NR **Country:** Germany
Inclusion	Routine newborn screening samples selected so at least 50% were false positive at first screen.

Exclusion	None reported
Study sample size	100
Test	1. **Immunoassay** (AutoDELFIA Neonatal 17-OHP from Perkin Elmer (Waltham, MA,USA)) 2. **LC-MS/MS assay: (-)MS3 screening method (optimised for high throughput).** For a quantitative 17-OHP screening from dried blood, the routine method ((+)MS2 routine, described below) was modified to maximize time efficiency while preserving specificity. The gradient profile was altered using the 25mm Chromolith® to provide very quick elution of 17-OHP and was 0-0.3 min 70% B, 0.3–1.0 min 70% to 100% B, 1.0–1.3 min 100% B, 1.3–1.5 min 70% B. General composition of the mobile phase, flow rates, and oven temperature were maintained. Detection using two-stage fragmentation via CID and resonance excitation in the ion trap in negative ionization ((−)MS3) replaced (+)MS2. The mass transitions m/z 329/285/123 for 17-OHP and 337/290/127 for 17- OHP-d8 were introduced. 3. **LC-MS/MS assay: (+)MS2 routine method.** A Prominence UFLC system from Shimadzu (Duisburg, Germany) was coupled to a QTRAP® 6500 from SCIEX (Framingham, MA, USA). A PAL HTS-xt autosampler from CTC Analytics (Zwingen, Switzerland) handled sample injection. Online solid phase extraction was performed on a POROS® column (30×2.1 mm) from Applied Biosystems (Foster City, CA, USA) at a flow rate of 3 mL/min. For chromatographic separation a Chromolith® High Resolution column (RP-18, endcapped, 25 x 4.6 mm) from Merck (Darmstadt, Germany) was used. The mobile phase consisted of 32.5% eluent A (0.2 mmol/L ammonium fluoride (NH4F) in water/methanol 97/3 v/v) and 67.5% eluent B (0.2 mmol/L NH4F in water/methanol 3/97 v/v) and was adjusted as follows: 0–1 min 32.5% B, 1–3.7 min 32.5% to 100% B, 3.7–4.2 min 100% B, 4.2–4.3 min 32.5% B. Flow rate was 1.5 mL/min and the column oven was set to 35 °C. Electrospray ionization was applied in positive and negative mode, leading to two injections of 20 µL (pos) and 100 µL (neg) per sample to cover all mentioned analytes. Detection was carried out using multiple reaction monitoring and one-stage fragmentation (MS2) via collision induced dissociation (CID).
Tier; day of test	1. First tier (intended use, but sample is a mix of first tier positive/negative) – day NR 2. First tier (intended use, but sample is a mix of first tier positive/negative) – day NR 3. Second-tier (intended use, but sample is a mix of first tier positive/negative) – day NR

Cut offs	NR. “7-OHP concentrations were evaluated regarding AutoDELFIA based CAH cut-offs accounted for gestational age, which are in consensus with the German Society for Neonatal Screening (DGNS).”
Reference Standard and target condition	Unclear - assume routine screening methodology in Germany Unclear which types of CAH diagnosed
Subgroup data of interest?	None

Outcomes

1. Immunoassay

***reviewer calculated**

TP	FP	FN	TN
3	61	0	36
Sensitivity	Specificity	PPV	NPV
100.00*	37.11*	4.69*	100.00*

2. LC-MS/MS assay: (-)MS3 screening method (optimised for high throughput)

TP	FP	FN	TN
3	4	0	93
Sensitivity	Specificity	PPV	NPV
100.00*	95.88*	42.86*	100.00*

3. LC-MS/MS assay: (+)MS2 routine method.

TP	FP	FN	TN

	3	0	0	97
	Sensitivity	Specificity	PPV	NPV
	100.00*	100.00*	100.00*	100.00*

Study Reference	**Boelen 2016(47)**
Study Design	**Design:** unclear if consecutive sample **Objective:** to develop a method to measure the steroid profile in DBS using LC–MS/MS, to determine the reference interval of the steroids measured in DBS based on GA and to evaluate the method by analysing DBS from newborns with a positive CAH screening **Dates:** NR **Country:** The Netherlands
Inclusion	CAH-positive heel prick test cards from the Dutch neonatal screening programme
Exclusion	None reported
Study sample size	92
Test	1. 21-deoxycortisol by UPLC–MS/MS assay. 2. Other biomarkers by UPLC-MS/MS assay. 11-deoxycortisol; 11-deoxycorticosterone; 17-hydroxyprogesteron; Δ4-androstenedion; Corticosterone; Cortisone; Cortisol.
Tier; day of test	1&2. Second-tier tests Day of test NR

Cut offs	21-deoxycortisol: <1nmol/L, for all gestational ages Other biomarkers cut-offs are reported in the study but not extracted here.
Reference Standard and target condition	21-Hydroxylase deficiency confirmed by mutation analysis, but unclear if all 92 cards were tested. 21-hydroxylase deficiency
Subgroup data of interest?	None
Outcomes ***reviewer calculated**	1. All Data read from figure 3 or inferred from reported specificity

TP	FP	FN	TN
8	0	0	84
Sensitivity	Specificity	PPV	NPV
100*	100*	100*	100*

None of the other tests were 100% specific. Other data displayed in Figure 3 of the study report, but unable to extract numerically

Study Reference	**Han 2019(46)**
Study Design	**Design:** Case-control **Objective:** We developed a candidate second-tier method using LCMS/MS for measuring 17-OHP and cortisol in the same DBS sample. We evaluated the method using DBS samples from clinically diagnosed CAH patients and QC/proficiency test (PT) materials.

	Dates: 2013 and 2014 **Country:** USA
Inclusion	The negative-control DBS samples were DBS samples collected by the NYS-NBS Program in 2015 and were among those that tested negative for CAH. The clinically diagnosed CAH patient DBS samples were from the NYS-NBS Program collected during 2013 and 2014.
Exclusion	None reported
Study sample size	45 (17 with CAH)
Test	**1. LC-MS/MS cutoff value of > 39.1 ng/mL for 17-OHP** Four 3-mm discs from each DBS for blank, double blank, calibrators, QC, and patient samples were suspended in 500 µL of distilled and deionized H2O, to which 10-µL aliquots of an internal standard solution containing [2,3,4-13C3]17-OHP and [2,3,4-13C3] cortisol were added, except for the double blanks, which did not receive internal standards. These mixtures were incubated for 20 min at room temperature, followed by addition of 1.4 mL MTBE and shaking (60 rpm) for 45 min at room temperature. The samples were centrifuged at 9000×g for 3 min. Each supernatant was aliquoted into two separate fractions: 900 µL was taken for 17-OHP analysis and 400 µL cortisol analysis. These fractions were evaporated to dryness under N2. The fraction for 17- OHP analysis was resuspended in 40 µL of methanol/ aqueous ammonium fluoride (25 µmol/L) (1:1), and the fraction for cortisol analysis was resuspended in 40 µL of 0.1% formic acid only. LC-MS/MS analysis: the LC-MS/MS system consisted of an Agilent 1100 liquid chromatograph interfaced with an AB Sciex API 2000™ triple-quadrupole mass spectrometer operated with Analyst® 1.6.2 software. For the analysis of 17-OHP, 30 µL of sample extract was injected and resolved at room temperature on a 50 × 2.1 mm (2.6-µm particle size) Kinetex C18 column (Phenomenex) protected by a Security Guard ULTRA Cartridge (UHPLC C18, 2.1 mm) (Phenomenex). Separation of 17-OHP was achieved using a gradient program consisting of an initial condition of 10% acetonitrile containing 25 µmol/L ammonium fluoride for 1 min, followed by an increase from 10 to 90% acetonitrile containing 25 µmol/L ammonium fluoride over 1 min, and from 90 to 100% acetonitrile containing 25 µmol/L ammonium fluoride over 6 min. The column was then re-equilibrated with 10% acetonitrile containing 25 µmol/L ammonium fluoride for 3.9 min. The flow rate was 200 µL/min. AValco valve, integral to the API 2000 MS/MS

	system, was set to direct the flow during 6.5– 7.5 min of each run to the ion source for MS/MS analysis, whereas the flow at other times was diverted to waste. 17- OHP analyses were carried out using electrospray ionization in the negative ion with multiple reaction monitoring. The ionization and MS/MS conditions were optimized using a flow-infusion analysis and are summarized as follows: declustering potential, - 55 V; entrance potential, - 8 V; collision energy, - 25 eV; focusing potential, - 350 V; collision cell exit potential, - 10 V; ion spray voltage, - 4500 V; source temperature, 375 °C; CUR, 40.0; GS1, 20.0; GS2, 11.0; and collision gas, 9.0. Detection of 17-OHP was achieved by monitoring the m/z 329 → 285 transition as the quantifier and the m/z 329 → 123 transition as the qualifier. The [2,3,4-13C3]17- OHP was analyzed by monitoring the m/z 332 → 126 transition. A dwell time of 300 ms was used for each transition. Cortisol was analyzed using the same C18 column and guard column that were used for 17-OHP, although a separate LC-MS/MS run was needed to quantify this steroid. The LC gradient consisted of an initial condition of 45% of methanol/ 1-propanol (3:1) containing 0.1% formic acid for 0.5 min, followed by an increase to 85% methanol/1-propanol (3:1) containing 0.1% formic acid over 1.0 min, and from 85 to 99% methanol/1-propanol (3:1) containing 0.1% formic acid over 2.3 min. The column was washed with 99% methanol/1- propanol (3:1) containing 0.1% formic acid for 1.2 min, followed by a re-equilibration with 45% of methanol/1- propanol (3:1) containing 0.1% formic acid for 2.9 min. The flow rate was 200 µL/min. The Valco valve was set to direct the flow during 0.8–1.4 min of each run to the ion source for analysis. The ionization and MS/MS conditions were declustering potential, 25 V; entrance potential, 6 V; collision energy, 40 eV; focusing potential, 350 V; collision cell exit potential, 10 V; ion spray voltage, 2500 V; source temperature, 400 °C; CUR, 10.0; GS1, 20.0; GS2, 70.0; and collision gas, 4.0. Detection of cortisol was achieved in the positive-ion multiple reaction monitoring mode by monitoring the m/z 363 → 121 transition as the quantifier and m/z 363 → 91 as the qualifier; the m/z 366 → 124 transition was used for detecting the internal standard, [2,3,4-13C3] cortisol. A dwell time of 200 ms was used for each transitionLC-MS/MS measurement of Androstenedione **in whole blood sample**. Positive if above the limit of detection. 2. LC-MS/MS ratio of 17-OHP/cortisol at the cut-off of 1.0 3. LC-MS/MS combination of 17-OHP > 39.1 ng/mL and the 17-OHP/cortisol at 1.0 as cut-off values
Tier; day of test	Second-tier tests Day of test NR

Cut offs	See “Test”
Reference Standard and target condition	Diagnosis of CAH was made by physicians based on diagnostic testing results and clinical symptoms. Unclear which CAH types
Subgroup data of interest?	
Outcomes ***reviewer calculated**	1. LC-MS/MS cutoff value of > 39.1 ng/mL for 17-OHP

TP	FP	FN	TN
NR	NR	NR	NR
Sensitivity	Specificity	PPV	NPV
100%	96.4%	NR	NR

2. LC-MS/MS ratio of 17-OHP/cortisol at the cut-off of 1.0

TP	FP	FN	TN
NR	NR	NR	NR
Sensitivity	Specificity	PPV	NPV
88.2%	75%	NR	NR

3. LC-MS/MS combination of 17-OHP > 39.1 ng/mL and the 17-OHP/cortisol at 1.0 as cut-off values

TP	FP	FN	TN
NR	NR	NR	NR

Sensitivity	Specificity	PPV	NPV
94.1%	100%	NR	NR

Study Reference	**Tang 2016(45)**
Study Design	**Design:** Regional screening cohort, assume consecutive **Objective:** The current Clinical and Laboratory Standards Institute standard recommends blood collection from 24 to 48 hours after birth for newborn genetic disorder screening. We used California population-level data to determine whether early specimens (collected from 12 to 23 hours) would also be considered satisfactory based on screening performance **Dates:** 2006–2013 (TP, FN); 2013 (FP, TN) **Country:** USA – California
Inclusion	All screened newborns
Exclusion	None reported

Study sample size	Unclear
Test	1. **Immunofluorescence and tandem mass spectrometry collected 12-23 hours** Dried blood spot specimens using the heel-stick procedure are collected and handled according to the CLSI guidelines with the exception that the time of blood collection can be as early as 12 hours after birth. CAH screening consists of two tiers. The first tier is screened with an immunofluorescence assay that measures 17-hydroxyprogesterone (17-OHP) levels with different cut-off ranges corresponding to different birth weight ranges. Newborns with highly elevated results are reported as positive and referred to a state-approved endocrine centre for diagnostic evaluation. If a specimen's 17-OHP value is moderately elevated but not high enough for immediate reporting (classified as "questionable"), then a second-tier test will be performed on the specimen using tandem mass spectrometry that measures 17-OHP, androstenedione, and cortisol to determine the positive status. Each birth weight group has its corresponding 17-OHP cut-offs. 2. **Immunofluorescence and tandem mass spectrometry collected 24-48 hours**
Tier; day of test	Two-tier testing Day of test: 12-23 or 24-48 of life
Cut offs	According to birth weight, cut-offs not reported
Reference Standard and target condition	All genetically confirmed disorders must be reported to the Genetic Disease Screening Program Newborn Screening Registry. A confirmed case identified as "missed by newborn screening" in the registry was defined as false negative. Although varying depending on the type of condition, the process of confirming a diagnosis after a case is referred to the appropriate specialty-care follow-up center is similar for all cases. For the false-positive analysis, we used initial screening interpretation results (positive or negative) and final resolution results (disease or no disease) to determine false-positive status and true-negative status of the newborns. Unclear which CAH types

Subgroup data of interest? By birth weight

Outcomes

1. Immunofluorescence and tandem mass spectrometry collected 12-23 hours

***reviewer calculated**

TP	FP	FN	TN
42	108	2	106277
Sensitivity	Specificity	PPV	NPV
95.5%*	99.9%*	Not calculable	Not calculable

NB: TP and FN data are for 2006-2013; FP and TN data are from 2013 only

2. **Immunofluorescence and tandem mass spectrometry collected 24-48 hours**

TP	FP	FN	TN
145	467	10	332176
Sensitivity	Specificity	PPV	NPV
93.5%*	99.86%*	Not calculable	Not calculable

NB: TP and FN data are for 2006-2013; FP and TN data are from 2013 only

1&2. **Immunofluorescence and tandem mass spectrometry collected 12-48 hours**

Birthweight	Time of sample	FP	TN	Specificity*	NPV*
1000	12-23	5	224	97.82	100.00
1000-1499	12-23	1	224	99.56	100.00
1500-2499	12-23	33	3287	99.01	100.00

2500+	12-23	55	102556	99.95	100.00
1000	24-48	3	886	99.66	100.00
1000-1499	24-48	9	1299	99.31	100.00
1500-2499	24-48	223	15947	98.62	100.00
2500+	24-48	108	314168	99.97	100.00

NB: TP and FN data are for 2006-2013; FP and TN data are from 2013 only

Appraisal for quality and risk of bias

Quality assessments of included studies are reported below.

Table 22. Quality assessment of studies included for Q1 (JBI Critical Appraisal Checklist for Studies Reporting Prevalence Data)

	Target population	Appropriate recruitment	Adequate sample size	Setting/subjects described	Data analysis coverage	Measurement criteria	Measurement reliability	Appropriate statistical analysis	Response rate
David 2019	Yes, national NBS programme	Yes, national NBS programme	Yes, national NBS programme	Unclear, limited details provided	Yes, national NBS programme	Unclear, limited details on identification methods	Unclear, false negatives from NBS programme should be picked up by clinical presentation and notified by paediatricians, but response rate unknown. Mild cases likely to be missed.	Unclear - limited statistical details provided	Yes, assumed
David 2018 (data overlap with David 2019)	Yes, national NBS programme (assumed)	Yes, national NBS programme (assumed)	Yes, national NBS programme (assumed)	Unclear, limited details provided	Yes, national NBS programme (assumed)	Unclear, limited details on identification methods	Unclear, false negatives from NBS programme should be picked up by clinical presentation and notified by paediatricians, but response rate unknown.	Unclear - limited statistical details provided	Yes, assumed

	Target population	Appropriate recruitment	Adequate sample size	Setting/subjects described	Data analysis coverage	Measurement criteria	Measurement reliability	Appropriate statistical analysis	Response rate
							Mild cases likely to be missed.		
Eshragh 2020	Yes, regional NBS programs	Unclear, if all live newborns included	Yes, large multi-regional study	Unclear, limited details provided	Yes (assumed)	Unclear, limited details on identification methods	Unclear, false negatives from regional screening should be picked up by clinical presentation and notified by paediatricians, but response rate unknown. Mild cases likely to be missed.	Unclear - if all cases identified	Yes, assumed
Fox 2020	Yes, regional NBS programs	Unclear, if all live newborns included	Yes, large regional study	Unclear, limited details provided	Yes (assumed)	Unclear, limited details on identification methods	Unclear, false negatives from regional screening should be picked up by clinical presentation and notified by paediatricians, but response rate unknown. Mild cases likely to be missed.	Unclear - if all cases identified	Yes, assumed

	Target population	Appropriate recruitment	Adequate sample size	Setting/subjects described	Data analysis coverage	Measurement criteria	Measurement reliability	Appropriate statistical analysis	Response rate
Heather 2015	Yes, national NBS programme	Yes, national NBS programme	Yes, national NBS programme	Yes	Yes, national NBS programme	Unclear, limited details on identification methods	Yes (identified via NBS [n=21] and clinically [n=23])	Unclear - if all cases identified, especially in those who may have missed screening (no data reported on this)	Yes, assumed
Held 2015	Yes, NBS programs from 7 US states	Unclear, if all live newborns included (NR for first screen; >85% of all newborns received mandated second screen at 8-14 days after birth)	Yes, large regional study	Unclear, limited details provided	Yes (assumed)	Unclear, limited details on identification methods	Unclear, false negatives from NBS programme should be picked up by clinical presentation and notified by paediatricians, but response rate unknown. Mild cases likely to be missed.	Unclear - if all cases identified, especially in those who may have missed screening (no data reported on this)	Yes, assumed
Iniguez 2018	Yes, multi-regional NBS programs	Unclear, if all live newborns included	Yes, multi-large regional study	Unclear, limited details provided	Unclear, no details provided	Unclear, no details on identification methods provided	Unclear, no details provided	Unclear - no details provided	Unclear, details NR
Khalid 2012	Yes - all UK children aged 0-16 between 2007-2009	Yes - all UK children aged 0-16 between 2007-2009	Yes - large national surveillance study	Yes	Yes - all children included	Unclear - criteria were clinical or objective	No - not all patients were tested	Partially - Incidence with 95% CI reported, but	Unclear whether response rate was

	Target population	Appropriate recruitment	Adequate sample size	Setting/subjects described	Data analysis coverage	Measurement criteria	Measurement reliability	Appropriate statistical analysis	Response rate
								the denominator was assumed from ONS data and it was assumed that all cases were identified	100% Knowles reports this as 92-94% for BPSU card return rate, and 95% for CAH case report questionnaire.
Knowles 2013	Yes - all UK children aged 1-15 within 24-month period	Yes - all UK children aged 1-15 within 24-month period	Yes - large national surveillance study	Yes	Yes - all children included	Unclear - criteria were clinical or objective	No - not all patients were tested	Partially - Incidence with 95% CI reported, but the denominator was assumed from ONS data and it was assumed that all cases were identified	Yes
Lai 2020	Yes, regional NBS programs	Unclear, if all live newborns included	Yes, large regional study	Unclear, limited details provided	Yes (assumed)	Yes	Unclear, false negatives from regional screening should be picked up by	Unclear - if all cases identified	Yes, assumed

	Target population	Appropriate recruitment	Adequate sample size	Setting/subjects described	Data analysis coverage	Measurement criteria	Measurement reliability	Appropriate statistical analysis	Response rate
							clinical presentation and notified by paediatricians, but response rate unknown. Mild cases likely to be missed.		
Pearce 2017 (data overlap with Pearce 2016)	No, seasonal data from NBS program from 1 US state	No, seasonal data from NBS program from 1 US state	No, seasonal data from NBS program from 1 US state	No, seasonal data from NBS program from 1 US state	No, seasonal data from NBS program from 1 US state	Unclear, limited details on identification methods	No, not all patients included (seasonal data)	No, not all patients included (seasonal data)	Unclear, details NR
Pearce 2016	Yes, NBS programs from 1 US state	Unclear, if all live newborns included	Yes, large regional study	Unclear, limited details provided	Yes (assumed)	Unclear, limited details on identification methods	Unclear, false negatives from state NBS programme should be picked up by clinical presentation and notified by paediatricians, but response rate unknown. Mild cases likely to be missed.	Unclear - if all cases identified, especially in those who may have missed screening (no data reported on this)	Yes, assumed
Speiser 2020	No, low birth weight infants excluded from some aspects	No, low birth weight infants excluded	Yes	No, data such as M/F missing, lack of clarity around	Yes	Some states used different methodologies, and their	No - not all patients were tested	Unclear, lack of clarity around the exclusion of	Not all states responded, so the

	Target population	Appropriate recruitment	Adequate sample size	Setting/subjects described	Data analysis coverage	Measurement criteria	Measurement reliability	Appropriate statistical analysis	Response rate
	of analysis, but not clear which	from some aspects of analysis, but not clear which		low birth weight infants		comparative accuracy is unclear		low birth weight infants	study can only be generalised to those that did. Unclear whether all data available from states that did respond.
van der Linde 2019	Yes, national newborn screening programme	Yes, national newborn screening programme	Yes, national newborn screening programme	Yes, national newborn screening programme	Yes, national newborn screening programme	Partial - false negatives from newborn screening programme should be picked up by paediatricians, but response rate unknown	Partial - false negatives from newborn screening programme should be picked up by clinical presentation and notified by paediatricians, but response rate unknown. Mild cases likely to be missed	Yes	Yes
Zetterstrom 2020	Yes, national NBS programs	Yes, >99.5% of all newborns screened	Yes, large national study	Unclear, limited details provided	Yes (assumed)	Unclear, limited details on identification methods	Unclear, false negatives from national l screening should be picked up by	Unclear - if all cases identified	Yes, assumed

	Target population	Appropriate recruitment	Adequate sample size	Setting/subjects described	Data analysis coverage	Measurement criteria	Measurement reliability	Appropriate statistical analysis	Response rate
							clinical presentation and notified by paediatricians, but response rate unknown (5FN detected between 2011 and 2019). Mild cases likely to be missed.		

Abbreviations: NBS, Newborn screening; NR, not reported; BPSU British Paediatric Surveillance Unit

Table 23. Quality assessment of studies included for Q2 (Murad *et al*. adapted tool)

Bomberg 2015/Sarafoglou 2014			
Domain	**Question**	**Judgement on risk of bias**	**Comment**
Selection	1. Do the patients represent the whole experience of the investigator (centre) or is the selection method unclear to the extent that other patients with similar presentation may not have been reported?	High	Only children with available data for height on at least 2 occasions during the 3 stages of growth were eligible
Ascertainment	2. Was the exposure* adequately ascertained?	Low	Exposure is diagnosis of CAH. No description of criteria for initial CAH diagnosis given but subtype classification (salt-wasting vs

			simple-virilizing) was adequately described - assigned by a pediatric endocrinologist at each participating institution and was based on clinical, hormonal, biochemical, and in some cases, molecular testing.
	3. Was the outcome adequately ascertained?	Unclear	Median age at diagnosis described, no further details of age at presentation of symptoms or nature of presenting symptom.
Causality	4. Was follow-up long enough for outcomes to occur?	N/A	N/A
Reporting	5. Are the cases described with sufficient details to allow other investigators to replicate the research or to allow practitioners to make inferences related to their own practice?	Unclear	Insufficient details on presentation, missing data
NB aim of study not to determine age of presentation of CAH. Overall unclear due to only age at diagnosis data, no further information on age at presenting symptoms. Population was highly selective.			

Fox 2020			
Domain	**Question**	**Judgement on risk of bias**	**Comment**
Selection	1. Do the patients represent the whole experience of the investigator (centre) or is the selection method unclear to the extent that other patients with similar presentation may not have been reported?	Unclear	Pre-screened cohort does not specify whether data includes all cases.
Ascertainment	2. Was the exposure* adequately ascertained?	Unclear	Exposure is CAH diagnosis. Clinical diagnosis (pre-screening),

			or by screening protocol. Unclear timescale for inclusion therefore unclear whether asymptomatic, mild, or late presenting cases were diagnosed.
	3. Was the outcome adequately ascertained?	Low	Age at screening, confirmation and reporting is clearly reported
Causality	4. Was follow-up long enough for outcomes to occur?	N/A	N/A
Reporting	5. Are the cases described with sufficient details to allow other investigators to replicate the research or to allow practitioners make inferences related to their own practice?	Low	Sufficient detail is given on diagnostic criteria and screening protocol
Overall Unclear risk			

Halper 2019			
Domain	**Question**	**Judgement on risk of bias**	**Comment**
Selection	1. Do the patients represent the whole experience of the investigator (centre) or is the selection method unclear to the extent that other patients with similar presentation may not have been reported?	Unclear	Selection method is unclear, unclear whether these are consecutive patients, or representative of all patients attending the clinic.
Ascertainment	2. Was the exposure* adequately ascertained?	Unclear	Diagnostic criteria not reported
	3. Was the outcome adequately ascertained?	Unclear	Median age of diagnosis, no further clarification on point at

			which age at diagnosis is made, no age by sub type of CAH
Causality	4. Was follow-up long enough for outcomes to occur?	N/A	N/A
Reporting	5. Are the cases described with sufficient details to allow other investigators to replicate the research or to allow practitioners make inferences related to their own practice?	High	All cases only described by treatment group, no overall descriptives
Overall judgement Unclear due to a lack of clarity in selection procedure and ascertainment			

Hsieh & White 2011			
Domain	**Question**	**Judgement on risk of bias**	**Comment**
Selection	1. Do the patients represent the whole experience of the investigator (centre) or is the selection method unclear to the extent that other patients with similar presentation may not have been reported?	Low	Chart review designed to capture all CAH cases within the sampling timeframe.
Ascertainment	2. Was the exposure* adequately ascertained?	Unclear	Diagnostic criteria not described
	3. Was the outcome adequately ascertained?	Low	Age of presentation of symptoms clearly reported
Causality	4. Was follow-up long enough for outcomes to occur?	N/A	N/A

Reporting	5. Are the cases described with sufficient details to allow other investigators to replicate the research or to allow practitioners to make inferences related to their own practice?	High	CAH cases not described as excluded from further study
Overall judgement Unclear due to lack of description of the excluded CAH sample			

Khalid 2012			
Domain	**Question**	**Judgement on risk of bias**	**Comment**
Selection	1. Do the patients represent the whole experience of the investigator (centre) or is the selection method unclear to the extent that other patients with similar presentation may not have been reported?	Unclear	The target population was all children presenting with symptoms in England, Scotland, and Wales during the 2 year study period. The study relies on notification by paediatricians and surveillance laboratories. It is unclear what % capture of cases the data represents or how many were not reported.
Ascertainment	2. Was the exposure* adequately ascertained?	Low	Exposure is CAH diagnosis. Authors sought to maximise ascertainment through physician and laboratory notification to BPSU. CAH final diagnosis was by expert panel using well-defined criteria, however it is not reported whether all criteria were satisfied in all cases.
	3. Was the outcome adequately ascertained?	Low	Incidence as defined by those notified as meeting the CAH diagnostic criteria within the UK annual birth rate was well-

			defined. Age at diagnosis is clearly reported.
Causality	4. Was follow-up long enough for outcomes to occur?	N/A	N/A
Reporting	5. Are the cases described with sufficient details to allow other investigators to replicate the research or to allow practitioners to make inferences related to their own practice?	Low	Reporting of cases is sufficiently detailed
Overall is unclear due to uncertainty in selection and ascertainment			

Knowles 2013			
Domain	**Question**	**Judgement on risk of bias**	**Comment**
Selection	1. Do the patients represent the whole experience of the investigator (centre) or is the selection method unclear to the extent that other patients with similar presentation may not have been reported?	Unclear	More detail is given on likely response rate from paediatricians and surveillance laboratories than is reported in Khalid. However the capture rate is nevertheless unclear.
Ascertainment	2. Was the exposure* adequately ascertained?	Low	Exposure is CAH diagnosis. Authors sought to maximise ascertainment through physician and laboratory notification to BPSU. CAH final diagnosis was by expert panel using well-defined criteria, however it is not reported whether all criteria were satisfied in all cases.
	3. Was the outcome adequately ascertained?	Low	Age at diagnosis is clearly reported.

Causality	4. Was follow-up long enough for outcomes to occur?	N/A	N/A
Reporting	5. Are the cases described with sufficient details to allow other investigators to replicate the research or to allow practitioners make inferences related to their own practice?	Low	Reporting of cases is sufficiently detailed
Overall judgement unclear due to uncertainty around capture rate			

Maccabee-Ryaboy 2016			
Domain	**Question**	**Judgement on risk of bias**	**Comment**
Selection	1. Do the patients represent the whole experience of the investigator (centre) or is the selection method unclear to the extent that other patients with similar presentation may not have been reported?	High	Selection is of all CAH patients identified by review of medical records, however the study focuses only on patients with 3 blood pressure readings
Ascertainment	2. Was the exposure* adequately ascertained?	Low	Diagnosis of CAH is sufficiently described
	3. Was the outcome adequately ascertained?	Low	Age at diagnosis sufficiently described and break down by CAH subtype reported
Causality	4. Was follow-up long enough for outcomes to occur?	N/A	N/A
Reporting	5. Are the cases described with sufficient details to allow other investigators to replicate the	Unclear	Focus is on factors associated with hypertension, little demographic information

	research or to allow practitioners to make inferences related to their own practice?		
Overall judgement Unclear risk due to high risk of selection bias and insufficient description of cases			

Pijnenburg-Kleizen 2019			
Domain	**Question**	**Judgement on risk of bias**	**Comment**
Selection	1. Do the patients represent the whole experience of the investigator (centre) or is the selection method unclear to the extent that other patients with similar presentation may not have been reported?	Low	All patients with CAH identified through retrospective review
Ascertainment	2. Was the exposure* adequately ascertained?	Low	Clear description of diagnostic criteria
	3. Was the outcome adequately ascertained?	Unclear	Age of diagnosis given, no description of age at presentation of symptoms
Causality	4. Was follow-up long enough for outcomes to occur?	N/A	N/A
Reporting	5. Are the cases described with sufficient details to allow other investigators to replicate the research or to allow practitioners to make inferences related to their own practice?	Unclear	Little description of demographic characteristics of cases
Overall judgement Low due to unselected cohort and low risk of ascertainment bias			

Table 24. Quality assessment of studies included for Q3 (QUADAS II)

See Appendix 4 for scoring criteria.

Domain/sub-questions	Boelen 2016	Gaudl 2019	Han 2019	Heather 2015	Pearce 2016	Tang 2016	van der Linde 2019
Domain 1 – patient selection							
A. Risk of Bias							
SQ1. Was a consecutive or random sample of patients enrolled?	Unclear - details limited	Unclear - details limited	Unclear - details limited	Yes - national cohort	Yes	Yes	Yes - national cohort with 99.7% coverage
SQ2.Was a case-control design avoided?	Unclear - details limited	No - recruited 50% false positive	No - recruited screen negative controls and clnically confirmed CAH cases	Yes	Yes	Yes	Yes
SQ3. Did the study avoid inappropriate exclusions?	Unclear - details limited	Unclear - details limited	Unclear - details limited	Unclear - unclear how non-severe cases were used in the analysis	Yes	High, excluded 2006-2012	Yes
Summary: Could the selection of patients have introduced bias?	Unclear	High	High	Unclear - were non-severe cases excluded?	Low	High - some analyses were of 2006-2013 whilst some only of 2013	Low

Domain/sub-questions	Boelen 2016	Gaudl 2019	Han 2019	Heather 2015	Pearce 2016	Tang 2016	van der Linde 2019
B. Concerns regarding applicability							
Is there concern that the included patients and settings do not match the review question?	Unclear	Unclear	Unclear	Unclear - were non-severe cases excluded?	Low	Low	Low
Domain 2: Index test(s)							
A. Risk of Bias							
SQ1. Were the index test results interpreted without knowledge of the results of the reference standard?	Unclear - details limited	Unclear if cases and controls were blinded	Unclear if cases and controls were blinded	Unclear - if patient had strong clinical presentation, did this influence interpretation?	Unclear - unclear what the reference standard was so can't judge	Unclear - if patient had strong clinical presentation, did this influence interpretation?	Unclear - if patient had strong clinical presentation, did this influence interpretation?

Domain/sub-questions	Boelen 2016	Gaudl 2019	Han 2019	Heather 2015	Pearce 2016	Tang 2016	van der Linde 2019
SQ2. If a threshold was used, was it pre-specified?	Yes	Unclear - cut offs not reported	No - ROC curves used to determine best cut-off	Yes	Yes	Yes	Yes
Summary: Could the conduct or interpretation of the index test have introduced bias?	Unclear	Unclear	High	Unclear	Unclear	Unclear	Unclear
B. Concerns regarding applicability							
Is there concern that the index test, its conduct, or interpretation differ from the review question?	Unclear - not reported	Unclear - not reported	Unclear - not reported	Unclear - after 48 hours of life	Unclear - 24–48 h after birth	Unclear - 12-48 hours after birth	Low - 3-7 days after birth
Domain 3. Reference standard.							

Domain/sub-questions	Boelen 2016	Gaudl 2019	Han 2019	Heather 2015	Pearce 2016	Tang 2016	van der Linde 2019
A. Risk of bias							
SQ1. Is the reference standard likely to correctly classify the target condition?	Yes - genetic testing (may miss some patients), 21-Hydroxylase deficiency CAH	Unclear - details NR	Unclear - unclear what tests performed and unclear which CAH types	Unclear - included long term follow-up, but did not define what "severe CAH" was	No - not clear what tests were done and included other forms of CAH	Unclear - included long term follow-up, but did not state which CAH	No - included LC-MS/MS and genetic confirmation, but also included other forms of CAH
SQ2. Were the reference standard results interpreted without knowledge of the results of the index test?	No - differential reference standard (genetic confirmation differentially applied)	Unclear - reference standard not described	Unclear - reference standard not described in enough detail	No - differential reference standard	No - differential reference standard	No - differential reference standard	No - differential reference standard
Summary: Could the reference standard, its conduct, or its interpretation have introduced bias?	High	Unclear	Unclear	High	High	High	High

Domain/sub-questions	Boelen 2016	Gaudl 2019	Han 2019	Heather 2015	Pearce 2016	Tang 2016	van der Linde 2019
B. Concerns about applicability							
Is there concern that the target condition as defined by the reference standard does not match the review question?	Low genetic test is for 21-hyroxylase deficiency	Unclear	Unclear if only 21-hydroxlyase deficiency	Unclear - "severe" CAH	High - includes non-21-hydroxylase deficiency CAH.	Unclear	High - other forms of CAH
Domain 4: Flow and timing							
A. Risk of bias							
SQ1. Was there an appropriate interval between index test(s) and reference standard?	Unclear when newborns were recruited and therefore how long follow-up was	Unclear - reference standard not described	Unclear - reference standard not described in enough detail	No (too short since cohort recruited)	No (too short since cohort recruited)	No (too short since cohort recruited)	Unclear how old included patients were at the time of the analysis - some cases may yet manifest.

Domain/sub-questions	Boelen 2016	Gaudl 2019	Han 2019	Heather 2015	Pearce 2016	Tang 2016	van der Linde 2019
SQ2. Did all patients receive a reference standard?	Unclear - does not state that all patients were genetically tested	Unclear - reference standard not described	Unclear - reference standard not described in enough detail	No - Differential verification dependent on known or unknown factors	Yes - avoided partial verification due to long term surveillance	Yes - avoided partial verification due to long term surveillance	Yes - avoided partial verification due to long term surveillance
SQ3. Did patients receive the same reference standard?	Unclear - does not state that all patients were tested	Unclear - reference standard not described	Unclear - reference standard not described in enough detail	No - Differential verification dependent on known or unknown factors	No - differential verification bias, unclear which	No - differential verification reference standard	No - Differential verification dependent on known or unknown factors, as clinical signs may have contributed to some patients getting referrals (states a proportion were identified clinically)
SQ4. Were all patients included in the analysis?	Yes	Yes	Yes	Unclear if severe excluded	Yes	No - only 2013 for TN & FP	No, missing data excluded
Summary of Q 1 to 4:	Unclear	Unclear	Unclear	High	High	High	High

Appendix 4 – QUADAS II risk of bias scoring schedule adapted to the specifics of this evidence summary, in accordance with Whiting *et al.* 2011

Domain 1: Patient selection	
A. Risk of bias	
Describe methods of patient selection:	
SQ 1: Was a consecutive or random sample of patients enrolled? *Score **yes** if states consecutive or random, or is all results from a national or regional screening programme* *Score **no** if states another method of patient sampling/selection* *Score **unclear** if unclear*	Yes/No/Unclear
SQ 2: Was a case-control design avoided? *Score **yes** if not case control* *Score **no** if case control* *Score **unclear** if unclear*	Yes/No/Unclear
SQ 3: Did the study avoid inappropriate exclusions? *Score **yes** if the study selected newborns, or newborns positive at first tier testing if the index test is a second-tier test*	Yes/No/Unclear
Summary: Could the selection of patients have introduced bias?	**RISK: LOW/HIGH/UNCLEAR**
B. Concerns regarding applicability	
Describe included patients (prior testing, presentation, intended use of index test and setting):	

Is there concern that the included patients do not match the review question? *Score **yes** if the study selected newborns, or newborns positive at first tier testing if the index test is a second-tier test* *Score **no** if the study has made inappropriate exclusions from the group it set out to select.* *Score **unclear** if it is unclear.*	**CONCERN: LOW/HIGH/UNCLEAR**
Domain 2: Index test(s).	
A. Risk of bias	
Describe the index test and how it was conducted and interpreted:	
SQ1: Were the index test results interpreted without knowledge of the results of the reference standard? *Where the reference standard includes clinical presentation, signs and symptoms,, this item may score unclear where clinical signs form part of the reference standard. Since they are evident before the index test is interpreted, they may influence interpretation of borderline tests. The extent to which this may influence test interpretation is unclear.* *Score **yes** if index test was interpreted blind to the reference standard or the index test was clearly interpreted before the reference standard was known* *Score **no** if results of reference standard were already known e.g. clinical signs and symptoms formed part of reference standard* *Score **unclear** if unclear*	Yes/No/Unclear
SQ 2: If a threshold was used, was it pre-specified? *Score **yes** if pre-specified cut of values were used (validation study)*	Yes/No/Unclear

*Score **no** if cut-off values were fitted to the data (derivation study)* *Score **unclear** if this is unclear*	
Summary: Could the conduct or interpretation of the index test have introduced bias?	**RISK: LOW/HIGH/UNCLEAR**
B. Concerns regarding applicability	
Is there concern that the index test, its conduct, or interpretation differ from the review question? *Tests performed within the same time period as DBS collection in England (4-8 days) can score **Low*** *Tests performed outside this period can score **Unclear,** since it is unclear whether time of sample collection influences diagnostic accuracy*	**CONCERN: LOW/HIGH/UNCLEAR**
Domain 3: Reference standard	
A. Risk of bias	
Describe the reference standard and how it was conducted and interpreted:	
SQ 1: Is the reference standard likely to correctly classify the target condition? *None of the reference standards available have 100% accuracy. The target condition is 21-hydroxylase deficiency CAH* *The following are acceptable, but their limitations should be noted:* **LC-MS/MS** – *some methodologies may produce false negatives and/or flase positives (classify people with the disease as negative, and classify people without the disease positive)* **Genetic testing** – *may produce false negatives (classify people with the disease as not having the disease) since not all mutations are known* **Long-term follow-up/CAH registry data** – *where this is based on a national register, may produce false negatives (classify people with the disease as not having the disease) unless participation is 100% and*	Yes/No/Unclear

follow-up is life-long (this may still miss mild cases or severe cases that die before diagnosis). *If the reference standard is a genuine composite (all tests performed in all patients and a scoring system is in place to determine outcome), this is also acceptable, but please note this against score.* *The reference standard should not include other forms of CAH, e.g. 11-B hydroxylase deficiency. Please note this against score if it is the reason for scoring no or unclear.* *Score **Yes** if one of the above or another reference standard that would correctly classify the tartget condition* *Score **No** if reference standard unlikely to correctly classify the target condition* *Score **Unclear** if its unclear*	
SQ 2: Were the reference standard results interpreted without knowledge of the results of the index test? *In the case of tiered testing, this is likely not to be the case.*	Yes/No/Unclear
Summary: Could the reference standard, its conduct, or its interpretation have introduced bias?	**RISK: LOW/HIGH/UNCLEAR**
B. Concerns regarding applicability	
Is there concern that the target condition as defined by the reference standard does not match the review question? *Score **low risk** if the target condition is 21-hydrosylase deficiency or one of its subtypes* *Score **high risk** if the target includes forms other than 21-hydroxylase deficiency.* *Score **unclear** if the target condition is unclear*	CONCERN: LOW/HIGH/UNCLEAR

Domain 4: Flow and timing	
A. Risk of bias	
Describe any patients who did not receive the index test(s) and/or reference standard or who were excluded from the 2x2 table (refer to flow diagram): **Describe the time interval and any interventions between index test(s) and reference standard:**	
SQ 1: Was there an appropriate interval between index test(s) and reference standard? *For a genetic reference standard, this can be taken at any time and will give the same result.* *For long term follow-up, It is unclear how late 21-hydroxylase deficiency CAH (e.g. non-classical) may present clinically. These patients can present in adulthood. Therefore follow-up less than 18 years, score No; follow-up 18-50 years, score Unclear; follow-up 50+ years score Yes. This is an arbitrary cut-off.*	Yes/No/Unclear
SQ 2: Did all patients receive a reference standard? *Score **yes** if all patients got a reference standard, even if these were different (see next question)* *Score **no** if a* **partial verification reference standard:** *only some participants get any reference standard, e.g. those who test negative at first tier testing don't get followed up or any further tests* *Score **unclear** if it is unclear who received the reference standard*	Yes/No/Unclear
SQ 3: Did patients receive the same reference standard? *The following score "no", please also include the category:* **Partial verification reference standard:** *only some participants get any reference standard, e.g. those who test negative at first tier testing don't get followed up or any further tests*	Yes/No/Unclear

Complete index test-dependent differential verification reference standard: *participants get a different reference standard according to the index test result, e.g. 1st tier negative get clinical follow-up, whilst 1st tier positive get diagnostic work-up/genetic testing* **Differential verification dependent on known or unknown factors:** *participants get a different reference standard according to some known or unknown factors, e.g. those with clinical signs or symptoms or 1st tier positive proceed to diagnostic work-up/genetic testing* *The following score "yes":* *All receive the same reference standard, e.g. all were tested genetically, regardlesss of index test result.*	
SQ 4: Were all patients included in the analysis?	Yes/No/Unclear
Summary: Could the patient flow have introduced bias?	**RISK: LOW/HIGH/UNCLEAR**

Appendix 5 – Deprioritised studies for question 3 on diagnostic accuracy

The following papers relevant to question 3 were deprioritised because they did not report sufficient diagnostic accuracy data. They were not formally extracted nor quality appraised. They were briefly summarised in a narrative synthesis in the main body of the evidence summary. They are reported for information in the table below.

Table 25. Deprioritised studies for question 3 on diagnostic accuracy

Author, year	Tier of testing	Test	Marker	Reference standard	Data available or calculable
De Hora, M. R., et al. (2020). New Zealand	Second-tier test	LC-MS/MS as 2nd tier	17-OHP	Unclear, included long term follow up. Notes some fn, but doesn't quantify	TP, FP, PPV, specificity
	Second-tier test	Immunoassay as 2nd tier	17-OHP		
Fox, D. A., et al. (2020). Canada	First-tier test	Immunoassay	17-OHP	Genetic test as clinically indicated	TP, FP, PPV
	Second-tier test	LC-MS/MS as 2nd tier	17-OHP and ratio of 17-OHP and 21-deoxycortisol/cortisol		
Held, P. K., et al. (2015). USA	One-tier testing	Immunoassay as first and only tier	17-OHP	Screening and cases volunteered by centres as missed cases (not comprehensive coverage)	TP, FN, Sens
	States with second screen (variable, unclear if second-tier or re-test, some resample)	Variable and unclear methodology	Unclear, possibly variable?		

Monostori, P., et al. (2015). Romania	First-tier test	Immunoassay	17-OHP	LC-MS/MS as next row	TP, FP, PPV
	Second-tier test	LC-MS/MS as 2nd-tier	21-dexoxycortisol and 11-deoxycortisol	Unclear	TP only
Speiser, P.W., et al 2020 USA	One-tier testing	Fluoroimmunoassay	17-OHP	Serum tests and confirmatory genetic testing	TP, FP, TN, FN assumed zero; sensitivity, specificity, NPV, PPV

Abbreviations: 17-OHP, 17α-hydroxyprogesterone; 21-OHP, FP, false positive, FN, false negative; LC–MS/MS, liquid chromatography with tandem mass spectrometry; NPV, negative predictive value; PPV, positive predictive value; SW, salt wasting; SV, simple virilising; TC, target condition; TP, true positive; TN, true negative

Appendix 6 – UK NSC reporting checklist for evidence summaries

All items on the UK NSC Reporting Checklist for Evidence Summaries have been addressed in this report. A summary of the checklist, along with the page or pages where each item can be found in this report, is presented in Table 2.

Table 26. UK NSC reporting checklist for evidence summaries

	Section	Item	Page no.
1.	TITLE AND SUMMARIES		
1.1	Title sheet	Identify the review as a UK NSC evidence summary.	Title page
1.2	Plain English summary	Plain English description of the executive summary.	4
1.3	Executive summary	Structured overview of the whole report. To include: the purpose/aim of the review; background; previous recommendations; findings and gaps in the evidence; recommendations on the screening that can or cannot be made on the basis of the review.	6
2.	INTRODUCTION AND APPROACH		
2.1	Background and objectives	Background – Current policy context and rationale for the current review – for example, reference to details of previous reviews, basis for current recommendation, recommendations made, gaps identified, drivers for new reviews	11
		Objectives – What are the questions the current evidence summary intends to answer? – statement of the key questions for the current evidence summary, criteria they address, and number of studies included per question, description of the overall results of the literature search.	13 14
		Method – briefly outline the rapid review methods used.	

2.2	Eligibility for inclusion in the review	State all criteria for inclusion and exclusion of studies to the review clearly (PICO, dates, language, study type, publication type, publication status etc.) To be decided *a priori*.	14
2.3	Appraisal for quality/risk of bias tool	Details of tool/checklist used to assess quality, e.g. QUADAS 2, CASP, SIGN, AMSTAR.	18
3.	SEARCH STRATEGY AND STUDY SELECTION (FOR EACH KEY QUESTION)		
3.1	Databases/ sources searched	Give details of all databases searched (including platform/interface and coverage dates) and date of final search.	14
3.2	Search strategy and results	Present the full search strategy for at least one database (usually a version of Medline), including limits and search filters if used. Provide details of the total number of (results from each database searched), number of duplicates removed, and the final number of unique records to consider for inclusion.	Appendix 1
3.3	Study selection	State the process for selecting studies – inclusion and exclusion criteria, number of studies screened by title/abstract and full text, number of reviewers, any cross checking carried out.	14-18
4.	STUDY LEVEL REPORTING OF RESULTS (FOR EACH KEY QUESTION)		
4.1	Study level reporting, results and risk of bias assessment	For each study, produce a table that includes the full citation and a summary of the data relevant to the question (for example, study size, PICO, follow-up period, outcomes reported, statistical analyses etc.). Provide a simple summary of key measures, effect estimates and confidence intervals for each study where available. For each study, present the results of any assessment of quality/risk of bias.	Study level reporting: 112-155 Quality assessment: 156-175

5.	QUESTION LEVEL SYNTHESIS		
5.1	Description of the evidence	For each question, give numbers of studies screened, assessed for eligibility, and included in the review, with summary reasons for exclusion.	20,33,47
5.2	Combining and presenting the findings	Provide a balanced discussion of the body of evidence which avoids over reliance on one study or set of studies. Consideration of four components should inform the reviewer's judgement on whether the criterion is 'met', 'not met' or 'uncertain': quantity; quality; applicability and consistency.	21,34,49
5.3	Summary of findings	Provide a description of the evidence reviewed and included for each question, with reference to their eligibility for inclusion. Summarise the main findings including the quality/risk of bias issues for each question. Have the criteria addressed been 'met', 'not met' or 'uncertain'?	44,60
6.	REVIEW SUMMARY		
6.1	Conclusions and implications for policy	Do findings indicate whether screening should be recommended? Is further work warranted? Are there gaps in the evidence highlighted by the review?	63
6.2	Limitations	Discuss limitations of the available evidence and of the review methodology if relevant.	64

References

1. Bacila I-A, Elder C, Krone NJAodic. Update on adrenal steroid hormone biosynthesis and clinical implications. 2019;104(12):1223-8.
2. Miller WL, Auchus RJJEr. The molecular biology, biochemistry, and physiology of human steroidogenesis and its disorders. 2011;32(1):81-151.
3. Fox DA, Ronsley R, Khowaja AR, Haim A, Vallance H, Sinclair G, et al. Clinical Impact and Cost Efficacy of Newborn Screening for Congenital Adrenal Hyperplasia. Journal of Pediatrics. 2020;220:101-8.e2.
4. El-Maouche D, Arlt W, Merke DP. Congenital adrenal hyperplasia. Lancet. 2017;390(10108):2194-210.
5. Tajima T, Fukushi M. Neonatal mass screening for 21-hydroxylase deficiency. Clinical Pediatric Endocrinology. 2016;25(1):1-8.
6. 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. 2018;103(11):4043-88.
7. Huynh T, McGown I, Cowley D, Nyunt O, Leong GM, Harris M, et al. The clinical and biochemical spectrum of congenital adrenal hyperplasia secondary to 21-hydroxylase deficiency. 2009;30(2):75.
8. White PC. Neonatal screening for congenital adrenal hyperplasia. Nature Reviews Endocrinology. 2009;5(9):490-8.
9. van der Kamp HJ, Oudshoorn CG, Elvers BH, van Baarle M, Otten BJ, Wit JM, et al. Cutoff levels of 17-α-hydroxyprogesterone in neonatal screening for congenital adrenal hyperplasia should be based on gestational age rather than on birth weight. 2005;90(7):3904-7.
10. Janzen N, Peter M, Sander S, Steuerwald U, Terhardt M, Holtkamp U, et al. Newborn screening for congenital adrenal hyperplasia: additional steroid profile using liquid chromatography-tandem mass spectrometry. 2007;92(7):2581-9.
11. Yau M, Khattab, A., Qamar, A., Nimkarn., S., Lin-Su K, Lekarev O, et al. BMJ Best Practice: Congenital Adrenal Hyperplasia. BMJ. 2017.
12. 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 & Metabolism. 2018;103(11):4043-88.
13. Wilson G. Screening for congenital adrenal hyperplasia. External review against programme appraisal criteria for the UK National Screening Committee UK NSC.; 2015.
14. NSC U. UK NSC: evidence review process 2017 [
15. Munn Z, Moola S, Riitano D, Lisy KJIjohp, management. The development of a critical appraisal tool for use in systematic reviews addressing questions of prevalence. 2014;3(3):123.
16. Murad MH, Sultan S, Haffar S, Bazerbachi FJBe-bm. Methodological quality and synthesis of case series and case reports. 2018;23(2):60-3.
17. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. 2011;155(8):529-36.
18. Khalid JM, Oerton JM, Dezateux C, Hindmarsh PC, Kelnar CJ, Knowles RL. Incidence and clinical features of congenital adrenal hyperplasia in Great Britain. Archives of Disease in Childhood. 2012;97(2):101-6.

19. Knowles RL, Khalid JM, Oerton JM, Hindmarsh PC, Kelnar CJ, Dezateux C. Late clinical presentation of congenital adrenal hyperplasia in older children: findings from national paediatric surveillance. Archives of Disease in Childhood. 2014;99(1):30-4.
20. Speiser PW, Chawla R, Chen M, Diaz-Thomas A, Finlayson C, Rutter MM, et al. Newborn screening protocols and positive predictive value for congenital adrenal hyperplasia vary across the United States. International Journal of Neonatal Screening. 2020;6(2).
21. Held PK, Shapira SK, Hinton CF, Jones E, Hannon WH, Ojodu J. Congenital adrenal hyperplasia cases identified by newborn screening in one- and two-screen states. Molecular Genetics & Metabolism. 2015;116(3):133-8.
22. Eshragh N, Van Doan L, Connelly KJ, Denniston S, Willis S, LaFranchi SHJHRiP. Outcome of newborn screening for congenital adrenal hyperplasia at two time points. 2020;93(2):128-36.
23. Pearce M, Demartino L, McMahon R, Hamel R, Maloney B, Stansfield DM, et al. Newborn screening for congenital adrenal hyperplasia in New York State. Molecular Genetics and Metabolism Reports. 2016;7:1-7.
24. Pearce M, Dauerer E, DiRienzo AG, Caggana M, Tavakoli NP. The influence of seasonality and manufacturer kit lot changes on 17alpha-hydroxyprogesterone measurements and referral rates of congenital adrenal hyperplasia in newborns. European Journal of Pediatrics. 2017;176(1):121-9.
25. Lai F, Srinivasan S, Wiley VJIjons. Evaluation of a two-tier screening pathway for congenital adrenal hyperplasia in the New South Wales Newborn Screening Programme. 2020;6(3):63.
26. Heather NL, Seneviratne SN, Webster D, Derraik JG, Jefferies C, Carll J, et al. Newborn screening for congenital adrenal hyperplasia in New Zealand, 1994-2013. Journal of Clinical Endocrinology & Metabolism. 2015;100(3):1002-8.
27. David J, Chrastina P, Vinohradska H, Al Taji E, Holubova A, Hlidkova E, et al. Neonatal screening in the Czech Republic: increased prevalence of selected diseases in low birthweight neonates. European Journal of Pediatrics. 2018;177(11):1697-704.
28. David J, Chrastina P, Peskova K, Kozich V, Friedecky D, Adam T, et al. Epidemiology of rare diseases detected by newborn screening in the Czech Republic. Central European Journal of Public Health. 2019;27(2):153-9.
29. Dulin Iniguez E, Ezquieta Zubicaray B. Newborn screening of congenital adrenal hyperplasia. 2018.
30. van der Linde AAA, Schonbeck Y, van der Kamp HJ, van den Akker ELT, van Albada ME, Boelen A, et al. Evaluation of the Dutch neonatal screening for congenital adrenal hyperplasia. Archives of Disease in Childhood. 2019;104(7):653-7.
31. Zetterström RH, Karlsson L, Falhammar H, Lajic S, Nordenström AJIJoNS. Update on the Swedish newborn screening for congenital adrenal hyperplasia due to 21-hydroxylase deficiency. 2020;6(3):71.
32. Statistics OfN. 2011 Census: Key statistics and quick statistics for local authorities in the United Kingdom. Office for National Statistics London; 2013.
33. Hoeffel E, Rastogi S, Kim MO, Shahid HJW, DC: US Census Bureau. The Asian population: 2010. 2010 Census Briefs, 2012. 2010.
34. Rudd B. Prevalence of adrenal 21-hydroxylase deficiency in neonates born in the West Midlands: a retrospective study. Practical Developments in Inherited Metabolic Disease: DNA Analysis, Phenylketonuria and Screening for Congenital Adrenal Hyperplasia: Springer; 1986. p. 155-6.
35. Wallace A. Review of CAH screening programmes and the Scottish experience. Practical Developments in Inherited Metabolic Disease: DNA Analysis, Phenylketonuria and Screening for Congenital Adrenal Hyperplasia: Springer; 1986. p. 135-41.

36. Wooding K, G., Gregory, J.W. Congenital adrenal hyperplasia. Research report of the Welsh Paediatric Surveillance Unit.; 2000.
37. McKiernan JJImj. Congenital adrenal hyperplasia in the Republic of Ireland, 1966-1982. 1985;78(1):6-7.
38. Standards for nhs newborn blood spot screening 2020 [Newborn blood spot screening: standards - GOV.UK (www.gov.uk)]
39. Bomberg EM, Addo OY, Kyllo J, Gonzalez-Bolanos MT, Ltief AM, Pittock S, et al. The relation of peripubertal and pubertal growth to final adult height in children with classic congenital adrenal hyperplasia. Journal of Pediatrics. 2015;166(3):743-50.
40. Sarafoglou K, Addo OY, Turcotte L, Otten N, Wickremasinghe A, Pittock S, et al. Impact of hydrocortisone on adult height in congenital adrenal hyperplasia-the Minnesota cohort. Journal of Pediatrics. 2014;164(5):1141-6.e1.
41. Halper A, Sanchez B, Hodges JS, Dengel DR, Petryk A, Sarafoglou K. Use of an aromatase inhibitor in children with congenital adrenal hyperplasia: Impact of anastrozole on bone mineral density and visceral adipose tissue. Clinical Endocrinology. 2019;91(1):124-30.
42. Hsieh S, White PC. Presentation of primary adrenal insufficiency in childhood. Journal of Clinical Endocrinology & Metabolism. 2011;96(6):E925-8.
43. 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.
44. Pijnenburg-Kleizen KJ, Thomas CMG, Otten BJ, Roeleveld N, Claahsen-van der Grinten HL. Long-term follow-up of children with classic congenital adrenal hyperplasia: suggestions for age dependent treatment in childhood and puberty. Journal of Pediatric Endocrinology & Metabolism. 2019;32(10):1055-63.
45. Tang H, Feuchtbaum L, Neogi P, Ho T, Gaffney L, Currier RJ. Damaged goods?: an empirical cohort study of blood specimens collected 12 to 23 hours after birth in newborn screening in California. Genetics in Medicine. 2016;18(3):259-64.
46. Han L, Tavakoli NP, Morrissey M, Spink DC, Cao ZT. Liquid chromatography-tandem mass spectrometry analysis of 17-hydroxyprogesterone in dried blood spots revealed matrix effect on immunoassay. Analytical & Bioanalytical Chemistry. 2019;411(2):395-402.
47. Boelen A, Ruiter AF, Claahsen-van der Grinten HL, Endert E, Ackermans MT. Determination of a steroid profile in heel prick blood using LC-MS/MS. Bioanalysis. 2016;8(5):375-84.
48. Gaudl A, Kratzsch J, Ceglarek U. Advancement in steroid hormone analysis by LC-MS/MS in clinical routine diagnostics - A three year recap from serum cortisol to dried blood 17alpha-hydroxyprogesterone. Journal of Steroid Biochemistry & Molecular Biology. 2019;192:105389.
49. Abbo O, Ferdynus C, Kalfa N, Huiart L, Sauvat F, Harper LH. Male infants with hypospadias and/or cryptorchidism show a lower 2D/4D digit ratio than normal boys. Archives of Disease in Childhood. 2015;100(7):643-7.
50. De Hora MR, Heather NL, Patel T, Bresnahan LG, Webster D, Hofman PL. Measurement of 17-hydroxyprogesterone by LCMSMS improves newborn screening for CAH due to 21-hydroxylase deficiency in New Zealand. International Journal of Neonatal Screening. 2020;6(1).
51. Monostori P, Szabo P, Marginean O, Bereczki C, Karg E. Concurrent Confirmation and Differential Diagnosis of Congenital Adrenal Hyperplasia from Dried Blood Spots: Application of a Second-Tier LC-MS/MS Assay in a Cross-Border Cooperation for Newborn Screening. Hormone research in paediatrics. 2015;84(5):311-8.