NHS RACE & HEALTH OBSERVATORY

Review of neonatal assessment and practice in Black, Asian, and minority ethnic newborns

Exploring the Apgar score, the detection of cyanosis, and jaundice

–

July 2023

Sheffield Hallam University | Centre for Applied Health & Social Care Research

The University of Nottingham

NHS Bradford Teaching Hospitals NHS Foundation Trust

Authors

Frankie Fair, Sheffield Hallam University
Amy Furness, Sheffield Hallam University
Gina Higginbottom, Nottingham University
Sam Oddie, Bradford Teaching Hospitals NHS Foundation Trust
Hora Soltani, Sheffield Hallam University

Corresponding Author

Professor Hora Soltani

Email

h.soltani@shu.ac.uk

Authors professions

Frankie Fair	Midwife
Hora Soltani	Midwife-Professor of Maternal and Infant Health
Gina Higginbottom	Health Visitor, Midwife, Professor Emerita Ethnicity and Health
Sam Oddie	Consultant Neonatologist

Contents

List of Figures

List of Tables

Acknowledgements

We would like to thank:

- The NHS Race and Health Observatory Maternal and Neonatal Health Advisory Group, particularly Arnie Puntis and Sam Rodger, for their support during this review.

- Ghazaleh Oshaghi and Kayla Baugh for their support in screening the citations by title and Hannah Thompson for support with data extraction within the healthcare professionals' systematic review.

- John Kirkby and Nick Waddington from Design Futures for developing the infographics.

- All participants who kindly gave of their time to take part in the interviews and provide such valuable information.

- Our advisory group for their support throughout different aspects of the project and all of the stakeholders who took part within the workshops to discuss the project findings. Advisory group members and workshop participants who agreed to be named are listed below.

Vicky Gilroy – *Institute of Health Visiting*

Janet Fyle – *Royal College of Midwives*

Sara Webb – Head of Midwifery Information and Research Services, *Royal College of Midwives*

Alessandra D'Angelo – *Royal College of Midwives*

Dr Kerrie Stevenson – NHS Public Health Doctor, *University College London*

Zenab Barry – Director at *National Maternity Voices*, Strategic Adviser at *NIHR Applied Research Collaboration (ARC) South London's Maternity and Perinatal Mental Health*, Maternal Health Advocate

Gemma Poole – Midwife and CEO of *Essential Baby company*

Josie Anderson – Policy, Research and Campaigns Manager, *Bliss*

Natasha Ratcliffe – *Couch Health*

Bridget Ogharanduku – Senior lecturer at *SHU*, Community member

Ghazaleh Oshaghi – ORAMMA Friendly Mother co-ordinator
(Migrant mothers/Community member)

Emma Ehizojie – Community member

Yasmin Iqbal – Community engagement worker

Elham Kohta – Community engagement worker

Natasha Mercier – Health Protection Engagement Co-ordinator,
Doncaster Council

Kath Nakielski – Course lead BSc Midwifery, *Sheffield Hallam University*

Victoria Schofield – Midwifery lecturer, *Sheffield Hallam University*

Rachel Best – Subject Group Lead for Midwifery, *Sheffield Hallam University*

Hannah Thompson – Midwifery lecturer, *Sheffield Hallam University*

Alison Salmon – Midwifery lecturer, *Sheffield Hallam University*

Sally Freeman – Senior Midwifery lecturer, *Sheffield Hallam University*

Foreword

Professor Jacqueline Dunkley-Bent
Co-Chair of The RHO Maternal and Neonatal Health Working Group

Dr Daghni Rajasingam
Co-Chair of The RHO Maternal and Neonatal Health Working Group

The NHS Race and Health Observatory Rapid Evidence Review, conducted in 2022, shed light on the existing ethnic inequalities within neonatal health outcomes. Until now, very little work has been undertaken to understand if current guidance for assessing a baby's condition after birth, and in particular the colour component of the APGAR score, is suitable for Black, Asian or minority ethnic babies. This prompted research to review neonatal assessment and practice in Black, Asian and minority ethnic newborns.

This review reveals areas in which existing guidance for assessing a baby's condition after birth is unsuitable for Black, Asian, or minority ethnic babies. This investigation examines and critiques terminology used to describe a baby's colour, particularly in medical diagnoses, where language and practice are based on a white normative standard. This biased assessment is exemplified by terms like 'pink' being used to describe a well-perfused baby, disregarding the diversity of skin colours within our population. Consequently, it raises concerns about the clinical accuracy of such assessments when applied to ethnically diverse populations.

These findings show that health inequalities among ethnic minority neonates underscore the lack of inclusivity in teaching resources that do not always adequately represent the diversity of Black, Asian, and minority ethnic communities.

One specific area of concern highlighted in the report is the subjective nature of clinical guidelines for assessing jaundice, heavily relying on changes in skin colour. Despite some clinical guidance and clinicians recognising the problematic and inaccurate nature of using skin colour to assess jaundice in Black and brown babies, this awareness has not translated into policy changes, updated guidance documents, or comprehensive healthcare education. This report emphasises the pressing need for more objective outcome measures

to mitigate the impact of racial bias when employing these assessments and serves as an urgent call to action, for transformative change in clinical guidance and policy; education of healthcare professionals; and public awareness of the assessment of hypoxia, cyanosis, and jaundice in babies from ethnically diverse populations.

By rectifying these anomalies that are present in our current practices, we can strive towards a more equitable healthcare system that upholds the health and well-being of all newborns, irrespective of their ethnic background. It is our collective responsibility to ensure that every baby receives the highest standard of care, free from biases and disparities.

Our approach to language

Ethnicity is recognised to be a complex, multidimensional concept, often defined by features such as a shared history, common cultural traditions, shared religion, a common geographical origin, language and literature. It is a highly subjective classification that an individual is usually required to articulate into a simple category. Recording of ethnicity is therefore poor, with only 85% of people in both the UK Hospital Episode Statistics and Clinical Practice Research Datalink having the same ethnicity recorded within both databases (Mathur et al., 2014).

Within this report we will be specific where possible about the ethnic groups we are referring to. Following NHS Race and Health Observatory guidance (NHS Race & Health Observatory, 2021), where there is a need to refer to more than one ethnic group at a time, the report will use the terms 'Black, Asian, and minority ethnic', or 'ethnic minority' or 'Black and minority ethnic' interchangeably, to reflect the varying views of stakeholders on language and representation. We acknowledge the focus of our report is on neonatal assessment and skin tone (phenotype) and that this does not necessarily represent ethno-cultural groups or nationality. We recognised that not all ethnic minorities have darker skin tones, but also recognise that the terms Black and Asian do not cover all of those with darker skin. For clarity, skin colour is used in the report when referring to assessing the skin for different colours such as pink, blue, yellow as part of a medical assessment. Skin tone is used when referring to skin pigmentation.

Further complexity arises as the North America literature, which builds a substantial proportion of the evidence base, commonly used the term 'race' rather than ethnicity. While we acknowledge that the term 'race' is contested (Nazroo & Williams, 2006) within the studies included within the reviews, the original ethnicity or race terms used to detail the included participants have been reported to ensure accurate representation of the studies' results. Additionally, direct quotations from the stakeholder interviews may use pre-existing terms that would not otherwise be the preferred terminology of the report authors nor of the NHS Race and Health Observatory.

Abbreviations

AGREE II Appraisal of Guidelines for Research & Evaluation II instrument

ABC Airway, breathing, circulation

AUC Area Under Curve

BAME Black, Asian, and minority ethnic[1]

BAPM British Association of Perinatal Medicine

CI Confidence interval

CINAHL Cumulated Index to Nursing and Allied Health Literature

COVID-19 SARS-CoV-2 (Coronavirus)

GMC General Medical Council

HCP Healthcare Professionals

HDI High Development Index

HEE Health Education England

HV Health visitor

iHV Institute of Health Visiting

MBRRACE Mothers and Babies: Reducing Risk through Audits and Confidential Enquiries across the UK

MMAT Mixed Methods Appraisal Tool

MW Midwife

NHS National Health Service

NICE National Institute for Health and Care Excellence

[1] Not the preferred term of authors of this review, but used by interviewees within quotations

NIPE Newborn and Infant Physical Examination

NNA Neonatal Nurses Association

NNP Neonatologist, paediatrician or neonatal nurse

NNU Neonatal Unit

NMC Nursing and Midwifery Council

OB Obstetrician

OHID Office for Health Improvement and Disparities

OR Odds ratio

PA Parent

PRISMA Preferred Reporting Items for Systematic Reviews and Meta-Analyses

RCM Royal College of Midwives

RCN Royal College of Nursing

RCOG Royal College of Obstetricians and Gynaecologist

RCPCH Royal College of Paediatrics and Child Health

ROC Receiver operating curve

SBr Total serum bilirubin

SaO2 Arterial oxygen saturation

SpO2 Pulse oximetry saturation

TCB Transcutaneous bilirubin level

UK United Kingdom

USA United States of America

WHO World Health Organization

Executive summary

Ethnic inequalities in maternal and neonatal healthcare provision are increasingly being recognised across the UK (Draper at al., 2022; Knight et al 2022). There are particular concerns around perinatal practices including Apgar scores and assessment of cyanosis and jaundice, all of which have been developed based on White European babies and normalised regardless of their applicability to diverse populations and neonates with varying skin tones (Kapadia et al., 2022).

This report presents the findings and recommendations from a review of neonatal assessments which, because of their focus on skin colour, have the potential to disadvantage babies with darker skin tones. The research involved a triangulation of approaches, including systematic reviews and qualitative research, to gather feedback and experiences from healthcare professionals and parents from Black, Asian, and minority ethnic groups. The study investigated three main areas of neonatal assessment, namely Apgar score, cyanosis, and jaundice. Additionally, the report examined existing policies relating to skin colour and neonatal assessment, with a focus on these same areas. This work has been conducted by academics and clinicians from Sheffield Hallam University, Bradford Teaching Hospitals and Bradford Institute of Health Research, Hull York Medical School, and the Royal College of Paediatrics and Child Health, alongside maternity and neonatal service user representatives.

Specific objectives of the review included:

- Reviewing the use of the Apgar score, with particular focus on the appearance component, as well as the more general detection of cyanosis in Black, Asian, and minority ethnic neonates within the relevant literature and among stakeholders.

- Reviewing the identification of jaundice in Black, Asian, and minority ethnic neonates within the relevant literature and among stakeholders.

- Examining the wider use of terminology used in policy and practice to ensure Black, Asian, and minority ethnic neonates receive equitable care and treatment.

- A broader analysis of potential inequalities for Black, Asian, and minority ethnic neonates within the UK.

A mixed methods approach was adopted to achieve the above objectives, comprising a desktop review of policy documents and systematic reviews, and stakeholder interviews. The data collection was carried out concurrently and results were combined with an equal weighting in a complementary paradigm.

Summary findings and recommendations are presented based on approach and topic area below.

Systematic review of healthcare professionals' experiences of providing care for Black or Asian or minority ethnic neonates in relation to neonatal assessment (Apgar, Cyanosis and Jaundice)

Out of 217 studies which were screened at full text, 56 studies were included as they reported comparative results between different ethnic minority backgrounds for the topics of interest. Six studies related to Apgar scores, three considered the detection of cyanosis, with the remainder considering the detection of jaundice. All but four of the studies were undertaken in developed countries. The impact of ethnicity and race in Apgar scoring remains unclear as studies also highlighted inconsistent scoring of the Apgar, with Black infants more likely to receive lower scores.

Detection of cyanosis from skin colour was noted to lead to many false positives and false negatives. Within one small study, the lips were considered the most reliable places to assess cyanosis across different ethnicities. A further study suggested that tongue colour being pink within the first ten minutes of life suggested that a neonate had an oxygen saturation above 70% regardless of ethnicity. Pulse oximetry is more reliable than observation of the skin, although one study has found occult hypoxemia, where pulse oximetry levels were 90% or more when arterial oxygen saturation was <85%, was slightly more common in Black infants (9.2%) compared to White infants (7.7%). Further exploration of the small but potentially clinically significant differences in arterial oxygen saturation compared to pulse oximetry in neonates from diverse race and ethnicity backgrounds is warranted.

Transcutaneous bilirubin (TCB) measurement has been found to be a more accurate method of detecting jaundice than visual assessment alone. However, when it comes to considering ethnicity, there is no consistent pattern regarding the correlation between TCB and serum bilirubin (SBr) or mean bias (TCB reading – SBr level). Some studies have found higher values in individuals with dark skin tones or from ethnic minority backgrounds, while other studies found lower values or no difference.

In terms of the utility of bilirubinometers, outstanding discrimination was seen across all ethnic and skin tone subgroups and across a range of SBr cut-offs, except for the Bilimed brand. Screening full-term neonates with transcutaneous

bilirubinometers may result in more blood testing needed in neonates of 'non-White' ethnicity or darker skin tones than in neonates of White or light skin.

Additionally, five of the seven studies that statistically assessed the limits of agreement around the mean bias (TCB reading -SBr level) found that variability was increased with increased pigmentation, with most of these TCB measurements noted to have been taken from the forehead. However, none of the included studies found the mean difference between TCB and SBr to be underestimated by more than 40µmol/l in neonates of any ethnicity.

Two studies considered healthcare professionals training. In one study, midwives and student midwives reported lack of education in clinical assessment of Black, Asian, and minority ethnic mothers and babies, with the study particularly focussing on assessment of the Apgar score. After the training, 96% of midwives felt the Apgar score was not the most appropriate way to assess all babies at birth due to the description of colour within it. The second study found maternity care assistants' knowledge and self-perceived confidence in detecting jaundice was not associated with actual ability to correctly estimate SBr levels from visual assessment.

Systematic review of the experiences of parents or carers seeking or receiving care for Black, Asian and ethnic minority neonates in relation to neonatal assessments (Cyanosis and Jaundice)

Out of 110 studies which were screened at full text, nine studies were included in this review. No studies conducted in high income countries considered the views of Black, Asian or minority ethnic parents or carers of the Apgar score or how parents from ethnic minority backgrounds may use skin colour to detect whether their baby is getting enough oxygen or subsequently seeking care. Just two studies reported incidences where mothers talked about their baby's struggles to breathe as part of a wider study of care experiences.

Only seven studies considered parents from Black or minority ethnic backgrounds, experiences of neonatal jaundice, of which only three studies considered the identification of jaundice. These three studies reported jaundice to be inadequately identified in their infant by healthcare professionals with concerns dismissed by a healthcare professional in one study and the woman not wanting to make a fuss by raising her concerns in another study. The other studies focussed on parents' inadequate understanding of jaundice when there was a language barrier, the anxiety they felt after a jaundice diagnosis, as well as how well they remembered the jaundice diagnosis over the following months.

Policy review to examine current policies and guidance in relation to their consideration of Black, Asian and ethnic minority neonates in common assessments (Apgar scoring, cyanosis and jaundice)

There were 18 policy and guidelines reports and one training resource in this review. Of the 18 policy documents, nine focused on general care of neonates, six exclusively detailed the assessment of cyanosis or hypoxia, one neonatal assessment using the Apgar score, and two exclusively focussed on the assessment of jaundice. Of the 18 policies, 15 were UK polices and three were international policies from WHO.

UK policy was assessed to understand whether ethnicity and race were appropriately considered in policy formulation regarding skin colour and neonatal examinations. The review showed that the impact of ethnicity was poorly considered during policy formulation and the development of most guidelines and training. Multiple policies referred to terms such as “pink” “blue” “pale” or “pallor” in reference to neonatal skin, without detailing how these skin colour descriptors may appear in ethnic minority neonates. These results further perpetuate the inequalities faced by those from Black, Asian, and ethnic minority backgrounds, by means of improper assessment or potential late diagnosis.

Stakeholder interviews

A focussed ethnography approach was used for stakeholder engagement, including semi-structured interviews of healthcare professionals (HCPs) and parents or carers. A systematic inductive approach was undertaken to analyse qualitative data using NVivo.

A total of 33 HCPs were interviewed, including thirteen midwives, eight health visitors, four paediatricians, three obstetricians, three neonatologists, and two neonatal nurses. Eleven HCP participants described themselves as Black, three as mixed ethnicity, two as Asian, and the remainder as White.

In total, 24 parents were interviewed of whom fifteen participants were Black (Black African n=12, Black Caribbean n=1, other Black background n=2), three were Asian (Indian background n=2, any other Asian background n=1), three were mixed ethnicity (Black Caribbean and White British n=2, African-Asian n=1) and two participants were from other ethnic groups, both being Arab. One mother was White and had a child of mixed ethnicity.

The interview results yielded a multitude of varied themes, some overlapping for the three conditions of focus: Apgar, cyanosis and jaundice. In-depth

discussions were had about changing pigmentation at birth, relevance of different skin colour descriptors (blue, pink, pale), suggestions for alternative practices, challenges in detection of these conditions, training needs, better listening skills, and improved communication between HCPs and parents.

There was a general consensus from HCPs and parents that the term “pink all over” in Apgar scoring is derived from observing White babies, and its relevance to babies from Black, Asian, and minority ethnic backgrounds was generally questioned. It was suggested that, due to the implicit bias within the Apgar score, continued use may be considered inherently racist in a multi-cultural society where infants hold diverse phenotypes.

However, for some, the question seemed to be where to look for “pink” or for “blue” colour as part of a whole range of other indicators for the health and wellbeing of the babies. Around the lips or the mouth mucosa were frequently cited by parents and HCPs to be an appropriate place for assessing cyanosis and for the appearance component of the Apgar score in all ethnicities. However, difficulties in detecting cyanosis in all infants, and particularly those from ethnic minority backgrounds, was noted.

Assessing the sclera and gums, rather than the skin, for jaundice were reported by HCPs to enable better detection in Black and minority ethnic neonates. However, visual assessment was noted to be poor with HCPs having a low threshold for testing TCB or SBr if any signs of jaundice were detected.

Some stakeholders felt there was a need for more appropriate training and educational materials for consideration of skin colour when undertaking neonatal assessments, as well as exposure to babies with variations in skin tone. To facilitate equitable maternity and neonatal care for all, some stakeholders suggested creating opportunities for exposure to multi-ethnic and racially diverse communities for all HCPs and establishing a databank of pictures from diverse communities and babies with various conditions.

The impact of ethnicity and race on care

Although the overall focus of this review was on neonatal assessments, challenges around the quality of care could not be ignored within the context of race and ethnicity. The systematic review of the experiences of parents or carers therefore also highlighted potential barriers (and facilitators) to accessing care for these communities. Similarly, within the stakeholder interviews, parents and healthcare professionals were asked about challenges or barriers to accessing or receiving care and any areas of good practice. These findings were then analysed, with several themes emerging.

The first theme identified was the communication barrier, where women faced problems in getting information from HCPs due to language differences or inadequate translation. Another issue was women being silenced within the system, where they felt dismissed, ignored, or belittled by HCPs. Fear of raising concerns was another reason why women remained silent, as they worried about being labelled as “difficult” or “too much trouble”.

The literature and interviews reviewed also revealed discrimination and racism in healthcare. Three specific areas of discrimination were identified, including stereotypes, lack of cultural competence, and inadequate care. In both the literature and in the interviews, women reported assumptions about their education and lifestyle and felt HCPs labelled Black, Asian, and minority ethnic women as “aggressive” or “difficult”. Parents and HCPs felt HCPs sometimes made unfounded racial assumptions about Black and minority ethnic women’s bodies and lacked cultural competence, resulting in over-medicalisation of childbirth and inadequate pain management. HCPs also overlooked potential difficulties due to women’s skin tone and failed to offer culturally tailored care. While few parents within the interviews reported feeling that their culture was not understood, the need for HCPs to understand a woman’s culture to better support her was recognised.

Inadequate care towards Black, Asian, and minority ethnic parents was reported in multiple studies and in the interviews. Racist microaggressions, including mispronouncing names and discriminatory comments, were reported in interviews with parents and within the studies included in the review. Lack of dignity and respect, neglectful care, and being left in vulnerable states were also commonly reported. Some mothers even engaged lawyers due to traumatic experiences and obstetric violence. HCPs noted that discrimination was present in complaints received from parents.

Systemic and organisational factors were also identified. These included difficulties around accessing appointments due to lack of knowledge about who and when to contact and a lack of flexibility within the system. Staffing was also raised as an issue, with workload pressures and lack of training thought to make discrimination more likely as overworked staff rely upon their biases more. Charging some migrants for care depending on status was also seen as an organisational barrier.

Social isolation, socio-economic status, and racism in society were identified as contextual issues impacting maternity care, as reported by both parents and HCPs. These issues included a lack of support network, inadequate housing, poor nutrition, mental health issues, and the racism they faced in general in society. There are inequalities in health outcomes, particularly for those from ethnic minority backgrounds, and racism affects parents’ choices and feelings about maternity care. Some parents changed their birth preferences due to concerns over discrimination and the additional risk of maternal and neonatal mortality for Black, Asian, and minority ethnic families.

Overall, the above highlights the need for HCPs to be more culturally competent and provide personalised, culturally safe, care, as well as the need for more support for parents from seldom-heard communities. Addressing racism and discrimination in healthcare, society, and decolonising the curriculum, practice, and policy is crucial for improving maternal and neonatal health outcomes, ensuring that all parents receive respectful and dignified care.

Recommendations for practice

1. Given poor visual detection of jaundice and cyanosis, particularly in Black and darker skin toned neonates, the following recommendations are made:

    - Jaundice: Exploration of wider availability and use of bilirubinometers is recommended to decrease health inequalities and ensure safe care for all. Collaboration with organisations such as BAPM, RCM and RCPCH, RCOG, iHV, CQC and other key stakeholders is recommended.

    - Cyanosis: Healthcare organisations should strongly consider use of pulse oximetry screening if there is any indication of concern over oxygenation. For this reason, the UK National Screening Committee should also strongly consider including routine pulse oximetry screening as a requirement within NIPE (the Newborn and Infant Physical Examination) to mitigate the health disadvantages experienced by those with darker skin tones.

2. NHS England to create a national data bank of open access images of Black, Asian and ethnic minority neonates to incorporate into training and education of HCPs and healthcare students, as well as to aid diagnosis in practice. These images should also be available for use in accessible resources for families. These should incorporate images of healthy neonates, as well as specific conditions such as cyanosis and jaundice, and other conditions suggested by participants, including skin rashes and sepsis. Hard copies and digital sources should be provided, including on platforms such as the online NHS Health A-Z pages. These images should be made available and used by all NHS provider organisations and the NHS should create an opportunity for people to upload pictures and share experiences of conditions in different skin tones.

3. There is an urgent need for regular education and training for healthcare professionals and healthcare students on undertaking clinical assessments on neonates from Black, Asian, or minority ethnic backgrounds, including within the yearly updates on neonatal resuscitation. Better education for families is also required. All training requires process evaluation to ensure effectiveness. Further action on training and education should be taken as follows:
    - RIn order to remind learners of the challenges that varying skin tones may introduce when assessing clinical signs, it is important for resuscitation

dolls in all Higher Education Institutions and NHS Trusts to include babies with Black or dark skin.

- Professional associations (including BAPM, iHV, NNA, RCOG, RCPCH, RCM, RCN) and regulatory bodies (NMC and GMC) should identify training requirements to ensure that those they represent are fully competent in assessments that include skin colour for Black, Asian, and ethnic minority neonates. In addition, they should ensure all HCPs and students are made aware that some ethnicities are at higher risk of neonatal jaundice.

- All healthcare students should have access to an actual or simulated placement to increase students' awareness, knowledge, and confidence in assessing Black, Asian, and minority ethnic neonates prior to qualification.

- An urgent review and update of written and digital materials provided to parents is required. These should be co-developed with parents or carers from diverse ethnicities. A particular focus should be around jaundice and detecting a deteriorating infant to ensure accessibility and relevance of pictures and information to those from Black, Asian, and minority ethnic backgrounds.

- Both community and hospital staff should receive training on anti-racist practice, alongside culturally safe, compassionate care, with good listening skills to meet the needs of our diverse, multi-ethnic society and to reduce current health inequalities.

Recommendations for policy

4. Guidelines that refer to neonatal assessment by skin colour should be immediately reviewed and updated to highlight the impact of race and ethnicity (BAPM, HEE, iHV, NICE, NNA, OHID, RCM, RCN, RCOG, RCPCH, Resuscitation Council, WHO). This should include guidelines around general care of the newborn, as well as specifically for jaundice, cyanosis, and Apgar scoring. Further action should be taken in the following areas:

    - Attention should be drawn to any potential differences in assessment techniques for Black, Asian, and ethnic minority neonates.

    - Guidelines or educational materials that currently refer to the terms pink, blue, or 'normal' colour need to detail how this would be assessed more objectively in Black and darker skinned neonates.

    - All guidance needs to highlight the limitations of visual assessment of the skin, particularly in those from ethnic minority backgrounds. A

comprehensive assessment of other areas including sclerae and gums for jaundice and mouth mucosa for cyanosis is recommended. If jaundice is suspected, additional TCB or SBr should be undertaken.

- When screening tools such as pulse oximeters or bilirubinometers are advised, rather than clinical visual inspection, particular attention should be given to any differences in reliability or accuracy for neonates with different skin pigmentation. When advising the use of a bilirubinometer, a minimum threshold below the treatment line when a confirmatory SBr test is required needs to be urgently established to ensure appropriate detection of jaundice regardless of ethnicity.

- All guidelines should contain or signpost to images and good descriptors of skin assessments in those of all skin tones to support recognition of conditions in those who do not regularly have exposure to Black, Asian, and minority ethnic neonates.

Recommendations for research

Regarding the Apgar score

5. The wording 'pink all over' was not considered appropriate within the UK by the majority of HCPs and parents, including for White neonates. Further exploration is required, including:

    - Determination of inter-rater reliability between HCPs when undertaking the Apgar score in Black, Asian, and minority ethnic neonates. In particular, a better understanding of how HCPs determine the Apgar score in darker skinned babies is required, with specific attention to the terminology of 'pink all over' and the location of pink areas.

    - A systematic review is required of alternative scores or assessments such as the Neonatal Resuscitation and Adaptation Score and Expanded Apgar, compared to the Apgar score including their reliability and validity in Black, Asian, and minority ethnic neonates.

Regarding jaundice

Research required around jaundice includes:

6. In depth case study/root cause analysis of severe cases of jaundice, including those admitted to neonatal units for intensive phototherapy, requiring an exchange transfusion, or with resultant kernicterus. Clear consideration of confounding factors is required, as well as timing and content of contact with HCPs. This could inform better identification, prevention, and timely management of severe jaundice.

7. A better understanding of what level jaundice is by the time it reaches the eyes or gums is required given concerns voiced by HCPs that the areas we currently look at to identify jaundice in Black or darker skinned neonates actually identify jaundice at an advanced stage.

8. Determining inter-rater variability in assessing jaundice to evaluate the impact of skin tone and race on HCP ability to detect jaundice.

9. Current guidelines within the UK vary in their recommendations for when a TCB reading should be followed up by a SBr. Further investigation and confirmation of a more accurate and consistent threshold for adequate identification of infants, particularly for those of varied race and ethnicities is required.

10. Given the wide availability of mobile phone technology, further development and exploration of mobile applications for detecting neonatal jaundice across all ethnic backgrounds is warranted.

Regarding cyanosis

11. In light of inequalities highlighted by the COVID-19 pandemic in adult studies, further UK based research to examine the small but potentially clinically significant differences in arterial oxygen saturation compared to pulse oximetry saturation in neonates from diverse race and ethnicity backgrounds is warranted.

Background

In the United Kingdom, there are well evidenced ethnic inequalities in maternal and neonatal health and care provision, with women from Black and Asian backgrounds having higher rates of maternal mortality (Knight et al., 2022), stillbirth, and neonatal mortality than those of White ethnicity (Draper et al., 2022). Additionally, women from ethnic minority backgrounds report poorer maternity care experiences (Henderson et al., 2013; Martin et al., 2016; Sigurdson et al., 2019). There are growing concerns around routine perinatal practices that are based on the normative White body which may therefore put babies with darker skin at a disadvantage (Kapadia et al., 2022). Numerous neonatal assessments require an element of assessment by skin colour, which is subject to observer and potentially racial bias (Dai et al., 1997; Moyer et al., 2000).

The Apgar score was developed in the 1950s (Apgar, 1952). It gives a score of 0 to 2 in 5 domains at 1 minute, 5 minutes and 10 minutes after birth to assess neonatal adaptation to extra-uterine life. The five domains are Appearance (colour), Pulse (heart rate), Grimace (reflex), Respiratory effort, and Activity (Tone) (Knight, 2020). Despite advances in the availability of medical technology since the 1950s, the Apgar score continues to be extensively applied by healthcare professionals globally as a standardised and convenient way of reporting the status of an infant at birth (Li et al., 2013). There are concerns, however, about the reliability of the overall Apgar score (Committee on Fetus and Newborn, 2006), as well as specifically in relation to the appearance component that assesses neonatal skin colour, which may mean disadvantaging babies with darker skin tones (Blake, 2010; Edwards et al., 2023). Researchers have therefore suggested a need to extend the assessment to more accurately assess wellbeing in African, Caribbean, and Asian neonates (Blake, 2010).

Other clinical assessments that use skin colour as an element of diagnosis have also been questioned. Neonatal hyperbilirubinemia is the most common reason for neonatal hospital readmission (Battersby et al., 2017). Jaundice, which is characterised by a yellowing of the skin and sclera, is the clinical manifestation of hyperbilirubinemia (Woodgate & Jardine, 2015). Diagnosis of hyperbilirubinemia frequently entails visual inspection of a neonate's skin colour to determine jaundice which is often subjective and can be inaccurate, with skin pigmentation presenting as a confounding factor (Lee et al., 2019; Keren et al., 2009). Missed or delayed diagnosis of jaundice is a preventable cause of brain injury in children (Ahmed et al., 2010). To avoid excessive blood sampling from neonates, which may increase the risk of infection and parental stress, transcutaneous bilirubin (TCB) monitoring devices are used to screen for jaundice (Dai et al., 1997; Jones et al., 2017). However, concerns have been

raised that racial bias may still affect the diagnosis of jaundice when using TCB devices, with some studies suggesting an overestimation of bilirubin in darker-skinned neonates (Ebbesen et al., 2012; Taylor et al., 2015).

Neonatal assessment of hypoxia can be characterised by a ‘blue’ or ‘pale’ discolouration of a neonate’s skin or lips (Kapadia et al., 2022). Diagnosing poor oxygenation in infants is difficult by means of visual assessment alone and varies among clinicians, including in infants of darker skin tones (Goldman et al., 1973). Guidelines for use by healthcare professionals therefore suggest that skin colour observation alone is not sufficient to determine oxygenation (Resuscitation Council, 2011, 2021). Routine newborn pulse oximetry screening is a non-invasive approach to assessing oxygenation and is used to screen infants for important conditions that may otherwise have been overlooked (Brown et al., 2020). Pulse oximetry devices have, however, previously been calibrated for use on White skin (Moran-Thomas, 2020). In an adult population, it has been suggested racial bias in pulse oximetry may place an increased risk on Black patients (Sjoding et al., 2020). In particular, the COVID-19 pandemic highlighted the disparities in healthcare provision and supporting technology, with pulse oximeter readings in dark skinned individuals found to be more likely to have inaccurate readings (NHS Race and Health Observatory, 2021b; Vesoulis et al., 2022).

It was therefore imperative to review neonatal assessment and practice in Black, Asian, and minority ethnic newborns from the point of view of both healthcare professionals and parents or carers, as well as within current policies and the evidence base.

Overall methodology

Overall aims and objectives

The overall aim of this work was to review neonatal assessment and practice in Black, Asian, and minority ethnic newborns.

This included:

- Reviewing the use of the Apgar score, with particular focus on the appearance component, as well as the more general detection of cyanosis in ethnic minority neonates within the relevant literature and among stakeholders.
- Reviewing the identification of jaundice in Black and minority ethnic neonates compared to White neonates within the relevant literature and among stakeholders.
- Examining the wider use of terminology used in policy and practice to ensure Black and Asian neonates receive equitable care and treatment.
- Undertaking a broad analysis of potential inequalities for Black, Asian, and minority ethnic neonates within the UK among relevant stakeholders.

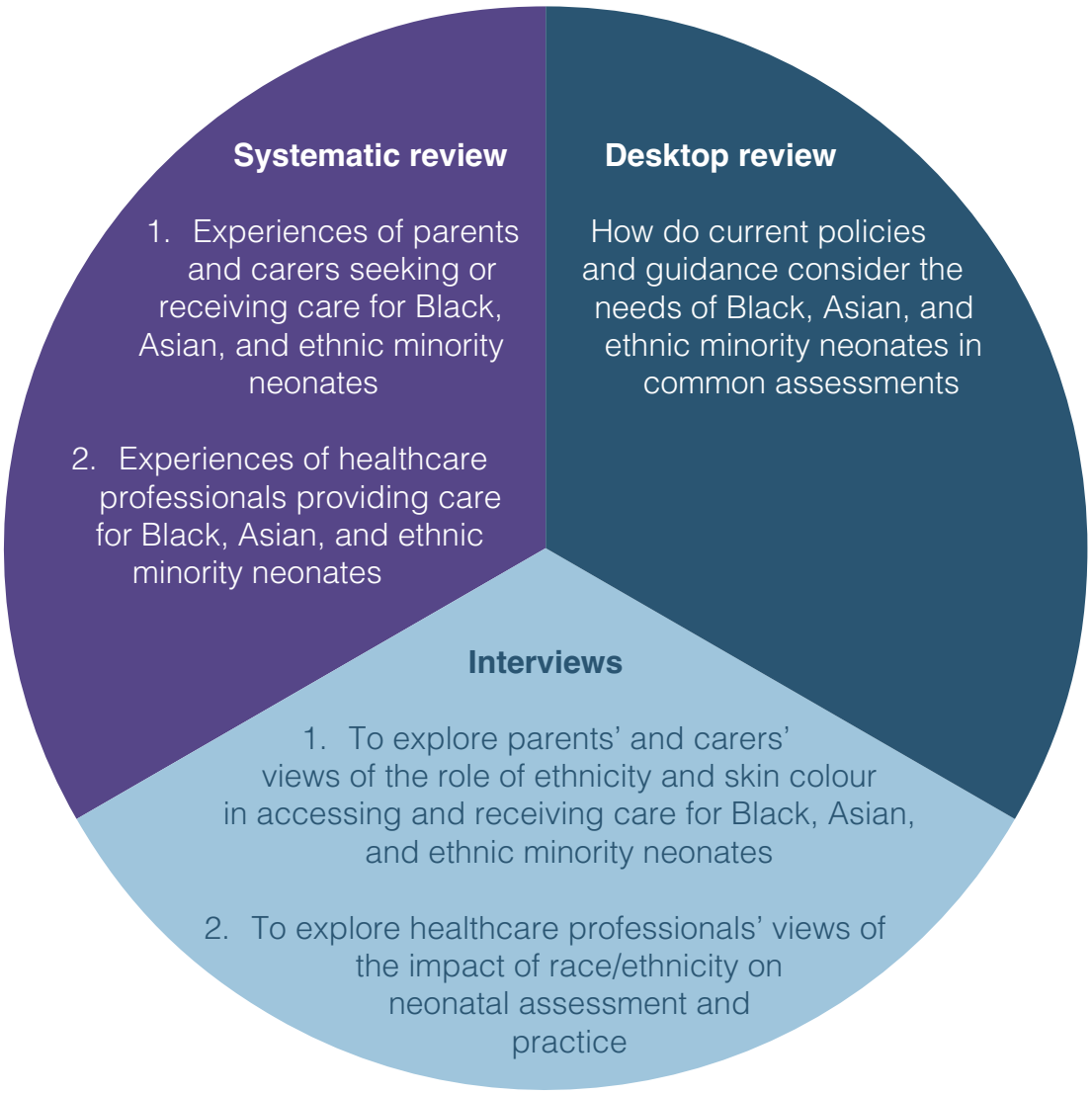
Methodological approach

A mixed methods approach was undertaken comprising a desktop review, systematic reviews and stakeholder interviews as shown in Figure 2.1. The desktop review was undertaken initially, with the systematic reviews and interviews undertaken in parallel. Full methodological details for each component are given within the relevant chapter of this report. Results from the different components were combined with an equal weighting in a complementary paradigm.

Stakeholder engagement

An advisory committee was developed at the start of the project including representatives from obstetrics, midwifery, health visiting, neonatal services and a charity, public health, maternity voices and other maternity service users. They were asked for any literature they were aware of in the area, and asked to provide comments on research documents such as interview schedules. Additionally, several stakeholder engagement workshops were held with a wide range of stakeholders to discuss the interim results and initial recommendations of the study.

Figure 2.1 Components within the mixed methods approach



Experiences of healthcare professionals: A systematic review

Aim

The aim of this systematic review was to identify the experiences of healthcare professionals providing care for Black or Asian or minority ethnic neonates in relation to neonatal assessment, specifically for Apgar scores (including the detection of cyanosis) and jaundice.

Methods

A systematic review was undertaken. The review protocol was published on the PROSPERO database (CRD42022344617).

Inclusion and exclusion criteria

The inclusion criteria for the review are defined below:

Population: Any relevant healthcare professionals (including but not limited to neonatologists, paediatricians, midwives, obstetricians, obstetric nurses, paediatric nurses, and health visitors).

Condition: Ethnic inequalities in neonatal health and care provision are of concern through the normalisation of tests and procedures which have been based on White European babies genetically, physiologically, and behaviourally. Concerns have been raised regarding limitations in routine perinatal practices such as Apgar scoring, detection of cyanosis and detection of jaundice.

Exposure: Healthcare professionals (HCP) experiences of providing care for a neonatal infant (less than 4 weeks old) where outcomes were specified according to Black, Asian, and minority ethnic background. A particular focus was placed on determining Apgar scores, cyanosis, and on the detection of jaundice in which skin colour is an element of the assessment which may mediate disadvantaging babies with darker skin.

Comparator: We excluded all articles where a 'White' or 'light' skinned control group was not used in comparison to other skin tones.

Study design: All study designs were considered including both qualitative and quantitative methodologies. These included:

- qualitative, quantitative, or mixed methods. Reviews of potential relevance, where not eligible for inclusion, were checked for further relevant citations.

- published in academic journals and unpublished, such as statutory body reports, consultation exercise reports that include quantitative or qualitative data components.

Where abstracts were identified, if sufficient information was provided to establish the trustworthiness of the study and the study authors confirmed that the information provided was the final analysis, the abstract was considered for inclusion.

Both published and unpublished studies in English were considered for inclusion. Studies were included from database inception to April 2022.

Search strategy and study selection

A comprehensive search was undertaken for both academic and grey literature. Peer reviewed literature was identified by searching MEDLINE, CINAHL and PsycINFO. Grey literature was searched using OpenGrey within the DANS EASY archive to identify unpublished studies that met the inclusion criteria. A combination of medical subject headings (MeSH) and key text words were used. An example search strategy is given in Appendix 1.

As well as the above search strategy, a group of expert stakeholders were contacted via email. These stakeholders included maternity user representatives, academics, healthcare professionals, and commissioners. Stakeholders were questioned regarding their awareness of further relevant literature for inclusion within the systematic review.

Retrieved citations were screened by title and abstract against the inclusion criteria, with each citation screen independently by two reviewers (AF for all citations and one out of FF, Ghazaleh Oshaghi or Kayla Baugh). Citations of

potential interest were then read at full text independently by two researchers for inclusion (AF, FF). Any disagreements were discussed within the whole research team to reach consensus. References of all included studies were screened.

Data extraction

Data was extracted by two reviewers using a pre-defined data extraction template to record relevant data such as title, author, year of publication, country of origin, study design, sample (size, ethnicity, neonatal age), methodology (sampling strategy, setting, data collection, data analysis), study results, potential confounding factors (such as education level, deprivation/socio-economic status, maternal age), source of study funding. Where possible, attention was given to the definition of the ethnicity or race categories used within each article. Where required, authors of the original study were contacted to clarify any information felt to be ambiguous within the published report.

Risk of bias assessment

Studies were critiqued using the Mixed Methods Appraisal Tool (MMAT) (Hong et al., 2018). The MMAT includes two screening questions. There are then five categories of study design including qualitative, quantitative (non-randomised, randomised and descriptive) and mixed methods. Within each category there are five methodological quality criteria that relate to the appropriateness of methodology, data collection techniques and data analysis techniques. Each included study was appraised by completing the five quality criteria corresponding to the appropriate study design. Included studies were critiqued independently by two researchers, with any disagreements discussed with a third researcher.

Data synthesis

A narrative synthesis of the results was performed, considering each assessment subgroup separately: Apgar, Jaundice and Cyanosis/Hypoxia. Results were reported considering the time period in which the study was undertaken, and the ethnicity of the neonates included. Other potential confounding factors were considered, such as education level, socio-economic status determined by level of deprivation or occupation) and maternal age. However, given the lack of reporting of results separately for these categories within the majority of the included studies no further sub-group analysis was possible.

Results

Database searches were undertaken at the end of April 2022. Of the 5704 citations obtained from the database searches and the 41 studies identified through other sources, 217 were screened at full text. In total 56 studies were considered eligible for inclusion within this review. A PRISMA flow diagram of the study search and selection process can be seen in Appendix 2 (Page et al., 2021).

Study characteristics

Study characteristics can be seen in Appendix 3. Six studies considered outcomes according to or education around the Apgar score (Chubb et al., 2022; Li et al., 2013; Mihoko Doyle et al., 2003; Serunian & Broman, 1975; Shankaran et al., 2004; Wolf et al., 1997). Three studies considered detection of cyanosis (Dawson et al., 2015; Goldman et al., 1973; Vesoulis et al., 2022), with the remainder considering the detection of jaundice. Studies were mainly undertaken in countries with a very high development index (HDI) ranging from 0.895 in Italy to 0.962 in Switzerland (United Nations Development Programme, 2023). Countries included Australia (n=3), Canada (n=4), Denmark (n=2), Various European countries (n=1), France (n=2), Israel (n=1), Italy (n=1), New Zealand (n=1), Netherlands (n=1), Norway (n=1), Spain (n=1), Sweden (n=1), Switzerland (n=3), UK (n=3), USA (n=24). One study was jointly undertaken in the USA and Bangladesh (HDI 0.661). The other four studies took place in countries with a lower HDI including Nigeria (n=1, HDI 0.535), Brazil (n=1, HDI 0.754), South Africa (n=1, HDI 0.713), Zimbabwe (n=1, HDI 0.593).

Study quality

Appendix 4. summarises the MMAT critical appraisal rating for each included study. All but one study (Bourchier et al., 1987) were deemed to have a clear research question and all but two studies (Bourchier et al., 1987; Wolf et al., 1997) to have collected data that fully allowed the research question to be addressed.

Within the Apgar studies one out of the six studies (Li et al., 2013) was judged to be of low risk of bias across all five domains. Within the cyanosis studies none of the three studies were judged to be of low risk of bias across all five domains. Within the jaundice studies, only eleven studies (Afanetti et al., 2014; Brits et al., 2018; Campbell et al., 2011; Ebbesen et al., 2002; Hannemann et al., 1982; Jones et al., 2017; Karen et al., 2009; Keren et al., 2008/2009; Lee et al., 2019; Maisels et al., 2004; Samiee-Zafarghandy et al., 2014) out of the 47 studies were judged to be of low risk of bias across all five domains.

Apgar score

Five included studies made a comparison between infants of White and any other skin tone and subsequent outcome sequelae from the Apgar score. One further study considered the impact of training healthcare professionals on implicit bias and clinical assessment of babies from Black and minority ethnic backgrounds. The impact of the Apgar score on neonatal mortality was considered in two large, linked birth-death datasets. The first (Mihoko Doyle et al., 2003) included 6,544,004 infants and the second (Li et al., 2013) included 25,936,357 infants. Both studies found that low Apgar score is predictive of neonatal death across different ethnicities. The study by Mihoko Doyle et al., (2003) found that after adjusting for birthweight the odds of neonatal death with a low Apgar score were lowest in Black infants (OR 20.40) and highest in Mexican American infants (OR 44.24), with the odds of neonatal death in White infants being in between (OR 36.21). The same pattern was noted for medium Apgar scores (4-6). However, Black neonates were more likely to receive a low 1-minute Apgar, with 10.6% receiving an Apgar score of 6 or lower, compared to 7.6% of White or Mexican American infants (Mihoko Doyle et al., 2003).

The second study by Li et al., (2013) similarly found that neonatal mortality was consistently higher in non-Hispanic White than non-Hispanic Black neonates at the same Apgar score, with the biggest difference noted for infants with a low Apgar score (1-3). This was evident in both term and preterm infants. See Appendix 5 for the graphical representation of results from the study. The lower risk of mortality in Black infants with a low Apgar score remained after adjusting for maternal education, marital status, smoking status during pregnancy and time of booking for antenatal care. However, this second study did not report whether Black babies were more likely to receive lower Apgar scores which would impact upon these findings (Li et al., 2013).

Three studies looked at longer term outcomes in accordance with Apgar score, all with small sample sizes ranging from 246 (Shankaran et al., 2004) to 391 (Serunian & Broman, 1975). The final study looked at 165 infants in Zimbabwe and compared them to 105 infants in the Netherlands and the Caribbean (Wolf et al., 1997). The studies provided inconsistent results about the long-term predictive value of the Apgar score. The Bayley mental development score was found to be significantly lower overall in those with low Apgar scores at 8 months (Serunian & Broman, 1975). However, when looking at the scores by ethnicity, the mental development score was only significantly lower in those of mixed ethnicity (a local combination group of mixed Black African and Portuguese) but not among White only or Black only infants (Serunian & Broman, 1975). In contrast the other study found Black infants with a low Apgar score and low birthweight ≤750g to have poorer mental development at 18-22 months corrected [OR 2.2 (95% CI 1.2-3.7)] (Shankaran et al., 2004). However Black race was no longer a significant predictor of mental development score after adjusting for multiple factors including household income and parental education (Shankaran et al., 2004).

The Bayley psychomotor development scores were found to be lower, at 8 months old, in those with a lower Apgar score (Serunian & Broman, 1975). Again, when looking at the scores by ethnicity, the psychomotor development score was only significantly lower in those of mixed ethnicity (a local combination group of mixed Black African and Portuguese) but not among White only or Black only infants (Serunian & Broman, 1975). Within the other study of infants with low Apgar scores and low birthweight, Black race was not significantly associated with the psychomotor development in univariate or multivariate analysis (Shankaran et al., 2004).

The final study looking at long term outcomes according to Apgar score found that infants born in Zimbabwe with an Apgar score <5 had more abnormal neurological classifications than infants born in Netherlands or the Caribbean (Wolf et al., 1997). However, there were too many confounding factors within this study to ascertain whether the differences were due to Black race or factors such as differences in maternal and neonatal care within the different countries.

The study looking at training healthcare professionals on implicit bias and clinical assessment of babies from Black, Asian, and minority ethnic backgrounds undertook pre- and post- training surveys (Chubb et al., 2022). The pre-training survey showed only 9.1% of staff (5/55) had received any prior training around caring for ethnic minority mothers and babies. Instead, they relied on colleagues, clinical experience, and self-directed learning. Midwives training in the previous 5 years were more likely to have been educated using Black mannequins (44%) compared to those trained 5-10 years ago (18%). However, only 2 of the 7 students who undertook the training had received resuscitation training on Black mannequins at university. Overall, only 12.7% (7/55) felt they had adequate prior training around Black and minority ethnic families (Chubb et al., 2022). After undertaking the training, 96% of midwives felt the Apgar score was not the most appropriate way to assess all babies at birth as they recognised that not all neonates may be pink at birth. Additionally, in the post-training survey 98% of midwives stated that they planned to change their clinical practice after the new knowledge gained. Within the free text responses within the post-training survey, midwives described being shocked by the levels of inequality and inequity within maternity care (Chubb et al., 2022).

Detection of cyanosis or hypoxia

Three studies were found that considered the detection of cyanosis in infants of White and any other skin tone (Goldman et al., 1973, Dawson et al., 2015, Vesoulis et al., 2022).

The first study, with 93 participants, considered the accuracy of visual assessment of the infants' skin in different areas to determine arterial oxygen saturation (Goldman et al., 1973). Within the whole sample, skin colour at

different anatomical sites was a crude assessment of arterial oxygen saturation in infants of all ethnicities. Hands, nailbeds, and around the mouth were found to have many false positive observations, where cyanosis was judged to be present when arterial oxygen saturation was over 90%, but very few false negatives where cyanosis was not observed when arterial oxygen saturation was below 80%. They felt that lips were the most reliable of all of the sites studied. Infants were defined as dark skinned if at least two of the three observers for cyanosis judged them to have dark skin. Dark skinned infants had fewer false positives, where the observer deemed the infant to have cyanosis when the arterial oxygen saturation was above 90%, when observing their hands, trunk and around their mouth (Goldman et al., 1973).

The second study considered whether tongue colour after birth provided a good indication of whether the neonate required supplemental oxygen (Dawson et al., 2015). They determined that a pink tongue is likely to mean that a neonate has an oxygen saturation above 70% so does not require supplemental oxygen. Ethnicity had little effect on the area under a Receiver Operator Characteristics Curve which was used to summarize overall diagnostic accuracy. The area under the curve (AUC) for not having a pink tongue to detect pre-ductal pulse oximetry saturation (SpO2) of <70% in a White neonate was 0.89 (95% CI 0l.84-0.95) and in a 'non-White' neonate was 0.94 (95% CI 0.87-1.00) (Dawson et al., 2015). While the study was only undertaken on a small sample of 68 infants and the number of 'non-White' infants included was not given, it showed that evaluation of the tongue was not less effective at detecting hypoxaemia in 'non-White' compared to White neonates.

The final study of 294 infants admitted to a neonatal intensive care unit examined the impact of ethnicity on pulse oximetry (Vesoulis et al., 2022). When pulse oximetry saturations were 95% or lower, there was a consistent difference between White and Black infants, with Black infants having higher pulse oximetry saturations at each arterial oxygen saturation level. The exact difference varied by SpO2 and can be visualised in the study graph presented in Appendix 6. Occult hypoxemia, where pulse oximetry levels were 90% or more when arterial oxygen saturation (SaO2) was <85% was more common in Black infants (9.2%) compared to White infants (7.7%), but this did not reach statistical significance (Vesoulis et al., 2022). The sensitivity of the pulse oximeter to detect true hypoxia (SpO2<90% when SaO2<85%) was similar for Black neonates (39% sensitive, 81% specific) and for White neonates (38% sensitive, 78% specific).

The studies found that visual assessment of cyanosis based on skin colour at different anatomical sites was not very accurate in determining arterial oxygen saturation in infants of all ethnicities. Tongue colour was found to be good indication of oxygen saturation in both White and non-White neonates. Pulse oximetry showed consistent differences between White and Black infants at low saturation levels and sensitivity to detect true hypoxia was similar for both groups

Jaundice

46 articles covering 45 studies considered the detection of jaundice in infants of White and any other skin tone. Additionally, one study considered practitioners knowledge and skills at identifying neonatal jaundice. Most studies assessed a transcutaneous bilirubinometer to assess the transcutaneous bilirubin level (TCB) against serum bilirubin (SBr). Bilirubinometer makes included BiliCheck, various models in the Minolta/Airshields JM series, the Kejian KJ-8000 and Bilimed. Some studies used a mix of transcutaneous bilirubinometers or did not identify which ones were used. Two studies assessed the utility of either a phone app or using mobile phone images to assess jaundice. The remainder assessed visual inspection including using an icterometer against either SBr or TCB. An icterometer is a type of ruler with increasingly yellow colour swatches to help estimate the level of jaundice. Given the wide variation in definition of different ethnic, racial or skin tone subgroups, as well as differences in gestational ages included within studies, it was not possible to meta-analyse the results.

Table 3.1 shows the correlation between transcutaneous assessment of jaundice and SBr or visual assessment of jaundice for different races, ethnicity or skin tone. There was no consistent pattern of correlation according to ethnicity when using the bilirubinometers to assess TCB, with TCB found to be more correlated with SBr in Black or African subgroups, Asian subgroups, or Hispanic subgroups than in White subgroups in some studies, but less correlated in other studies. Within the study that assessed multiple brands of TCB devices, TCB and SBr were less correlated when using the Bilimed bilirubinometer than other devices (Raimondi et al., 2012). One study evaluated both visual assessment and TCB against SBr (Szabo et al., 2004b), with visual assessment being less well correlated with SBr levels than TCB evaluation. Additionally in both studies using visual assessment correlation with SBr (Szabo et al., 2004b) and with TCB (Keren et al., 2008, 2009) was lower in the ‘non-White’ subgroups showing visual assessment for jaundice was less effective in those from a ‘non-White’ background.

Table 3.1. Correlation of assessment of jaundice using the skin compared to serum bilirubin †

	First author	Gestation at birth	Sample size	Overall correlation	White	Black/ African American/West African	Asian/ Asian American	Māori	Hispanic/ Latino	Mixed ethnicity/ non-White	Medium skin tone	Dark skin tone	Significance
BiliCheck	Bhutani et al., (2000)	Term and late preterm (≥35 wks)	490	0.91	0.907	0.912	0.897^		0.932				NR
	Slusher et al., (2004)	Term and preterm	127	0.92	0.91Y						0.94	0.87	NR
	Szabo et al., (2004b)	Term	140	0.889 (f)	0.949 (f) 0.943 (s)					0.933 (f) 0.900 (s)			Yes
	Holland & Blick (2009)	Term (≥36 wks)	343		0.94 (f) 0.93 (s)	0.88 (f) 0.89 (s)			0.83 (f) 0.92 (s)				NS
	Campbell et al., (2011)	Term and late preterm (≥35 wks)	430	0.83 0.81 ∫	0.82 0.78 ∫	0.80 0.79 ∫	0.84 0.81 ∫		0.86 0.85 ∫				NS
	Raimondi et al., (2012)	Term and late preterm (≥35 wks)	289	0.86		0.88							NR

	First author	Gestation at birth	Sample size	Overall correlation	White	Black/ African American/West African	Asian/ Asian American	Māori	Hispanic/ Latino	Mixed ethnicity/ non-White	Medium skin tone	Dark skin tone	Significance
JM-101	Hegyi et al., (1981)	Term	60		0.77 (f) 0.90 (s) 0.88 (a) 0.87 (knee) 0.72 (sole) 0.87 (elbow) 0.72 (palm) 0.87 (upper back) 0.86 (lower back)	0.92 (f) 0.95 (s) 0.94 (a) 0.95 (knee) 0.91 (sole) 0.93 (elbow) 0.90 (palm) 0.94 (upper back) 0.89 (lower back)							NR
	Goldman et al., (1982)	Term and preterm	125		0.705 (>38 wks) 0.516 (33-38 wks) 0.32 (<33 wks) ◊	0.52 (>38 wks) 0.566 (33-38 wks) 0.905 (<33 wks) ◊							Yes < 33 wks
	Hannemann et al., (1982)	Term and late preterm	161		0.90 (term) 0.88 (34-37 wks)	0.71 (≥34 wks)							NR
	Palmer et al., (1982)	Term and preterm	100		Term: 0.676 (s) 0.698 (a) Y						Term: 0.740(s) 0.657(a)		NS
	Kenny et al., (1984)	Term	53		0.846 (f)	0.962 (f)	0.919 (f)						NR
	Bourchier et al., (1987)	Term	2277 (729 known race)		0.58 (f) δ 0.67 (s)			0.57 (f) 0.62 (s)					NR

	First author	Gestation at birth	Sample size	Overall correlation	White	Black/ African American/West African	Asian/ Asian American	Māori	Hispanic/ Latino	Mixed ethnicity/ non-White	Medium skin tone	Dark skin tone	Significance
JM-102	Szabo et al., (2004b)	Term	140	0.906	0.949					0.954			NR
JM-103	Maisels et al., (2004)	Term and late preterm (≥35 wks)	849	0.915	0.949	0.822	0.9257 § 0.941 (Asian-American) 0.866 (Middle Eastern)						NR
	Thomson et al., (2008)	Term and late preterm (<35 wks)	235		0.75	0.90	0.88						NR
	Wainer et al., (2009)	Term only	938 (744 in analysis)	0.91 0.88 ∫	0.85 Y∫						0.91 ∫		NR
	Raimondi et al., (2012)	Term and late preterm (≥35 wks)	289	0.85		0.92							NR
	Afanetti et al., (2014)	Term and preterm	86	0.923	0.916					0.934			Yes
	Samiee-Zafarghandy et al., (2014)	Term and late preterm (≥35 wks)	451	0.93	0.95 Y						0.94	0.96	NR

	First author	Gestation at birth	Sample size	Overall correlation	White	Black/ African American/West African	Asian/ Asian American	Māori	Hispanic/ Latino	Mixed ethnicity/ non-White	Medium skin tone	Dark skin tone	Significance
JM-103/ 105	Aune et al., (2020)	Term	342	0.91	0.93					0.84			NR
JM-105	Maya-Enero et al., (2021)	Term and preterm (≥31 wks)	1359		0.935 Y						0.924 (medium-light) 0.908 (medium-dark)	0.956	NR
KJ-8000	Starowicz et al., (2020)	Term and preterm	201	0.8	0.84					0.71			NR
Bilimed	Karen et al., (2009)	Term and preterm (≥28 wks)	150		0.63					0.69			NS
	Raimondi et al., (2012)	Term and late preterm (≥35 wks)	289	0.70		0.74							NR
JAISY	Norman et al., (2022) ϕ	Term and late preterm (≥35 wks)	141	0.94 (f) 0.94 (s)	0.95 (f) 0.96 (s)					0.92 (f) 0.93 (s)			NS
Phone image	Aune et al., (2020)	Term	342	0.84*	0.87*					0.75*			Yes

	First author	Gestation at birth	Sample size	Overall correlation	White	Black/ African American/West African	Asian/ Asian American	Māori	Hispanic/ Latino	Mixed ethnicity/ non-White	Medium skin tone	Dark skin tone	Significance
Phone app	Taylor et al., (2017)	Term and late preterm (≥35 wks)	530	0.91	0.92	0.90	0.88		0.91				NR
Undefined	Hannemann et al., (1979)	Term	103		0.831 (six wave-lengths) 0.778 (three wave-lengths)	0.877 (five wavelengths) 0.865 (three wavelengths)							NR
	Brucker & MacMullen, (1987)	Term	20	0.69	0.77								NR
Visual - Kramer	Szabo et al., (2004b)	Term	140	0.728 (nurse)	0.860 (nurse) 0.837 (paediatri-cian)					0.843 -nurse 0.806 - paediatri-cian			Yes
	Keren et al., (2008, 2009) ϕ	Term and late preterm (≥35 wks)	812 (522 had paired TCB)	0.51ρ	0.55ρ √	0.45ρ							No

† - terminology in relation to race, ethnicity or skin tone is used as presented in included studies

Correlation coefficient – Pearson's unless stated otherwise

(f) – forehead measurement

(s) – sternum measurement

(a) - abdomen

NS – not significant

wks - weeks

NR – not reported

Y - Study comparison was "light/ fair skin pigmentation" rather than White per se

∫ - Lin concordance correlation coefficient

* - correlation lower in all cases than for TCB vs SBr

§ - included Asian, Indian Pakistani, Middle Eastern and Hispanic

ρ - spearman's rho

√ - all "non-Black infants"

^ - included Asian and 'other'

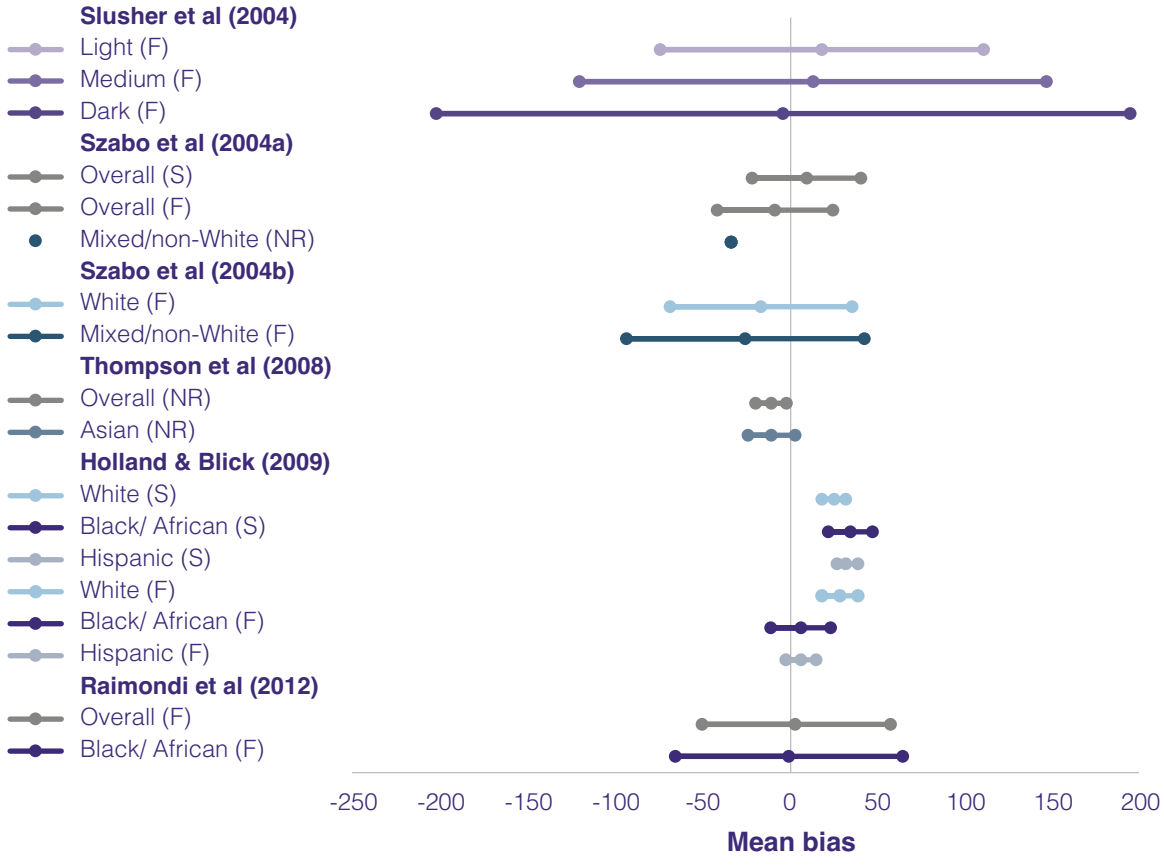
◊ - included Hispanic and non-Hispanic Whites with White infants and Hispanic Black in Black infants

δ - authors suggested correlations lower than other studies as jaundiced infants only, so few low SBr values

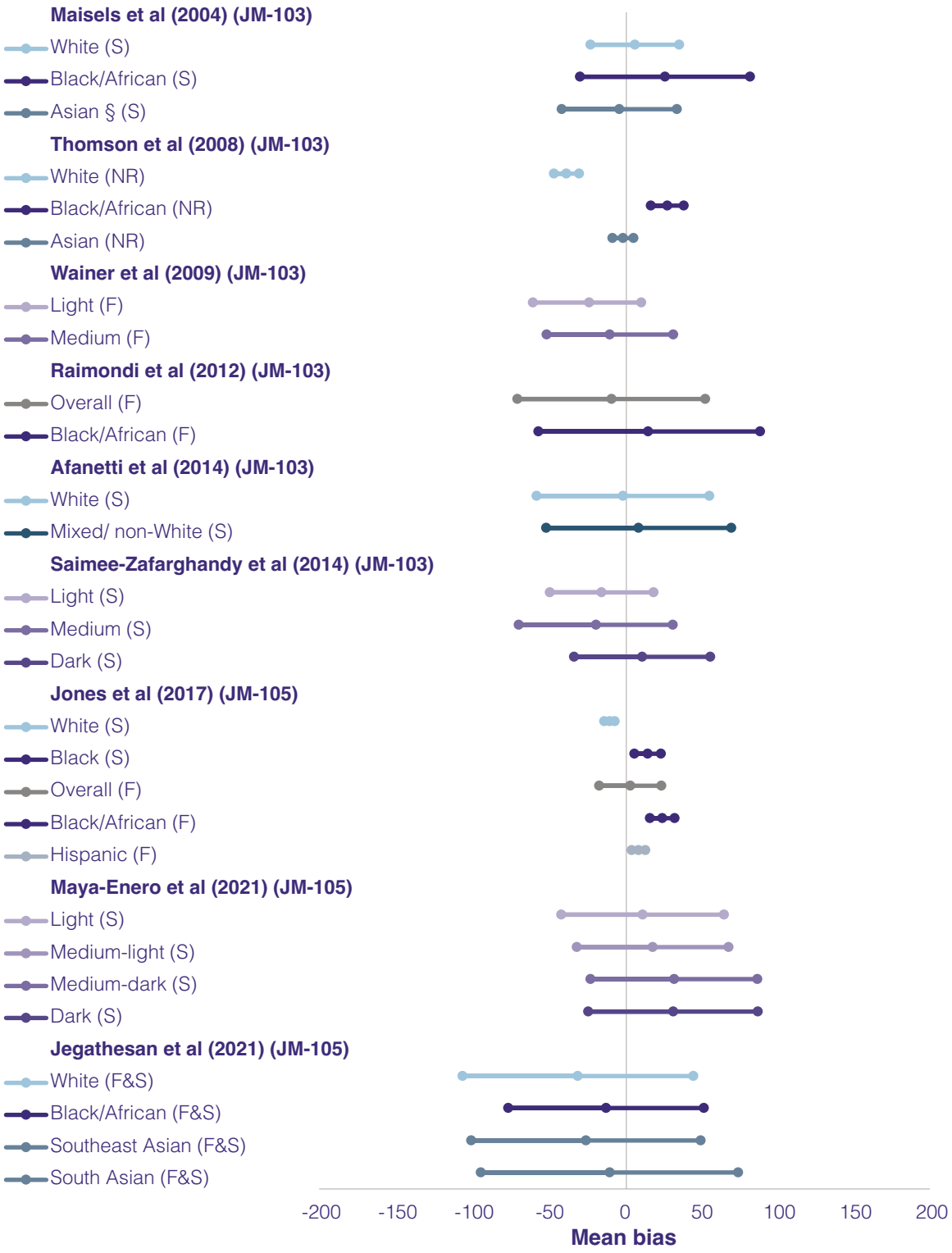
ϕ –compared to TCB not SBr

Figure 3.1. Graphical representation of mean bias and imprecision (95% CI/ limits of agreement) for the Bland-Altman results according to Bilirubinometer machine and ethnicity

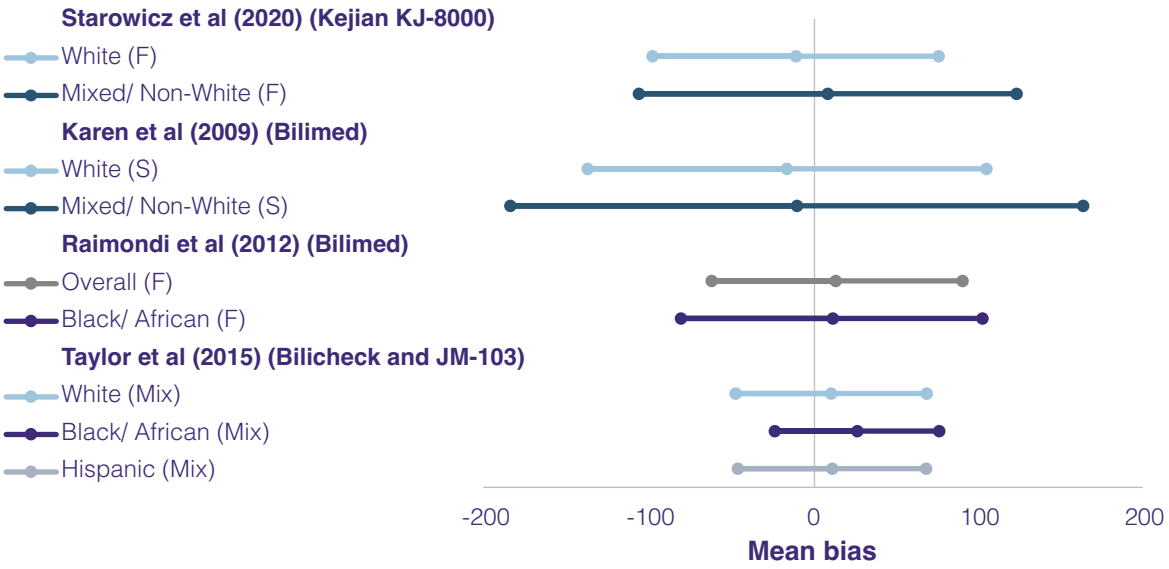
Bilicheck



Mintola / Airshields



Other bilirubinometers



Szabo et al., (2004a) – mean bias only given within article and not imprecision

(F) – TCB reading taken from forehead

(S) – TCB reading taken from sternum

(NR) – not reported where TCB taken from

(mix) – TCB taken from different anatomical sites including sternum, forehead and other

§ - included Asian, Indian Pakistani, Middle Eastern and Hispanic

Figure 3.1 graphically represents the mean bias and imprecision for the Bland-Altman analyses within the different studies. Full results can be seen in Appendix 7. Only seven studies reported statistically testing for differences in the imprecision between different ethnicities/races/skin tones. Variability (wider limits of agreement) significantly increased with increased pigmentation in five of the studies (Slusher et al., 2004; Szabo et al., 2004b for the BiliCheck; Holland and Blick 2009 when assessing the forehead; Starowicz et al., 2020; Karen et al., 2009), with TCB measured on the forehead in all but one study (Karen et al., 2009) where increased variability was noted. Variability was found to be increased when using the BiliCheck (3 studies), the Kejian-8000 (1 study) and the Bilimed (1 study). Only one of the studies using the Minolta/Airshields device reported the significance of the variability, however within several studies the same trend for less variability of the results in those who were White was noted. Only two studies assessing this statistically found no significant increase with increased pigmentation however neither of these studies reported the results in full (Karon et al., 2008; Szabo et al., 2004b for the JM-102).

Within the nine studies that assessed the significance of the mean bias with different ethnicity, race, or skin tone subgroups, four studies found TCB was significantly overestimated in those with dark skin or from ethnic minority backgrounds (three studies used JM-103 or JM-105 {Thomson et al., 2008;

Samiee-Zafarghandy et al., 2014; Jones et al., 2017} and one used Bilimed {Starowicz et al., 2020}). TCB was also found to be underestimated in those with light or medium skin tones or from a White background (Thomson et al., 2008; Samiee-Zafarghandy et al., 2014; Jones et al., 2017- sternum only). Another study reported bias to significantly increase from light to medium to dark skin tones (Maya-Enero et al., 2021). In contrast, two studies found TCB to significantly underestimate SBr in 'non-White' neonates, however they compared White to everyone else including Asian, Indian, and Black rather than one ethnic group per se (Szabo et al., 2004a, Szabo et al., 2004b). There was reported to be no difference in mean bias across different race /ethnicity subgroups in within Jegathesan et al., (2021) who used the JM-105 and in Thompson et al., (2008) and Ahmed et al., (2010) when using the BiliCheck device. However, it should be noted that both Jegathesan et al., (2021) and Ahmed et al., (2010) only included preterm neonates. Within no study was the mean bias more than 40μmol/l below the SBr level.

Table 3.2. Area under the curve within studies assessing jaundice

	First author	Gestation at birth	Sample size	All infants	White	Medium skin tone	African/ Black/ Dark	Signif.
BiliCheck	Raimondi et al., (2012)	Term and late preterm (≥35 wks)	289	SBr>240µmol/l 0.95 (95% CI 0.92-0.97)			SBr>240µmol/l 0.98	NR
JM-103	Maisels et al., (2004)	Term and late preterm (≥35 wks)	849	SBr>170 µmol/L 0.962; SBr>222 µmol/L 0.963; SBr>255 µmol/L 0.975	SBr>170µmol/L 0.972; SBr>222µmol/L 0.937; SBr>255µmol/L 0.958 √		SBr>170µmol/L 0.972; SBr>222µmol/L 0.984; SBr>255µmol/L 0.989	NR
	Wainer et al., (2009)	Term	938 (744 samples for analysis)	SBr>150µmol/l 0.953 (95%CI 0.937–0.969); SBr>200µmol/L 0.987 (0.979–0.996); SBr>250µmol/L 0.993 (0.987–0.999)	SBr>150µmol/l 0.966 (0.950–0.983); SBr>200µmol/L 0.991 (0.980–1.00); SBr>250µmol/L 0.999 (0.998–1.00) Y	SBr>150µmol/l 0.961 (0.939–0.984); SBr>200µmol/L NA; SBr>250µmol/L 0.989(0.979–0.999)		NS
JM-103	Raimondi et al., (2012)	Term and late preterm (≥35 wks)	289	SBr>240µmol/l 0.91 (95% CI 0.88-0.95)			SBr>240 µmol/l 0.94	NR
	Samiee-Zafarghandy et al., (2014)	Term and late preterm (≥35 wks)	451		SBr 110µmol/L 0.96; SBr 170µmol/L 1.0; SBr 230µmol/L NA Y	SBr 110µmol/L 0.98; SBr 170µmol/L 0.97; SBr 230µmol/L 0.94	SBr 110µmol/L 0.99; SBr 170µmol/L 0.96; SBr 230µmol/L 0.99 ψ	Yes – but not clinically meaningful
Bilimed	Raimondi et al., (2012)	Term and late preterm (≥35 wks)	289	SBr>240µmol/l 0.75 (95% CI 0.70-0.79)			SBr>240µmol/l 0.75	NR

NS – not significant
NR – not reported
NA – not applicable (insufficient sample)

Y - Study comparison was "light/ fair skin pigmentation" rather than White per se
√ - all "non-Black infants"
ψ - Study comparison was "dark skin pigmentation

When considering the receiver operating curve (ROC) analysis (Table 3.2), Bilimed had an area under the curve of 0.75 in all infants and those in the Black/ African subgroup suggesting acceptable discrimination for SBr>240µmol/l. For the other bilirubinometers, BiliCheck and JM-103, the area under the curve was consistently >0.90 across all ethnic and skin tone subgroups and across a range of SBr cut-offs suggesting outstanding discrimination.

Full sensitivity, specificity, positive predictive value, and negative predictive value results are given in Appendix 8. Of note is that, within term infants, fewer blood tests were avoided in neonates of 'non-White' ethnicity or of darker skin tones when using TCB than in neonates of White or light skin.

Multiple other statistical tests undertaken are summarised in Table 3.3, with full results given in Appendix 9. The majority of these tests assessed whether skin tone, ethnicity, or race were independent predictors of SBr or of TCB when controlling for other factors. Skin tone, race, or ethnicity was consistently not an independent predictor when using the BiliCheck. However, within seven out of the ten studies undertaking additional statistical tests using the Minolta/Airshields device, skin tone or ethnicity or race was found to be a significant factor. The significant impact of ethnicity/ race was seen in older versions (JM-101) as well as the newer version (JM-105) and in both term and preterm infants. When using the Kramer visual scale there was also a significant impact of race/ethnicity on the accuracy of jaundice assessment.

A final study undertook a survey of Dutch maternity care assistants who provide postnatal care under the direction of a community midwife (van der Geest et al., 2021). Within the 1465 survey responses the median correctly answered questions were 5 out of 6, with significantly better knowledge seen in those who had received training on neonatal hyperbilirubinemia within the last year. A total of 82% reported that they felt capable or very capable of assessing jaundice and 90% felt their knowledge was sufficient or more than sufficient regarding neonatal hyperbilirubinemia. Despite this, 63% still said they would like to learn more about it. Within the survey, three case scenarios were presented (two White infants and one of a 'non-White' infant) with the exact ethnicity of the infant not described further within the study. Accuracy of estimation of jaundice from skin colour within these three case studies was poor and prone to underestimation, with 62% not estimating the SBr range correctly in any of the three cases. Adequacy of SBr estimation did not vary by knowledge (p=0.067) and self-perceived capability was not associated with actual ability to correctly estimate SBr range (p=0.794). More respondents estimated SBr above the actual range in the 'non-White' case (n=245) than in the White cases (infant 1 n=24, infant 2 n=19). Twelve out of 1144 respondents noted within their comments that assessment for jaundice was harder in the 'non-White' neonate (van der Geest et al., 2021).

Table 3.3. Other statistical tests used to examine impact of skin tone on accuracy of jaundice assessment

	First author	Gestations at birth	Sample size	Statistical test	Significance of ethnicity/ race/skin tone
BiliCheck	Rubaltelli et al., (2001)	Term and preterm (>30 weeks)	210	ANCOVA	NS
	Ebbesen et al., (2002)	Term and preterm	488	Multiple linear Regression	NS - healthy NS - NNU
	Robertson et al., (2002)	Term (but mean GA 37.7 ± 2.2 weeks so includes late preterm	101	Linear Regression	NS
	Leite et al., (2007)	Term and preterm	200	ANCOVA	NS
	Ebbesen et al., (2012)	Preterm	133	Multiple Regression	NS
	Rubio et al., (2017)	Preterm	167	Multivariate Multilevel Logistic Regression	NS
JM 101	Hegyi et al., (1981)	Term	60	Regression	Yes
	Hannemann et al., (1982)	Term and late preterm	161	Regression	NR
	Kenny et al., (1984)	Term	53	Regression	Yes, for Asian
JM-102	Robertson et al., (2002)	Term (mean GA 37.7±2.2 weeks so includes late preterm	101	Linear Regression	Yes
	Szabo et al., (2004a)	Preterm (34-36+6 weeks)	69	Mann Whitney	NS
JM-103	Maisels et al., (2004)	Term and late preterm (≥35 weeks)	849	Proportion with TCB-SBr difference >3mg/dl	Yes
	Schmidt et al., (2009)	Preterm only	90	Chi-Square	NS
	Wainer et al., (2009)	Term only	938 (744 samples analysed)	Multivariate Linear Regression	Yes
	Ebbesen et al., (2012)	Preterm	133	Multiple regression	Yes
	Afanetti et al., et al., (2014)	Term and preterm	86	Multiple Linear Regression	NS

	First author	Gestations at birth	Sample size	Statistical test	Significance of ethnicity/ race/skin tone
JM-105	Maya-Enero et al., (2021)	Term and preterm (31 weeks +)	1359	Multiple Linear Regression	Yes, with and without adjusting for GA
Bilimed	Karen et al., (2009)	Term and preterm (≥28 weeks)	150	Multiple stepwise linear regression	No
Mixed*	Taylor et al., (2015)	Term and late preterm (≥35 weeks)	769	Multivariate regression analysis Chance of TCB-SBr difference ≥2mg/dl	Yes Yes
Undefined	Brucker & MacMullen, (1987)	Term	20	NR	No
Visual - Kramer	Szabo et al., (2004a)	Preterm (34-36+6 weeks)	69	Mann Whitney	Yes
	Keren et al., (2008, 2009) ϕ	Term and late preterm (≥35 weeks)	812 (522 paired samples)	Logistic regression	Yes
Visual	Brits et al., (2018) ϕ	Term	96	Only 10% of Black infants diagnosed with jaundice looked clinically jaundice compared to 42% of non-Black infants	NA
Icterometer	Merritt & Coulter, (1994)	Preterm	90	Too small numbers to analyse fully	Trend
	Madlon-Kay, (1997)	Mainly term (mean GA 39 weeks)	171	T-tests and ANOVA	NS
	Madlon-Kay, (2001)	NR	164	ANOVA	NS
	Lee et al., (2019) ϕ	Term and preterm	790	Mean TCB at each Biliruler score	No

*Combination of BiliCheck and JM-103
GA - gestational age
ϕ - compared to TCB not SBr

Summary

Two large cohort studies have shown low Apgar score to be linked to increased neonatal mortality within all ethnicities. The comparative impact of ethnicity and race in Apgar scoring remains unclear as studies also highlighted inconsistent scoring of the Apgar, with Black infants more likely to receive lower scores. Additionally, although largely considered outside of the remit of the Apgar score, the ability of the Apgar to predict long term outcomes has been explored in a few studies with small sample sizes. These studies found inconsistent results.

Detection of cyanosis from skin colour was noted to lead to many false positives and false negatives. Within one small study the lips were considered the most reliable places to assess cyanosis across different ethnicities. A further study suggested that tongue colour being pink within the first 10 minutes of life suggested that a neonate has an oxygen saturation above 70% regardless of ethnicity. Pulse oximetry is more reliable than observation of the skin, although occult hypoxemia where pulse oximetry levels were 90% or more when arterial oxygen saturation was <85% was slightly more common in Black infants (9.2%) compared to White infants (7.7%) in one study. To determine the significance of these small, but potentially clinically significant, differences in arterial oxygen saturation compared to pulse oximetry saturation in neonates from diverse race and ethnicity backgrounds, further exploration is warranted.

TCB monitoring has been shown to be more accurate at detecting jaundice than visual assessment alone. When considering ethnicity, no consistent pattern has been shown when considering either correlation between TCB and SBr or mean bias (TCB reading – SBr level), with some studies finding higher values in those with dark skin tones or from ethnic minority backgrounds and other studies finding lower values or no difference. Area under the curve was consistently >0.90 across all ethnic and skin tone subgroups and across a range of SBr cut-offs suggesting outstanding discrimination in all TCB devices except the Bilimed. There was some suggestion that, within term infants, neonates of 'non-White' ethnicity or of darker skin tones required more blood tests when using TCB than neonates of White or light skin. Additionally, five of the seven studies statistically assessing the limits of agreement around the mean bias (TCB reading -SBr level) found variability was increased with increased pigmentation, with most of these TCB measurements noted to have been taken from the forehead. When considering general imprecision, none of the included studies found the mean difference between TCB and SBr to be underestimated by more than 40µmol/l regardless of the ethnicity of the neonate.

Two studies considered HCP training. In one, midwives and student midwives reported lack of education in clinical assessment of Black, Asian, and minority ethnic mothers and babies, with the study particularly focussing on assessment of the Apgar score. After the training, 96% of midwives felt the Apgar score was not the most appropriate way to assess all babies at birth due to the description of colour within it. The second study found maternity care assistants' knowledge and self-perceived confidence around detection of jaundice was not associated with actual ability to correctly estimate SBr levels from visual assessment.

The experiences of parents or carers: A systematic review

Aim

To identify the experiences of parents or carers seeking or receiving care for Black, Asian, or minority ethnic neonates in relation to neonatal assessments that include an element of skin colour, specifically for Apgar scores (including the detection of cyanosis) and jaundice.

Methods

A systematic review was undertaken, with the protocol published on the PROSPERO database (CRD42022341604).

Inclusion and exclusion criteria

The inclusion criteria for the study were defined using the PEOS acronym.

Population: Mothers, fathers, or other family members or guardians of neonates from Black, Asian, and minority ethnic backgrounds. Articles were included that exclusively recruited participants from ethnic minority backgrounds, or where the majority of the sample was from an ethnic minority background. Articles were limited to those published in countries with a very high development index (0.9 or above) (Human Development Office, 2020).

Exposure: Accessing or receiving health service care for a neonate (less than 4 weeks old) of Black, Asian or minority ethnic background within the last 12 months.

Outcome: Our primary aim was to synthesise the experiences of parents or carers of providing or seeking/receiving care for a neonatal infant (less than four weeks old) of Black or Asian, or ethnic minority background specifically for care that required observation of the skin, including Apgar scores, jaundice, and cyanosis or hypoxia.

Additionally potential barriers and facilitators to accessing care were identified in relation to overall care and also specifically for care that required observation of the skin including Apgar scores and jaundice.

Study design: All study designs were considered for inclusion including published and unpublished reports, as well as qualitative and quantitative methodologies. Reviews of potential relevance were not eligible for inclusion but were checked for further relevant citations. Where abstracts of potential relevance were identified with no corresponding full text article, if sufficient information was provided to establish the trustworthiness of the study then the abstract was considered for inclusion. Otherwise, study authors were contacted for further information.

Studies were limited to those published in the English language. Studies were included from database inception to April 2022.

Search strategy and study selection

Peer reviewed literature was identified by undertaking a comprehensive search of MEDLINE, CINAHL and PsycINFO databases. Grey literature was searched using OpenGrey within the DANS EASY archive and in the Patient Experience Library to identify unpublished studies that met the inclusion criteria. A combination of medical subject headings (MeSH) and key text words were used around ethnicity, Apgar, jaundice and hypoxia, as well as terms around neonate. Boolean logic was applied to combine the search terms with an example search strategy given in Appendix 10.

In addition to the above search strategy a group of expert stakeholders, including maternity user representatives, academics, healthcare professionals and commissioners were contacted regarding their awareness of further relevant literature for inclusion within the systematic review.

Retrieved citations were screened by title and abstract against the inclusion criteria by two researchers independently (AF and FF). Citations of potential interest were then read at full text independently by two researchers for inclusion

(AF and FF), with disagreements regarding inclusion discussed with a third researcher (HS).

References of all included studies were screened. All articles that have cited included studies or other studies of interest were also searched for further relevant literature.

Data extraction

Data was extracted from articles by two reviewers using a pre-defined data extraction template to record relevant information under the following headings: title, author, year of publication, country of origin, study design, sample (size, occupation/ maternity user status, ethnicity), methodology (sampling strategy, setting, data collection, data analysis), experiences, barriers, facilitators, study funding. Where possible, attention was given to the definition of the ethnicity or race categories used within each article. Where required any ambiguous information was checked with the authors of the original study.

Risk of bias assessment

The quality of included studies was assessed using the Mixed Methods Appraisal Tool (MMAT) (Hong et al., 2018). The MMAT includes two screening questions, followed by five quality criteria with each of five study design categories including qualitative, quantitative (non-randomised, randomised, descriptive) and mixed methods. These criteria relate to the appropriateness of methodology, data collection techniques and data analysis techniques. Each included study was appraised independently by two reviewers, with any disagreements discussed with a third reviewer.

Data synthesis

A narrative synthesis of the results was performed. This included a narrative discussion of the results of the included studies considering each neonatal assessment subgroup separately: Apgar, Jaundice and Cyanosis. Additionally potential barriers and facilitators to providing or accessing care were discussed in relation to overall care and also specifically for care that requires observation of the skin including Apgar scores and jaundice.

Consideration was given to other potential confounding factors for further subgroup analysis, such as education level, socio-economic status and maternal age, however information was not categorised by these factors within any of the studies. Consideration was also given to determine any differences in experiences with respect to the ethnicity of the neonates or the parent or carer, however this was not practical given the limited number of included studies.

Results

All database searches were undertaken on 10th May 2022, which identified a total of 8525 articles. An additional six articles were identified from other sources. Following title and abstract screenings 110 articles met the inclusion criteria for full-text review, of which a total of nine were eligible for inclusion within this review. A PRISMA flow diagram of the study search and selection process can be found in Appendix 11 (Page et al., 2021).

Characteristics of the included studies

The characteristics of the nine included studies (Birthrights, 2022; Casey et al., 1992; Fivexmore, 2022; Hannon et al., 2001; Hurst, 2001; Lee & Weiss, 2009; Palau et al., 2019; Tarnow-Mordi & Pickering, 1983; Witt et al., 2022) can be seen in Table 4.1. Six studies were conducted in the USA (n= 345 mothers or fathers) and three in the UK (n= 2,653 mothers or birthing people). Two studies exclusively included parents of preterm infants (Hurst, 2001; Witt et al., 2022), two studies included parents of both term and preterm infants (Casey et al., 1992; Tarnow-Mordi & Pickering, 1983), three studies mention preterm infants but did not report gestational age in full (Birthrights, 2022; Lee & Weiss 2009; Palau et al., 2019). In the final two studies, gestational age was not reported (Fivexmore, 2022, Hannon et al., 2001). Six of the studies obtained qualitative data through interviews and/or focus groups (Birthrights, 2022; Casey et al., 1992; Hannon et al., 2001; Hurst, 2001; Lee & Weiss, 2009; Witt et al., 2022), with one study using participant observation alongside the interviews (Hurst, 2001) and one doing a survey alongside the interviews/focus groups (Birthrights, 2022). Of the remaining studies one conducted a survey (Fivexmore, 2022), one described two separate case studies (Tarnow-Mordi & Pickering, 1983) and one compared structured interview responses with medical notes (Palau et al., 2019). Maternal age, education, age of neonate, language spoken and race/ethnicity varied substantially between studies (see Table 4.1).

The results of the critical appraisal using the MMAT are provided in Appendix 13. Four studies were quantitative descriptive studies, four qualitative studies and one mixed methods. All studies except one (Tarnow-Mordi & Pickering, 1983) were deemed to have had a clear research question and to have collected data that would allow the research question to be addressed. Of the four qualitative studies only two were deemed to have low risk of bias across all five domains (Hannon et al., 2001; Hurst, 2001). None of the four quantitative descriptive studies or the mixed methods study were deemed low risk of bias within all five domains.

Table 4.1. Characteristics of studies included in the systematic review of parents' experiences.

Author, year	Country	Study aims	Methodology	Sample characteristics	Ethnicity of sample	Data Analysis	Key findings
Birthrights, 2022	UK	To shine a spotlight on racism, the harms it causes, which human rights are in jeopardy and to provide concrete solutions. The focus was therefore the experiences of maternity care for Black, brown and Mixed Ethnicity women and birthing people	Mixed methods research comprising of surveys, focus groups, interviews and an online poll to assess the main findings.	Call for evidence survey: Women and birthing people n=187 HCP n= 57 Focus group: Women n=50 Midwives= 5 Interview: LGBTQ+ birthing people/ partners n= 3 Women n=2 Midwife n=1 Clinical negligence solicitors/barristers n=8 Online Poll: Women n= 1,069 GA: Mentions prematurity but no details given	Self-reported. Call for evidence survey: Black African 19%; Black Caribbean 20%; Other Black background 6%; Indian 10%; Pakistani 12%; Bangladeshi 2%; Chinese 0%; Other Asian 5%; White and Black Caribbean 9%; White and Black African 4%; White and Asian 2%; Any other mixed background 5%; Arab 0%; Not disclosed 6% Online Poll: 556 White; 145 Black/Black British; 227 Asian or Asian British; 141 Mixed.	Themes identified, tested and confirmed in qualitative data Descriptive statistics for the poll	Jaundice missed in infant with darker skin tones, resulting in hospitalisation and life-long consequences for the infant. Dismissal of mothers' concerns about jaundice by HCPs. Barriers included being ignored or disbelieved, racism and microaggressions, dehumanisation, lack of physical safety, poor communication preventing choice and consent, structural barriers including lack of diverse workforce, poor governance and policies not evidence based.

Author, year	Country	Study aims	Methodology	Sample characteristics	Ethnicity of sample	Data Analysis	Key findings
Casey et al., 1992	USA	To evaluate the accuracy of maternal reports of events during pregnancy, birth and the perinatal period.	Structured interviews using open-ended questions.	Mothers n=69 Mean age 25 years. 29% complete 12 years education. Infant n=69 GA: 17 were ≤36 wks; 48 were ≥37 wks Age of neonate: 29 days (range 5-80 days) Male/Female: 33/36	84% African American	Sensitivity and specificity statistics.	There was good agreement between hospital records and maternal recall for jaundice, infant birthweight and mode of birth, but not for other factors such as prematurity, foetal distress during labour and neonatal illnesses.

Author, year	Country	Study aims	Methodology	Sample characteristics	Ethnicity of sample	Data Analysis	Key findings
Fivexmore, 2022	UK	To gain insight into Black and Black mixed women's attitudes towards and experiences of maternity care in the UK.	Online survey deigned with support from midwives, doctors, health visitors, GPs and women's maternal organisations/ community groups and distributed on social media and through organisations.	Mothers n= 1340 Under 18 <1%; 18-25 19%; 26-35 65%; 36-45 15%; Over 45 <1% Education: Degree 52%; postgrad 25%; A level/ post 16 diploma 17%; GCSE 4%; no qualification 1% 53% care was given in London GA: NR	Self-identified: Black Total 76%: Black Caribbean 50% Black African 45% Black African & Caribbean 4% Other Black 1% Black mixed Total 24%: Black Caribbean & White 71% Black African & White 16% Other mixed 13%	Quantitative data: descriptive and inferential statistics Qualitative data: Thematic content analysis.	Mothers reported HCPs' poor understanding of clinical presentation of conditions in babies of non-European descent, including thinking natural skin tone was jaundice. Barriers included HCPs attitude, HCPs lack of knowledge of Black/Black/ Mixed anatomy, assumptions around pain and life circumstances and lack of choice of treatment. A diverse workforce made women feel more comfortable

Author, year	Country	Study aims	Methodology	Sample characteristics	Ethnicity of sample	Data Analysis	Key findings
Hannon et al., 2001	USA	To explore whether mothers currently express concerns about neonatal jaundice and perceive it as a serious condition; if so, to identify factors influencing these perceptions and to elicit maternal recommendations for improved healthcare interactions.	Ethnographic interviews using grounded theory methods. Audiotaped data were transcribed and analysed for themes using a qualitative data analysis software program	Mothers n= 47 Age, y [n (%)] 16-19 [6 (13%)] 20-24 [15 (32)%] 25-29 [12 (25%)] 30-38 [14 (30%)] Parity n(%) Multiparous 25 (53%) Primiparous 22 (47%) Marital status n(%) Married 31 (66%) Unmarried 16 (34%) Education High school graduate 23 (49%) High school graduate 9 (19%) High school education 15 (32%) GA: NR	Ethnicity [n (%)] Latina [34 (72%)] White, non-Hispanic [6 (13%)] African American [5 (11%)] Other [2 (4%)] Language of interview [n (%)] English [25 (53%)] Spanish [22 (47%)] Birthplace United States [23 (49%)] Other [24 (51%)]	Themes developed through an iterative process	Jaundice caused mothers anxiety. They perceived it as a dangerous condition and worried about short- and long-term repercussions for the baby. Mothers frequently felt they were to blame for the jaundice. They expressed misunderstandings due to cultural disparities and language hurdles and asked for more information about jaundice. Mothers' responses to information were influenced by their interactions with HCPs and other mothers who had first-hand experience with jaundice.

Author, year	Country	Study aims	Methodology	Sample characteristics	Ethnicity of sample	Data Analysis	Key findings
Hurst, 2001	USA	To understand how mothers described and understood their experience of a hospitalised, premature baby, including the mothers' actions in the NNU and the factors affecting their descriptions, interpretations and actions.	Ethnographic approach with participant observation at a Tertiary-level NNU (448 hours in total) and taped open-ended interview techniques.	Mothers n= 12 Maternal ages 20 to 43 years. Infants n=14 GA: Preterm: ≤34wks. Mean GA: 30 (±2.5) wks. Mean birthweight (g): 1503g (±584.9) Sex: 10 male 4 female. Length of time in NNU mean 48 days (range 9-118)	Ethnicity of mother: Euro-American n=7 Latina n= 4 African American n=1	Thematic content analysis	One mother was dismissed by HCPs when suggesting her child was having a false oxygen desaturation as the child had not changed colour. Barriers to care included power differential, inadequate staffing ratios, frequent HCP changes preventing relationships being built, lack of accessible information, fear of asking questions and HCPs misinterpreting maternal actions.

Author, year	Country	Study aims	Methodology	Sample characteristics	Ethnicity of sample	Data Analysis	Key findings
Lee & Weiss, 2009	USA	To explore the experiences of first-generation Chinese American parents while their infants are cared for in intensive care units (ICUs).	Phenomenological study conducting interviews in Mandarin using open-ended questions. Translated into English by two bilingual translators using detailed interview notes.	Families n=25 [56% of the families were nuclear families] Mothers n=25 Fathers n= 21 Grandmother n= 1 The mean age was 32.6 years for mothers and 33.1 years for fathers. GA: prematurity mentioned but no details given	First-generation Chinese American families.	Phenomenological reflection	Families were confused about their child's condition due to poor translation, especially of the word 'jaundice'. Barriers identified were lack of confidence to care for their child, blame from others and themselves, lack of social support, communication issues and cultural differences preventing parents from raising questions. HCPs of the same ethnicity were appreciated.

Author, year	Country	Study aims	Methodology	Sample characteristics	Ethnicity of sample	Data Analysis	Key findings
Palau et al., 2019	USA	To explore disparities in communication between English and Spanish-speaking parents and their neonatal intensive care unit (NNU) HCPs.	Prospective, observational study admissions records screened structured interview in primary language.	132 parents (mother or father) GA: prematurity mentioned but no details given	English (n = 88) or Spanish (n = 44) as their primary language.	Statistical tests – Chi square r Fisher exact for categorical, t-test for continuous. Multivariate logistic regression.	Spanish-speaking parents are four times more likely to incorrectly identify their child's diagnosis and had a lower understanding of laboratory tests. More than half of Spanish-speaking parents did not know why their child was admitted to NNU Spanish-speaking parents were updated in their language 39% of the time by doctors and 44% by nurses, despite an interpreter being on-site in both hospitals for 12 hours each day, with communication the most common improvement parents would like

Author, year	Country	Study aims	Methodology	Sample characteristics	Ethnicity of sample	Data Analysis	Key findings
Tarnow-Mordi & Pickering, 1983	UK	To report on two infants whose jaundice was diagnosed and treated late.	Case study/ report of two infants requiring exchange transfusion.	Infants n=2 Case 1- GA 33 wks Case 2- GA 38 wks	Case 1- West Indian parents. Case 2- Nigerian	Not reported.	A mother saw discolouration in her infant's eyes five days before a doctor noticed, but she did not want to trouble the staff. The infant subsequently displayed delayed development.
Witt et al., 2022	USA	To characterise the lived experiences of stress associated with having a preterm infant hospitalised in the NNU among Black and Hispanic mother.	Secondary analysis of in-depth interviews with data collected as part of two prior studies.	Mothers n= 39 Data set one: Maternal age median: 28 years GA: 24-36 weeks Data set two: Maternal age median: 34 years GA: 29-36 weeks	All self-identified as Black or Hispanic. Non-Hispanic Black: n= 19 (49%) Hispanic Spanish-speaking: n= 16 (41%) Hispanic English-speaking: n= 4 (10%)	Content analysis	Watching their baby go pale when struggling to breathe made mothers fearful. Stressors for mothers were identified as individual factors, hospital factors, language barriers, logistical issues (i.e., transportation) and community factors (i.e., immigration issues and lack of social support).

UK – United Kingdom
USA- United States of America
GA – gestational age
wks – weeks
HCP – healthcare professionals
NNU – neonatal unit

Conditions assessing skin colour

The findings of the study are presented below under each subgroup based on the different neonatal examinations involving skin colour: Apgar score, cyanosis, and jaundice. Barriers and facilitators to accessing care identified within the included studies are presented alongside the stakeholder interview results in the chapter on the impact of ethnicity and race on care.

The Apgar score:

No studies were found that described women's experiences of the Apgar score assessment in their newborn baby.

Cyanosis:

Two studies described women's experiences of cyanosis in the neonate (Hurst, 2001; Witt et al., 2022).

Mothers recalled observing colour changes in their infants when they were experiencing cyanosis or hypoxia (Hurst, 2001; Witt et al., 2022). Colour changes in their infants incited fear in the mothers, with one participant suggesting watching her child 'go pale' (Witt et al., 2022). Another mother observed the lack of 'colour change' in her infant when monitors were alarming to suggest hypoxia. The mother unsuccessfully tried to explain to the nurse that due to the lack of colour change and the baby's heart rate remaining good that she felt the infant was therefore not having an oxygen desaturation (Hurst, 2001). It was unclear within this study however whether this was an experience with a Black, Asian or minority ethnic neonate as the participants were of mixed ethnicity.

Jaundice:

Seven studies considered women's experiences of jaundice (Birthrights, 2022; Casey et al., 1992; Fivexmore, 2022; Hannon et al., 2001; Lee & Weiss, 2009; Palau et al., 2019; Tarnow-Mordi & Pickering, 1983). These studies explored concerns around the detection of jaundice, as well as concerns around communication and understanding and the challenges of treatment.

Ability to detect jaundice:

Multiple studies expressed concerns about healthcare professionals (HCP) ability to detect jaundice in infants with darker or different skin tones (Birthrights, 2022; Fivexmore, 2022; Tarnow-Mordi & Pickering, 1983). One mother recounted an instance where an HCP was 'adamant' her child of mixed ethnicity had jaundice and could not understand the clinical appearance of conditions of infants that weren't of European descent.

> *'She didn't listen or understand that mixed-race babies come in different shades.' (Fivexmore, 2022)*

In two instances mothers noticed the yellowing of their child's eyes which was still not noticed until later by HCPs (Birthrights, 2022; Tarnow-Mordi & Pickering, 1983). Another mother described the persistence required to convince a health visitor to test her baby's jaundice levels. Even after a 'super high' reading came back, the HCP still insisted nothing was wrong with the infant, which necessitated further persistence from the mother to achieve a referral to hospital. Shortly after the baby was hospitalised, a further HCP suggested there was nothing to indicate 'severe jaundice' (Birthrights, 2022). A further instance was described by a clinical negligence solicitor where failure to listen to the mother's concerns over jaundice resulted in an infant developing cerebral palsy (Birthrights, 2022).

> *'The White staff did not recognise jaundice in a Black baby.' (Birthrights, 2022)*

Some mothers were hesitant to raise concerns about their infant having jaundice (Tarnow-Mordi & Pickering, 1983), while other mothers who did raise concerns to HCPs described feeling 'totally dismissed'. (Birthrights, 2022; Fivexmore, 2022)

> *'His mother had noticed this on the third day but had not wanted to trouble the staff.' (Tarnow-Mordi & Pickering, 1983)*

Poor communication and lack of understanding

One study found good agreement between maternal recall of neonatal jaundice and a diagnosis within the medical records (Casey et al., 1992). However, in several other studies, mothers were reported to have a lack of knowledge and understanding about jaundice, leading them to misunderstand their child's condition or to be uncertain of the cause of their child's jaundice (Hannon et al., 2001; Lee & Weiss, 2009; Palau et al., 2019). For many the lack of understanding was due to poor communication.

Poor communication

Women were embarrassed when English was not their first language, especially around their limited knowledge of medical terms which made them feel uncomfortable learning about jaundice in English and prevented them from asking questions to English-speaking staff (Hannon et al., 2001). These language barriers often contributed to mothers' misunderstanding of their child's jaundice (Hannon et al., 2001). In particular, mistranslation of the term jaundice led mothers to misunderstand their baby's condition and its aetiology (Lee & Weiss, 2009; Palau et al., 2019). One instance saw a mother believing jaundice had something to do with her infant's cholesterol (Lee & Weiss, 2009). In addition, one Spanish family interpreted their child's diagnosis of 'stridor and hyperbilirubinemia' as 'his throat is loose or something.' (Palau et al., 2019).

> *'In Mandarin, bilirubin sounds as 'tan-hung-su' and cholesterol sounds as 'tan-ku-chun.' (Lee & Weiss, 2009)*

Further issues arose among Spanish speakers due to the translation of 'bilis' meaning anger in Spanish. This accentuated mothers' self-blame for their infant's condition.

> *'Bilirubin, they are like bilis when one gets mad. I felt that when I was pregnant and had so much stress, that's what caused it.'*
> *(Hannon et al., 2001)*

Poor communication with HCPs and inadequate explanations led women in one study to ask interviewers for more information of jaundice causes (Hannon et al., 2001). While mothers reportedly liked having interpreters present, they preferred having direct conversations with an informed person as they feared information being lost in translation (Hannon et al., 2001).

> *'I want to directly speak to her [the physician] and I want to ask my questions … I am left with doubts.' (Hannon et al., 2001)*

Despite the difficulties with communication, medical staff were the most respected source of information about jaundice, with friends and family also used to obtain information (Hannon et al., 2001). Most women expressed a preference for being informed about jaundice in the antenatal period, although some were happy to only receive this information if their infant experienced jaundice. Women would prefer information to be provided in written leaflets and/or videotapes, one-on-one interactions, or small-group discussions (Hannon et al., 2001).

Lack of understanding

Parents had varied beliefs about the cause of neonatal jaundice. Some parents understood biomedical explanations for the onset of jaundice including that the infant had high bilirubin levels, that their blood cells were 'not functioning' or that the infant was 'reacting to mothers' cells' (Hannon et al., 2001). Other explanations proposed by parents included an assumption that babies were born with jaundice, that it was a 'development problem' or that breastfeeding caused jaundice due to the quality or quantity of breastmilk (Hannon et al., 2001). Not being given an explanation as to what jaundice was, led some mothers to make assumptions such as hepatitis was causing the jaundice (Hannon et al., 2001). Additionally, some mothers blamed themselves for causing their infants jaundice, due to what they did or didn't do during pregnancy, when breastfeeding, or due to their current illnesses and medications (Hannon et al., 2001).

> *'They haven't told me very well what happened.' (Hannon et al., 2001)*

> *'I just wanted to be reassured that was it me causing the jaundice or was it something natural? Just let me know so I can correct it.' (Hannon et al., 2001)*

The severity of jaundice was perceived differently by parents. Some mothers described professionals emphasising the danger of high levels of bilirubin and the risk of brain damage, whilst others were left to their own thoughts. Parents were concerned when blood monitoring ceased and visual assessments were carried out in isolation that the jaundice may return and have long-term effects on their child's health. In contrast, some parents perceived jaundice not to be serious, comparing the illness to a 'cold'. Parents felt assured by the doctor's reassurance and as their infant's skin colour returned to normal.

> *'And every time I called the hospital to see if I could pick him up, it [bilirubin level] just got higher. And that's what really upset me. What was wrong? Why was it getting higher? I was afraid something might happen to him. I hear they get brain-damaged if they're not treated?' (Hannon et al., 2001)*

> *'If you bring the baby home and he begins to turn yellow, you get scared; then you think it's something serious.' (Hannon et al., 2001)*

> *'One mother, whose neonate had mild hemolytic jaundice, stated: 'This second pregnancy I read about what causes yellow jaundice. When the nurse told me [that the second infant had jaundice], I was more clear. I knew it wasn't serious.' (Hannon et al., 2001)*

Treatment challenges

Mothers indicated that the blood testing process required to determine bilirubin levels was difficult to watch and painful for both the mother and child. They noted the frequency of blood drawing and lack of success with the first attempt. Mothers used the terms 'screamed' or 'suffered' to describe the reactions of their neonates and 'poked', 'bruised', 'tormented', or 'tortured' to describe the actions of medical staff (Hannon et al., 2001).

> *'Hearing him crying; nothing that I can do. Doctors sticking your baby. But as a mother, crying and only afterwards holding him.' (Hannon et al., 2001)*

Mothers also found separation from their infant during treatment or testing to be difficult. Feelings of guilt arose and lessened their ability to bond with their baby, with these feelings heightened if mothers were discharged from the maternity ward without their child (Hannon et al., 2001). By asserting themselves, some mothers negotiated extended maternity stays or home phototherapy to minimize the separation from their infant. However, the offer of home phototherapy, without medical supervision, also left mothers feeling pressured to watch their infants carefully (Hannon et al., 2001).

> *'Right after they took her away, I didn't want to go see her because I knew I was gonna want to hold her, and I didn't want to take her out of the light.' (Hannon et al., 2001)*

Mothers' concerns continued whilst their babies underwent treatment (Hannon et al., 2001). The phototherapy lights scared some mothers and others raised concerns about their baby overheating, the effectiveness of the treatment and the impact of the blindfold on the neonate.

> *'I knew it [phototherapy] would help the level go down and the lights looked harmless, but it was awful, horrible. The blindfold. I worried my baby will freak out and cry.' (Hannon et al., 2001)*

Summary

No studies conducted in high income countries considered the views of Black, Asian or minority ethnic parents or carers on the Apgar score, or how parents from ethnic minority backgrounds may use skin colour to detect concerns about whether their baby is getting enough oxygen or around subsequently accessing care. Just two studies reported incidences where mothers talked about their baby's struggles to breathe as part of a wider study of care experiences.

Only seven studies were found that considered parents from Black or minority ethnic backgrounds experiences of neonatal jaundice. Only three studies considered the identification of jaundice. These three studies reported jaundice to be inadequately identified in their infant by HCPs with concerns dismissed by an HCP in one study and the woman not wanting to make a fuss by raising her concerns in another study. The other studies focussed on parents' inadequate understanding of jaundice when there was a language barrier, the anxiety they felt after a jaundice diagnosis, as well as how well they remembered the jaundice diagnosis over the following months.

Policy review

Aim

The aim of this review was to examine current policies and guidance in relation to their consideration of Black, Asian, and minority ethnic neonates in common assessments such as Apgar scoring (including assessment of oxygenation) and the detection of jaundice.

Methods

Inclusion and exclusion criteria

Documents considered eligible for inclusion were clinical or practice guidelines, frameworks for practice (including draft frameworks), quality standards, good practice points, and relevant learning tools.

Any policy or guideline covering care of the neonate, including routine care as well as specific practices such as Apgar scores, cyanosis, and jaundice were eligible for inclusion.

Guidelines or policies looking exclusively at pharmacological practices were excluded.

Search strategy

A desktop search was undertaken and included searches of guidance produced by the National Institute for Health and Care Excellence (NICE), Office for Health Improvement and Disparities (OHID), the Royal College of Obstetricians and Gynaecologists (RCOG), The Royal College of Midwives (RCM), the Neonatal Nurses Association (NNA), the British Association of Perinatal Medicine (BAPM), the Royal College of Paediatrics and Child Health (RCPCH), the Resuscitation Council UK and the Institute of Health Visiting (iHV). These searches focused on United Kingdom (UK) guidance. Further searches of the Cochrane database and World Health Organization (WHO) looked for any further international guidance applicable in the UK context. Example search strategies can be found in Appendix 13. Only the most recent version of any guidance was included, with

guidance only considered if published from 2010 onwards to ensure relevance. Searches were also made of DansEasy and Google for documents that met the inclusion criteria. Additionally, experts in the field and project advisory committee members were asked to identify any policies or guidance they thought relevant for inclusion.

Study selection

Retrieved guidance was screened by two researchers independently against the inclusion criteria. Any disagreements were discussed with a third researcher. Where possible, the references on which guidance was based regarding applying common practices in Black and Asian neonates, were also retrieved. Additional materials signposted in the included documents were viewed and subsequently analysed for their content.

Data extraction

All guidelines and publications were analysed for content relating to the use of skin colour as a means of diagnosis or identification of the conditions mentioned above. Keywords were identified relevant to the research question by two researchers and confirmed within the team. Keywords were searched within the included documents to identify relevant sentences and paragraphs in the text. These keywords were: 'Apgar', 'pale', 'blue', 'colour', 'color', 'cyanosis', 'hypox', 'oxy', 'red', 'yellow', 'pink', 'jaundice', 'Black', 'Asian', 'Cauc', 'ethnicity', 'pigment', 'skin', 'dark' and 'ethnic'. This data was extracted onto a pre-determined framework.

Critical appraisal

A detailed assessment of the guidelines according to the AGREE II (Appraisal of Guidelines for Research & Evaluation II) instrument (Brouwers et al., 2017) was deemed to be outside of the scope of this review, however an overall assessment was made regarding the relevance of the guidelines and the impact of ethnicity on neonatal assessment.

Data synthesis

A narrative synthesis of the results was performed, detailing how practices and recommendations were to be applied to Black, Asian, and minority ethnic neonates, alongside language used particularly for guidance or policies that described care that requires observation of the skin including Apgar score, jaundice, and cyanosis.

Results

We identified 18 guidelines/frameworks for practice/quality standards developed for the use of healthcare professionals (see Table 5.1) as well as learning tools developed by one organisation (see Table 5.2). Of the 18 policies, guidelines, and training resources reviewed, nine guidelines focused on general care of neonates, six exclusively detailed the assessment of cyanosis or hypoxia, one neonatal assessment using the Apgar score, and two exclusively focussed on the assessment of jaundice. Of the 18 policies, 15 were UK polices and three were international policies.

Table 5.1. Characteristics of identified guidance

No	Year of publication	Organisation	Title	Type of file
1	2022a	British Association of Perinatal Medicine (BAPM)	Postnatal care of the late preterm Infant	Draft Guideline
2	2022b	British Association of Perinatal Medicine (BAPM)	Deterioration of the newborn	Draft Guideline
3	2015	British Association of Perinatal Medicine (BAPM)	Newborn early warning trigger and track (NEWTT)	Framework for practice
4	2022	Institute of Health Visiting (iHV)	Updated Good Practice Point – Babies who have neonatal jaundice	Good Practice point
5	2010 Guideline (updated 2016) & 2014 quality standard.	NICE (National Collaborating Centre for Women and Children's Health) & RCOG (Royal College of Obstetricians & Gynaecologists)	Jaundice in newborn babies under 28 days	Guideline
6	2019	NICE	Specialist neonatal respiratory care for babies born preterm	Guideline
7	2017	NICE	Intrapartum care for healthy women and babies	Guideline
8	2021	NICE and RCOG	Postnatal care	Guideline

No	Year of publication	Organisation	Title	Type of file
9	No date a	Neonatal Nurses Association (NNA)/ University of Hertfordshire	Assessment of the neonate	Resources for Nursing Practice
10	No date b	NNA/University of Hertfordshire	Monitoring vital signs in the neonate.	Resources for Nursing Practice
11	2021	Office for Health Improvement and Disparities (OHID)	Newborn infant examination	Guidance
12	2011	Resuscitation Council UK	Air/oxygen blenders and pulse oximetry in resuscitation at birth	Quality standard
13	2021	Resuscitation Council UK	Newborn resuscitation and support of transition of infants at birth Guidelines	Guideline
14	2012	Royal College of Midwives (RCM)	Immediate care of the newborn	Guidelines
15	2017	Royal College of Nursing (RCN)	Standards for assessing, measuring and monitoring vital signs in infants, children and young people	Guideline
16	2022a	World Health Organization (WHO)	WHO recommendations on postnatal care of the mother and newborn	Guideline
17	2022b	WHO	Early essential newborn care	Clinical practice guide
18	2012	WHO	Basic newborn resuscitation	Guideline

Table 5.2. Health Education England (HEE, 2022) e-learning packages

Module title	Topics included
Avoiding Term Admissions into Neonatal units (ATA)	Physiology of Jaundice. Term Newborn Babies at Risk of Jaundice First Hour Care of the Term Newborn Infant Respiratory Distress
NHS Newborn Infant Physical Examination (NIPE) Programme	Screening and the Newborn and Infant physical examination (NIPE) Screening examination of the eyes Screening examination of the cardiovascular system Further information for NIPE practitioners
Midwifery Identification, Stabilisation and Transfer of the Sick Newborn (MIST); Midwifery Identification, Stabilisation and Transfer of the Sick Newborn	Colour (Anaemia and Cyanosis) Feeding and Abdominal Concerns Prematurity Hypoxia and Encephalopathy

Routine care

General skin colour

Some assessment of skin colour as part of a routine care was covered within 11 guidelines (BAPM, 2015; BAPM, 2022a; BAPM, 2022b; NICE, 2019; NICE, 2021; NNA/University of Hertfordshire, no date (N.D.)a; OHID, 2021; RCM, 2012; RCN, 2017; Resuscitation Council, 2021; WHO, 2022b) and one training resource (HEE, 2022). Guidelines suggest that an initial assessment of skin colour within the first few hours of life should be used to assess adaptation of the newborn to extra-uterine life (RCM, 2012; Resuscitation Council, 2021; BAPM, 2022a; BAPM, 2022b; NICE, 2021; WHO, 2022b), including for preterm infants (NICE, 2019). For example, observations should include:

> *"Colour, tone, breathing and heart rate." (RCM, 2012) or*
>
> *"Colour, activity, temperature, heart rate and respiratory rate for the first 24 hours." (BAPM, 2022a)*
>
> *A "normal baby should be centrally pink, though the extremities of hands and feet are usually tinged blue". (HEE, 2022)*

After the neonate has detached from the breast it should be assessed to ensure it is "breathing well, pink and warm." (WHO, 2022b)

Discoloured peripheries and alternative locations

Other guidelines identified that discoloured peripheries could indicate a change in the infant's condition which should be observed, particularly while the infant is in skin-to-skin contact.

"The baby should be assessed by looking at the whole of the baby's body as the limbs can often be discoloured first. Subtle changes to colour indicate changes in the baby's condition." (BAPM, 2022b)

Colour is one of the key components within the neonatal early warning trigger and track (NEWTT) chart (BAPM, 2015). Within the original guidance, a cause for concern was considered to be an infant that was "pale or blue or had an oxygen saturation < 90%" while a colour of "pink" or oxygen saturation above 95% was considered normal (BAPM, 2015). Within the new draft guidance, the skin colour components have been updated to "very pale or blue" as a cause for concern and to "pink or normal" as no cause for concern (BAPM, 2022b).

Additionally, skin colour is one of the components used to assess adequate respiration in infants (RCN, 2017). As well as considering skin colour when initially assessing the neonate, the NNA/ University of Hertfordshire advises assessing mucous membranes, with pink membranes seen as normal and blue mucous membranes viewed as an abnormal sign that requires action (NNA/ University of Hertfordshire, N.D.a). An infant's skin colour not being sufficient to assess oxygenation is also highlighted within the HEE MIST e-learning (HEE, 2022), with colour change to an infant's tongue, gums or lips being considered most reliable.

To assess infant wellness, a Newborn and Infant Physical Examination (NIPE) is conducted within the first 72 hours (OHID, 2021). The NIPE screening examination guideline advises healthcare professionals to observe the infants "general tone & central and peripheral colour" when considering for signs of congenital heart abnormality, as well as looking for signs of concern such as:

"Episodes of apnoea lasting longer than 20 seconds or associated with colour change" (OHID, 2021) or "central cyanosis." (HEE, 2022; OHID, 2021) or "poor colour". (HEE, 2022)

Additionally, when screening for heart problems the healthcare professionals should ask parents if they have observed any colour changes in their infant (HEE, 2022; OHID, 2021).

> *"Parents should be asked if their baby ever gets breathless or changes colour at rest or while feeding." (OHID, 2021)*

Within all of these policies for routine neonatal care and assessment using skin colour, there is no mention of how ethnicity may impact the assessment of skin colour.

Apgar score

Two guidelines included for review provided advice on Apgar scoring (NICE, 2017; NNA/University of Hertfordshire, N.D.a). Additionally, the training resource (HEE, 2022) noted that the Apgar score should be undertaken. The first guideline recommended recording the Apgar score routinely at 1 minute and 5 minutes, but only provided guidance on the components included and not on how to determine the score (NICE, 2017). The other, made by the University of Hertfordshire (signposted to by the Neonatal Nurse Association), described colour assessment of neonates for determining the Apgar Score as:

> *"White for a score of 0, Blue for a score of 1 and Pink centrally for a score of 2". (NNA/University of Hertfordshire, N.D.a).*

This, however, differs from the standard Apgar score definition of "pale or blue" for a score of 0, "body pink and extremities blue" for a score of 1 and "completely pink" for a score of 2 (Knight, 2020).

There was no mention of the consideration of ethnicity or possible variations in relation to skin colour assessment for the Apgar score within either guideline or training resource (HEE, 2022; NICE, 2017; NNA/University of Hertfordshire, N.D.a). However, the HEE (2022) training acknowledged that colour assessment was "subjective" and "should not be relied upon in isolation".

Jaundice

The detection of jaundice was considered within seven guidelines (NCC-WCH & RCOG, 2010; NNA/University of Hertfordshire, N.D.a; WHO, 2022a; WHO, 2022b; BAPM, 2022a; BAPM, 2022b; iHV, 2022), as well as in the HEE training resource (HEE, 2022). The guidelines detailed the key manifestation when assessing neonates for jaundice to be a "yellow" colour of the skin, (HHE, 2022; iHV, 2022; WHO, 2022b; NCC-WCH & RCOG, 2010/2016; NICE, 2014; NICE, 2016) as well as yellowing of the sclerae (HHE, 2022; iHV, 2022; NCC-WCH & RCOG, 2010; NICE, 2016) and the palate or mucous membranes (HEE, 2022; NCC-WCH & RCOG, 2010). The description of jaundice as the "yellowing of the skin or whites of the eyes" was also noted within the example guide to provide to parents of late preterm infants (BAPM, 2022a).

Other guidelines suggest assessing the peripheries for yellow discoloration with “yellow palms and soles at any age” being of concern (WHO, 2022b; WHO, 2022a). When assessing for jaundice the importance of good lighting is emphasised with jaundice looking worse in artificial light and potentially being missed in poor lighting (WHO, 20222b; NCC-WCH & RCOG, 2010; NICE, 2014; NICE, 2016). Additionally, the institute of Health visitors provided detailed advice around the detection of prolonged jaundice from stool and urine colour, with immediate referral to a paediatrician mandated if abnormal in colour (iHV, 2022).

Several guidelines on assessment of the condition of the neonate advised that skin should be assessed for signs of jaundice at every contact (NNA/University of Hertfordshire, N.D.a; BAPM, 2022b; BAPM, 2022a), especially in the first 72 hours (BAPM, 2022b; HEE, 2022). BAPM (2022b) noted that bilirubin should be monitored with any concerns regarding jaundice, although assessment using transcutaneous bilirubinometers was not recommended within the first 24 hours after birth. HEE (2022) recommends the use of either transcutaneous bilirubinometer or another laboratory assessment if an infant is visibly jaundiced and advises healthcare professions not to “guess” the jaundice levels by the depth of the yellow discoloration. One guideline (BAPM, 2022b) noted babies of Asian ethnicity to be at increased risk of jaundice. No guidelines except for NICE considered how jaundice may present within neonates of different ethnicities, or how to mitigate the difficulties associated with different skin pigmentations and phenotype. The NICE guidelines and quality standard (NCC-WCH & RCOG, 2010; NICE, 2014; NICE, 2016) describe potential issues with diagnosis in neonates with different skin tones.

> *“Babies with very pale skin can appear ‘suntanned’ rather than yellow.” (NCC-WCH & RCOG, 2010).*
>
> *“Detection of jaundice in babies with dark skin tones can be almost impossible.” (NCC-WCH & RCOG, 2010).*
>
> *“Clinical recognition and assessment of jaundice can be difficult, particularly in babies with darker skin tones.” (NICE, 2016).*

As a result, the NCC-WCH & RCOG guideline (2010) and quality standard (NICE, 2014) state examination of the “sclerae, gums and blanched skin is useful across all skin tones.” This has since been updated (NICE, 2016) where the terminology “all skin tones” has been removed and replaced with directions of how to assess “blanched skin”.

> *“Examine the sclerae and gums and press lightly on the skin to check for signs of jaundice in ‘blanched’ skin.” (NICE, 2016)*

The e-learning provided by HHE (2022) also considered ethnicity. They used a Pakistani infant as a case study and asked participants to describe the “risk factors” of the infant. They noted that jaundice becomes visible at about 80 µmol/L in infants with “pale skin” and suggested that infants from ethnic minority

groups might require “closer monitoring”. In darker skin tones they suggested “Assessment of the sclera and mucus membranes may be more reliable” than skin colour assessment (HEE, 2022).

The NICE guidance recognised that visual assessment for jaundice is not recommended to be used in isolation.

> *“Do not rely on visual inspection alone to estimate the bilirubin level in a baby with suspected jaundice.” (NCC-WCH & RCOG, 2010, NICE, 2014, NICE, 2016).*

However, they also advised against measuring bilirubin routinely in all infants (NCC-WCH & RCOG, 2010; NICE, 2016):

> *“Do not measure bilirubin levels routinely in babies who are not visibly jaundiced” (NCC-WCH & RCOG, 2010; NICE, 2016).*

In contrast, the latest WHO postnatal care guidance (WHO, 2022a) recommends universal screening for neonatal hyperbilirubinemia by transcutaneous bilirubinometer (TCB) at health facility discharge. While recognising total serum bilirubin as the most accurate method of estimation, it was also recognised that a heel prick test requires access to laboratory assessment which is not globally available. There was insufficient evidence for or against universal serum bilirubin screening at health facility discharge (WHO, 2022a).

The above policies for assessment of jaundice are all largely skin colour dependant. There was some mention of ethnicity potentially hindering the detection of jaundice within some of the guidance. Additionally, one guideline (WHO, 2022a) detailed the potential for transcutaneous bilirubinometer assessment to overestimate bilirubin levels in newborns with dark skin tones, but that current evidence is conflicting.

Cyanosis or hypoxia

The detection of cyanosis and/or hypoxia is mentioned within eight guidelines/policies (NICE, 2017; NICE, 2019; NICE, 2021; NNA/University of Hertfordshire, N.D.b; OHID, 2021; RCN, 2017; Resuscitation Council, 2011; Resuscitation Council, 2021) and one training resource (HEE, 2022).

When an infant is born at higher risk (e.g. through passage of meconium in utero or prolonged rupture of the membranes) the NICE intrapartum guidelines suggests central cyanosis should be observed for and “confirmed by pulse oximetry if available” (NICE, 2017). Similarly, one sign seen as suggestive of congenital heart disease in the neonate was “central cyanosis” (OHID, 2021).

When observing for respiratory issues, observations of “skin colour, pallor, mottling, cyanosis and any traumatic petechiae around the eyelids, face and neck” should be undertaken (RCN, 2017). An infant that appears “mottled”, “blue” or “pale” in colour was seen within the training resource to potentially be a sign of respiratory distress (HEE, 2022). Another guideline noted an abnormal respiratory assessment criterion for a neonate would be “cyanosis” (NNA/ University of Hertfordshire, N.D.b).

One guideline specifically addressed how to assess for cyanosis in a neonate (NICE, 2021). They referred to skin colour assessment, with the healthcare professional prompted to observe for the neonate “appearing pale, ashen, mottled or blue” (NICE, 2021). None of these guidelines explicitly state how cyanosis or central cyanosis may be observed in relation to different skin tones, phenotypes, or ethnic characteristics. However, the online resource (HEE, 2022) is more explicit suggesting “any change from the established centrally pink colour in a newborn is always abnormal” and that central colour change was most “reliably noted in the lips, gums and tongue” and can be “paleness or dusky blue”.

An additional guideline made it very explicit that skin colour alone should not be used in the recognition of cyanosis.

> *“Colour is a poor means of judging oxygenation as cyanosis can be difficult to recognise.” (Resuscitation council, 2021).*

The training resource (HEE, 2022) went one step further suggesting “skin colour also varies according to genetic determinants and therefore is not a good means of assessing oxygenation”. For this reason, they suggested any baby with a “dusky appearance” should have their pulse oximetry checked (HEE, 2022).

Regarding hypoxia, one guideline noted that if an infant is floppy, breathing inadequately with a very slow heart rate of less than 60 beats per minute, where they may also be “pale” in colour it is suggestive of significant hypoxia (Resuscitation council, 2021). However, it was again noted that visual observation alone should not be relied upon:

> *“One cannot accurately assess oxygenation, and particularly hyperoxia, by colour alone.” (Resuscitation council, 2011).*

For this reason, the Resuscitation Council suggest that only in the “rare circumstance” of an absence of “both a pulse oximeter and an air/oxygen blender” that 100% oxygenation should be given and assessed by observing the colour and heart rate, however, this should be avoided, wherever possible or “rectified as soon as possible” (Resuscitation council, 2011).

The difficulty of assessing hypoxia in a preterm neonate was particularly noted in one guideline (NICE, 2019). It recognised that there was currently insufficient

evidence of the accuracy of pulse oximetry or transcutaneous measurement of partial pressure of arterial oxygen compared to arterial oxygen levels (NICE, 2019).

None of the guidelines or training resources around hypoxia or cyanosis specifically mentioned issues with skin colour assessment or issues of accuracy when assessing oxygen saturation in neonates of different ethnicities or skin pigmentation.

Evidence used in guideline developments:

Some of the guidelines, frameworks, and quality standards detailed evidence for the basis of the recommendations. Where included the evidence for the recommendations given in the guidelines, policies, and frameworks is discussed below.

Jaundice

NICE evidence summary

The NICE guideline (NCC-WCH & RCOG, 2010; NICE, 2014; NICE, 2016) evaluated evidence suggesting darker skin tone can hinder the assessment of neonatal jaundice. The evidence provided in the evidence summary of the 2010 document, suggested that the value of the correlation coefficient was much less for preterm babies and babies with dark skin tones compared with babies with light skin tones and term babies (NCC-WCH & RCOG, 2010). The guideline development group also recognised that in one study parental assessment of jaundice was more accurate than that of the healthcare professional (Maldon-Kay, 1997).

The guideline acknowledged the difficulty of estimating jaundice using visual assessment methods suggesting it to be “moderately correlated” accuracy, but they stated that bilirubin levels should not be routinely measured in “babies who are not visibly jaundiced”.

The evidence was deemed to adequately demonstrate the accuracy of transcutaneous bilirubinometers (Minolta JM-103 and BiliCheck) in term babies with low bilirubin levels (bilirubin < 250 µmol/litre) (NCC-WCH & RCOG, 2010). They also suggested that the BiliCheck produced more accurate results than the Minolta JM-102 or JM-103 in those with dark skin tones (NCC-WCH & RCOG, 2010). However, while more accurate than visual inspection, the BiliCheck was noted to be less accurate in those of darker skin tones (NCC-WCH & RCOG, 2010). However, the guideline stated that more research is needed into “transcutaneous bilirubin screening” and “risk factors”, which include babies with “dark skin tones” to assess diagnostic accuracy between serum bilirubin and transcutaneous bilirubin devices (NCC-WCH & RCOG, 2010). In addition,

they felt that there was a lack of good-quality evidence regarding the use of icterometers in babies with dark skin (NCC-WCH & RCOG, 2010).

WHO recommendations

The WHO updated guideline for postnatal care acknowledged potential difficulties with transcutaneous bilirubin measurements overestimating serum bilirubin levels in neonates with darker skin tones but noted that the evidence is conflicting (WHO, 2022a). While recommending that transcutaneous bilirubin assessment should be undertaken universally at discharge, the recommendations considered there to currently be insufficient evidence to recommend universal screening of total serum bilirubin at health facility discharge given the large associated costs and that the feasibility and acceptability of the test varied markedly (WHO, 2022a).

Postnatal care

NICE evidence summary

The 2021 review and drafting of recommendations were completed "afresh". The previous 2006 guideline included "Healthy babies should have normal colour for their ethnicity," (Dermott et al., 2006) however the different appearance of healthy skin colour in different ethnicities was not mentioned within the newer guidance (NICE, 2021). Within the latest review, all studies associated with scoring systems for illness in babies and signs and symptoms of serious illness in babies were suggested to be of very low to moderate quality and have a paucity of evidence. Ethnicity is not addressed in any of the evidence in either summary.

Specialist neonatal respiratory care for babies born preterm: NICE evidence summary.

When developing the NICE "Specialist neonatal respiratory care for babies born preterm" guidance, no studies were found that considered the diagnostic accuracy of pulse oxygen saturation or colour assessment of the neonate (NICE, 2019).

Summary

UK policy was assessed to understand whether ethnicity and race were appropriately considered in policy formulation regarding skin colour and neonatal examinations. The results showed that the quality of the guidelines varied widely, with some guidelines providing comprehensive and evidence-based recommendations, while others were limited in scope and lacked rigour. The impact of ethnicity was poorly considered during policy formulation and the development of guidelines and training. Multiple policies referred to terms such

as “pink” “blue” “pale” or “pallor” in reference to neonatal skin, without detailing how these skin colour descriptors may appear in ethnic minority neonates. These results further perpetuate the inequalities faced by those from Black, Asian, and ethnic minority backgrounds, by means of improper assessment or potential late diagnosis.

Stakeholder interviews

Aims

1. To explore stakeholders' views of the Apgar score, detection of cyanosis and detection of jaundice with a particular focus on Black, Asian, and minority ethnic neonates.

2. To explore stakeholders' perspectives of relevant clinical guidelines.

3. To explore with stakeholders the wider use of terminology within healthcare and the impact of this on Black and Asian neonates receiving equitable treatment and care.

4. To explore stakeholders' views of ethnic inequalities in neonatal care.

Methods

Methodology

A focussed ethnography approach was used. This used the process of "learning about people by learning from them" (Roper & Shapira, 2000). It included both explicit and covert dimensions of culture (Higginbottom et al., 2013), allowing researchers to immerse themselves into the social world of participants to grasp what they know, believe and what they do (Roper & Shapiro, 2000). Focussed ethnography is recognised as a pragmatic way to gather data on the topic of importance and to determine ways of improving care or care processes (Higginbottom et al., 2013). It is increasingly used within the field of nursing to enhance and understand practice (Cruz & Higginbottom, 2013).

Participants

Eligibility

A range of stakeholders were included within the interviews, including healthcare professionals (HCPs) and parents or carers (parents). HCPs from a wide range of disciplines including midwives, obstetricians, health visitors, neonatologists, paediatricians, or neonatal nurses were eligible for inclusion. Parents or carers were considered eligible if they were caring for a Black, Asian, or ethnic minority child born within the last five years.

Recruitment

A purposive sampling strategy was used during study recruitment. The study was advertised via poster and shared within professional organisations, local networks and via LinkedIn, Facebook, and Twitter. Interested participants contacted the research team to register their interest in being interviewed. Every effort was made to ensure maximum phenomenon variation by recruiting HCPs of different professions, different ethnicities, working in areas of high or low ethnic diversity and with different amounts of experiences since qualification. For the interviews with parents or carers, different ethnicities from different geographical areas were targeted. Participants were screened prior to arranging an interview to ensure demographic diversity within the group. Participation was encouraged from those who did not have English as a first language, with interpreters offered for the interviews if wanted by the parent or carer. Recruitment was undertaken between August 2022 and January 2023. Recruitment continued until data saturation was achieved as no new topics were emerging from the data.

Data collection

Interviews were semi-structured. The interview schedule contained basic demographic questions followed by primarily open-ended questions to elicit in depth responses from participants. For the HCPs, questions gathered information about their experiences of providing care and, in particular, assessing the Apgar score, cyanosis and jaundice in Black, Asian, and minority ethnic neonates. Parents were asked about their experiences of accessing care in the UK for their ethnic minority neonates. A focus was placed on their awareness of how to assess for jaundice or a lack of oxygenation in their newborn. Within all interviews, interviewees views around the challenges faced when providing or accessing care for these infants and the role that they believed that ethnicity or their skin tone did or did not play in the care of neonates was also explored.

The interview schedule was developed in collaboration with a range of stakeholders to confirm comprehensiveness, acceptability and clarity of the

schedule. Following this the interview schedule was piloted amongst two HCPs and two parents, with minor amendments made before finalising the interview schedule.

Given the wide acceptance of digital technology post pandemic, interviews were undertaken via Zoom and audio recorded then transcribed verbatim.

Data analysis

Qualitative data analysis was managed using NVivo software. An inductive approach to analysis was undertaken using the systematic approach described by Roper and Shapira (2000). After familiarisation with the data, the 5 analytical steps were followed: a) open coding of transcripts line by line with descriptive labels, these codes were kept close to the text; b) sorting initial codes according to patterns emerging; c) identification of outliers or disparate cases; d) generalisation to identify constructs and theories and e) memoing reflections of the researchers emerging thinking (Roper & Shapira, 2000). One transcription was coded independently by two researchers and then compared to ensure coding credibility and transferability. Subsequently coding of all transcripts was undertaken independently by two researchers, with a subset of the transcripts also sent to the other team members. The whole team then discussed the patterns emerging, along with the constructs and theories emerging. The ethnic diversity of the research team, as well as including maternity user group representatives, was seen as a strength when undertaking this process.

Within the analytical process it was important for the researcher to eliminate all personal biases and preconceptions, however the extent to which this is truly possible is questioned as all research is socially constructed and interpreted (Cruz & Higginbottom, 2013). The researchers however undertook the process of reflexivity through the practice of memoing throughout all stages of data interpretation to identify their potential influence on the data interpretation. Member checking and participant validation was employed to enhance trustworthiness of the findings, with two stakeholder workshops also undertaken to discuss interim results of the study. The analysis is presented narratively with extensive direct quotations presented to illustrate and confirm the researchers' interpretations.

Ethical considerations

Ethical approvals were obtained from Sheffield Hallam University Research Ethics committee before undertaking the interviews. Prior to the interview participants were emailed a participant information sheet and consent form. Participants signed the consent form and returned it electronically to the research team. Consent was also confirmed verbally at the start of the interview.

Parents were given a £20 gift voucher to compensate them for the time they had given to participate within the interview. A voucher was also sent to HCPs if time compensation was requested.

The confidentiality of the participants was maintained by using codes within the transcripts and with the illustrative quotations provided below.

Results

Healthcare professionals' demographics

A total of 33 HCPs were interviewed. Interviews lasted an average of 46 minutes and ranged from 21 to 96 minutes. HCP interviewee demographics are given in Table 6.1. The majority of HCP participants were female, with just five male participants. Thirteen participants were midwives, eight health visitors, four paediatricians, three obstetricians, three neonatologists and two neonatal nurses. Eleven HCP participants described themselves as Black, of whom six classified themselves as Black Caribbean and four as Black African and one as any other Black background. Three participants stated they were Mixed ethnicity two of whom were Black Caribbean and White and one was any other mixed background. Two HCP participants stated they were Asian, one Indian and one Pakistani. The remaining participants were White, the majority of whom were White British (n=15), but two were any other White background. Five had received their primary qualification in a non-UK country including Ghana (n=2), Greece (n=1), Jamaica (n=1) and South Africa (n=1).

Regarding years in practice, thirteen (39.4%) had practised for ten years or less, eight (24.2%) for 11-20 years and twelve (36.4%) for more than 20 years. HCPs worked in a service with between <1% and 100% of infants being of Black and minority ethnic background. Locations where HCP currently or had previously worked within the UK included Birmingham, Blackpool, Bolton, Bradford, Cardiff, Darlington, Doncaster, Glasgow, Halifax, Hampshire, Leeds, London, Manchester, Middlesex, North Scotland, Nottingham, Sheffield, Swansea, Surry, Walsall, Wakefield, Wessex, West Wales, Wolverhampton and across Yorkshire and Humber. The countries where participants had worked outside of the UK included Ghana, Jamaica, South Africa, St Lucia, and the United Arab Emirates, with one HCP also working across multiple locations in Africa and Asia.

Table 6.1. Healthcare provider interviewees' demographics

Participant	Gender	Race	Years in practice	Proportion of Black, Asian, and minority ethnic infants in their service	Country of primary qualification
MW01	F	Mixed	≤5	25%	UK
MW02	F	Black	11-20	60%	UK
MW03	F	Black	>25	40%	UK
MW04	F	Asian	11-20	32%	UK
MW05	F	Mixed	6-10	90%	UK
MW06	F	Black	≤5	65-70%	UK
MW07	F	Black	21-25	30%	UK
MW08	F	White	6-10	40%	UK
MW09	F	White	≤5	2%	UK
MW10	F	White	11-20	90%	Non-UK
MW11	F	White	6-10	38-40%	UK
MW12	F	White	6-10	25%.	UK
MW13	F	White	>25	50%	UK
NNP01	M	Black	11-20	100%	Non-UK
NNP02	F	Black	≤5	50%	Non-UK
NNP03	F	White	11-20	50%	UK
NNP04	M	Mixed	11-20	50%	UK
NNP05	F	White	≤5	10%	UK
NNP06	F	White	≤5	1-5%	UK
NNP07	M	White	>25	50%	Non-UK
NNP08	F	White	>25	30-40%	UK
NNP09	F	Black	6-10	40%	Non-UK
OB01	M	White	>25	<1% Location 1 20% Location 2	UK
OB02	F	Black	6-10	40% Location 1 3% Location 2	UK
OB03	M	Asian	>25	61%	UK
HV01	F	White	>25	98%	UK
HV02	F	Black	21-25	60%	UK
HV03	F	White	21-25	25%	UK
HV04	F	White	11-20	5%	UK
HV05	F	Black	6-10	40%	UK
HV06	F	White	>25	50%	UK
HV07	F	White	21-25	20%	UK
HV8	F	Black	11-20	65-70%	UK

MW=Midwife; **HV**=Health visitor; **OB**=obstetrician; **NNP**=neonatologist, paediatrician or neonatal nurse.

Parents' demographics

In total, 24 parents were interviewed. Parents interviews lasted on average 30 minutes, ranging from 16 to 57 minutes. Basic demographics are given in Table 6.2; broad categories have been used to avoid identification of participants. Twenty mothers and four fathers were interviewed. As well as being a mother, one woman also had a relative's child placed under her care. One of the mothers and one of the fathers were a couple and chose to be interviewed together. Parents ranged in age from 25 to 41 years. The age of their youngest child ranged from 8 weeks to 5 years. Not all parents gave their geographical location. Those that did were from Birmingham, Bradford, London, Manchester, Sheffield, Swindon, Portsmouth and York, indicating a good geographical spread of respondents within England. When considering race, fifteen participants were Black (Black Africa n=12, Black Caribbean n=1, other Black background n=2), three were Asian including two from an Indian background and one from another Asian background. Three participants were mixed ethnicity, two being Black Caribbean and White British and one African-Asian. Two participants were from other ethnic groups, both being Arab, one being of Iraqi descent and another of Yemini descent. One mother was White with a child of mixed ethnicity.

Table 6.2. Parent interviewees' demographics

Participant code	Number of children	Relation to child	Race	Age	Country of Birth
PA01	1	Mother	Other	25-29	UK
PA02	1	Mother	Asian	≥35	Non-UK
PA03	2	Father	Mixed	30-34	Non-UK
PA04	3	Mother	Black	30-34	UK
PA05	1	Father	Black	25-29	Non-UK
PA06	2	Mother	Black	30-34	Non-UK
PA07	2	Mother	Black	30-34	Non-UK
PA08	2	Mother	Black	25-29	Non-UK
PA09	2	Mother	Black	25-29	Non-UK
PA10	1	Mother	Black	30-34	Non-UK
PA11	2	Father	Black	30-34	Non-UK
PA12	1	Mother	Black	25-29	Non-UK
PA13	2	Mother	Asian	≥35	UK
PA14	1	Mother	White	30-34	UK
PA15	1	Mother	Black	25-29	UK

PA16	2	Mother	Asian	≥35	UK
PA17	3	Mother	Black	≥35	Non-UK
PA18	1	Mother	Black	30-34	Non-UK
PA19	2	Mother	Black	30-34	Non-UK
PA20	1	Mother	Black	≥35	UK
PA21	1	Mother	Other	25-29	UK
PA22	1	Mother	Mixed	30-34	UK
PA23	3	Mother	Black	≥35	UK
PA24	1	Father	Mixed	30-34	UK

Eleven parents had been born in the UK and 13 elsewhere including Ghana (n=4), Kenya (n=1), the Netherlands (n=1), South Africa (n=4), Sri Lanka (n=1). Two participants chose not to give specific countries of birth stating Africa (n=1) and Western Africa (n=1). Parental education level is given in Figure 6.1. The majority of the participants had completed graduate or postgraduate education (83%). All but one parent reported themselves to be employed at the time of the interview, with the other recently having lost their job but providing occupational data (Figure 6.2). Fourteen had managerial or professional occupations including teacher (n=6), researcher or scientist (n=3), university lecturers (n=2), social worker (n=1) child and youth domestic violence advocate (n=1) and an international development officer (n=1). Six had intermediate occupations including three in administration, one self-employed, one restaurant manager and one civil servant. Four with routine or manual occupations included two cleaners, one waitress and one residential care worker.

Figure 6.1. Parental education level (n=24)

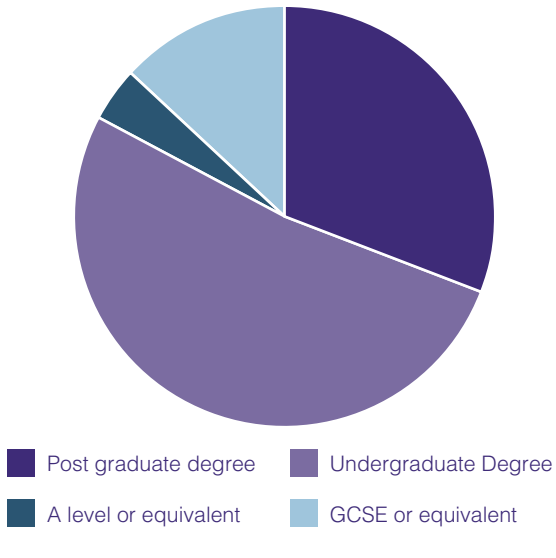
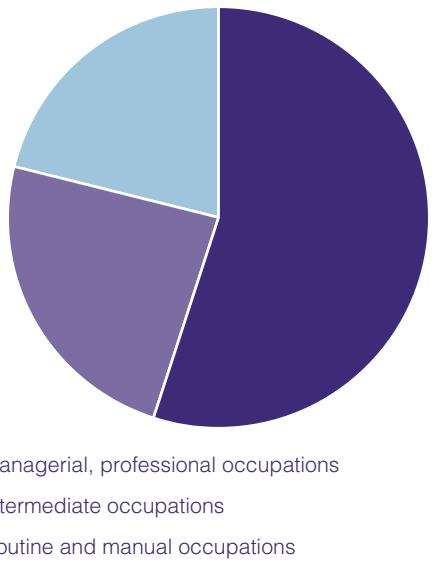


Figure 6.2. Parental employment (n=24)

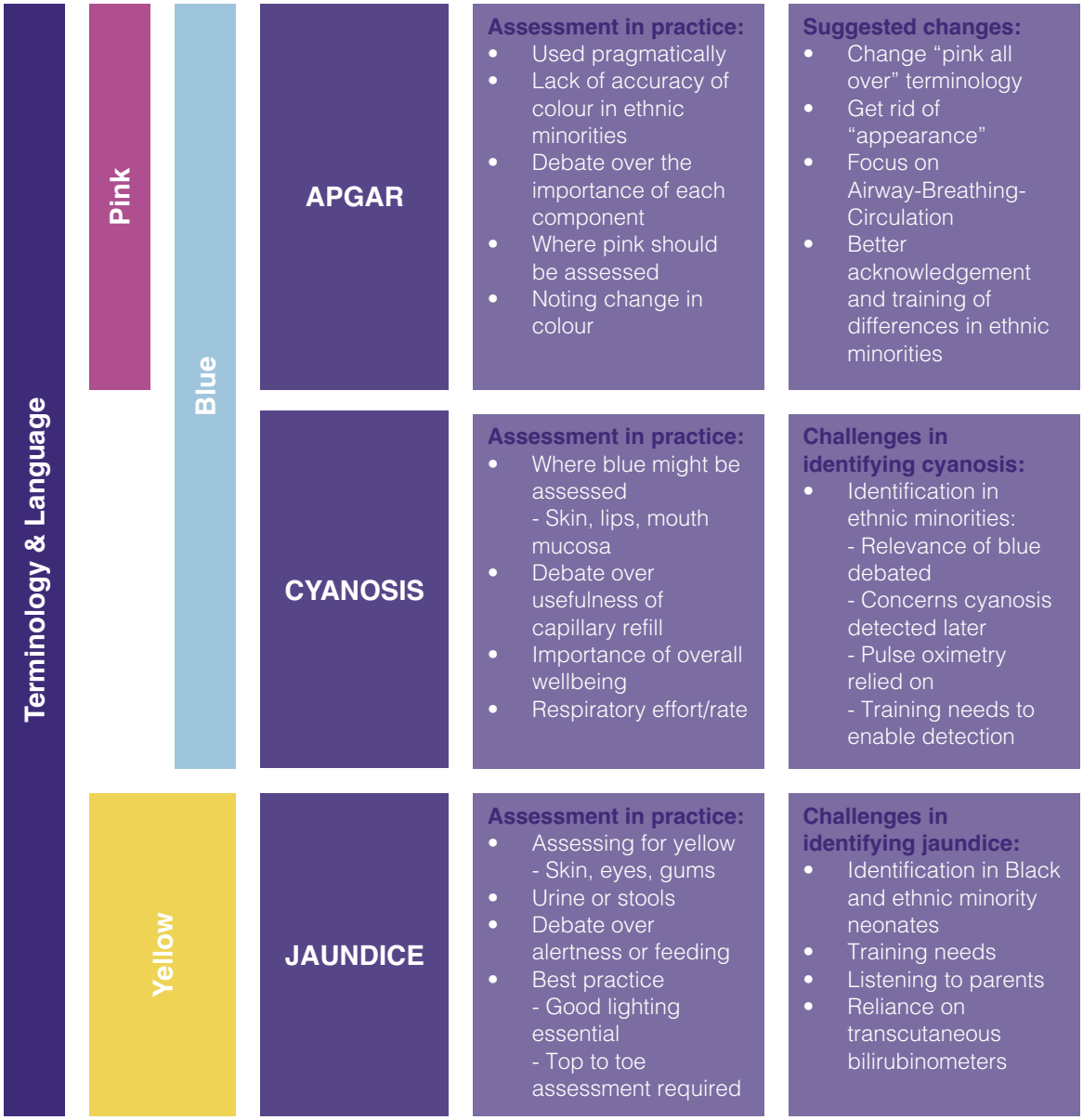


Within the interview analysis direct quotations from parents are labelled as PA, from midwives as MW, from health visitors as HV, from obstetricians as OB, from neonatologist, paediatricians or neonatal nurses as NNP. Participants are referred to as the overall category of Black, Asian, Mixed, Other or White to protect confidentiality given the limited numbers within further subcategories of ethnicity.

Assessment of skin colour

A summary of the findings from this section are presented in Figure 6.3.

Figure 6.3. Summary of interview findings regarding assessment of skin colour in neonates



Parental awareness

At the time of the interview, seven parents were unaware that skin colour can be used to assess the wellbeing of a baby and seventeen said they were aware, although one parent described not really being sure how it is used in practice. One woman's awareness of skin colour assessments was due to her husband being a doctor, however she was unaware that any assessments of her child's skin had been undertaken. Other parents became aware during the pregnancy or after the birth, for example being told in antenatal classes that skin colour is used when giving the baby a score at birth or due to their children having jaundice. One father described how his awareness had developed due to being fearful before having his first child, so had deliberately asked family members and completed his own research to know what to expect. One had only learnt about Apgar scores after birth as she tried to understand the events that surrounded her child's birth.

> *"I only knew through an antenatal course I did. That they look at skin colour when the baby is born to give a score." (PA21 – Other)*

The majority of parents (n=17) had not been told that assessments were more difficult due to the colour of their baby's skin, four parents were told an assessment was more difficult, with other parents being aware things could look different due to their skin tone.

> *"She [the midwife] did tell me that our skin colour would quite be a little bit um, should I say tough for us to care for and I didn't really get what she meant by that." (PA11 – Black)*

Differences in pigmentation at birth

Nine HCPs and three of the parents talked during the interview about differences in the colour of an infant at birth, with suggestions pigmentation developed and changed over the first few days. While most infants were suggested to get darker as they got older, one mother noted her baby had got progressively lighter with time. The skin tone of a White infant was also noted by one HCP to change with time.

> *"But when a Black baby's first born, they're not always, their pigment is not always fully developed then, so you can see clearly." (MW07 – Black)*

> *"She [3rd child] got brown the quickest out of all my other children. The other two I generally thought they were albinos. So, I was like, I know we come out quite light, but I was just like oh wow and they took a very long time to get darker." (PA04 – Black)*

Apgar scores

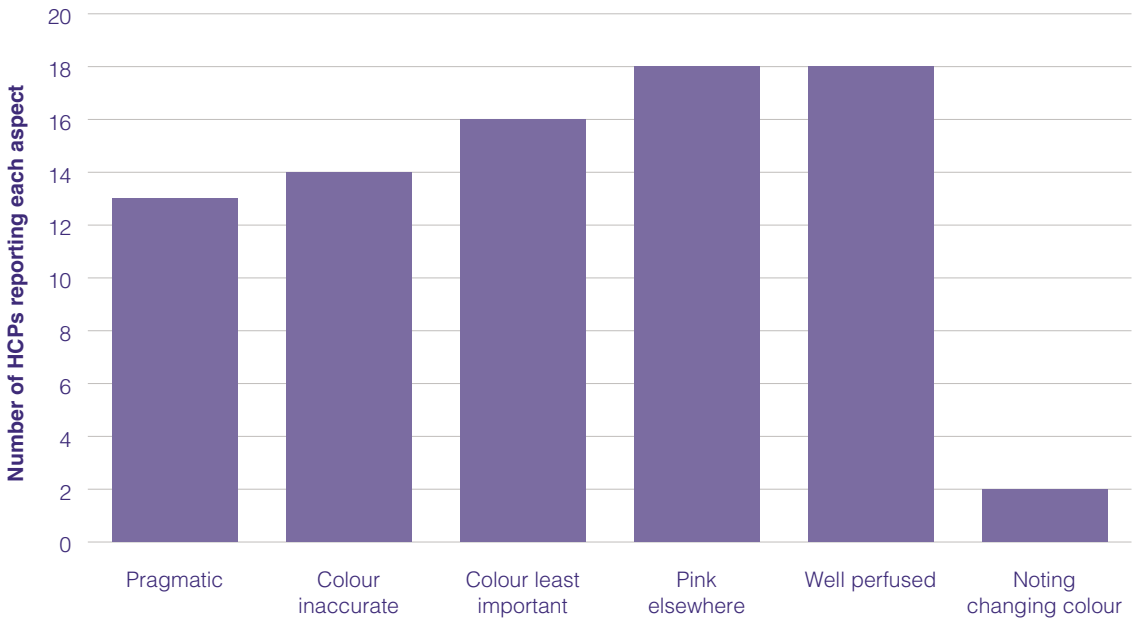
When asked how often they assessed Apgar score in Black, Asian, and minority ethnic neonates over the last two years, four health visitors had never assessed it as they were not midwives prior to becoming a health visitor and the other four health visitors had not assessed an Apgar score recently as they were no longer practicing midwives. Four other HCPs stated it was not part of their role, with one of the obstetricians saying the Apgar score was undertaken by midwives and paediatricians, however, they liked to do a quick assessment for comparison. One obstetrician with 35 years of experience was now retired and five midwives were in managerial or other roles which meant that they had no experience (n=3) or limited experience in the last two years (n=2) and another midwife could "count on one hand" the number of times she had assessed the Apgar score in ethnic minority neonates in the last 2 years due to very limited exposure in her geographical location. Two other midwives didn't directly assess Apgar scores but were still actively involved in other neonatal assessments. Of the other HCPs, four assessed Apgar scores in Black and minority ethnic neonates approximately monthly and eight on a weekly basis.

> *"I do it quickly, then you hand them over to the paeds and then I say, 'oh can I just check the Apgar scores if it matches with what I've got?" (OB02 – Black)*

Assessment in practice

Within the HCP interviews, each interviewee described their own way of currently assessing the Apgar score, with limited consensus among HCPs on how this was achieved. The main categories are detailed in Figure 6.4 and described below. Many HCPs stated they used the Apgar score pragmatically, without detailed reference to the actual criteria. The term pink in the appearance component of the Apgar was felt to lack relevance by most HCPs, with most not receiving additional training on how to assess this in Black, Asian, and ethnic minority neonates. This led the HCPs to find alternative ways to assess this component, with not all described methods being evidence based. HCPs frequently described using multiple methods to assess the Apgar score. Several HCPs noted the importance of good lighting when undertaking any assessment.

Figure 6.4. Number of HCPs describing each aspect when assessing an Apgar score



Used pragmatically

Numerous HCPs described an Apgar score as something done in retrospect for the notes and some described a "pragmatic" use of the Apgar score, in the sense that practitioners used their professional knowledge and overall views rather than the official Apgar guidance. Once qualified for a few years, HCPs described "instinctively" or "automatically" recognising abnormalities, making the Apgar a tool of secondary importance. Well babies were described as automatically being given a score of 9 without the HCP physically assessing each aspect, with the individual components not given much consideration if the baby was breathing spontaneously. The main time the full Apgar score was seen as important was within the educational environment and when a baby required resuscitation at birth with the Apgar score seen as a way to assess the effectiveness of resuscitation attempts.

> *"To be honest, we don't really do [it], I mean we, if the baby's fine, you just kind of say it's 9 and if they're in trouble, then you kind of sort them out and then, in retrospect, you assign an Apgar score … It's one of those things you have to do for the notes you know … you don't go through all of this if they don't need resuscitating." (NNP08 – White)*

> *"We don't look at the colour as such, but we do know what an abnormal colour looks like." (MW13 – White)*

> *"It's when they come out and you know those ones that they're not crying immediately. Then everybody stops and really counts the Apgar properly." (OB02 – Black)*

This pragmatic use was noted by some to make the Apgar score a crude assessment, with several noting that it was not really a clinical assessment but based largely on guesswork. Effective auscultation of the heart rate in particular was felt to be poor. The Apgar score was seen by two HCPs to simply be a concept that could be used within research to make comparisons between groups. Other variations in practice were also noted within the interviews with one midwife stating that she would look at parental complexion to assist them when assessing the neonate's colour, but another midwife felt it would be inaccurate due to the variations noted within mixed ethnicity families.

> *"I think an argument that the Apgar score is not the best method of assessing well-being, you know. And sometimes it's quite, it's quite controversial in two or one. Well, the difference between you know, one and none or one and two in the difference with Apgar scores." (NNP03 – White)*

Several HCPs (n=3) recognised that their practice differed from others with some having lower thresholds for action than others. Other HCPs were uncertain over how others would undertake the Apgar assessment, with one HCP in particular feeling that the Apgar score would be assessed very differently in a country with a majority Black population compared to majority White population.

> *"I don't know what midwives do, whether they look at the soles of the feet or the palms of the hands, or what they do, whether they just think this baby's fine, they'll just tick the number 2 on the appearance." (NNP08 – White)*

> *"I would probably argue that if you assess Apgar scoring and the way persons assess it in countries that are probably predominantly of African descent, it's going to be a completely different. Well, this is what I'm assuming I haven't assessed the actual data myself, but I would think that it's going to be completely different compared to, if you assess, if it's assessed in a country where Caucasian is the majority of patients." (NNP09 – Black)*

Accuracy of the colour component in ethnic minority neonates

The descriptors pink, pale, and blue are all used within the Apgar score to assess the appearance component. The descriptors pink and pale are discussed below, with the term blue discussed within the separate section around cyanosis.

Relevance of the term pale within the Apgar score

Most HCPs (n=12) felt that the term pale was applicable to infants of all skin tones. One Black Caribbean HCP recognised that when cold, she would go pale therefore it is an acceptable descriptor. However, if an infant was pale it was considered to be in very poor condition and 'kind of lifeless' (MW10 – White). Although not directly asked about the term pale, several parents (n=3) also

described being aware that a baby could look pale or described instances when their baby had been pale.

> *"People tend not to understand that Black babies can still be pale." (MW02 – Black)*

> *"I think even with babies of colour they can look paler if they need resuscitation and they can kind of have a little bit more of a blue, like a cyanoid and appearance." (MW09 – White)*

> *"I do think that you can tell in Black and Asian minority babies if they are pale as such because the skin changes in all of us." (MW08 – White)*

Additional HCPs (n=5) felt it would be very difficult for them or their colleagues to assess for pallor in a darker pigmented neonate, or that it was easier to see in some babies than others. Assessing a neonate for the first time was viewed as particularly difficult by one HCP as they then had no comparison.

> *"It would be very difficult to ascertain in a baby that was dark skinned. How would you clarify if that baby was pale or blue and that would be really difficult to do." (HV04 – White)*

> *"I guess using the word blue or pale we may still see that to some extent in a BAME baby." (MW11 – White)*

Relevance of the descriptor pink within the Apgar score

The term pink was felt to be relevant by two HCPs and eight parents, with three parents considering their infant as healthy when their skin was red or pink. One HCP felt all well perfused babies could be described as 'pink', with well perfused assessed using other areas such as mucous membranes in addition to skin assessment. Another HCP felt 'pink all over' could be relevant to specific groups such as very preterm Black babies, who could look quite pink due to the thinness of their skin and that even in older babies with darker skin there could still be a pink undertone when healthy.

> *"To decide on the Apgar score of this baby, you can see whether a baby's colour is, whether baby's skin is not well oxygenated. But they do talk about whether baby is centrally pink, centrally blue and you can see that whether the baby is darker skinned or not." (NNP07 – White)*

> *"I suppose that under the skin there's a kind of pink undertone. Yeah, and that would kind of tell me that the baby is like well." (HV05 – Black)*

The vast majority of HCPs (n=23) and parents (n=15) however felt that the term pink was not relevant to ethnic minority neonates, with one parent noting that she would just see the colour Black. Additionally, one HCP felt that it would be hard

to detect pink in some neonates and another HCP felt that 'pink' was simply a proxy for not blue or pale. The term pink was deemed by at least five HCPs and one parent to be irrelevant to neonates of all skin tones including White, except for preterm neonates. If any baby looked pink all over, one HCP stated that they would be concerned that it was plethoric and at a higher risk of jaundice. Another HCP noted that if a baby was pink that they would be worried it had a fever, with two parents also commenting that if their child was pink, they would think it was ill or take in to see the doctor. Some HCPs were concerned whether the current terminology of 'pink all over' meant that Black, Asian, and minority ethnic neonates had poorer outcomes. Others suggested the terminology made ethnic minority neonates more vulnerable to an unnecessary cascade of events if the tick box of 'pink' couldn't be checked.

> *"Pink to me is a bit of an oxymoron, really. It's like, you don't get pink Black or Asian babies, do you?" (MW02 – Black)*
>
> *"Certainly, pink is never going to happen in a Black African baby." (MW13 – White)*
>
> *"You probably would find the average Black staff or brown staff won't call a baby pink, if they're Black. They won't do it because we don't think they're pink." (MW07 – Black)*
>
> *"We do say pink, but we just mean not blue. When we're saying the baby is pink and well perfused, we say it frequently in babies that are not Caucasian, what we're saying is the baby is not cyanosed and not anaemic. Basically, it's what we're trying to say. Yeah, it's probably not actually accurate because the baby is not pink." (NNP04 – Mixed)*
>
> *When asked how they might assess 'pink' in a Black, Asian or minority ethnic baby one midwife replied: "I don't know, that is my honest answer." (MW11 – White)*
>
> *When asked whether pink was relevant to their baby one mother responded: "No!! No no, no not remotely." (PA23 – Black)*
>
> *"I think the pink was probably like the first couple of days and then she started to get brown." (PA04 – Black)*
>
> *"Even a White baby doesn't come out glowing pink and so I don't think that's accurate anyway." (MW09 – White)*
>
> *"I mean is pink even a word that you describe a White person these days?" (PA14 – White)*

The terminology used within the Apgar score was noted by several HCPs (n=4) and one parent to imply that it had been developed for White infants and was

therefore inherently biased. Another HCP was unsure whether the Apgar score had considered Black or Asian babies during its development and wanted further clarification of this to understand it's continued relevance. One parent was also aware that assessments were based on European skin tones, which could disadvantage those with darker skin tones.

> *"I'm not a neonatologist, but in my opinion the thing what you're sharing [Apgar score] is written for White babies." (OB03 – Asian)*

> *"I kind of knew that there can be a difference in terms of how skin tone views and how because the majority of the assessments are based on the European skin tone that can be detrimental to darker skin tones." (PA23 – Black)*

When shown the Apgar score on the screen several HCPs were surprised at the terminology:

> *"Oh yeah appearance, colour, look at that … Haaa, pink goodness gracious that's terrible isn't it!" (HV03 – White)*

> *"I'd forgotten that said completely pink … I mean it's a bit odd, isn't it?" (NNP08 – White)*

Inaccuracy of assessing colour within the Apgar score in ethnic minority neonates

Some HCPs (n=4) had not considered, prior to interview, how colour would be assessed in a Black, Asian, and minority ethnic neonate. Many other HCPs, whether they felt 'pink' was a relevant term or not, noted difficulty or uncertainty in assessing colour or scoring the appearance component of the Apgar score in relation to infants with darker skin pigmentation (n=12). Difficulty in accurately assessing colour was especially noted for staff who were not from an ethnic minority background themselves. Several parents also acknowledged that 'pink' would not be relevant to all babies as it could be hard to detect in some ethnicities but were unsure how else it could be described. HCPs therefore felt extra vigilance was required in a neonate with darker skin pigmentation. One HCP felt exposure was necessary to become confident and competent in assessing colour in ethnic minority neonates, with several recognising that additional training may be required for some HCP to feel confident in using the Apgar (n=2). The colour component was described by some to be guessed or assumed to be satisfactory for Black and minority ethnic neonates (n=3). However, another HCP felt that whether an infant scored 1 or 2 with regards to being 'pink all over' was immaterial.

> *"It's [pink] going to be less obvious in a baby of ethnicity." (HV08 – Black)*

> *"Uhm, my honest answer is I don't really know and that makes me think about what potentially I've done in the past when I've been looking after*

these babies. I imagine if I've had no concerns about the baby, I've probably scored a 2, but thinking back that probably wasn't correct because I wasn't able to say that baby was pink." (MW11 – White)

"I think they coded newborn Black babies and Asian bases incorrectly up by giving them a higher Apgar score when I kind of have argued that it was a lesser score ... so they've given an Apgar score of 9 or 10 and I'd be more saying, well, I think it's more 7 or 8 because of the colour." (MW02 – Black)

"You get good at things because you see it over and over, so it's hard. But I think we just have to have different pathways reason to be aware that the Apgar score is not as accurate in persons who are not trained in, you know assessing darker skin baby." (NNP09 – Black)

"We've got no system, direct system, in place for Black and Asian babies. It's just a universal thing, that we assess, you know, … we check the colour as best we can in whatever situation that we're in. Yeah, yeah. But as I said, sometimes it's not always accurate." (MW02 – Black)

One mother also reported feeling that her daughter's one minute Apgar score had been inaccurate. She had been given a score of 4 despite crying and feeding straight away, which the mother felt was based on her daughter's darker skin tone. Other colours parents used to describe their baby immediately after birth included muddy and a blue, purply shade.

"She was fine, she came out and screamed and peed all over me and then started feeding within seconds. But I think because again, there's the slight change in tone, isn't there. So, she came out almost looking a little bit blue, but I think that's just because her skin tones a bit darker. So, the pinkish hue was perhaps changed slightly." (PA23- Black)

Relative importance of individual Apgar score components

Assessment of the appearance component of the Apgar score was seen as subjective (n=3) with only the heart rate seen as an objective measure of neonatal wellbeing (n=2). The appearance component was therefore cited by two HCPs to be the least important aspect of the Apgar score. One HCP described the appearance as just one piece of the assessment puzzle to determine neonatal wellbeing, with several HCPs describing how they would undertake a comprehensive assessment instead (n=2). HCPs considered other components of the Apgar score to be more important than colour including all of the other components (n=7), the neonate's tone or alertness (n=5), their oxygenation and breathing (n=4), their heart rate and breathing or oxygenation (n=2) or whether the neonate was warm, cold to touch or clammy (n=2). The essential aspects of the Apgar score were seen by one HCP to be the same as doing an ABC (airway, breathing, circulation) assessment.

"So, its colour, tone, heart rate, respiration and response to stimulus. Those are the five pieces in the Apgar. So, a baby who's not oxygenating well, won't have good tone, won't be breathing well and so it's [colour] just one part of that assessment puzzle" (NNP07 – White)

"It's not always accurate. And sometimes, unfortunately, we compensate with the other, with the other measurements and with the heart rate, with the breathing, with the tone, with the reflexing." (MW02 – Black)

"If my baby still got slow heart rate and not breathing, I don't really care what the tone is doing." (NNP03 – White)

If "the Black baby was born and it is floppy then the chances are it's also cyanosed as well." (MW03 – Black)

"You know how to sort of rate it based on the whole picture maybe, so all that baby is crying and alert and sort of showing normal reflexes. So, you'd give it a, you might give it a 10." (MW12 – White)

"I would be looking at the pulse. I'd be looking at their respiratory rate, their tone or attitude in this one, they're grimace and their heart rate as well. If that was indicated at the time, so that's how I performed that assessment. I would notice if the baby was very pale, or if the baby was blue or had like a cyanoid appearance to it. But yes, I would be looking at the other parameters of my assessment." (MW09 – White)

Where pink assessed

Many HCP recognised that skin colour alone was not always enough to assess neonatal wellbeing particularly "pink all over" and they preferred to look at other areas. Different areas of the body assessed by HCPs for pink or pallor included the lips (n=8), mucous membranes (n=7), inside the mouth or the gums (n=4), the tongue (n=3) and the conjunctiva or lower eyelids (n=3). Two parents also said that they would assess for the colour 'pink' in their child's lips. The mucous membranes were noted to be something that could be looked at quickly, although it was recognised that not all HCP have been trained to assess mucosa colour. Dry or dull skin was also considered to indicate lack of wellbeing (n=1).

"I think when we talk about being pink, it's not just the skin it's looking other uhm, sort of mucous membranes. And that's where I look for signs of cyanosis because of this issue, but it's not just the skin that you can see signs of cyanosis" (NNP04 – Mixed)

When thinking about areas where someone may be pink, palms, soles, fingers, toes, or nail beds were all discussed (HCPs n=11, parents n=2), although the limitations of the usefulness of these peripheral areas was recognised in the first

few days after birth (n=8). Other locations parents suggested they may be able to see the colour pink on their child included cheeks (n=2), eyelids (n=1), the chest (n=1), the stomach (n=4), legs (n=3) and the whole body (n=1).

> *"The peripheries you can use it in Black babies because they, you can actually pick that [up]. And again is very difficult because the peripheries when you don't have a good circulation can be a bit of a problem." (NNP01 – Black)*

> *"For me it's pink cheeks like I know when he's got pink cheeks, he's like well." (PA21 – Other)*

Well perfused

Many HCPs talked about looking at whether an infant was well perfused to assess wellbeing (n=17). Another HCP also equated well perfused with an infant being 'pink'. When asked further about the term well perfused several were uncertain how to assess if an infant was well perfused. Other HCPs however said they would assess if a neonate was well perfused through capillary refill time, with four other HCPs considering capillary refill was something that could be done in an older infant, rather than to assess the Apgar score.

> *"I guess we'd just convert being completely pink to being well perfused and normal looking colour." (NNP08 – White)*

> *"In terms of telling how well a baby was perfused, I'm not sure that I would feel comfortable knowing how to do that." (HV04 – White)*

Noting colour change

One midwife felt that colour changes occur in all humans, the issue is whether HCPs are trained how to identify it in different ethnicities. Another midwife also felt the change in colour within the first few minutes as a baby's source of oxygen changed should be evident to any HCP assessing properly, especially central colour change.

> *"The thing is when the baby is first born, you know straight away that that baby hasn't got good circulation and it's within a minute that that starts to change. So, if you're looking properly at your baby, you will know that your baby has changed in that minute, which you should really be looking at." (MW04 – Asian)*

Suggested changes to the Apgar score

Only two HCPs felt that no changes were required to the way appearance is recorded in the Apgar score. However, for one of these HCPs it was noted that they didn't describe assessing the appearance component of the Apgar score in accordance with the textbook. A further HCP did not feel qualified to say what might be more appropriate.

> *"I think that one is looking at the colour of oxygenated blood. You aren't looking at skin colour. Yeah, well you are looking at skin colour and mucous membranes and fingernails and palms. But I don't think that the use of the word pink urm is discriminatory." (NNP07 – White)*

Change only acceptable if objective

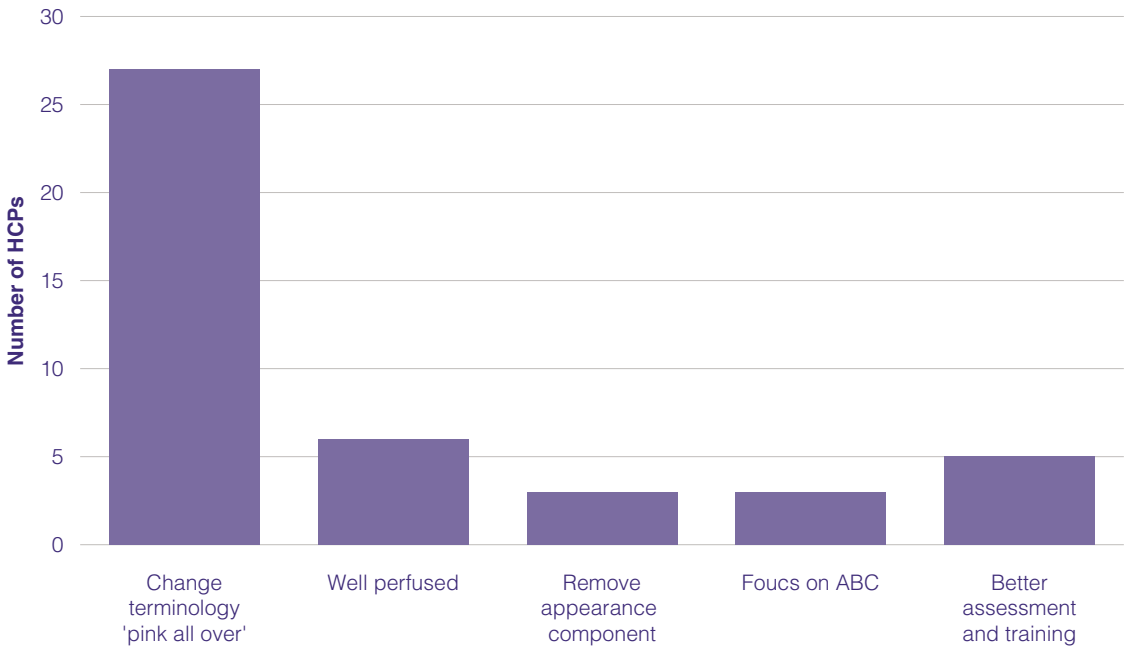
Several HCPs noted that changing the Apgar could only be done if a purely objective and accurate assessment could be found.

> *"It's going to be difficult to find something that's going to be 100% objective so it's almost like kind of the best thing that we have maybe we just have to find ways to modify it?" (OB02 – Black)*

> *"I think it will be confusing and I think it will be catering to an unnecessary need … I'm not sure what we can change it to that is going to be accurate. Because it must be accurate." (NNP07 – White)*

However other HCPs described several potential ideas, which are discussed below and presented in Figure 6.5. However, it should be noted that suggestions were anecdotal and not evidence based.

Figure 6.5. Number of HCPs proposing each of the suggested changes to Apgar score



Changes to the terminology of 'pink all over'

Most HCPs (n=23) felt that changes were necessary to the terminology within the Apgar score, to move away from the term 'pink' to better account for the multi-cultural society in which we live and to consider that not all babies are White. Using the term 'pink' was felt by one HCP to just be habitual, whereas others raised concerns that the term 'pink' was racist (n=1) and could offend families (n=2). It was however felt by some that the terms blue and pale should remain.

> *"I feel the Apgar score is kind of, … it's archaic now. I think we need to move on to something a bit more modern and up to date. Say you know something that recognises a multicultural society that we currently live in because I feel as I've kind of stated before that the Apgar score was brought in for White European babies where you know, we're far from that now. We're a bit of a, you know, multicultural society now." (MW02 – Black)*

> *"I think we need to design it so that it takes into account that not all babies are necessarily going to be White … At the moment we aren't, we just aren't trained in that way, so you're not even aware of the alternative ways that you can assess [Black, Asian, and minority ethnic neonates]." (HV04 – White)*

> *"I think when we start from the beginning, you know take away the word pink and think of something else to put there more relevant to generic … I think probably pale or blue should be there perhaps, but the pink I don't know." (MW07 – Black)*

For several HCPs they had not really considered the terminology within the Apgar score or how they would apply it to an ethnic minority neonate prior to the interview. When asked if they thought any changes were required to the Apgar score, they replied:

> *"I do now that I've met you! Because I've never thought about it before, which is awful. But I've never really paid attention to the fact that it says pink in the assessment and like and that makes it really difficult to assess, you know a coloured baby." (MW11 – White)*

> *"I've never really thought about it and that's sort of, you know, naivete on my part." (MW12 – White)*

When asked what terms 'pink' could be replaced with, terms such as good or acceptable colour (n=1) and 'normal' colour were suggested (n=4). Another HCP suggested that signs of cyanosis or not was all that was required. The term skin colour was also recommended to be changed to skin tone by one HCP.

> *"We quite often will document 'baby is pink alert with good tone' and that's our general kind of documentation about all babies if we're saying that they're healthy … but personally I changed that to saying a baby is normal in colour and that with good tone." (MW09 – White)*

> *"Maybe even if they just said good colour other than pink, do you know what I'm saying or uhm, you know something like along the lines of colour acceptable." (MW02 – Black)*

> *"When we're looking at pink and alert, you know we need to change that. We need to change it to skin tone." (MW08 – White)*

Two midwives talked about previously trying to raise concerns about the colour 'pink' being used to describe skin when not all people have pink skin. They reported either being met by aggression or had been laughed off by other HCPs. They therefore noted the difficulties involved in changing engrained practice.

> *"It's someone in authority that teaches you that way of learning, you accept it as normal. And you might actually feel like it's not correct, but it's difficult to challenge." (MW04 – Asian)*

Several HCPs discussed simply reclarifying where to look for pink, especially focusing on the mucous membranes (n=3) as they were seen as something easy to look at. Other areas that could be assessed included the bottom eyelid (n=1) or the lips (n=3), although one HCP noted that assessment of the lips alone would be an inadequate assessment.

"So, I think the "pink and well perfused" statements should be scrapped completely and we change that to, you know, assessing the mucous membranes." (NNP09 – Black)

Well perfused

While eight HCPs wanted to see a move away from pink towards 'perfused' or 'well perfused', two HCPs, one of whom had trained and worked in a majority Black country, queried the applicability of the term to Black, Asian, and minority ethnic neonates. It was questioned whether it was just another way of saying pink and they expressed concerns that it was not actually possible to accurately assess in someone with darker pigmentation.

"The term Black, pink and well perfused is pretty much for Caucasian persons and it's not something that I learned practising in medicine in the Caribbean … We have to comment on this differently because you're missing a big subset of persons in that statement." (NNP09 – Black)

Get rid of the appearance component

Given questions over its accuracy and relevance, scrapping the appearance component of the Apgar completely was suggested by three HCPs. However, one did then go on to voice concerns that getting rid of that component may make it more likely to miss a neonate being cyanosed.

"In my world, I don't think that appearance means an awful lot, so I I'd quite like to get rid of that." (MW13 – White)

Others suggested moving to a simple dichotomous score, for example having a score that stated whether the infant was in a good or poor condition at birth or whether they needed resuscitation or not. Another suggested the interventions a neonate required during resuscitation would accurately depict their condition at birth.

Focus on breathing and circulation

One HCP felt that the Apgar could be abandoned completely as only a proper assessment of the heart rate is necessary as all other aspects stem from this.

"For years now I've realised that the Apgar score is total rubbish. Because and it all depends on a good circulation." (OB01 – White)

The above HCP and another one wanted to potentially see oxygen saturations monitored more regularly through pulse oximetry. A final HCP felt that the Apgar

was of lesser importance than undertaking a basic ABC (airway, breathing, circulation) assessment.

> *"What you do you, you know, [when] you've got a baby who's comes out who doesn't look good and doesn't immediately cry? Yeah, you dry them. You put, that you do airway neutral position. You assess them, then you give them breathing support. Then you assess them and then you give them circulation support. So actually, what you do is an ABC assessment and maybe that's, that's more important." (NNP03- White)*

Better acknowledgement and training of assessment in those from ethnic minority backgrounds

The current Apgar was felt to be inadequate for a multi-ethnic population and two HCPs suggested a separate scoring system for different ethnicities, yet the need to keep the score standardised across the globe was also recognised. Therefore, the need for further training to ensure accurate identification in those from ethnic minority backgrounds was identified (n=4). This was seen as particularly important as most HCPs described a lack of training in how to effectively undertake an Apgar assessment in a Black, Asian, and minority ethnic neonate.

> *"I don't recall having any training or any sort of additional education for babies within that BAME category." (MW11 – White)*

> *"We have to change up the examination, so I think moving from this pathway that 'All neonates and all persons are examined in the same way'. I think we actually have to reassess that and say, 'Oh maybe we should be doing things a bit differently for persons of darker skin colour" (NNP09 – Black)*

> *"I do think that you can assess. You just need to be taught properly to assess it." (MW08 – White)*

Cyanosis

When asked how often they had assessed neonates from a Black, Asian, and minority ethnic background for cyanosis within the last two years, twelve HCPs had either never assessed it, or it was not part of their current remit. Two HCPs did not report how often they had assessed for cyanosis in the last two years. Five HCPs had assessed for cyanosis in ethnic minority neonates a few times in the last two years and the remaining fourteen had regularly assessed for cyanosis in ethnic minority neonates at least once a month over the last two years.

The majority of parents (n=18) who were interviewed had never had any concerns over whether their infant was getting enough oxygen, with some continuing to describe their baby as 'breathing fine' (n=2) or 'everything's normal' (n=1). Parents who did have concerns about their neonate getting adequate oxygen generally described a specific episode where they had been concerned. These included their child being on oxygen in NNU (n=1), noticing multiple oxygen desaturations when their child was in NNU due to their prematurity (n=1), their child stopping breathing during a feed on NNU (n=1) and due to fast breathing (n=2) which was more noticeable for one parent due to the baby being mucusy post birth and due to what was described as 'bad' breathing when their baby caught covid at 2 weeks old (n=1). One parent didn't have a specific episode which had caused concern, but they were very conscious about their child's oxygen levels due to a family history of babies dying.

> *"So in in hospital I do remember feeling like his breathing was quite quick and I did call the help button, one of the midwives or healthcare assistance came to check … he was quite mucusy anyway, so it was quite loud so I could hear his breath, so it's kind of like brought to my attention a bit more." (PA21 – Other)*

For those that had concerns, one mother recalled being seen straight away by doctors even when their baby had tested positive for COVID and being reassured that their infant's chest sounded clear. She was then shown videos by the doctor about sternal recession and told to call an ambulance if that happened, however, she would also have liked to be given a leaflet of what action to take for different concerns. Another concern she raised was the doctor assessed her baby's oxygen levels using an adult sized pulse oximeter. Another mother described being shown what to do if her infant had a desaturation while feeding and had been sent on a first aid course prior to her child being discharged from the neonatal unit (NNU).

Detection of cyanosis

Fourteen parents stated that they did not know how they would tell if their child was not getting enough oxygen, although three of them went on to describe at least one thing they may look at. Several mothers commented that no one specifically told them anything when they had their baby, one assumed that their husband who was a doctor would pick up if there were any concerns with two others assuming that other doctors would notice. Figure 6.6 shows how parents and HCPs would assess for adequate oxygenation, either on its own or as part of the Apgar score. These revolved around assessment of colour on the skin, the lips and the mouth mucosa, or looking at capillary refill, overall wellbeing or respiration. As noted above this is what parents and HCPs described doing, rather than necessarily evidence-based ways to assess for cyanosis. Each aspect is described in detail below.

Assessing for the colour blue

Skin colour

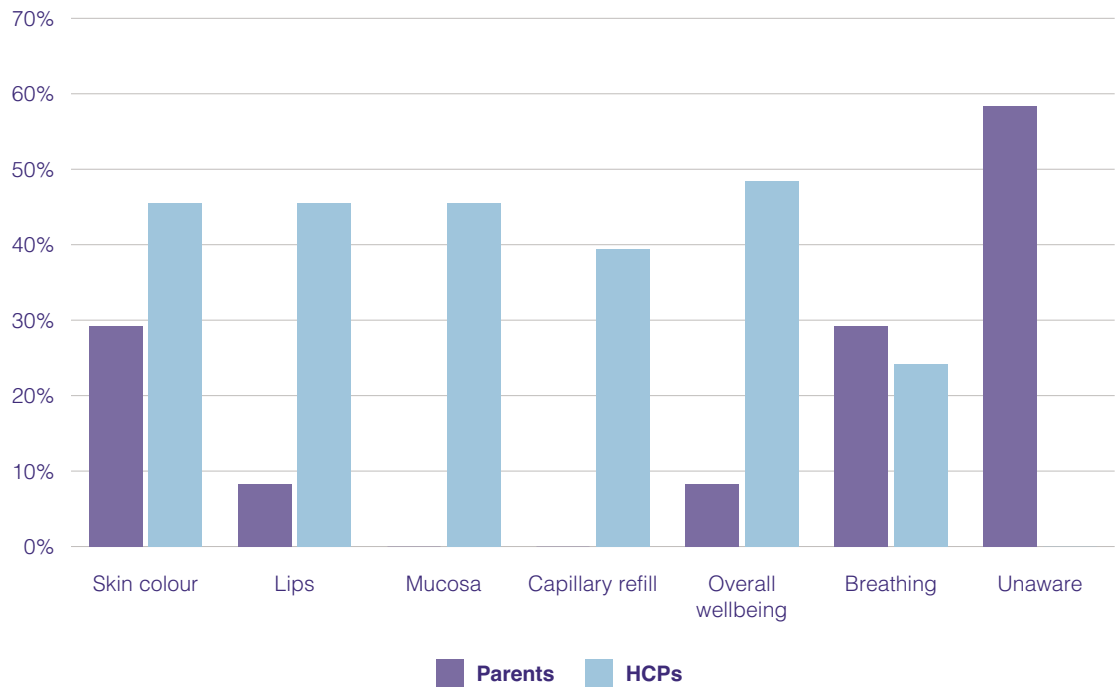
It was frequently reported that the skin colour blue would be looked for during an assessment. One midwife stated that she would look at skin colour on the abdomen to detect cyanosis. Other HCPs would look in very specific places such as the hands and/or feet or nailbeds or fingertips for colour (n=10), although immediately after birth extremities were expected to be blue and cold (n=2). Around the neonate's nose and mouth would also be assessed by four HCPs, although one midwife noted that cyanosis would not always be apparent there. A neonate that looked translucent (n=1) or pale (n=5) was also seen as signs of concern. HCPs however agreed that you can't rely on skin colour for assessment (n=3), that skin colour doesn't really form part of the assessment (n=1), or that skin colour would be low on the list of priorities in an unwell infant (n=4). However, this was not known universally amongst HCPs, with one midwife describing conversations with other HCPs:

> *"I can think of some discussions that I've had with neonatal registrars and when I've been training with the neonatal teams and the fact is that they haven't even realized how limiting using colour as a sign of deterioration is, until we've had that conversation. And it just goes to show how deep rooted the kind of taught mechanism is and how long it's going to take to unpick it." (MW04 – Asian)*

Many parents reported that they would look at skin colour. Some would look for blue generally on their baby's skin (n=5), others would look specifically at the face (n=12). One thought the baby may look green but acknowledged that they didn't really know. One mother also recalled an instance where her son was choking in which his face went red rather than blue. Two mothers noted that if they saw any blueness to their child that they would be worried and call the doctor.

> *When asked what signs they may look for to see if their child was not getting enough oxygen: "I assumed they would go slightly blue, but of course that also might be what happens with White babies and I don't know what happens with mixed or Black babies, so it's a really good question." (PA22 – Mixed)*

Figure 6.6. How interviewees would assess for cyanosis or concerns over oxygenation



Lips

Four parents thought that their baby would have blue lips if they were not getting enough oxygen, with one mother feeling that their child's lips would look blue even if their skin didn't. Fifteen HCPs would also look at the neonate's lips for a blue, pale or aubergine colour, but five suggested that it might be difficult to pick up in Black and minority ethnic neonates and that they might not look blue even when a neonate was severely hypoxic.

Mouth mucosa

None of the parents described that they would look inside their baby's mouth. However numerous HCPs would look at the mucous membranes (n=6), generally inside the mouth (n=5), to see whether the tongue was pale or blue (n=7), or at the gums (n=3). One HCP however acknowledged that this may still appear differently according to ethnicity.

Capillary refill

Additionally, capillary refill to assess perfusion was commonly cited as a method HCPs would use to assess adequate oxygenation by assessing perfusion or good circulation (n=13).

> *"Even dark skinned babies are showing that this skin is well oxygenated because it, because when perfusion isn't good, you can see that … well perfused skin, it transcends colour." (NNP07 – White)*

Overall wellbeing

Sixteen HCPs and two parents would think about the overall condition of the neonate, with this considered to be more important than skin colour by one midwife. Areas that would be considered included whether the neonate was responsive and alert or lethargic (nine HCPs and one mother) and whether it was of good tone or floppy (seven HCPs and one mother). Five HCPs also said they would consider whether the neonate was feeding adequately and one whether it was content. Dehydration, lacking skin tone (three HCPs) or dull skin (one HCP) were also considered to be signs of cyanosis, although emollients used especially in Black neonates were noted to affect this (one HCP). To assess overall wellbeing stripping the baby off fully and considering it as an individual were seen to be essential (one HCP).

Respiratory effort/ rate

Breathing difficulties were the second most commonly cited sign that parents would look for (n=5) with two parents also thinking that their baby would cry if they weren't getting enough oxygen. Breathing difficulties were felt to be easier to observe than skin colour changes by one parent. Respiratory effort was also considered by eight HCPs including assessing the rate of breathing for tachypnoea (n=6), for sternal recession (n=2) and grunting (n=1).

Other signs

Two parents incorrectly thought that their baby may have a rash if they weren't getting enough oxygen. Other things assessed by HCPs included listening to the heart rate (n=1) or assessing for bradycardia (n=1), looking at bottom eyelid colour (n=2), or looking for clubbing (n=1). In addition, two HCPs noted the importance of listening to parents' concerns.

> *"More importantly speaking to the parents, people tend to forget and I always say this to my mum's if you feel that your baby is unwell and you know, you're kind of a perceived to be a paranoid mum. And I said, you sing it from the rooftops … they know when their babies unwell." (MW02 – Black)*

Challenges in identifying cyanosis

Identification in ethnic minority neonates

Relevance of the term blue

HCPs and parents were very split over the relevance of 'blue' to Black, Asian, and minority ethnic neonates. Blue was felt to be relevant and important by eight HCPs and twelve parents felt blue was pertinent to their own baby. One HCP

described that blue colour with cyanosis was visible on darker skin in the same way the discoloration of a blue grey slate nevi was visible. For two HCPs, the relevance of blue was specific to the first few days after birth as the pigment was not yet fully developed making it easier to assess blue or purple.

> *"You still have to look for blue, you know … And that is, that's very important because a Black baby can be blue." (MW07 – Black)*

> *"You can certainly see if they are looking pale or if they're looking blue because they look darker and they look bluer. There's maybe less contrast, uhm but you can certainly assess whether they are bluer." (NNP07 – White)*

Most HCPs (n=15) and ten parents voiced concerns that noticing blue could be more difficult in Black, Asian, and minority ethnic neonates, acknowledging that the blue colour may have to be much deeper before it would be picked up. One parent felt it would be very difficult to assess for any skin colour changes in their infant. Additionally, two HCPs felt that cyanosis would be easy to miss in most people, even those with light coloured skin.

> *"Black babies don't look blue." (MW03 – Black)*

> *"What I'm seeing cyanosis looks like in a White skin tone you're going to miss it in a Black, Asian, you know Asian skin tone." (NNP03 – White)*

> *"I think cyanosis this is one of the things where unless you have lots of experience, it can be very difficult to spot." (NNP01 – Black)*

> *"As a Black baby it is very, very difficult for you to be able to detect if something is wrong with the skin colour because the skin colour does not change." (PA17 – Black)*

In contrast, one HCP didn't feel that detecting cyanosis was an issue in Black and minority ethnic neonates so described not feeling a need to teach it to junior colleagues and that no cases of cyanosis were missed as a result.

> *"I can't remember any case where a baby has been missed because they're cyanosed because they are a baby of colour and I don't find myself teaching saying that UM, look out for babies of colour because you might miss the cyanosis." (NNP07 – White)*

The type of blue a neonate may go if they were not healthy was described by one father as the same type of blue as a White child. However other parents and HCPs felt that Black, Asian, and minority ethnic neonates would go blue in a 'different way'. The most common descriptor was grey (seven HCP and three parents), however two HCPs noted that once an infant is grey it is seriously ill. Two HCPs and five parents felt the neonate would look a darker blue, with

another HCP and two parents describing it as ‘very dark to black as opposed to blue’ (NNP09 – Black). In contrast, four parents felt their child would look lighter blue, especially around their lips. Other descriptors put forward included purple (four parents), mottled (three HCP), dusky (two HCPs), gloomy (one HCP), ashen (one HCP) and a blue green undertone to the skin (one HCP). One mother noted assessing skin colour would be especially difficult in the first few days after birth as they were unaware of their infants’ ‘normal’ colour at that point.

“They [Black, Asian or minority ethnic neonates] go, they go blue in a different way.” (NNP03 – White)

One mother when asked if she though blue was relevant: “He’s a mixed baby, his skin is more white. So, my baby, yes. But now I’m thinking about my nephews, my family members, probably not, because if you have like a darker complexion identifying blue could be a bit tricky.” (PA02 – Asian)

“I think for the Black babies, it might be kind of difficult to identify that … And if he actually developed that, it would have been kind of difficult for me to assess.” (PA15 – Black)

“It’s going to actually be the same type of blue colour yeah, because the yellow colour in my child eye was quite similar to that of a White child.” (PA11 – Black)

“It tends to look just very dark and lacking in colour so it I would say they probably will never look blue.” (NNP09 – Black)

“I would assume it would be quite light to start with, which also will make it difficult to identify. The difficult thing is actually is of course, it’s compared to normal and when your babies a day old, you don’t really know what normal is.” (PA22 – Mixed)

Increased severity and collapse in Black, Asian, and ethnic minority neonates

Due to the differences in presentation of cyanosis in those from Black, Asian, and ethnic minority backgrounds, several HCP who themselves were from ethnic minority backgrounds gave examples of where they had picked up cyanosis that had been missed by their colleagues. Concerns were also raised by one paediatrician that the action of colleagues could vary from being overly concerned when an infant was fine, to not noticing when an infant with darker skin pigmentation was quite blue.

“I would say that probably in the UK, I would say it’s almost in a way, it’s hit and miss. In that some babies you know I’ve seen some SHOs show that are very worried because a baby doesn’t pink up … and I go to review the baby and the baby is clinically fine they’re just darker skin … Whereas I’ve seen other cases where … they call me for another reason … and I look at them

and I say, actually, this baby is quite blue, so and we put on the sats and it's quite low." (NNP09 – Black)

There were concerns that this lack of knowledge of how to detect cyanosis and the difficulty in identifying it in Black, Asian, and minority ethnic neonates lead to increased severity and poorer outcomes once identified (n=5) as only the more extreme cases were identified (n=1).

"Sometimes it's quite difficult because you might have a child who is severely unwell and because of the dark skin, you might not see the cyanosis there. The lips that are going blue, you might not see that." (HV02 – Black)

"What happens is if people are not confident in detecting it, things you know you could get, you could have a really, really unwell baby that's been unwell for hours and nobody knows because 'ohh but the baby's colour looks fine'". (HV08 – Black)

Pulse oximetry

Given the inaccuracies and difficulties of identifying cyanosis, especially through skin colour, several HCPs noted the importance of pulse oximetry. Two HCPs noted that increased use of routine pulse oximetry at birth within their area of practice meant issues around assessing the appearance component within the Apgar had decreased. Seven HCPs felt oxygen saturation should be checked either routinely or with a very low threshold if there were any concerns. Five also reported oxygen saturation monitors to be relied on for all infants in NNU which made one paediatrician confident that no infants had missed a diagnosis of cyanosis. One HCP appreciated that oxygen saturations were now routinely used within their area of practice as part of the NIPE examination and another felt that the use of pulse oximetry had superseded the need to recognise cyanosis visually. Others however reported limitations of pulse oximetry both from their own experience (n=1) and due to knowledge that cardiac anomalies can still be missed (n=1). Reliance on pulse oximetry was also questioned in Black, Asian, and minority ethnic neonates given the underestimation noted in adults with darker skin pigmentation (n=6).

"I don't know many people who would be confident enough to say this person cyanosed I don't need a sats monitor, we would always use that as a sort of a backup." (NNP04 – Mixed)

Assuming all ill infants are in hospital

There appeared to be an assumption by two HCPs that all ill infants are in hospital, so they didn't need to be assessing for cyanosis as most of the babies they see are well.

"It's not something that I see often because most of the babies that I see they're at home, they're well, you know." (HV08 – Black)

Difficulty in HCP with colour blindness

Detecting subtle skin colour changes can present significant challenges, particularly in situations where ambient light may affect the accuracy of assessments. Additionally, two HCPs reported that their own colour blindness further complicated the skin colour assessment process with both feeling it was almost impossible to accurately detect subtle changes in skin colour.

Training needs for detecting cyanosis in ethnic minority neonates

Many HCPs (n=9) reported that they had not received any training on how to detect cyanosis in Black, Asian, and minority ethnic neonates, despite receiving regular updates. As a result, several HCPs (n=4) felt underconfident in their ability to detect cyanosis in Black, Asian, and minority ethnic individuals or unaware of what to look for to determine 'normality' (n=1). One HCP wondered if she herself had been Black whether she would have a better understanding of how to assess an infant. Their perceived under-confidence had led one HCP to undertake their own learning activities. Only one midwife described being adequately trained around recognition of cyanosis in an ethnic minority population. Additional training was also wanted to know the best way to assess cyanosis (n=9). Even one parent called for better training for HCPs about how things can look different in skin tones.

"Considering that we get a yearly update of neonatal resuscitation, BAME is not part of that at all, really reflecting on it. And assessing babies from BAME backgrounds does not get covered at all. It is just a bit of a one stop shop for everybody." (MW12 – White)

"I really don't think I've got a good, you know a good working knowledge of what I should be looking for in terms of skin colour." (HV03 – White)

"We need more training on that to understand what is normal colour for babies, right? … I wouldn't be able to really describe that to you off the top of my head … Can I say what I think is normal in colour for a Black or brown baby? No, because it's not something that I'm exposed to." (MW09 – White)

"Learning about what to look for in different skin tones is basic. And it should be taught same way you know if you're doing hairdressing, I would expect you to know how to do all types of hair." (PA04)

Two interviewees had questioned others about detecting cyanosis in Black, Asian, and minority ethnic neonates. When one of the interviewees had undertaken training, she described questioning the trainer about detection of cyanosis and despite the trainer working in NNU for over 10 years they confessed they had not considered how cyanosis detection may differ in those with darker skin. The other had been met with the response that she had asked a good question, but then received no further training.

> *"The answer is that was a really good question. It's never followed up with an action or any action points. So, when in regards to your question, it's not very much training that that happens at all, in fact there's none." (MW01 – Mixed)*

Others talked about learning to recognise cyanosis in infants from ethnic minority backgrounds becomes easier with experience (n=4) and for one HCP was assisted by seeing the colour change process in reverse as oxygen was given to an infant with low oxygen saturations.

Jaundice

HCPs were asked how often they had assessed a neonate for jaundice in the last two years. It was not part of the role of the participating obstetricians, seven other HCPs had not undertaken an assessment in the last two years, six had done so but only infrequently and seventeen had done on at least a monthly basis.

> *"In the last two weeks I've admitted 4 babies with jaundice, 2 Asian, 2 Black babies with jaundice you know, it's a regular occurrence with me." (MW02 – Black)*

HCPs were aware that they didn't always have much exposure to neonates from Black, Asian, and minority ethnic backgrounds who had jaundice.

> *"To be honest and I can't think of any babies with very very, very dark skin that I've been concerned about." (HV04 – White)*

Detection of jaundice

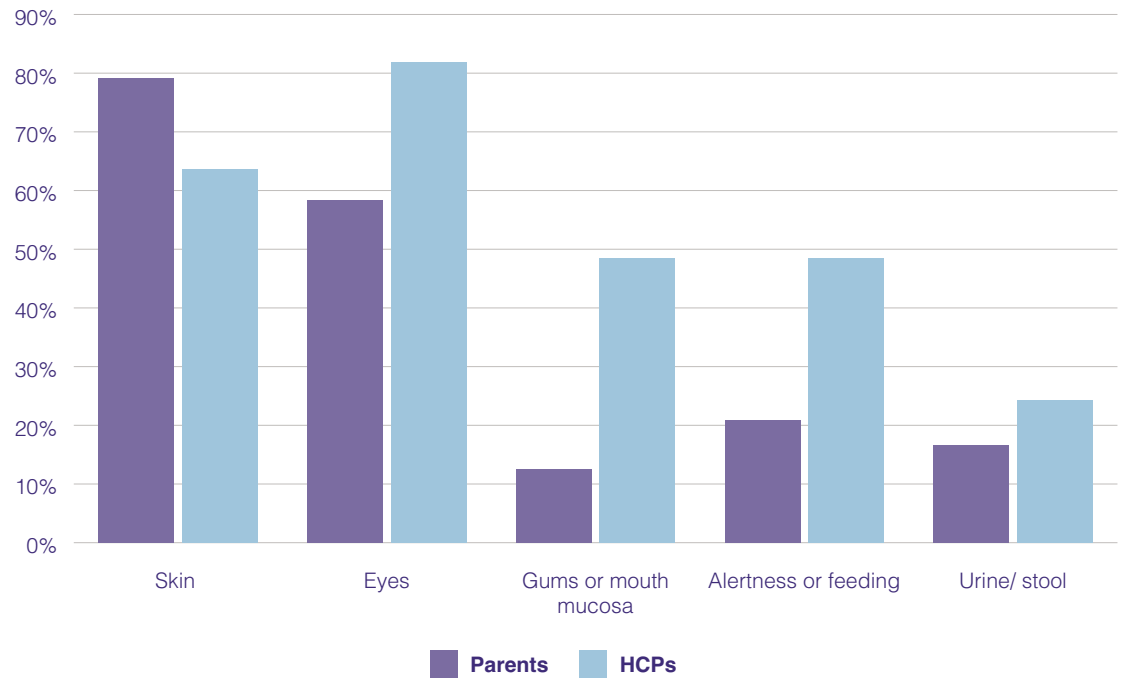
When parents were asked if they knew how to assess if their baby had jaundice, eight parents said they were not aware of any signs, although six of these parents went on to guess that they might be able to see it in their baby's skin. Two reported that they had been unaware before they had been given guidance by HCPs in hospital and four that they had been unaware before they had a child with jaundice. Two parents said they were aware but gave no further details, three were aware due to having a relative who had had neonatal jaundice, one

mother had read about it, three stated that they had only been aware to look at skin colour before their own child had jaundice and another mother said they were slightly aware as had jaundice themselves, but much more aware after having a child with jaundice.

> *"I didn't know that you had to look at the eyes and … the gums … the wee. I didn't know any of that until I'd actually looked it up myself. Yeah, nobody even in the hospital, no midwife, no support worker, no doctors, no nobody had pointed that out." (PA14 – White)*

HCPs and parents described that they would look at numerous signs to assess for jaundice. These revolved around looking for the colour yellow in various locations including the skin, eyes, gums, or mouth mucosa or considering the feeding pattern, tone or the infant's urine or stools. These are shown in Figure 6.7 and described below.

Figure 6.7. How interviewees would assess for jaundice



Looking for the colour yellow

Skin

Looking at skin colour was the most common way parents knew about to detect jaundice in their infants. While most parents (n=16) would look for the skin being yellow, one mother talked about darker looking skin and the parents of two babies who had shown signs of jaundice outside of the neonatal period, described seeing blemishes on the baby's skin that looked yellow. However, not all parents who said they would probably look for jaundice on the infant's skin were completely sure about this.

"I'd probably look at his face in his eyes, his eyes, the whites of his eyes. But I guess then any of his skin I don't. But then I'm just, I'm guessing here I don't really know." (PA16 – Asian)

Among healthcare professionals, twenty-one said they would assess the neonatal skin in some way. Two HCPs felt that yellow was recognisable in all skin tones, while others noted it just looked different in Black and minority ethnic neonates, especially in those with severe jaundice. Two HCPs felt that in neonates with darker pigmentation red may be evident rather than yellow or that the skin may sometimes look flushed before becoming yellow.

"Every single baby still has a yellow tinge under the skin." (HV05 – Black)

Specific places on the skin where HCPs or parents would look for yellow colouration included the soles of the feet or palms of hands (7 HCPs and 5 parents), or the nail beds (one HCP). However, it was noted by one paediatrician that, if jaundice was visible in the feet, the neonate would need an immediate blood test. One HCP and one parent said that they would assess the neonate's arms and/or legs more generally. Other areas parents described that they would check for yellow colouration included the stomach (n=3) and the back (n=1), with one HCP also saying they would check the body in general. Four parents said they would look at their babies face and another specifically the forehead, with two HCPs also saying they would look at the forehead. However, for one of the HCPs that would include pressing on the forehead to see the colour when the skin was blanched. Five HCPs and one parent stated that they would assess for jaundice on the neonate's nose, with all of the HCPs stating they would do this by pressing on the nose. A further three HCPs said that they would press in the sternum. Five more HCPs talked about assessing capillary refill or pressing on the skin but did not explain what location on the body they would do this.

To assist them while assess a baby for jaundice, one HCP said they would compare the baby's skin tone to that of its parents. However, that HCP and two others noted that assuming a neonate's skin colour is due to their parent's ethnicity may also lead to missing cases of jaundice.

"Always look at the parent if the parent and the child look very similar, you're probably on around the right road yeah. When the child looks completely different to any of the parents then you think somethings not quite right here … It's a simple thing, but certainly not fool proof." (HV01 – White)

"I have actually had one recently for a baby who was from a different ethnicity and he appeared to have jaundice and I was pretty certain that he had jaundice, but his parents were quite sallow skinned so you're not always aware of what's kind of normal and what is concerning." (HV04 – White)

The importance of assessing more than just the skin, particularly in Black, Asian, and minority ethnic neonates, was raised by 16 HCPs due to their awareness that jaundice is not always visible on skin, with five of these HCPs seeing an overall assessment of the neonate as more important than looking at its skin colour.

> *"Jaundice is one of the issues, because you can never be like 100% sure on OK, this baby looks yellow or it doesn't. And it has happened, because I'm quite thorough, it has happened that I was like I don't think so, but let's do a test and then came into like the highest treatment level kind of thing." (MW10 – White) (On a Black baby)*

> *"I suppose in a way because I don't feel I've got any expertise in skin colour; I would be probably wanting to make sure that the other signs and symptoms that I know to be important I really do explore with the mum. Whereas if again if it was a neonate of White origin and you know, I could see that there was no sign of jaundice from just looking I probably would be less likely to." (HV03 – White)*

Eyes

Fourteen parents and 27 HCPs said they would look for yellow in the whites of the eyes to detect jaundice. However, several HCPs were concerned that by the time jaundice is visible in the eyes jaundice is already severe (n=5) or that it doesn't always help to look there (n=1).

Gums

Two parents whose child had been hospitalised with severe jaundice were aware that jaundice could be detected by looking at the gums. Eleven HCPs suggested they would look at the gums which were considered better than looking at the skin. However, again one HCP raised concerns that jaundice is more severe once it can be identified in the gums. Seven HCPs also stated that they would look at the mouth mucosa, with one mother also saying that they would maybe look in the mouth although they weren't totally sure about this.

Alertness or feeding issues

Fifteen HCPs and three parents stated that they would consider whether the baby was feeding adequately when assessing for jaundice, with failure to gain weight seen as a sign of inadequate feeds (n=4). Whether an infant was sleepy or alert would be assessed by ten HCPs, with one other HCP saying they would look at the neonate's behaviour in general and another that they would consider the neonate's tone. However, the limitations of focussing on alertness or feeding were well recognised by one couple whose baby had remained alert and feeding despite high bilirubin levels. As a result, his jaundice had been dismissed by a HCP as not serious, when in actuality his bilirubin levels were over the exchange transfusion line.

Urine or stools

Lots of dirty nappies were seen by three HCPs to be a good indication of whether the baby was feeding well. Six HCPs also stated that they would look at stool colour, with a jaundiced neonate expected to have pale and chalky stool. Urine colour or quantity would also be assessed by four parents and five HCPs, with darker urine expected in a jaundiced neonate.

Misconceptions

One midwife suggested that they would associate vomiting with neonatal jaundice and three parents suggested incorrect symptoms of jaundice including abdominal swelling, infection, and itchy skin.

Best practice during assessment

Multiple HCPs (n=13) highlighted the necessity for any jaundice assessment to be undertaken in bright or natural light or there would be a danger of diagnosing jaundice later. Top to toe assessment by fully undressing the baby was also described by seven HCPs. After the pandemic, the decrease in face-to-face appointments was noted by two health visitors to make any assessment of jaundice almost impossible.

> *"We're seeing babies undressed as well, when you're looking at the body, you know in the natural light you will see where the tide of jaundice is stopping." (HV05 – Black)*

> *"I would not say it's at all advisable to do jaundice assessments as a virtual thing." (HV01 – White)*

Risk factors

Although interviewees were not specifically asked about risk factors for jaundice several specifically discussed risk factors. The risk factor most frequently mentioned was that certain ethnic minority backgrounds are at increased risk of high levels of jaundice, with some feeling this risk factor was too frequently not recognised. Other risk factors mentioned included preterm infants or those born with low birthweight, instrumental birth, the mother having diabetes during pregnancy and a history of other children within the family having jaundice. One midwife also felt that within her practice that neonates born by Caesarean were more likely to get jaundice, although she recognised that this wasn't traditionally seen as a risk factors. Given the lack of recognition of risk factors by HCPs One neonatologist talked of her desire to have some sort of risk score for jaundice, so if any signs are present with risk factors the baby would be recognised as needing immediate further investigation rather than adopting a wait and see approach.

Challenges in detecting jaundice

Identification in Black and minority ethnic neonates

Many parents (n=10) felt that detecting jaundice wouldn't be any harder for them as parents of their ethnic minority baby, with several parents saying they would look for other signs rather than skin colour. Many parents (n=9) also felt that detecting jaundice in a Black, Asian, and minority ethnic baby wouldn't be any harder for HCPs or they hoped it wouldn't be (n=2) as they expected doctors to be experts. Only one HCP however felt this to be the case, stating that jaundice was very clear to see in all Black and minority ethnic neonates.

In contrast, other parents (n=13) felt that their baby being from an ethnic minority would make it harder for them to identify jaundice, especially as they are unaware of what their child's normal colour would be when they were first born (n=4). Many also felt that identifying jaundice may be harder (n=5) or would be harder for HCPs (n=8), especially if they weren't used to seeing jaundice in ethnic minority neonates. One parent recalled being asked by a paediatrician what she thought of her child's skin colour but finding the question difficult to answer given the above concerns reported by parents that they don't know what is normal for their child initially.

> *"There was times when it was difficult to tell because of obviously the skin colour because urm, yeah, because of the tone of my kids' skin tone, it was at times difficult to tell." (PA13 – Asian)*

> *"One of the doctors did once ask me what I thought about his skin colour related to the jaundice, but I was a bit like, "well I don't really know what he's supposed to look like, it's a baby."" (PA16 – Asian)*

Most HCPs also felt that it was more difficult to identify jaundice in Black, Asian, and minority ethnic neonates (n=20), with one also noting it to be particularly difficult to detect in children of dual heritage. HCPs felt jaundice was frequently missed in Black and minority ethnic neonates (n=11), with the level of jaundice being higher once identified (n=4) and therefore Black, Asian, and minority ethnic neonates being over-represented within readmissions to NNU (n=4). An additional HCP was uncertain whether cases of jaundice were more likely to be missed in those from ethnic minority backgrounds but hoped not. Within the parental interviews, three separate cases of jaundice were reported to have been missed at some point by at least one HCP.

> *"It can be difficult, you know, especially like the darker the skin the more difficult it can be." (HV05 – Black)*

> *"I know that's our job in community to identify the babies who are unwell, but I think a lot of these could have been identified earlier." (MW02 – Black)*

"I would hate to hear that we are missing more jaundice in Black and ethnic Asian babies possible difficulty in understanding how they may present with jaundice. I mean, I and I don't know that." (HV03 – White)

"The most severe cases that are re-admitted and tend to be non-White. Umm that's probably a bit of a broad generalisation, but the ones that need to because they come to the unit and have multiple lights, tend not to be White." (NNP03 – White)

"She [mentor as a student] was an English, you know, she was White. A mature, you know, 40 year midwife. She still failed to identify that jaundice in the Asian baby and that baby needed to be re-admitted. Luckily didn't have transfusion but it went on the triple phototherapy on the Neonatal unit." (MW02 – Black)

"My partner is a lot darker than I am so possibly because baby came out a mixed colour probably of myself and my partner, so it probably could have been missed for that reason." (PA21 – Other)

Given the different presentations of jaundice in Black, Asian, and minority ethnic neonates, with yellow skin colour not always being evident, the need for different assessments (n=3) and a higher level of vigilance (n=3) were highlighted. This high level of suspicion was felt to have been absent by one parent, with too much emphasis placed on the baby being active meaning that other obvious features of jaundice such as yellow eyes, yellow skin and poor feeding were ignored and meant the child was over the exchange transfusion line by the time the jaundice was formally identified. One midwife also described how she had changed her practice in light of a missed case of jaundice within her work area so that she now had a higher level of awareness to look for jaundice within the NIPE examination.

"One has essentially to be on the alert and to be more aware of being able to recognise jaundice in a baby who has a darker skin." (NNP07 – White)

"I think you have to be aware that things look different. I think you have to I guess almost not trust your eyes, because you're actually missing, potentially you can miss jaundice." (NNP03 – White)

"In terms of my routine NIPE examination, even though it says that the jaundice is not part of the examination, is something that I've routinely begun to do on the back of this recent risk that we had the baby where it wasn't picked up, but it was picked up far too late." (MW03 – Black)

HCP themselves being from an ethnic minority background were seen to assist with the identification of jaundice in Black and minority ethnic neonates (n=6). While colleagues from White backgrounds were noted to have missed it (n=4). Within the parental interviews, one mother felt that the reason the HCP who

visited them may have had no problems identifying their child's jaundice may have been due to the HCP being from an ethnic minority background herself. In contrast, another parent reported their child's jaundice to have been missed by an HCP from an ethnic minority.

> *"The midwife who came, also wasn't White, so maybe that's why she picked it up as well. I can't remember where she was from, but she had quite similar skin tone to me and maybe that's why she also noticed it." (PA22 – Mixed)*

The lack of timely identification was noted to increase the intensity of the treatment required, such as needing exchange transfusion (n=2) and to potentially have a long-term impact on the infant's health (n=3). Not picking jaundice up within the hospital environment was felt to be demoralising for the woman if readmission was later required (n=1), as well as having an impact on bonding and breastfeeding if more intensive levels of treatment were therefore needed (n=2). For one couple, this impact on both breastfeeding and bonding was noted when they talked about a paediatrician coming in and taking their child mid-feed to place him under phototherapy immediately given the severity of their child's jaundice when identified, they had found the incident quite scary. The long-term impact on women's mental health and trust in general health services and HCPs was also noted to be of concern (n=1).

This lack of trust in health services was noted in one of the parent interviews, where the father expressed a concern that their earlier care may have been wrong due to their ethnicity. Additionally, the couple whose child's jaundice was severe by the time it was identified were informed that the paediatrician had put a complaint in on their behalf. As a result of this complaint, the parents had been informed that the hospital was going to put various measures in place to prevent a similar incident from happening again, including buying bilirubinometers for the community and ensuring staff received annual training on jaundice. However, over three years later they were aware that the measures had not actually been implemented which made them lose trust in the health services.

> *"I was quite fearing and I was trembling. Yeah, I was like did the doctor do something wrong or did the hospital do something wrong because of my ethnicity?" (PA11 – Black)*

Training needs

Many HCPs (n=13) either didn't receive, or could not remember having ever received, training in the identification of jaundice in Black, Asian, and minority ethnic neonates. As a result, some midwives had done their own reading to ensure they could identify jaundice in all infants within their practice. Four HCPs described learning about detecting jaundice in Black and minority ethnic neonates from colleagues with more experience in ethnically diverse areas. Four other HCPs reported that they themselves specifically training others to

ensure effective detection of jaundice in ethnic minority neonates, with one of these paediatricians noting that she had to teach more senior colleagues in the UK what to look for due to their experience of detection in Black babies in a Caribbean country. A recently produced resource for students around the detection of jaundice in Black neonates was also welcomed by one midwife. Additional training was wanted to know the best way to identify jaundice in Black, Asian, and minority ethnic neonates (n=13). This was seen as critical given that current guidelines say to test bilirubin levels if the HCP has any concerns, but people need training on when to be concerned. Better training around the risk factors for jaundice was also mentioned, to increase HCPs awareness that those from an ethnic minority may be more at risk of jaundice, but also at risk of not being identified in a timely manner which is why a high index of suspicion is required in Black, Asian, and minority ethnic neonates. Better understanding that bilirubinometers are not precise measures of bilirubin, but a guide as to who required further testing was also mentioned.

> *"I would say the jaundice … is a big thing. You're not taught even now not really taught at university in relation to skin colour and babies on what is normal what's not normal." (MW03 – Black)*

> *"I have not received any teaching whatsoever in the UK specific to assessing darker skin neonates." (NNP09 – Black)*

> *"I remember at the XXXX local hospital where I work one of the consultant neonatologists specifically telling me don't look at the skin, look at the sclera and look at the gums." (OB03 – Asian)*

> *"I think the training in terms of identifying cyanosis and jaundice and pallor in children who have different skin colours is really important." (NNP04 – Mixed)*

Eleven of the parents reported that their baby did not show any signs of jaundice, ten that their child or child they cared for had shown signs of jaundice in the neonatal period and three whose child had had jaundice outside of the neonatal period, several months after birth in two cases and at six months old in the other.

Of the 12 cases where the infant showed signs of jaundice it was noticed by the parents first in six infants, by paediatricians in the hospital for two infants, by the community midwife/health visitor for two infants, by a grandma who was a nurse by background in one of these cases and by the mother, her husband and grandma in the last case. Several parents reported that jaundice had initially been missed by HCPs in the hospital or community (n=3) and in one instance when the parent had raised concerns about jaundice they were dismissed by the HCP.

> *"I didn't realise it at the time I didn't. It might, it could have been the skin colour that's why it was hidden, but I didn't realise at the time, but I think it*

was a second appointment or first appointment that the health visitor came out or midwife came out. They did say that he had mild jaundice." (PA21 – Other)

Missed cases were put down to a lack of staff training in identification by one of the midwives. Additionally, one parent specifically felt that HCPs weren't adequately trained to identify jaundice in infants from an ethnic minority. Several HCPs felt that identification of jaundice in neonates from Black and minority ethnic backgrounds became easier with experience (n=4), with another HCP also attributing her lack of confidence in identifying jaundice in Black, Asian, and minority ethnic babies to the fact that she worked in an area of low ethnic diversity. Some (n=2) observed that cases were missed as it was too easy to assume an infant's colour was due to their ethnicity rather than considering jaundice.

"So, I think the perception of what is a normal colour for an ethnically Chinese sort of Asian person, the perception that it's normal to be yellow, which of course it isn't can be life threatening really. And yes, I mean to get admitted to NICU [neonatal intensive care unit] with a bilirubin on the exchange line and then to hear that the midwife when the parents are worried said "oh it's just your race", it's pretty terrifying." (NNP08 – White)

Listening to parents

Given the reported difficulties in identifying jaundice in Black, Asian, or minority ethnic neonates, some HCPs (n=6) noted the importance of listening to parents' concerns, with one of these HCPs and four others, therefore, mentioning the need to ask the parents for a full history. Three HCPs relayed stories of where parents' concerns had been dismissed by HCPs resulting in treatment delays. This concern was also raised within the parent interviews with both parents of one child who ended up with severe jaundice recalling not being listened to which resulted in significant delays in obtaining a diagnosis of and treatment for jaundice. However, two HCPs also recalled situations where parents hadn't noticed their child's jaundice and remarked that parental awareness cannot solely be relied upon.

"Listen to the parents if the parents are worried, just do the bloods. Yeah, you know, if you've got a parent saying to you "I think my baby jaundice" I mean, they're probably right." (NNP03 – White)

Reliance on transcutaneous bilirubinometers and serum bilirubin

The known difficulties in identifying jaundice in those with darker skin pigmentation meant numerous HCPs (n=8) reported having a very low threshold for checking either a TCB or an SBr. In the community, if they didn't have access

to a bilirubinometer, this required any concerns to be referred to a hospital or clinic for proper assessment of bilirubin levels. Although one midwife did not feel that this was done frequently enough. Within the NNU, there was a tendency to solely rely on SBr given lower bilirubin level would require treatment in a preterm infant (n=3).

> *"You've just got to use different strategies and have a lower threshold to check a formal bilirubin otherwise, you run the risk of missing it." (NNP04 – Mixed)*

> *"I have seen babies that hardly look jaundice and they have quite high levels when you test them. And I was lucky because I was in a hospital environment that I could test them and they were lucky also because they were over 24 hours so I could use the transcutaneous machine." (MW10 – White)*

Opinions were split over the use of TCB machines. Some were very grateful for them, feeling they helped avoid missing cases (n=3) and unnecessary referrals to the hospital due to being able to assess the neonate at home. One midwife reported that she was trying to improve access to them within her community area. While the cost implications of buying TCB machines were voiced by three HCPs, they were felt to be cost-effective compared with missing cases. However, seven HCPs, including paediatricians, neonatologists, midwives and an obstetrician, all voiced concerns over the accuracy of TCBs, particularly in darker pigmented neonates with both under and over-estimation of serum bilirubin levels noted. As a result, one midwife noted that they should not be seen as definitive. Additionally, concerns were raised about whether all HCPs properly understood that the TCB reading isn't exactly representative of SBr levels.

> *"The TCB monitor that we use tends to overestimate in darker coloured neonates …I think the TCB works really well on babies who are lighter skinned, … and when these children come in [to hospital] then it's much closer to the line. So, I would say I probably see a lot more children who are of darker skin coming in to be assessed for jaundice because the TCB is not working as well." (NNP09 – Black)*

> *"A mother was worried the Billi metre gave a low reading and the midwife did a bilirubin and it was above treatment level, so the baby came in. And Billi Metre had read over a hundred below that." (NNP03 – White)*

> *"Our bilirubinometer, if it's over 250 you have to do an SBR … I did a bilirubin, it said it was 160. I still wasn't happy because this baby was bright yellow was OK and they did it [SBr], it was over treatment … it was a good bilirubinometer, it was calibrated. My point is, you know, there is room for error." (MW02 – Black)*

The low threshold for checking bilirubin levels was noted by one neonatologist to lead to more blood tests in Black, Asian, and minority ethnic neonates, with

a further paediatrician feeling that there was an over-reliance on SBr. This increased number of blood tests was also voiced by two parents. One mother whose premature child was in NNU recalled them having regular blood tests to assess bilirubin. While she found it reassuring that doctors weren't solely basing their diagnosis of jaundice on the colour of her child's skin, she also recognised that her being from a mixed background may have impacted how many blood tests they required. Another parent also felt that her child was tested more due to his darker skin and her nephew had had a similar experience.

> *"I think that general rule that it's harder to assess in infants who have darker skin, so do a blood test. So, I think we have a lot of unnecessary blood tests being done." (NNP09 – Black)*

> *"So, he had jaundice when he was born, but not severely enough to have treatment. But of course, his skin color is darker so it's harder to tell. So, some of the checks we thought you know they went on a little too long because he was fine, he was just darker than a White baby, basically. And that actually happened to my brother's baby as well." (PA22 – Mixed)*

Action required

Additional suggestions for training

Within several areas it was noted that all of the resuscitation dolls used for training were White, which didn't naturally lead to thinking about how conditions may look different in babies from ethnic minority backgrounds, so there was a call for mannequins to better represent the diverse nation that we live in (n=2). Knowing what to look for was viewed as central to ensure HCPs could care for everyone effectively, whatever their skin tone, to enable current inequalities to be reduced (n=5). It was suggested that all training should start with the data around current inequalities, such as increased maternal and neonatal mortality, to demonstrate why we need to improve our services. Exposure was viewed as the best way to learn, but it was recognised that this was not always possible. There was therefore a particular call for training for HCPs who worked in areas of low ethnic diversity (n=6), as they get less exposure and so have less knowledge and experience. Others however wanted to see more widespread training, for example to include all university students on HCP courses (n=4), junior midwives and doctors (n=1) or more general mandatory training with regular updates for all HCPs around differences in Black, Asian, and minority ethnic presentation (n=4). One paediatrician had really appreciated the training they had received at the start of one rotation about the main differences they would need to look out for as the hospital served a largely ethnic minority demographic.

"We won't fix it [disparities] until everybody is singing from the same hymn sheet. And unfortunately, in my almost 20 years I haven't seen that much of an improvement. And that's the sad thing about it all." (MW02 – Black)

"I think the training needs to always start with the data. The data that certain people are dying unnecessarily. It's our job to stop mothers and babies dying or being unwell or have life changing conditions and therefore it's our job whether we like it or not to improve our service and improve the way we work." (NNP08 – White)

"I've rotated through maybe 10 hospitals so far in the UK maybe that's a ballpark figure. And only one of them taught you know specifically address this and that was because it was in an area with mostly minorities. I think all hospitals should be addressing that in neonates you know. These are the main differences that we're looking out for. When in doubt, ask a senior or have a low index of suspicion. So, I think it should be part of induction when we rotate through each hospital." (NNP09 – Black)

Unconscious bias training was also seen as beneficial by two HCPs, with one of these appreciating the diversity training they had received in the UK when nothing of that manner had been provided in their home country.

"No one works to hurt anybody or upset anybody. But sometimes if you just don't know, we just don't know, do you? So, if you have unconscious bias training just to put that out in the forefront. And people think twice before saying certain things or doing certain things." (MW07 – Black)

Establishing a databank of images of babies from diverse communities

Numerous HCPs recognised that pictures in current textbooks, guidelines, or information leaflets are generally of White babies. There was a really strong call among HCPs therefore for a national data bank of pictures covering a diverse range of conditions and what the condition may look like in a Black, Asian, and minority ethnic neonate including jaundice, sepsis, rashes, and cyanosis (n=14). This resource was called for both for HCPs (n=13) and for parents (n=6). Some HCPs also liked the idea of videos as well as pictures. HCPs felt that having this resource would make it simpler to determine in practice whether something was normal and wanted clear action points to follow if there were any concerns. HCPs mainly talked about it being an online resource, with one also liking the idea of having a small pocketbook to refer to. Such a databank of pictures was also seen as a useful resource so that relevant images could be added to textbooks and the curriculum. Several HCPs (n=3) talked about the fact that it was unnecessary to start something from scratch, with liaison with other countries across the world to share resources, especially those with majority Black or Asian populations, seen as particularly beneficial. Several HCPs also mentioned

the ‘Mind the Gap’ project which has already started to gather ethnically diverse pictures of medical conditions into the public domain.

> *“Really simple pictures of different skin conditions and how they look on darker skin we’re just not very aware of that, widely aware I don’t think.” (HV04 – White)*

> *“There should be more training, even the basic stuff of images and pictures and photographs and videos should include various ethnic minorities just so that it’s easier to identify people and babies and who are affected.” (MW05 – Mixed)*

> *“Why are we trying to make a new directory when we have a global majority that probably already has a directory of images and we just need to reach out to them and see if we can make a better, a representative resource.” (MW04 – Asian)*

While most felt pictures could be a really helpful tool, some did acknowledge that just looking at photos or having a lecture would make it harder to assimilate the knowledge than seeing something face-to-face in a real-life learning situation (n=3). However, the limited exposure in practice to things such as cyanosis, means that videos and photos become invaluable.

Educating women and families

HCPs wanted appropriate advice for women including around how to detect jaundice (n=7), how to recognise an unwell baby or a baby whose condition in worsening (n=7), with details provided of who to contact when (n=4). This was seen as being of greater importance given the increasing proportion of early discharges from hospital and that visits by HCPs in postnatal period have been reduced. One HCP had also undertaken a study showing minimal knowledge around jaundice among mothers. HCPs wanted to see information provided to women in appropriate languages and in a culturally sensitive way, with pictures seen as ideal as they are easier to understand than written words for many.

> *“Parents do need training because they’re the ones that with early discharges nowadays they’re the ones that flag up concerns and I do think it needs to be focused on parents of darker skinned babies. Because they it’s just as difficult for them as it is for us.” (NNP08 – White)*

> *“We often say to women, look out for yellow for jaundice and that is just a, you know, a run off the tongue things so we’re not actually tailoring it to the people that are receiving this information.” (MW01 – Mixed)*

"I think it's the parents that need it [training] as much as anybody because they're the people who see that child from birth all the way through." (MW13 – White)

Guidelines

Current guidelines were felt to be too generic with a 'one size fits all' mentality, assuming that everyone is White and so not adequately considering whether practices disadvantage those from an ethnic minority (n=14). One HCP even checked their local jaundice policy during the interview, discovering that it didn't talk about detection in Black and minority ethnic neonates, although they recognised that may be due to them working in an area of low ethnic diversity but noted that that was still not an excuse for it not to be considered. HCPs called for guidance to be more inclusive and reflective of the whole population particularly around the conditions discussed during the interview (n=19), but also around more general guidance (n=13), with any impact for those from ethnic minority backgrounds considered proactively as any guidance was written or updated, rather than waiting for serious incidents to initiate any change. This was also highlighted by the parents of one child who had a bilirubin level over the exchange transfusion line, as they found out after the event that their local Trust guidance hadn't been updated to reflect wider guidance which noted that closer monitoring may be required with different skin tones.

"I think that unfortunately, most policymakers are White male, so you know we are still, we still have White privilege that dominates our communities, our science, our politics." (MW13 – White)

"There is notes in our NIPE trust guidelines to be observant of ethnic and cultural difference but there isn't anything specific." (MW05 – Mixed)

"I think only in that in the last few years are we hearing a talk about skin colour and even recognising that it makes it, it does change the course of someone's clinical journey. So, I think once you begin to realise that, oh yeah, hang on there is a difference only then we start to ask the questions. OK, well what is the difference and what can we do about it?" (OB02 – Black)

"It needs to be clear in all guidelines that, it might just be blindingly obvious, but saying that when assessing a BAME baby, take into consideration skin colour, et cetera. So really making a point of it straight away … even just mentioning that it is difficult to assess colour for a of a baby from a BAME background, so because of this, we recommend doing XY and Z." (MW12 – White)

"I think it [skin tone] should be considered in all policies and guidelines and just to have that acknowledgement that there is a difference and we need to consider the difference and whether there's a difference or not." (MW05 – Mixed)

HCPs wanted national guidance to be rolled out. Interestingly, while some HCPs felt that guidelines were depended on by HCPs (n=2), others cited a lack of time to read things like guidelines (n=3) and so relied on clinical educators to pass on any relevant information. This was backed up by twenty-one HCPs stating they were unaware of guidance or were not up to date with the latest guidance, including thirteen around jaundice guidance and thirteen around cyanosis guidance. Given clinicians lack of time, any new guidance needs to be kept simple, for example by using a flow diagram. One HCP also felt that it needs to be well signposted, as he felt many people write good equality impact assessments, that could be more widely circulated to support others.

"Most of what we do we can't remember off the top of our heads, especially when we newly qualified. So we do depend on our guidelines and our policies, so having that there is it resource is really important." (MW09 – White)

Specific guidance wanted around cyanosis included that they should say to look harder for cyanosis in ethnic minority neonates; to advise looking at mouth, tongue, and mucous membrane colour rather than skin colour; and to use pulse oximetry as this was seen as better than assessing skin colour. Regarding jaundice guidelines, one health visitor noted that their newly updated guidance no longer focused on skin colour as a key factor, but on other areas for assessment such as feeding and stools which they felt was a positive step forward. Other HCPs called for more detail within other guidance on ways to detect jaundice that didn't involve skin colour. Some of the areas felt important to include were to look at the mucous membranes and eyes and to do a TCB if there were any concerns.

Other conditions of importance regarding neonatal assessment and skin colour

Within the interviews, parents and HCPs were provided with an opportunity to highlight other areas of difficulty. The most commonly raised concern was around identifying skin rashes including eczema (14 HCPs and 4 parents). Assessment of infection was raised by two HCPs, as well as one parent who had been denied antibiotics for an infected nipple after her son bit it as her nipple didn't look red (she had Black skin). Six HCPs also raised concerns about the identification of sepsis in Black, Asian, and minority ethnic children and wanted to see further education for both them and parents around this.

Slate grey nevi birthmarks were also discussed by 13 HCPs. Although HCPs acknowledged increased awareness and the significance of documenting to prevent false allegations from families, three parents raised concerns. These concerns included missed symptoms on initial examinations or instances where the parents felt the need to ensure repeated and comprehensive documentation to prevent potential safeguarding complications later on.

Three paediatricians or neonatologist highlighted the issue of Black, Asian, and minority ethnic babies being harder to canulate. One mother also wondered whether the difficulties staff had when canulating her during labour may have been due to her skin tone.

> *"When you're taught to look for things you're taught, like oh a red blotch or oh it'll be raised, or the cheeks will be rosy red, but not everyone is the same complexion, you know, even amongst Caucasian people, they're not all the same. So you can't just say oh it will be red and it will look like this." (PA04 – Black)*

> *"Whilst it's easy to get cannula in a Caucasian baby it might not be easy in a Black or Asian baby." (NNP01 – Black)*

> *"I mean I don't know if this was to do with skin colour or not. I know that I have notorious from my veins are difficult to find, but I really struggled with some having the … cannula put in. It was, I think it was two midwives tried and then they called in for an anaesthetist and then eventually that also caused a problem and they had to switch hands and an anaesthetist had to be called again." (PA20 – Black)*

Summary

- There is a general consensus that "pink all over" in Apgar scoring is derived from observing White babies and its relevance to babies from Black, Asian, and minority ethnic backgrounds was questioned. The question voiced by many was "where to look" for "pink" or "blue" as part of a whole range of other indicators for the health and wellbeing of the babies. The continued use of the Apgar score in the current form may be perceived to be an inherently racist practice.

- The need to provide adequate training and educational resources for accurately assessing neonates, regardless of their skin tone, was emphasised. It was noted that exposure to babies with diverse skin tones is important in achieving this goal.

- Around the lips or the mouth mucosa were frequently cited by parents and/or HCPs to be an appropriate place for assessing cyanosis and for the appearance component of the Apgar score in all ethnicities. However, difficulties in detecting cyanosis in all infants and particularly those from ethnic minority backgrounds was noted.

- Assessing the sclera and gums rather than the skin for jaundice were reported by HCPs to enable better detection in Black and minority ethnic neonates, however visual assessment was noted to be poor, with HCPs having a low threshold for testing TCB or SBr if any signs of jaundice were detected.

- Creating opportunities for exposure to multi-ethnic and racially diverse communities for all HCPs and establishing a databank of pictures from diverse communities and babies with various conditions impacting skin colour for training purposes in both hard copies and digital sources are important and would facilitate equitable maternity care for all.

The impact of ethnicity and race on care

Although the overall focus of this review was on neonatal assessments, challenges around quality of care could not be ignored within the context of race and ethnicity. As part of the systematic review of parents' experiences, data around potential barriers (and facilitators) to accessing care was extracted. Similarly, within the stakeholders' interviews, parents and healthcare professionals were asked about challenges or barriers to accessing or receiving care, as well as any areas of good practice. Additionally, they were asked about any areas they were aware of regarding discrimination or inequalities within care. The decision was therefore taken to further analyse this evidence. Five themes were identified within the data and are presented in Figure 7.1

Figure 7.1 Themes representing the impact of ethnicity and race on care.

Not given a voice

The first theme identified was around women not being given a voice. This was due to communication issues, being deliberately silenced within the system, and through lack of consent and choice.

Communication issues

A common concern identified from literature and interviews was poor communication and misinformation from healthcare professionals.

Some communication issues stemmed from language barriers between patients and staff (n= 19 HCPs; Hannon et al., 2001; Lee & Weiss, 2009; Palau et al; Witt et al., 2022). Language barriers were thought to add to parental difficulties within appointments, understanding medical jargon, colloquialisms, or the information given to them and with obtaining consent. Within the stakeholder interviews, several HCPs identified the need for appropriate language resources (n=10). However, HCP interviewees also identified issues with interpreters (n=7), such as interpreters not being used, relatives used for translation (n=5), and worries about information being lost in translation (n=3). Similarly, within the review, mixed views were identified about the use of interpreters and translators, with some parents suggesting it was 'brilliant' and others indicating they were not forthcoming or adequate (Birthrights, 2022; Palau et al., 2019; Witt et al., 2022), or that women described relying on their partner to translate for them, even if they understood 'a little' (Witt et al., 2022).

> *"Seeing that as a right for those parents to have information given to them in a language they understand." (NNP04 – Mixed)*
>
> *In one instance in the literature, a woman was denied an interpreter for her appointment as it was viewed by the HCP as a "waste of public money" (Birthrights, 2022).*
>
> *Discussing this issue, a midwife recalled an instance where a patient's husband was used as a translator and his "English was awful… and they (HCPs) were using him to translate medical terms… there's no informed consent there." (MW01 – Mixed)*

Within the interviews, several HCPs (n=3) made sure to check parents understanding before leaving, to ensure parental comprehension. However, several studies within the systematic review identified parents' misunderstandings of their child's condition and their tests due to poor translation and staff misinterpretation of medical terms (Lee & Weiss, 2009; Palau et al., 2019). One concern expressed by both parents and HCPs in the interviews was the potential difficulty for those whose first language was not English to be

able to ask HCPs appropriate questions to "put forward an argument" or to be "assertive" to receive the care they needed.

> *"What you could ask is depending on ethnicity or language proficiency… are some groups more or less likely to ask critical questions about care?" (OB03 – Asian)*

In the literature and interviews, language issues included resources supplied to parents not being provided in alternative languages which may hinder access to care (n=3 HCPs; Birthrights, 2022). Other communication issues identified by HCPs and parents (n=4) stemmed from educational barriers, occupational status, and parental literacy levels.

> *"We have mothers and fathers from all ethnicities who need more or get more attention and time because of their education levels and their confidence in asking." (NNP08 –-White)*

Another issue in communication raised in the systematic review was around the lack of availability of nurses to answer questions (Hurst, 2001). Effective communication and willingness to answer questions put mothers at ease and made them feel empowered (Birthrights, 2022; Hurst, 2001). Within the interviews, parents similarly suggested poor communication and a lack of support offered to them by HCPs (n=7) and being made to feel like an inconvenience when requesting more support (n=3).

> *One mother said "The nurses are great. They try to answer your questions, but they really are busy." (Hurst, 2001)*

Women silenced within the system

Dismissed

Many parents reported not feeling listened to in the literature and in interviews (n=12 parents; Birthrights, 2022; Fivexmore, 2022; Hurst, 2001). Parents voiced being dismissed by staff about their labour progression (n=4), pain (n=6), their own health, or the infant's health (n=6 parents; Birthrights, 2022; Fivexmore, 2022; Hurst, 2001). In interviews and literature, ethnic minority women felt this dismissal by their HCPs led to adverse clinical situations arising and unnecessary pain for mothers (Fivexmore, 2022). During their interview, four HCPs also acknowledged a dismissive approach among co-workers.

> *"I'm not really listened to until there's a human head." (PA04 – Black)*

One parent recounted a particularly traumatic experience but had chosen to undertake the interview so that her voice was heard.

"They actively didn't want me to have a voice… It's good to have it [the interview] and it's I feel also that this is me being heard." (PA22 – Mixed)

A further mother talking about this issue, suggested she felt belittled by HCP and was treated like "a problem that they needed to deal with as quickly as possible." (PA20 – Black)

Fear of raising concerns

Mothers feared raising concerns to HCPs (n=1 parent; Hurst, 2001; Lee & Weiss, 2009; Tarnow-Mordi & Pickering, 1983). They worried raising their concerns would mean they would be labelled as "difficult" or "too much trouble" (Hurst, 2001; Tarnow-Mordi & Pickering, 1983).

"I can't raise any questions. I am also afraid of asking inappropriate questions. I don't want to upset the doctors and nurses." (Lee & Weiss, 2009)

Lack of consent and choice

Parents in the systematic review and within the interviews felt their consent was not requested for procedures such as giving medications, injections, or before operations (n= 4 parents; Birthrights, 2022; Fivexmore, 2022). Parents in the interviews suggested they felt coerced into consenting, or they were not given enough information to make an informed decision (n=3).

"She said she wanted to see how dilated I was, but also carried out a cervical stretch without my prior knowledge or permission." (Fivexmore, 2022)

"How can you ever possibly illicit informed consent for me….because I know that to get us onto labour ward that I have to allow you to do this [vaginal examination] so it's not proper consent then is it." (PA23 – Black)

"No one told me that I had a choice to be, to say I don't want to be examined or whatever." (PA13 – Asian)

Some parents felt their choices were not respected (n=3) and when concerns were voiced in the form of formal complaints, all participants who submitted complaints had no resolution (n=3).

System and organisational factors

The second theme identified was around structural and organisational barriers including accessibility, workforce issues, and logistical issues.

Accessibility

Parents (n=7) and HCPs (n=5) in the interviews reported issues with accessing appointments. Two mothers recounted instances of appointment mix-ups, with others reporting last-minute appointment notices or cancellations. The lack of access to community care was raised, which was felt particularly strongly by first-time mothers. HCPs recounted feeling that services were inflexible and not designed around the needs of women (for example due to childcare difficulties). They also talked about instances where parents hadn't received the required follow up as they were unaware that it was needed, as well as parents' difficulties in navigating online booking systems and lack of understanding of which HCP to contact when.

> *"I would have expected a bit more care, as in once you give birth. And like checking on the baby a lot more. And because I am a first-time mum, I seriously don't know what I'm doing." (PA01 – Other)*

The system was viewed by one parent to "be a system that is not supportive or promotes individual rights." (PA23 – Black). In contrast, another mother reported being explicitly told by the midwife in her second pregnancy about her rights and wondered whether that was due to a policy change or due to the midwife being Asian herself.

Workforce issues

Staffing

Staffing ratios were described as inadequate by mothers (n=4), HCPs (n=3) and within the review (Hurst, 2001; Witt et al., 2022). Parents identified a high staff turnover and staff suggested that their high workload meant they could not provide the expected level of care (Birthrights, 2022; Witt et al., 2022).

> *"I definitely don't think I should have been left alone for that for periods of time, especially during like the end of labour, where I'm about to actually push the baby out." (PA07 – Black)*

> *"Nowadays there's not enough of us to do what we're expected to do in the time available." (HV01 – White)*

One mother whose child was on NNU reported:

> *"Imagine that, I have never found a nurse there with him, looking after him. Every time I came, he was always alone. Always." (Witt et al., 2022)*

One midwife suggested the danger of overworked staff was that "under pressure staff are going to rely on their biases more often" (MW04 – Asian).

Some HCPs identified workforce diversity as a facilitator for positive care (n=8). Four parents also suggested having a Black, Asian, or minority ethnic HCP positively impacted the care they or their baby received. However, two mothers also reported that their care was negatively affected when cared for by an HCP from a similar background.

> *"I feel like he was expecting me to just take his word for it because obviously we're like, you know, similar backgrounds, whatever." (PA01 – Other)*

Staff attitude

Issues with staff attitude were addressed by HCPs within the interviews. They identified a lack of appetite for change (n=6) and the fact that everyone needs to be motivated for change to happen (n=3). Some HCPs also reported feeling dismissed by other staff whilst raising issues of ethnicity, which left them feeling attacked (n=5).

> *"There's a very real fear on both sides on marginalised members, staff and non-marginalised members of staff who fear having these open conversations within the workspace." (MW04 – Asian)*

Logistical issues

Parents in the review described logistical issues, particularly if they had an infant in NNU. Staff interviews and the systematic review identified issues that arose with transportation, childcare, and the cost of parking (HCPs (n=2); Witt et al., 2022).

> *"I used to pay 30, 31 dollars [on car parking]. It was hell, because I spent a lot more money. And it was a bit difficult." (Witt et al., 2022)*

Other logistical issues that were identified as barriers to receiving adequate care included the poor distribution of services (HCPs n=3; Birthrights, 2022). It was recognised that deprived areas, which were more likely to have an over-representation of ethnic minority families, were more likely to have high staff turnover within the service and for the services to generally received poorer funding.

"The more that the population's health needs are, the less likely they are to get it funded." (OB03 – Asian)

Additionally, both HCPs (n=2) and parents (n=1) talked about the impact of the NHS charging certain immigrants for care and the impact it had not only on those who were charged, but anyone without British citizenship. One woman recounted a shocking instance of having to prove their right to access NHS services when arriving at the hospital having a miscarriage.

"The kind of administrator wanted me to fill out the form to prove that I had access, I had the right to use the NHS … like seeing your passport number and of course in the moment of going to the hospital that wasn't my priority … [There's] this sense that you are not, you're not welcome, you don't belong, even in this moment of incredible vulnerability and pain, it's still more important for us to figure out whether you have a legal entitlement" (PA20 – Black)

Within the review, woman also talked about being told they had to pay for procedures upfront or threatened with deportation if they did not pay, which stopped women from trying to access healthcare even when problems arose (Birthrights, 2022).

"If you are not going to pay the money then probably you will be in trouble. Maybe the Home Office will come to know about this and you will have some problems, they will send you back." (Birthrights, 2022)

Discrimination

The theme of discrimination recurred throughout this study. Twelve parents reportedly felt their care did not differ due to their skin tone. However, many parents made mitigations to ensure their or their infants' care was not impacted. Three parents suggested treatment of those from ethnic minority backgrounds varied between institutions, with one father talking about undertaking background research of the hospitals in his area to ensure the one they chose for the birth meant their level of care was good. Another interviewee stated that they felt they did not experience any differences in care due to their ethnicity because they lived in a location which has "a diverse ethnic community" (PA21 – Other).

"My advice to people of ethnic minorities is that they should try as much as possible to know the type of hospital that their people are administered with here." (PA11 – Black)

Three specific areas of discrimination were identified including stereotypes, lack of cultural competence, and inadequate care.

Stereotyping

Lack of education

Two parents within the interviews felt that it was assumed that they were uneducated. A further mother felt that it was presumed that she didn't speak English. Both within the interviews and the review women talked about making sure HCPs knew their profession (Parents n=2, Fivexmore, 2022), with some women reporting preconceptions made by HCPs changed after they learned of their profession. Two women also felt that they benefited from sounding 'British', with one also having an English sounding name which meant on a "telephone conversation with someone, you probably wouldn't be able to identify my ethnicity based on how I sound." (PA23 – Black)

> *"I'd turn up in a tracksuit and be spoken to in a certain (dismissive) way until they learnt I was a lawyer…and they would be more respectful overall in my experience." (Fivexmore, 2022)*

> *"I will make it quite clear from the beginning that I work in safeguarding team, you're not going to necessarily be able to suggest certain things to me without me challenging that you know." (PA23 – Black)*

Assumptions were also made about the mother's lifestyle. If she attended appointments alone it was presumed the pregnancy was unplanned and if she attended with the father of the current pregnancy, it was assumed it was a different father to the previous pregnancy or the paternity of her child was questioned (Parent n=1, Fivexmore, 2022).

> *"First visit a nurse said she was shocked I knew who the father was. As people like me usually don't know." (Fivexmore, 2022)*

> *"There were questions about the paternity of my child. They wanted to know … if the dad was an African man like me, or you know if the dad was White." (PA06 – Black)*

Labelled as difficult

Interview HCPs identified that Black, Asian, or minority ethnic women were labelled by some co-workers as "aggressive" or "difficult" (n=4). One HCP recognised that Afro-Caribbean women, in particular, were treated differently and seen as difficult.

> *"Routinely Afro-Caribbean mothers are treated differently in delivery. Routinely, they're spoken to differently, routinely they're labelled as being aggressive. They're labelled as being difficult to the point when someone told me that a patient or mother was being difficult, I knew the patient was going*

to be Black because I knew that that was almost a synonym for, you know, 'this patient is being really difficult'." (NNP09 – Black)

Viewed as medically different

Concerns were expressed about assumptions made by HCPs that women of Black and minority ethnic descent were "high-risk bodies" (One parent; Birthrights, 2022), which led to their childbirth experience being "over-medicalised" due to their skin tone rather than individually ascertained risk.

"There's also this assumption that we're going to have these problems even if we don't." (Birthrights, 2022)

HCPs demonstrated unfounded racial assumptions and lack of knowledge about Black and Black mixed women's anatomy, suggesting that procedures in labour were affected due to their ethnicity. One Black woman was told her epidural had failed as the anaesthetic had to "work harder" as Black women had a bigger spine curvature and another was told she was "more stretchy" due to her ethnicity (Fivexmore, 2022).

"One midwife when doing the sweep said that the reason for dilation taking so long for me was 'probably due to an African pelvis'." (Black woman) (Fivexmore, 2022)

While women were incorrectly viewed as different in many areas, other areas of potential difficulty such as recognising sepsis due to women's skin tone were overlooked by HCPs (Birthrights, 2022).

Pain relief

An invisibility/hypervisibility paradox was noted for people from ethnic minority backgrounds, where they were presumed to be high risk but were still not listened to (Birthrights, 2022). Additionally, in both interviews and in the literature, pain thresholds were assumed to differ between women of different ethnicities with pain therefore not taken seriously (Parents n=6; HCPs n=6; Birthrights, 2022; Fivexmore, 2022). Parents and HCPs when interviewed suggested that HCPs made assumptions about the level of pain women were experiencing and how they should be managing it.

"I was in a lot of pain so I was denied any pain relief, uh, well they offered me paracetamol, but that was it and said that's all I could have." (PA22 – Mixed)

"Black and Asian women are kind of maybe ignored somewhat with regards to their care and they're seen as being strong and tough and have good pain thresholds." (MW05 – Mixed)

Lack of cultural competence

Another problem reported both within the HCP interviews and the review was the lack of cultural understanding, not offering women culturally tailored care or asking women about their culture (HCPs n=15; Birthrights, 2022; Lee & Weiss, 2009). Numerous cultural practices were mentioned by HCPs, most notably concerns about being mindful of consanguinity, co-sleeping, and certain ethnicities not giving colostrum to newborns. Additionally, societal differences within some cultures were suggested, for example that in some cultures doctors should not be challenged (n=1). Some HCPs suggested that the cultural or religious practices of parents are not always respected (n=2) and others raised concerns about how to manage certain cultural practices, for instance Female Genital Mutilation (FGM) and male circumcision (n=3). HCPs also talked about different community structures within other cultures for example women being submissive to their husbands or to elders within the family, which could impact on the choices they themselves could make about the care their neonate received.

> *"Women in other cultures have a bit of a submissive role… where you might say oh, can I do this? They'll say things like I have to ask my husband." (HV08 – Black)*

Professionals in the interview identified that sensitivity is needed around different cultures (n=2) and recognised the importance of cultural awareness (n=14), with the review also highlighting that practices considered normal within the UK are not always superior to other practices (Birthrights, 2022). The need for HCPs to understand a woman's culture to be better able to support her was also demonstrated within the review by the self-blame that some women felt after a preterm birth, for example due to not following traditional prenatal care practices.

> *"I remember we had, you know saying to a Jamaican mother, 'you're talking too loud on the NICU' and she said everyone in Jamaica talks." (NNP07 – White)*

> *"We just give people choices in a completely non-judgmental way and we just say look this is the situation we're in, it's nobody's fault, there are choices for you and your family going forward." (NNP04 – Mixed)*

> *"Western ideas are the best and other choices are inferior." (Birthrights, 2022)*

> *"I feel very guilty. I think if I had eaten traditional food (abstinence from cold food), my baby might not have any problems." (Lee & Weiss, 2009)*

Interestingly, few mothers within the interviews felt that their culture was not understood (n=3). Issues were raised by mothers regarding, judgment around

co-sleeping (n=2) and presumptions around religion (n=1). However, another woman also commented that she was not treated differently due to her culture as she no longer practices her culture in the UK.

Inadequate care

Racist microaggressions

One concern expressed within the review was the use of racial microaggressions by caregivers. For example, women's names not being pronounced correctly or declared to be too difficult to say (Birthrights, 2022; Fivexmore, 2022). This could have profound consequences with one woman having the wrong scan performed due to her name not being double-checked. The sonographer transferred the blame to the patient by informing them that their "name is difficult, that was why the problem happened." (Birthrights, 2022)

Within the interviews, seven parents felt they had been treated differently or not as 'nicely' as their White counterparts, with seven parents also reporting not being seen as quickly. One woman also noted that HCPs spoke differently to her White partner than to her. Two other parents also felt like they were being treated differently because of their ethnicity or skin tone but were not entirely sure.

> *"It's hard, I think when you are of colour. You sometimes don't know you, kind of you don't know if you are being treated different." (PA13 – Asian)*

Additionally, within the reviews, women reported racial discrimination and racist assumptions whilst on the postnatal wards, for example HCPs making comments such as "this is not how we do things over here" despite women being born in the UK (Fivexmore, 2022). Others described inappropriate comments made by HCPs, including remarks regarding the skin, hair, or eye colour of their neonates (Parents n=2; Birthrights, 2022; Fivexmore, 2022).

In one instance the woman herself had accusations of racism recorded in her notes due to the fact she requested a "Black midwife for the delivery of her baby" (Birthrights, 2022).

Lack of dignity and respect

Black, Asian, and minority ethnic parents in many studies included within the review and within the interviews (Parents n=9; Birthrights, 2022; Fivexmore, 2022; Witt et al., 2022) were more likely to describe receiving poor care from their HCP, believing it to be because of their ethnic or racial background or darker skin tone. Five HCPs were also aware that the complaints received from parents showed that discrimination was present. The poor care described included

instances of neglectful care, not being respected by their HCPs, and being left in vulnerable states and conditions. One mother described not feeling 'like a human being' due to the way she was cared for (PA20 – Black), with one obstetrician also very aware that women report a lack of dignity and respect in the care they received. Within the interviews and review, mother's poor experience with staff were noted to subsequently result in concern about the care of their children (Parents n=1; HCP n=2; Witt et al., 2022).

One mother described waiting for an HCP whilst "lying there with my legs and stirrups with my vagina uncovered and I wasn't able to hold my baby breastfeed etc while I was waiting." (PA20 – Black)

> *"The way in which women have been spoken to based on feedback that they've given afterwards, leading to a loss of dignity." (OB03 – Asian)*

> *"I listen to women and their experiences and what I hear, in terms of the difference of care that they receive and the babies receive, skin colour plays a part in that difference." (MW03 – Black)*

> *"If they treat me like this, then they will treat my daughter the same way in the other area [NNU]." (Witt et al., 2022)*

Within the reviews some women's traumatic experiences led them to engage lawyers. One lawyer described "very serious" incidences of "obstetric violence" towards women and instances of coercion and bullying (Birthrights, 2022).

> *"During my deliveries I have had midwives carry out procedures... I was left battered and bruised and no one did anything to protect me." (Birthrights, 2022)*

Contextual issues

Social isolation

Within the interviews and reviews (HCPs n=1; Lee & Weiss, 2009; Witt et al., 2022), parents of newborn babies suggested they felt isolated and in need of a support network due to living far away from their families or friends but reported being largely unaware of social resources and support groups. Within the interviews one woman noted the impact of Covid restrictions had made her feel isolated, while two others felt that HCPs seemed to rely on the woman having a good support network which must be difficult for women who don't have family around.

"I cried every day during the first week but then I could not cry anymore because my tears had dried out . . . there is really no one in here I can talk to." (Lee & Weiss, 2009)

Socio-economic status

Five HCPs were aware that Black and minority ethnic families were over-represented in deprived communities. They, therefore, were more likely to live in inadequate housing and have poor nutrition, with fewer local community resources available. The impact of deprivation on mental health was also highlighted by HCPs.

"Some families do make it and they settle in and they assimilate into the community. But there's plenty remain on the periphery. They've got high levels of poverty, living in poor accommodation, poor nutrition, poor understanding of the way that the systems work." (HV01 – White)

Racism in society

The care women received was frequently interpreted through the lens of the discrimination they faced in society in general. Four parents talked about being treated differently in society, including being called names and being dismissed when wearing traditional clothing. HCPs (n=4) also noted the prevalence of racism in society, including in school where Black and minority ethnic children were seen as having to work harder to achieve equivalent results. Furthermore, some HCPs talked about themselves or their friends not always disclosing their full ethnic identity to avoid discrimination.

Inequalities in health outcomes

Parents (n=8) and HCPs (n=12) acknowledged the prevalence of ethnic inequalities in healthcare outcomes, particularly around maternal and infant mortality and stillbirths, but also wider inequalities highlighted by the Covid pandemic. One parent was also aware that her ethnicity was linked to higher risk of gestational diabetes. One father stated that "we live in a society where most minority people tend to get low-quality care." (PA11 – Black). Additionally, two HCPs and one mother who had had an infant admitted to NNU noted that there were more ethnic minority than White neonates admitted. One HCP and the mother both directly attributed this to inadequate care during pregnancy or labour.

"That's part of our role to try and reduce inequalities, but it's very difficult when you're set up in a system that promotes these inequalities." (HV4 – White)

"If mums are not being adequately supported and cared for in their pregnancy in their birth period, then of course that's going to directly impact the chances of the neonate being admitted." (MW04 – Asian)

Impact of racism on care

The concern over racist behaviour impacted parents' choices, behaviour, and feelings during their maternity care (n=10). Women's experiences of racism and discrimination during birth also negatively impacted their desire to have more children or led to a distrust of NHS services (Parents n=2; Birthrights, 2022; Fivexmore, 2022).

"My whole birthing experience and aftercare has put me off having another child." (Fivexmore, 2022)

Mothers described changing their birth preferences due to concerns over hospital, staffing, and racism. Two interviewees decided to have a homebirth to mitigate concerns over discrimination or due to previous negative birth experiences, with one of these women also feeling the need to employ a doula. As described above, one father viewed three hospitals before deciding on an appropriate institution and another parent studied the local hospital ratings.

"I knew that there are you know higher rates of maternal and infant morbidity and mortality and so my whole intention of choosing homebirth was trying to avoid that." (PA20 – Black)

Another interviewee reported that Black mothers are fearful to give birth in the UK and have more stressful births meaning "some Black women wish to go back to their homeland for the birth because they felt they are more understood there." (PA19 – Black)

Parents also voiced concerns over HCPs, reporting safeguarding concerns over their children (n=2) one mother suggested it "felt like there was an extra layer of fear that they would report what I said or I'd be marked as not adhering to medical advice and that could lead to further interventions" (PA20 – Black). Others worried about their child's birthmarks as a potential safeguarding concern and insisted they were recorded by every HCP at each visit. One mother verbalised that the NHS is struggling for everyone at the moment, but when the additional layer of discrimination is added, that it is when in becomes riskier for Black, Asian, and minority ethnic families.

"I think the NHS is just purely in firefighting mode and they're waiting for things to get really bad before they intervene rather than doing preventive stuff, which is where it gets more expensive and more difficult and more risky and dangerous for people. And then potentially more risky, dangerous for not White people, if there are prejudices in there as well." (PA22 – Mixed)

Summary

The study involved a systematic review of literature on the experiences of parents when accessing healthcare, as well as interviews with healthcare professionals and parents or carers to identify challenges and barriers to accessing or receiving care, areas of good practice and any instances of discrimination or inequalities within care. The results of both the systematic review and interviews were combined and analysed together.

The first theme identified was the communication barrier, where women faced problems in getting information from HCPs due to language differences or inadequate translation. Another issue was women being silenced within the system, where they felt dismissed, ignored, or belittled by HCPs. Fear of raising concerns was another reason why women remained silent, as they worried about being labelled as difficult or too much trouble.

The literature and interviews revealed discrimination in healthcare. Three areas of discrimination were identified, including stereotypes, lack of cultural competence and inadequate care. Some women reported assumptions about their education and lifestyle, while some HCPs labelled Black, Asian, or minority ethnic women as "aggressive" or "difficult". It was suggested that HCPs often made unfounded racial assumptions about Black and minority ethnic women's bodies and lacked cultural competence, resulting in over-medicalization of childbirth and inadequate pain management. HCPs were also shown to overlook potential difficulties due to women's skin tone and failed to offer culturally tailored care. While few parents within the interviews reported feeling that their culture was not understood, the need for HCPs to understand a woman's culture to better support her was recognised.

Inadequate care towards Black, Asian, and minority ethnic parents was reported in multiple studies and the interviews. Racist microaggressions, including mispronouncing names and discriminatory comments, were reported by both parents and within the studies included in the review. Lack of dignity and respect, neglectful care and being left in vulnerable states were also commonly reported. Some mothers even engaged lawyers due to traumatic experiences and obstetric violence. HCPs noted that discrimination was present in complaints received from parents.

System and organisational factors were also identified. These included difficulties around accessing appointments due to lack of knowledge of who and when to contact and a lack of flexibility within the system. Staffing was also raised as an issue, with workload pressures and lack of training noted to make discrimination more likely as overworked staff have to rely upon their biases more. Charging for care to some migrants depending on status was also seen as an organisational barrier.

Social isolation, socio-economic status, and racism in society were identified as contextual issues impacting maternity care, as reported by both parents and healthcare professionals. These issues included a lack of support network, inadequate housing, poor nutrition, mental health issues, and the racism they faced in general in society. There are inequalities in health outcomes, particularly for those from ethnic minority backgrounds, with racism affecting parents' choices and feelings about maternity care. Some parents changed their birth preferences due to concerns over discrimination and the additional risks for Black, Asian, and minority ethnic families.

Overall, this highlights the need for HCPs to be more culturally competent and provide personalised culturally safe care, as well as the need for more support for parents from seldom-heard communities. Addressing discrimination in healthcare, society, and decolonising the curriculum, practice and policy is crucial for improving maternal and neonatal health outcomes, ensuring that all parents receive respectful and dignified care.

Discussion

Current UK policy, guidelines, and training resources regarding neonatal testing do not aptly address differences due to skin pigmentation. Yet for many neonatal assessments, observations of an infant's skin colour are still mentioned as signs of certain conditions. An implication here is the possibility that assumptions made by skin colour assessment are only relevant to those with White skin and may not be inclusive or representative of the diverse communities that make up our multi-ethnic and racially diverse society. The subjectivity of skin colour assessment was highlighted within only three out of the 18 reviewed policies as well as in the training resource, which indicated that skin colour assessment should not be used in isolation to assess a neonate. However, for some conditions, a visual assessment by healthcare professionals is still the first screening step and therefore relied upon before further testing, which may disadvantage Black, Asian, and minority ethnic neonates (Keren et al., 2009; Riskin et al., 2008). Three major neonatal assessments in which skin colour was a key element of screening - Apgar score, cyanosis and jaundice - have been the subject of our study and the results are discussed below.

Apgar score

The Apgar score was devised as a means of rapid scoring of an infant (Apgar, 1952). However, the subjectivity of many components within the score was highlighted within this study (Blake, 2010). The appearance component that assesses the colour of a neonate's skin has long been called into question, with appearance being the component least correlated with cord pH, arterial carbon dioxide (pCO2) and base excess (Crawford et al., 1973). Additionally, others have only found the heart rate and respiratory effort components to be predictive of neonatal morbidity and mortality (Ensing et al., 2015).

Differences have been shown in how the Apgar score is rated both between professional groups such as midwives, neonatologists and obstetricians; and between hospitals (Arri et al., 2018). Others have also found differences in Apgar scores between different European countries., This suggests differences in clinical training and convention when scoring the Apgar between countries (Siddiqui et al., 2017).

It has particularly been suggested that the terminology "pink" causes confusion to staff (Blake, 2010; Adams et al., 2014; Chubb et al., 2022), particularly the

terms “pink all over” or “all pink”. Indeed, within the guidelines and training resource reviewed, no consideration was given to how “pink” may be assessed in Black, Asian, and ethnic minority neonates (HEE, 2022; NICE, 2017; NNA/ University of Hertfordshire, N.D.a) and the majority of HCP interviewees described a lack of training around how the assessment of colour may differ in neonates from ethnic minority backgrounds. As a result of this lack of clarity, many HCPs developed their own way of assessing for “pink”, without it always being evidence based.

Within our interviews, and in a previous survey of midwives within one trust, most HCPs and parents felt that the term “pink” was irrelevant to Black, Asian, and minority ethnic neonates. There was also uncertainty over the relevance of the term “blue” which is also used within the appearance component of the Apgar score. Alternatives were suggested such as replacing the term pink with “normal colour” or “good colour”, although issues would still remain over how this would be assessed in Black, Asian, or minority ethnic neonates.

Another alternative suggestion was to look at whether the neonate was “well perfused”, however perfusion and cyanosis are not synonymous with it being possible to be cyanosed while also well perfused. Assessing pink in alternative locations such as the mucous membranes was also suggested, however the evidence base around this, especially in those from ethnic minority backgrounds, is lacking. Additional suggestions included using the same ABC (airway, breathing, circulation) approach as for any other resuscitation. The unique situation of resuscitation of a newborn is however noted, due to both the potential for an infant to have had a prolonged hypoxic insult prior to birth and the subsequent resuscitation event (Douthwaite & Kennea, 2021), as well as due to the extra-uterine adaptations that the neonate is undergoing to its cardiovascular system as the lungs become functional in the immediate period after birth (Michaelides, 2017).

Within the literature, alternatives to the Apgar score have also been suggested. One alternative that has been tested is the Neonatal Resuscitation Assessment and Adaptation Score (NRAS). Like the Apgar score, the NRAS has five components that are scored between zero and two, but the NRAS does not consider skin colour (Witcher et al., 2018; Villota et al., 2021). Pilot studies have suggested that the NRAS correlates well with the Apgar score (Jurdi et al., 2015; Witcher et al., 2018; Elglil et al., 2020; Villota et al., 2021). While the score has been tested in small samples within several populations, no studies have considered the predictiveness within neonates from different ethnicities and the developers note that it is too early to recommend the wider use of the NRAS. Additional alternatives have also been proposed such as the Expanded Apgar whose use is encouraged by the American College of Obstetricians and Gynaecologists in infants requiring resuscitation (Watterberg et al., 2015). A systematic review of existing literature comparing Apgar score alternatives with Apgar is warranted.

Cyanosis

Guidelines and training resources suggest observing a neonate for change in colour to identify cyanosis, respiratory distress, or heart abnormalities (HEE, 2022; RCN, 2017, OHID, 2021; NNA/University of Hertfordshire, N.D.b; NICE, 2021). However, it is noted that colour change may not be as apparent in infants with darker skin (Goldman, 1973; Schott & Henley, 2000). Therefore, it may remain unobserved by parents and professionals, with the deterioration of a Black, Asian, or minority ethnic neonate subsequently identified later in the absence of a pulse oximeter. Some of the guidelines and training resources reviewed suggested colour was not an appropriate assessment tool to assess oxygenation in isolation (HEE, 2022; Resuscitation council, 2011), with the term “going blue” also seen as inappropriate when assessing cyanosis in Black, Asian, and ethnic minority neonates necessitating new parameters and guidance for this group (NHS RHO, 2021b). Visual assessment of cyanosis should therefore only be used in the absence of pulse oximetry, particularly in any neonatal resuscitation scenario (Resuscitation council, 2011). Tongue colour was suggested in one small study to be a good indication in the delivery room of the need for supplemental oxygen, regardless of ethnicity (Dawson et al., 2015). However, areas around the mouth and lips provide more immediate access and may therefore be more desirable. This was also suggested within the HCP interviews, where they identified the mouth mucosa and lips as the best place to assess for cyanosis in infants of all ethnicities when pulse oximetry was not available.

Interviewed HCPs suggested the need for colour assessment decreased with the increased use and accessibility of pulse oximetry, which is commonly relied upon in practice. Inequalities highlighted by the COVID-19 pandemic however suggested a bias in pulse oximetry in Black, Asian or minority ethnic adults (Cabanas et al., 2022; NHS RHO, 2021b; Sjoding et al., 2020). One recent study has also suggested that pulse oximetry may not be as accurate in Black and minority ethnic preterm neonates with more incidences of occult hypoxaemia (Vesoulis et al., 2022). While pulse oximetry was still more accurate than visual inspection, the authors suggested the avoidance of saturations at the lower end of the normal range in Black preterm neonates to reduce the risk of adverse outcomes within that group. Further research to examine the small but potentially clinically significant differences in how pulse oximeters might operate in neonates from diverse races and ethnicities is warranted.

Jaundice

The detection of jaundice in neonates is primarily done via visual assessment, which has been recognised to be subjective, particularly in infants with Black or dark skin (Dionis et al., 2021; HSIB, 2023). The National Institute for Health and Care Excellence (NICE) guidance has stated that the visual detection of jaundice in darker skin tones is “almost impossible” (NICE, 2010). This issue is of particular importance as ethnic minority neonates are at an increased risk of jaundice (Slusher et al., 2017) and are over-represented in those requiring admission to neonatal units for jaundice (Kapadia et al., 2022). As such, those at the highest risk of developing jaundice are the most likely to be missed by means of visual assessment.

The literature shows that there is no consistent pattern between transcutaneous bilirubinometry (TCB) over or underestimating results in different ethnicities, with a mean bias for TCB-serum bilirubin (SBr) within ±40μmol/l in every study, regardless of the ethnicity subgroup examined. Current guidelines within the UK vary in their recommendations for when a TCB reading should be followed up by an SBr. Therefore, a minimum threshold of 40μmol/l could be argued to ensure adequate identification of infants of all ethnicities as the mean bias was less than this in all included studies. Additionally, HCPs should be able to refer for an SBr regardless of TCB reading if clinically concerned, given the tendency for the results to be more variable in those with darker skin.

Interviewed HCPs acknowledged the cost implications for their trusts with wider distribution of TCB devices, but also voiced that they still could be cost-effective compared to missing cases. Moreover, it was highlighted within the HSIB report that compensatory cases of harm to babies as a result of jaundice over the last 10 years was at a cost to the NHS of approximately £150.5 million, with costs likely to continue to rise (Rennie et al., 2019; HSIB, 2023). Other studies have suggested that the use of TCB reduces the cost and the associated pain for the neonate of SBr heel prick tests (McClean et al., 2018; Paul et al., 2021) and reduces the time taken by nurses, as well as providing an immediate result (McClean et al., 2018).

Regarding parental detection of jaundice, few HCPs felt that yellow skin was identifiable in darker skin tones, predominantly suggesting that yellow could be assessed in the infant’s eyes. While some parents have reported that they have recognised changes in their infant’s skin colour before HCPs, it would also be pertinent to better educate and equip parents around the detection of jaundice through other methods such as looking at the sclera or gums.

Based on the results of this study and previous research, it is recommended that NICE reviews and updates its guidance if required regarding the reliability of visual assessment of jaundice particularly in Black and ethnic minority neonates, as well as ensuring all risk factors for jaundice are clearly identified. Additionally,

given the ready accessibility of mobile phone technology across most of the globe, further investment in developing applications that could accurately detect neonates of all ethnicities that require additional monitoring for jaundice could be beneficial.

The impact of ethnicity on care

Racial discrimination is unlawful. The Equality Act 2010 states people must not be discriminated against for their race, colour, nationality, ethnic origin and ethnic or racial group (Equality and Human Rights Commission, 2020). Under the act, the 'general equality duty' was established to impose a duty on public authorities to give due consideration to equality (Equality and Human Rights Commission, 2023). Facilitators and barriers to care for Black and minority ethnic neonates are well documented and understood. The public sector should therefore aim to mitigate discrimination, which should be reflected in policy design and development but also NHS practice and education.

Women from minority ethnic backgrounds are known to have a poorer experience of maternity services (Henderson et al., 2013), with Black parents reporting more dissatisfaction with neonatal care than their White counterparts (Martin et al., 2016). As identified by HCPs and parents in interviews, accessing care for those from ethnic minority backgrounds may be hindered for a multitude of reasons. This is reflected in previous literature, with known disparities in timely access to healthcare, the number of antenatal appointments, ultrasound scans, and pain relief provided in labour (Henderson et al., 2013). Barriers to care have previously been detailed (Fair et al., 2020), with the interviews within the current study further highlighting the urgent need to put mitigations in place such as improving communication, listening to mothers and therefore involvement in decisions and obtaining truly informed consent, as well as supporting parents' understanding of care provision.

The majority of HCPs within this study stated that they had received no additional training in identifying jaundice, cyanosis, or undertaking an Apgar score in Black and minority ethnic neonates. Indeed, a recent survey has similarly reported that training in identifying clinical signs in Black and Asian skin was undertaken in less than one third of UK maternity service providers (Ledger et al., 2021). While training was more prevalent in areas of high ethnic diversity, even within these areas only just over half of services offered this type of training (Ledger et al., 2021).

Urgent staff training on identifying common conditions in those from Black, Asian, and minority ethnic backgrounds is required to ensure disparities and inadequate care of parents and infants no longer occurs. This is particularly pertinent given that some HCPs within the interviews had never considered

how some assessment may need to differ to accurately identify conditions in ethnic minority neonates. Perceived capability to detect jaundice has previously been shown not to be associated with actual skill (van der Geest et al., 2021), highlighting the need to train all HCPs regardless of self-perceived competence. Additionally, there is a call for more training around culturally competent care (Fair et al., 2021), with a recent Mind the Gap report into maternity staff training calling for this to be mandatory for all maternity staff in the UK (Ledger et al., 2021).

Given the concerns raised over lack of training in assessing the skin in Black, Asian, and minority ethnic neonates such as colour changes related to cyanosis, jaundice or assessing the appearance component of the Apgar score, HCPs called for access to a database of relevant images. They wanted these images to depict conditions such as cyanosis and sepsis in a range of skin tones for training purposes and to enable effective identification.

The Healthcare Safety Investigation Branch has similarly called for an open-access bank of resources for both HCPs and the public to include images of medical conditions in a range of skin tones (HSIB, 2021). The requirement for appropriate resources for parents was also further highlighted within this study by the general lack of knowledge of how to identify jaundice other than by skin colour in the majority of parents, despite them being a highly educated sample. Many of the parents who were aware of other signs of jaundice reported only being so as a result of their child having had jaundice, rather than knowing beforehand.

Textbooks and training have often been modelled on White skin in the UK, as well as across the globe, with little questioning (Rimmer, 2020; Schott & Henley, 2000). Gaps in healthcare professionals' knowledge of dermatological assessment in different skin tones has especially been identified due to the limited diversity of publications and educational resources (Shah et al., 2022; Wilson et al., 2021). In acknowledgement of the fact that many textbooks largely contain pictures of White patients, Mukwende et al., (2021) established Minding the Gap – a handbook of clinical signs in Black and brown skin. In addition, Don't forget the Bubbles and the Royal London Hospital are also developing an online bank of images as part of their skin deep project. While these advancements are welcomed, further decolonisation of training, curriculum, and resources for the neonatal period are still required to provide culturally safe care.

Several HCPs reported being uncertain over the effectiveness of medical equipment for neonates with darker skin tones. The evidence base used to support the development of neonatal assessment guidelines therefore needs to clearly identify the ethnicity of participants within all identified studies. This will enable HCPs to be reassured when adequate evidence is available on neonates of all ethnicities, as well as allow better identification of the gaps where the impact of ethnicity is not currently known. This will lead to increased transparency regarding the relevance of the current evidence to all ethnicities.

Conclusions

Significant concerns have been raised within this review around the assessment of skin colour in Black, Asian, and minority ethnic neonates. The use of terminology in current neonatal assessment criteria such as “pink all over” within Apgar scores was no longer seen to be relevant in a diverse society. Similarly, limited reliability of visual assessment for cyanosis and jaundice was highlighted. Looking for colour changes in locations such as around the lips, mouth and the tongue were highlighted as alternative methods for detecting cyanosis alongside increased use of pulse oximetry. For jaundice it was suggested that HCPs assess for signs of yellow in the sclera and gums in Black, Asian, and minority ethnic babies. Given the inadequacy of assessing jaundice levels visually, any baby showing signs of jaundice should be assessed with a transcutaneous bilirubinometer or have a serum bilirubin blood test. Lack of appropriate training for healthcare professionals and parents in identifying the above conditions within the constructs of our diverse and multi-ethnic populations were identified.

Recommendations

Recommendations for practice

1. Given poor visual detection of jaundice and cyanosis, particularly in Black and darker skin toned neonates, the following recommendations are made:

    - Jaundice: Exploration of wider availability and use of bilirubinometers is recommended to decrease health inequalities and ensure safe care for all. Collaboration with organisations such as BAPM, RCM and RCPCH, RCOG, iHV, CQC and other key stakeholders is recommended.

    - Cyanosis: Healthcare organisations should strongly consider use of pulse oximetry screening if there is any indication of concern over oxygenation. For this reason, the UK National Screening Committee should also strongly consider including routine pulse oximetry screening as a requirement within NIPE (the Newborn and Infant Physical Examination) to mitigate the health disadvantages experienced by those with darker skin tones.

2. NHS England to create a national data bank of open access images of Black, Asian and ethnic minority neonates to incorporate into training and education of HCPs and healthcare students, as well as to aid diagnosis in practice. These images should also be available for use in accessible resources for families. These should incorporate images of healthy neonates, as well as specific conditions such as cyanosis and jaundice, and other conditions suggested by participants, including skin rashes and sepsis. Hard copies and digital sources should be provided, including on platforms such as the online NHS Health A-Z pages. These images should be made available and used by all NHS provider organisations and the NHS should create an opportunity for people to upload pictures and share experiences of conditions in different skin tones.

3. There is an urgent need for regular education and training for healthcare professionals and healthcare students on undertaking clinical assessments on neonates from Black, Asian, or minority ethnic backgrounds, including within the yearly updates on neonatal resuscitation. Better education for families is also required. All training requires process evaluation to ensure effectiveness. Further action on training and education should be taken as follows:

- In order to remind learners of the challenges that varying skin tones may introduce when assessing clinical signs, it is important for resuscitation dolls in all Higher Education Institutions and NHS Trusts to include babies with Black or dark skin.

- Professional associations (including BAPM, iHV, NNA, RCOG, RCPCH, RCM, RCN) and regulatory bodies (NMC and GMC) should identify training requirements to ensure that those they represent are fully competent in assessments that include skin colour for Black, Asian, and ethnic minority neonates. In addition, they should ensure all HCPs and students are made aware that some ethnicities are at higher risk of neonatal jaundice.

- All healthcare students should have access to an actual or simulated placement to increase students' awareness, knowledge, and confidence in assessing Black, Asian, and minority ethnic neonates prior to qualification.

- An urgent review and update of written and digital materials provided to parents is required. These should be co-developed with parents or carers from diverse ethnicities. A particular focus should be around jaundice and detecting a deteriorating infant to ensure accessibility and relevance of pictures and information to those from Black, Asian, and minority ethnic backgrounds.

- Both community and hospital staff should receive training on anti-racist practice, alongside culturally safe, compassionate care, with good listening skills to meet the needs of our diverse, multi-ethnic society and to reduce current health inequalities.

Recommendations for policy

4. Guidelines that refer to neonatal assessment by skin colour should be immediately reviewed and updated to highlight the impact of race and ethnicity (BAPM, HEE, iHV, NICE, NNA, OHID, RCM, RCN, RCOG, RCPCH, Resuscitation Council, WHO). This should include guidelines around general care of the newborn, as well as specifically for jaundice, cyanosis, and Apgar scoring. Further action should be taken in the following areas:

    - Attention should be drawn to any potential differences in assessment techniques for Black, Asian, and ethnic minority neonates.

    - Guidelines or educational materials that currently refer to the terms pink, blue, or 'normal' colour need to detail how this would be assessed more objectively in Black and darker skinned neonates.

- All guidance needs to highlight the limitations of visual assessment of the skin, particularly in those from ethnic minority backgrounds. A comprehensive assessment of other areas including sclerae and gums for jaundice and mouth mucosa for cyanosis is recommended. If jaundice is suspected, additional TCB or SBr should be undertaken.

- When screening tools such as pulse oximeters or bilirubinometers are advised, rather than clinical visual inspection, particular attention should be given to any differences in reliability or accuracy for neonates with different skin pigmentation. When advising the use of a bilirubinometer, a minimum threshold below the treatment line when a confirmatory SBr test is required needs to be urgently established to ensure appropriate detection of jaundice regardless of ethnicity.

- All guidelines should contain or signpost to images and good descriptors of skin assessments in those of all skin tones to support recognition of conditions in those who do not regularly have exposure to Black, Asian, and minority ethnic neonates.

Recommendations for research

Regarding the Apgar score

5. The wording 'pink all over' was not considered appropriate within the UK by the majority of HCPs and parents, including for White neonates. Further exploration is required, including:

    - Determination of inter-rater reliability between HCPs when undertaking the Apgar score in Black, Asian, and minority ethnic neonates. In particular, a better understanding of how HCPs determine the Apgar score in darker skinned babies is required, with specific attention to the terminology of 'pink all over' and the location of pink areas.

    - A systematic review is required of alternative scores or assessments such as the Neonatal Resuscitation and Adaptation Score and Expanded Apgar, compared to the Apgar score including their reliability and validity in Black, Asian, and minority ethnic neonates.

Regarding jaundice

Research required around jaundice includes:

6. In depth case study/root cause analysis of severe cases of jaundice, including those admitted to neonatal units for intensive phototherapy, requiring an exchange transfusion, or with resultant kernicterus. Clear consideration of confounding factors is required, as well as timing and content of contact with HCPs. This could inform better identification, prevention, and timely management of severe jaundice.

7. A better understanding of what level jaundice is by the time it reaches the eyes or gums is required given concerns voiced by HCPs that the areas we currently look at to identify jaundice in Black or darker skinned neonates actually identify jaundice at an advanced stage.

8. Determining inter-rater variability in assessing jaundice to evaluate the impact of skin tone and race on HCP ability to detect jaundice.

9. Current guidelines within the UK vary in their recommendations for when a TCB reading should be followed up by a SBr. Further investigation and confirmation of a more accurate and consistent threshold for adequate identification of infants, particularly for those of varied race and ethnicities is required.

10. Given the wide availability of mobile phone technology, further development and exploration of mobile applications for detecting neonatal jaundice across all ethnic backgrounds is warranted.

Regarding cyanosis

11. In light of inequalities highlighted by the COVID-19 pandemic in adult studies, further UK based research to examine the small but potentially clinically significant differences in arterial oxygen saturation compared to pulse oximetry saturation in neonates from diverse race and ethnicity backgrounds is warranted.

References

Adams, B.N., Grunebaum, A. (2014). Does “pink all over” accurately describe an Apgar color score of 2 in newborns of color? Obstetrics & Gynecology, 123, 36S.

Afanetti, M., dit Trolli, S.E., Yousef, N., Jrad, I., Mokhtari, M. (2014). Transcutaneous bilirubinometry is not influenced by term or skin color in neonates. Early Human Development, 90(8), 417-420.

Ahmed, M., Mostafa, S., Fisher, G., Reynolds, T.M. (2010). Comparison between transcutaneous bilirubinometry and total serum bilirubin measurements in preterm infants <35 weeks gestation. Annals of clinical biochemistry, 47(1), 72-77.

Apgar, V. (1952). A proposal for a new method of evaluation of the newborn. Classic Papers in Critical Care, 32(449), 97.

Arri, S.J., Bucher, H.U., Merlini, M., Fauchère, J.C. (2018). Inter-Observer Variability of the Apgar Score of Preterm Infants between Neonatologists, Obstetricians and Midwives. Journal of Neonatology and Clinical Pediatrics, 5, 024.

Aune, A., Vartdal, G., Bergseng, H., Randeberg, LL., Darj, E. (2020). Bilirubin estimates from smartphone images of newborn infants’ skin correlated highly to serum bilirubin levels. Acta Paediatrica, 109(12), 2532-2538.

Battersby, C., Michaelides, S., Upton, M., Rennie, J.M. (2017). Term admissions to neonatal units in England: a role for transitional care? A retrospective cohort study. BMJ Open, 7(5), e016050.

Bhutani, V.K., Gourley, G.R., Adler, S., Kreamer, B., Dalin, C., Johnson, L.H. (2000). Noninvasive measurement of total serum bilirubin in a multiracial predischarge newborn population to assess the risk of severe hyperbilirubinemia. Pediatrics, 106(2), e17-e17.

Birthrights. (2022). Systemic racism, not broken bodies. An inquiry into racial injustice and human rights in UK maternity care. Retrieved from: https://www.birthrights.org.uk/campaigns-research/racial-injustice/ Accessed 12th Jan 2023.

Blake, D. (2010). Do we assess ‘colour appropriately using the Apgar score? Journal of Neonatal Nursing, 16(4), 184-187.

Bourchier, D., Cull, A.B., Oettli, P.E. (1987). Transcutaneous bilirubinometry: 22 months experience at Waikato Women's Hospital. The New Zealand Medical Journal, 100(832), 599-600.

British Association of Perinatal Medicine (BAPM). (2015). Newborn Early Warning Trigger & Track (NEWTT) – a Framework for Practice (2015). British Association of Perinatal Medicine: London. Retrieved from https://www.bapm.org/resources/38-newborn-early-warning-trigger-track-newtt-a-framework-for-practice-2015 Accessed 12th Jan 2023.

British Association of Perinatal Medicine (BAPM). (2022a). DRAFT framework: Deterioration of the Newborn (NEWTT 2). British Association of Perinatal Medicine: London. Retrieved from https://www.bapm.org/resources/deterioration-of-the-newborn-newtt-2-a-draft-framework-for-practice Accessed 12th Jan 2023.

British Association of Perinatal Medicine (BAPM) (2022b). DRAFT Framework: Early Postnatal Care of the Moderate-Late Preterm Infant. British Association of Perinatal Medicine: London. Retrieved from https://www.bapm.org/resources/framework-early-postnatal-care-of-the-moderate-late-preterm-infant Accessed 12th Jan 2023.

Brits, H., Adendorff, J., Huisamen, D., Beukes, D., Botha, K., Herbst, H., Joubert, G. (2018). The prevalence of neonatal jaundice and risk factors in healthy term neonates at National District Hospital in Bloemfontein. African Journal of Primary Health Care and Family Medicine, 10(1), 1-6.

Brouwers, M.C., Kho, M.E., Browman, G.P., Burgers, J.S., Cluzeau, F., Feder, G., & Zitzelsberger, L., for the AGREE Next Steps Consortium. (2017 update). AGREE II: Advancing guideline development, reporting and evaluation in healthcare. Retrieved from: https://www.agreetrust.org Accessed 12th Jan 2023.

Brown, S., Liyanage, S., Mikrou, P., Singh, A., Ewer, A. K. (2020). Newborn pulse oximetry screening in the UK: a 2020 survey. The Lancet, 396(10255), 881.

Brucker, M.C., Macmullen, N.J. (1987). Neonatal jaundice in the home: assessment with a noninvasive device. Journal of Obstetric, Gynecologic & Neonatal Nursing, 16(5), 355-358.

Cabanas, A.M., Fuentes-Guajardo, M., Latorre, K., León, D., Martin-Escudero, P. (2022). Skin Pigmentation Influence on Pulse Oximetry Accuracy: A Systematic Review and Bibliometric Analysis. Sensors, 22(9), 3402.

Campbell, D.M., Danayan, K.C., McGovern, V., Cheema, S., Stade, B., Sgro, M. (2011). Transcutaneous bilirubin measurement at the time of hospital discharge in a multiethnic newborn population. Paediatrics & Child Health, 16(3), 141-145.

Casey, R., Rieckhoff, M., Beebe, S.A., Pinto-Martin, J. (1992). Obstetric and perinatal events: The accuracy of maternal report. Clinical Pediatrics, 31(4), 200-204.

Chubb, B., Cockings, R., Valentine, J., Symonds, E., Heaslip, V. (2022). Does training affect understanding of implicit bias and care of Black, Asian, and minority ethnic babies? British Journal of Midwifery, 30(3), 130-135.

Committee on Fetus and Newborn. (2006). ACOG Committee Opinion. Number 333, May 2006 (replaces No. 174, July 1996): The Apgar score. Obstetrics and Gynecology, 107(5), 1209–1212. https://doi.org/10.1097/00006250-200605000-00051

Crawford, J.S., Davies P., Pearson, J.F. (1973). Significance of the individual components of the Apgar score. British Journal of Anaesthesia, 45(2), 148-158.

Cruz, E., Higginbottom, G.M.A. (2013). The use of focused ethnography in nursing research. Nurse Researcher, 20(4), 36-43.

Dai, J., Parry, D.M., Krahn, J. (1997). Transcutaneous bilirubinometry: its role in the assessment of neonatal jaundice. Clinical Biochemistry, 30(1), 1-9.

Dawson, J.A., Ekström, A., Frisk, C., Thio, M., Roehr, C.C., Kamlin, C.O.F., ... & Giraffe Study Group. (2015). Assessing the tongue colour of newly born infants may help to predict the need for supplemental oxygen in the delivery room. Acta Paediatrica, 104(4), 356-359.

Dermott, K., Bick, D., Norman, R., Ritchie, G., Turnbull, N., Adams, C., … Taylor, C. (2006). Clinical Guidelines and Evidence Review for Post Natal Care: Routine Post Natal Care of Recently Delivered Women and Their Babies. National Collaborating Centre For Primary Care And Royal College Of General Practitioners: London.

Dionis, I., Chillo, O., Bwire, G.M., Ulomi, C., Kilonzi, M., Balandya, E. (2021). Reliability of visual assessment of neonatal jaundice among neonates of Black descent: A cross-sectional study from Tanzania. BMC Pediatrics, 21(1), 1-383. doi:10.1186/s12887-021-02859-x

Douthwaite, A., Kennea, N. (2021). Resuscitation of the newborn. Obstetrics, Gynaecology & Reproductive Medicine, 31(4), 95-102.

Draper, E.S., Gallimore, I.D., Smith, L.K., Matthews, R.J., Fenton, A.C., Kurinczuk, J.J., … & Manktelow, B.N. (2022). MBRRACE-UK Perinatal Mortality Surveillance Report, UK Perinatal Deaths for Births from January to December 2020. The Infant Mortality and Morbidity Studies, Department of Health Sciences, University of Leicester: Leicester.

Ebbesen, F., Rasmussen, L.M., Wimberley, P.D. (2002). A new transcutaneous bilirubinometer, BiliCheck, used in the neonatal intensive care unit and the maternity ward. Acta Paediatrica, 91(2), 203-211

Ebbesen, F., Vandborg, P.K., Trydal, T. (2012). Comparison of the transcutaneous bilirubinometers BiliCheck and Minolta JM-103 in preterm neonates. Acta Paediatrica, 101(11), 1128-1133.

Edwards, S.E., Wheatley, C., Sutherland, M., Class, Q.A. (2023). Associations between provider-assigned Apgar score and neonatal race. American Journal of Obstetrics & Gynecology, 228(2), 229.e1-9.

Elglil, A.M.A., Ibrahim, A.M., Shihab, N.S., El Mashad, A.E.R.M. (2020). Study of the Neonatal Resuscitation and Adaptation Score (NRAS) Compared to the Apgar Score in Neonatal Resuscitation. Journal of Advances in Medicine and Medical Research, 32(24), 100-110.

Ensing, S., Engelen, M.C., Tamminga, P., Ravelli, A.C.J., Mol, B.W.J., Abu-Hanna, A. (2015). Risk factors and prognostic models for perinatal asphyxia at term. (Doctoral Thesis). University of Amsterdam Digital Academic Repository: Amsterdam.

Equality and Human Rights Commission. (2020). Race discrimination. Retrieved from https://www.equalityhumanrights.com/en/advice-and-guidance/race-discrimination#h4

Equality and Human Rights Commission. (2023). Public Sector Equality Duty. Retrieved from https://www.equalityhumanrights.com/en/advice-and-guidance/race-discrimination#h4

Fair, F., Raben, L., Watson, H., Vivilaki, V., van den Muijsenbergh, M., Soltani, H. & the ORAMMA team (2020). Migrant women's experiences of pregnancy, childbirth and maternity care in European countries: A systematic review. PLoS ONE, 15(2): e0228378

Fivexmore (2022). The Black maternity experiences survey. A nationwide study of Black women's experience of maternity services in the United Kingdom. Retrieved from: https://www.fivexmore.com/blackmereport

Goldman, H.I., Maralit, A., Sun, S., Lanzkowsky, P. (1973). Neonatal cyanosis and arterial oxygen saturation. The Journal of Pediatrics, 82(2), 319-324.

Goldman, S.L., Peñalver, A., Peñaranda, R. (1982). Jaundice meter: evaluation of new guidelines. The Journal of Pediatrics, 101(2), 253-256.

Hannemann, R.E., Dewitt, D.P., Hanley, E.J., Schreiner, R.L., Bonderman, P. (1979). Determination of serum bilirubin by skin reflectance: effect of pigmentation. Pediatric Research, 13(12), 1326-1329.

Hannemann, R.E., Schreiner, R.L., DeWitt, D.P., Norris, S.A., Glick, M.R. (1982). Evaluation of the Minolta bilirubin meter as a screening device in White and Black infants. Pediatrics, 69(1), 107-109.

Hannon, P.R., Willis, S.K., Scrimshaw, S.C. (2001). Persistence of maternal concerns surrounding neonatal jaundice. An exploratory study. Archives of Pediatrics & Adolescent Medicine, 155, 1357-1363.

Health Education England (HEE). (2022). E-learning for healthcare. Retrieved from https://www.e-lfh.org.uk/ Accessed 12th Jan 2023.

Health Safety Investigation Branch (HSIB). (2021). Recognition of the acutely ill infant. I2018/024. Retrieved from: https://www.hsib.org.uk/investigations-and-reports/recognition-of-acutely-ill-infant/ Accessed 22/02/2023

Health Safety Investigation Branch (HSIB). (2023). Detection of jaundice in newborn babies. NI-003729. Retrieved from: https://www.hsib.org.uk/investigations-and-reports/detection-of-jaundice-in-newborn-babies/ Accessed 22/02/2023

Hegyi, T., Hiatt, I.M., Indyk, L. (1981). Transcutaneous bilirubinometry. I. Correlations in term infants. The Journal of Pediatrics, 98(3), 454-457.

Henderson, J., Gao, H., Redshaw, M. (2013). Experiencing maternity care: The care received and perceptions of women from different ethnic groups. BMC Pregnancy and Childbirth, 13(1), 1-14.

Higginbottom, G., Pillay, J., Boadu, N. (2013). Guidance on performing focused ethnographies with an emphasis on healthcare research. The Qualitative Report, 18(17), 1–16.

Holland, L., Blick, K. (2009). Implementing and validating transcutaneous bilirubinometry for neonates. American journal of clinical pathology, 132(4), 555-561.

Hong, Q.N., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., ... & Pluye, P. (2018). The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. Education for information, 34(4), 285-291.

Hurst, I. (2001). Vigilant watching over: mothers' actions to safeguard their premature babies in the newborn intensive care nursery. The Journal of Perinatal & Neonatal Nursing, 15(3), 39-57.

Institute of Health Visiting (iHV) (2022). Babies who have neonatal jaundice. Good practice points for health visitors. Institute of Health Visiting: London. Retrieved from https://ihv.org.uk/news-and-views/news/updated-gpp-babies-who-have-neonatal-jaundice/ Accessed 12th Jan 2023.

Jegathesan, T., Campbell, D.M., Ray, J.G., Shah, V., Berger, H., Hayeems, R.Z., Sgro, M. (2021). Transcutaneous versus total serum bilirubin measurements in preterm infants. Neonatology, 118(4), 443-453.

Jones, D.F., McRea, A.R., Knowles, J.D., Lin, F.C., Burnette, E., Reller, L.A., … & Lohr, J.A. (2017). A prospective comparison of transcutaneous and serum bilirubin within brief time intervals. Clinical Pediatrics, 56(11), 1013-1017.

Jurdi, S.R., Jayaram, A., Sima, A.P., Hendricks Muñoz, K.D. (2015). Evaluation of a Comprehensive Delivery Room Neonatal Resuscitation and Adaptation Score (NRAS) compared to the Apgar score: a pilot study. Global Pediatric Health, 2, 2333794X15598293.

Kapadia, D., Zhang, J., Salway, S., Nazroo, J., Booth, A., Villarroel-Williams, N., ... & Esmail, A. (2022). Ethnic Inequalities in Healthcare: A Rapid Evidence Review. NHS Race and Health Observatory: London

Karen, T., Bucher, H.U., Fauchère, J.C. (2009). Comparison of a new transcutaneous bilirubinometer (Bilimed®) with serum bilirubin measurements in preterm and full-term infants. BMC Pediatrics, 9(1), 1-7.

Karon, B.S., Teske, A., Santrach, P.J., Cook, W.J. (2008). Evaluation of the BiliChek noninvasive bilirubin analyzer for prediction of serum bilirubin and risk of hyperbilirubinemia. American journal of clinical pathology, 130(6), 976-982.

Kenny, M., Metzger, G., Landicho, D., Chow, J. (1984). Transcutaneous bilirubin monitoring of newborns. Annals of the New York Academy of Sciences, 428, 251-262.

Keren, R., Luan, X., Friedman, S., Saddlemire, S., Cnaan, A., Bhutani, V.K. (2008). A comparison of alternative risk-assessment strategies for predicting significant neonatal hyperbilirubinemia in term and near-term infants. Pediatrics, 121(1), e170-e179.

Keren, R., Tremont, K., Luan, X., Cnaan, A. (2009). Visual assessment of jaundice in term and late preterm infants. Archives of Disease in Childhood-Fetal and Neonatal Edition, 94(5), F317-F322.

Knight, M., Bunch, K., Patel, R., Shakespeare, J., Kotnis, R., Kenyon, S., … & Kurinczuk, J.J. (2022). Saving Lives, Improving Mothers' Care - Lessons learned to inform maternity care from the UK and Ireland Confidential Enquiries into Maternal Deaths and Morbidity 2018-20. MBRRACE-UK. National Perinatal Epidemiology Unit, University of Oxford: Oxford.

Knight, M. (2020). Resuscitation of the healthy baby at birth: the importance of drying, airway management and establishment of breathing. In Marshall, E., Raynor, M. (Eds) Myles' textbook for midwives. 17th edition. Elsevier: Edinburgh

Ledger, S., Hindle, G., McKee, A., Smith, T. (2021). Mind the Gap. An investigation into maternity training for frontline professionals across the UK (2020/2021). Baby Lifeline: UK. Retrieved from: https://www.babylifelinetraining.org.uk/the-mind-the-gap-report-by-baby-lifeline/ Accessed 22/02/2023

Lee, A.C., Folger, L.V., Rahman, M., Ahmed, S., Bably, N.N., Schaeffer, L., ... & Baqui, A.H. (2019). A novel icterometer for hyperbilirubinemia screening in low-resource settings. Pediatrics, 143(5).

Lee, S-Y., Weiss, S.J. (2009). When East meets West: intensive care unit experiences among first-generation Chinese American parents. Journal of nursing scholarship, 41(3), 268-275.

Leite, M.dG.dC., Granato, V.dA., Facchini, F.P., Marba, S.T.M. (2007). Comparison of transcutaneous and plasma bilirubin measurement. Jornal de Pediatria, 83, 283-286.

Li, F., Wu, T., Lei, X., Zhang, H., Mao, M., Zhang, J. (2013). The Apgar score and infant mortality. PLoS ONE, 8(7), e69072.

Linder, N., Regev, A., Gazit, G., Carplus, M., Mandelberg, A., Tamir, I., Reichman, B. (1994). Noninvasive determination of neonatal hyperbilirubinemia: standardization for variation in skin color. American journal of perinatology, 11(03), 223-225.

Madlon-Kay, D.J. (1997). Recognition of the presence and severity of newborn jaundice by parents, nurses, physicians, and icterometer. Pediatrics, 100(3), e3.

Madlon-Kay, D.J. (2001). Home health nurse clinical assessment of neonatal jaundice: comparison of 3 methods. Archives of Pediatrics & Adolescent Medicine, 155(5), 583-586.

Maisels, M.J., Ostrea Jr, E.M., Touch, S., Clune, S.E., Cepeda, E., Kring, E., ... & Huang, R. (2004). Evaluation of a new transcutaneous bilirubinometer. Pediatrics, 113(6), 1628-1635.

Martin, A.E., D'Agostino, J.A., Passarella, M., Lorch, S.A. (2016). Racial differences in parental satisfaction with neonatal intensive care unit nursing care. Journal of Perinatology, 36(11), 1001-1007.

Mathur, R., Bhaskaran, K., Chaturvedi, N., Leon, D.A., vanStaa, T., Grundy, E., … & Smeeth, L. (2014). Completeness and usability of ethnicity data in UK-based primary care and hospital databases. Journal of Public Health, 36(4), 684–692.

Maya-Enero, S., Candel-Pau, J., Garcia-Garcia, J., Duran-Jordà, X., López-Vílchez, M.Á. (2021). Reliability of transcutaneous bilirubin determination based on skin color determined by a neonatal skin color scale of our own. European Journal of Pediatrics, 180, 607-616.

McClean, S., Baerg, K., Smith-Fehr, J., Szafron, M. (2018). Cost savings with transcutaneous screening versus total serum bilirubin measurement for newborn jaundice in hospital and community settings: a cost-minimization analysis. CMAJ Open, 6(3), E285–E291. https://doi.org/10.9778/cmajo.20170

Merritt, K.A., Coulter, D.M. (1994). Application of the Gosset icterometer to screen for clinically significant hyperbilirubinemia in premature infants. Journal of Perinatology: Official Journal of the California Perinatal Association, 14(1), 58-65.

Michaelides S. (2017). Physiology, assessment and care of the newborn. In Macdonald S, Johnson G (Eds) Mayes' Midwifery. Fifteenth Edition. Elsevier: London.

Mihoko Doyle, J., Echevarria, S., Parker Frisbie, W. (2003). Race/ethnicity, Apgar and infant mortality. Population Research and Policy Review, 22, 41-64.

Moran-Thomas, A. (2020). How a popular medical device encodes racial bias. Boston Review, 8(5), 2020.

Moyer, V.A., Ahn, C., Sneed, S. (2000). Accuracy of clinical judgment in neonatal jaundice. Archives of pediatrics & adolescent medicine, 154(4), 391-394.

Mukwende, M., Tamony, P., Turner, M. (2021). Mind the Gap: A handbook of clinical signs in Black and Brown skin. 2021. Retrieved from: https://www.blackandbrownskin.co.uk/mindthegap Accessed 22/03/2023

National Collaborating Centre for Women's and Children's Health (NCC-WCH) and Royal College of Obstetricians & Gynaecologists (RCOG.) (2010). Neonatal Jaundice. Royal College of Obstetricians & Gynaecologists: London.

National Institute for Health and Care Excellence (NICE). (2017). Intrapartum care. Care of healthy women and their babies during childbirth. Version 2 (NICE Clinical Guideline No. CG190). National Institute for Health and Care Excellence: London.

National Institute for Health and Care Excellence (NICE). (2019). Specialist neonatal respiratory care for babies born preterm (NICE Guideline No. NG124). National Institute for Health and Care Excellence: London.

National Institute for Health and Care Excellence (NICE) & Royal College of Obstetricians & Gynaecologists (RCOG) (2021). Postnatal care (NICE Guideline No. NG194). National Institute for Health and Care Excellence: London.

National Institute for Health and Care Excellence (NICE). (2016 update). Jaundice in newborn babies under 28 days (NICE Clinical Guideline No. CG98). National Institute for Health and Care Excellence: London.

National Institute for Health and Care Excellence. (2014). Jaundice in newborn babies under 28 days. (Quality standard No. QS57). National Institute for Health and Care Excellence: London.

Nazroo, J.Y., Williams, D.R. (2006). The social determinants of ethnic/racial inequalities in health. In Marmot, M.G., Wilkinson, R.G. (Eds) Social Determinants of Health. New York: Oxford University Press

Neonatal Nurses Association/University of Hertfordshire (no date a). Assessment of the neonate. Retrieved from https://www.herts.ac.uk/study/schools-of-study/health-and-social-work/course-subject-areas/nursing/childrens-nursing/neonatal-nursing Accessed 12th Jan 2023.

Neonatal Nurses Association/University of Hertfordshire. (no date b). Monitoring vital signs in the neonate. Retrieved from https://www.herts.ac.uk/study/schools-of-study/health-and-social-work/course-subject-areas/nursing/childrens-nursing/neonatal-nursing Accessed 12th Jan 2023.

NHS Race and Health Observatory. (2021a). The Power of Language. A consultation report on the use of collective terminology at the NHS Rach & Health Observatory. Retrieved from https://www.nhsrho.org/wp-content/uploads/2021/11/NHS_RaceHealthObservatory_Terminology-consultation-report-NOV-21-1.pdf. Accessed 22/02/2022.

NHS Race and Health Observatory. (2021b). Pulse oximetry and racial bias: Recommendations for national healthcare, regulatory and research bodies. Retrieved from https://www.nhsrho.org/wp-content/uploads/2021/03/Pulse-oximetry-racial-bias-report.pdf Accessed 12/01/2023.

Norman, M., Aytug, H., Celebi, H.B. (2022). Evaluation of a new transcutaneous bilirubinometer in newborn infants. Scientific Reports, 12(1), 5835.

Office for Health Improvement & Disparities (OHID) (2021). Newborn and infant physical examination (NIPE) screening programme handbook. Office for Health Improvement & Disparities: London. Retrieved from https://www.gov.uk/government/publications/newborn-and-infant-physical-examination-programme-handbook/newborn-and-infant-physical-examination-screening-programme-handbook Accessed 12th Jan 2023.

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffman, T.C., Mulrow, C.D., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. International Journal of Surgery, 88, 105906.

Palau, M.A., Meier, M,R., Brinton, J.T., Hwang, S.S., Roosevelt, G.E., Parker, T.A. (2019) The impact of parental primary language on communication in the neonatal intensive care unit. Journal of Perinatology, 39, 307-313.

Palmer, D.C., Zenner, E.M., Drew, J.H. (1982). Transcutaneous bilirubinometry: use in Australia. Journal of Paediatrics and Child Health, 18(4), 273-276.

Paul, H.A., Adams, B.J., Venner, A.A. (2021). Improving quality of transcutaneous bilirubin measurements: Value of in-house developed quality control. Practical Laboratory Medicine, 24, e00206.

Raimondi, F., Lama, S., Landolfo, F., Sellitto, M., Borrelli, A. C., Maffucci, R., ... & Capasso, L. (2012). Measuring transcutaneous bilirubin: a comparative analysis of three devices on a multiracial population. BMC Pediatrics, 12, 1-5.

Rennie, J.M., Beer, J., Upton, M. (2019). Learning from claims: hyperbilirubinaemia and kernicterus. Archives of Disease in Childhood-Fetal and Neonatal Edition, 104(2), F202-F204.

Resuscitation Council (2011). Air/oxygen blenders and pulse oximetry in resuscitation at birth. Resuscitation Council United Kingdom (RCUK): London. Retrieved from https://www.resus.org.uk/library/quality-standards-cpr/airoxygen-blenders-and-pulse-oximetry-resuscitation-birth Accessed 12th Jan 2023.

Resuscitation Council. (2021). Newborn resuscitation and support of transition of infants at birth guidelines. Resuscitation Council United Kingdom (RCUK): London. Retrieved from https://www.resus.org.uk/library/2021-resuscitation-guidelines/newborn-resuscitation-and-support-transition-infants-birth Accessed 12th Jan 2023.

Rimmer, A. (2020). Presenting clinical features on darker skin: Five minutes with ... Malone Mukwende. BMJ (Online), 369, m2578. doi:10.1136/bmj.m2578.

Riskin, A., Tamir, A., Kugelman, A., Hemo, M., Bader, D. (2008). Is Visual Assessment of Jaundice Reliable as a Screening Tool to Detect Significant Neonatal Hyperbilirubinemia? The Journal of Pediatrics, 152(6), 782-787.

Robertson, A., Kazmierczak, S., Vos, P. (2002). Improved transcutaneous bilirubinometry: comparison of SpectR x BiliCheck and Minolta jaundice meter JM-102 for estimating total serum bilirubin in a normal newborn population. Journal of Perinatology, 22(1), 12-14.

Roper, J.M., Shapira, J. (2000). Ethnography in nursing research. Sage Publications: London

Royal College of Midwives (RCM) (2012). Immediate Care of the Newborn. Evidence based guidelines for Midwifery-Led Care in labour. RCM: London.

Royal College of Nursing (RCN) (2017). Standards for Assessing, Measuring and Monitoring Vital Signs in Infants, Children and Young People. Royal College of Nursing: London. Retrieved from https://www.rcn.org.uk/professional-development/publications/pub-005942 Accessed 12th Jan 2023.

Rubaltelli, F.F., Gourley, G.R., Loskamp, N., Modi, N., Roth-Kleiner, M., Sender, A., … & Vert, P. (2001). Transcutaneous bilirubin measurement: a multicenter evaluation of a new device. Pediatrics, 107(6), 1264-1271.

Rubio, A., Epiard, C., Gebus, M., Deiber, M., Samperiz, S., Genty, C., ... & Debillon, T. (2017). Diagnosis accuracy of transcutaneous bilirubinometry in very preterm newborns. Neonatology, 111(1), 1-7.

Samiee-Zafarghandy, S., Feberova, J., Williams, K., Yasseen, A. S., Perkins, S. L., Lemyre, B. (2014). Influence of skin colour on diagnostic accuracy of the jaundice meter JM 103 in newborns. Archives of Disease in Childhood-Fetal and Neonatal Edition, 99(6), F480-F484.

Schmidt, E.T., Wheeler, C.A., Jackson, G.L., Engle, W.D. (2009). Evaluation of transcutaneous bilirubinometry in preterm neonates. Journal of Perinatology, 29(8), 564-569.

Schott, J., Henley, A. (2000). Health-care equity: how to recognise clinical signs in skin. British Journal of Midwifery, 8(5), 271-273.

Serunian, S.A., Broman, S.H. (1975). Relationship of Apgar scores and Bayley mental and motor scores. Child Development, 696-700.

Shah, S., Arbuiso, S., Gorgy, M., Cline, A., Moy, J. (2022). Assessment of diversity in skin colour in dermatology medical education resources. Clinical and Experimental Dermatology, 47(11), 2038-2040. doi:10.1111/ced.15328.

Shankaran, S., Johnson, Y., Langer, J. C., Vohr, B.R., Fanaroff, A.A., Wright, L.L., … & Poole, W.K. (2004). Outcome of extremely-low-birth-weight infants at highest risk: gestational age≤ 24 weeks, birth weight≤ 750 g, and 1-minute Apgar≤ 3. American Journal of Obstetrics and Gynecology, 191(4), 1084-1091.

Siddiqui, A., Cuttini, M., Wood, R., Velebil, P., Delnord, M., Zile, I., … & Zeltin, J. (2017). Can the Apgar score be used for international comparisons of newborn health? Paediatric and Perinatal Epidemiology, 31(4), 338-345.

Sigurdson, K., Mitchell, B., Liu, J., Morton, C., Gould, J. B., Lee, H. C., ... & Profit, J. (2019). Racial/ethnic disparities in neonatal intensive care: a systematic review. Pediatrics, 144(2).

Sjoding, M.W., Dickson, R.P., Iwashyna, T.J., Gay, S.E., Valley, T.S. (2020). Racial bias in pulse oximetry measurement. New England Journal of Medicine, 383(25), 2477-2478.

Slusher, T.M., Angyo, I.A., Bode-Thomas, F., Akor, F., Pam, S.D., Adetunji, A.A., ... & Stevenson, D.K. (2004). Transcutaneous bilirubin measurements and serum total bilirubin levels in indigenous African infants. Pediatrics, 113(6), 1636-1641.

Slusher, T.M., Zamora, T.G., Appiah, D., Stanke, J.U., Strand, M.A., Lee, B.W., … & Olusanya, B.O. (2017) Burden of severe neonatal jaundice: a systematic review and meta-analysis. BMJ Paediatrics Open, 1(1), e000105.

Starowicz, O., Edwards, P., Schmidt, P., Birch, P. (2020). Evaluation of the Kejian KJ-8000 bilirubinometer in an Australian setting. Journal of Paediatrics and Child Health, 56(2), 283-288.

Szabo, P., Wolf, M., Bucher, H.U., Haensse, D., Fauchere, J.C., Arlettaz, R. (2004a). Assessment of jaundice in preterm neonates: comparison between clinical assessment, two transcutaneous bilirubinometers and serum bilirubin values. Acta Paediatrica, 93(11), 1491-1495.

Szabo, P., Wolf, M., Bucher, H.U., Fauchere, J.C., Haensse, D., Arlettaz, R. (2004b). Detection of hyperbilirubinaemia in jaundiced full-term neonates by eye or by bilirubinometer? European Journal of Pediatrics, 163, 722-727.

Tarnow-Mordi, W.O., Pickering, D. (1983) Missed jaundice in Black infants: a hazard? British Medical Journal, 286(6363), 463-464.

Taylor, J.A., Burgos, A.E., Flaherman, V., Chung, E.K., Simpson, E.A., Goyal, N.K., ... & Better Outcomes through Research for Newborns Network. (2015). Discrepancies between transcutaneous and serum bilirubin measurements. Pediatrics, 135(2), 224-231.

Taylor, J.A., Stout, J.W., de Greef, L., Goel, M., Patel, S., Chung, E.K., ... & Larson, E.C. (2017). Use of a smartphone app to assess neonatal jaundice. Pediatrics, 140(3), e20170312.

Thomson, J., Culley, V., Monfrinoli, A., Sinha, A. (2008). Transcutaneous bilirubinometers and ethnicity. Archives of Disease in Childhood-Fetal and Neonatal Edition, 93(6), F474-F474.

United Nations Development Programme. (2023). Human Development Index (HDI). United Nations. Retrieved from https://hdr.undp.org/data-center/human-development-index#/indicies/HDI Accessed 22/02/2023

van der Geest, B.A., Theeuwen, I.M., Reiss, I.K., Steegers, E.A., Been, J.V. (2021). Assessing knowledge and skills of maternity care professionals regarding neonatal hyperbilirubinaemia: a nationwide survey. BMC Pregnancy and Childbirth, 21, 63.

Vesoulis, Z., Tims, A., Lodhi, H., Lalos, N., Whitehead, H. (2022). Racial discrepancy in pulse oximeter accuracy in preterm infants. Journal of Perinatology, 42(1), 79-85.

Villota, E.C., Pasquel, D.P., Cuenca, F.A., De Los Monteros, R.E. (2021). Noninferiority assessment of the “neonatal resuscitation and adaptation score” versus the Apgar score. Revista Ecuatoriana de PediatrÍa, 22(3), 20.

Wainer, S., Rabi, Y., Parmar, S. M., Allegro, D., Lyon, M. (2009). Impact of skin tone on the performance of a transcutaneous jaundice meter. Acta Paediatrica, 98(12), 1909-1915.

Watterberg, K.L., Aucott, S., Benitz, W.E., Cummings, J.J., Eichenwald, E.C., Goldsmith, J., ... & Wharton, K.R. (2015). The Apgar Score. Pediatrics, 136(4), 819-822.

Wilson, B.N., Sun, M., Olayinka, J., Murrell, D.F., Murase, J. (2021). Assessment of skin of color and diversity and inclusion content published in the Journal of Investigative Dermatology: An analysis and call to action. Journal of Investigative Dermatology, 141(5), S97.

Witcher, T.J., Jurdi, S., Kumar, V., Gupta, A., Moores Jr, R.R., Khoury, J., … & Rozycki, H.J. (2018). Neonatal Resuscitation and Adaptation Score vs Apgar: newborn assessment and predictive ability. Journal of Perinatology, 38, 1476-1482.

Witt, R.E., Colvin, B.N., Lenze, S.N., Shaw Forbes, E., Parker, M.G.K., Hwang, S.S., … & Colson, E.R. (2022). Lived experiences of stress of Black and Hispanic mothers during hospitalization of preterm infants in neonatal intensive care units. Journal of Perinatology, 42, 195-201.

Wolf, M.J., Beunen, G., Casaer, P., Wolf, B. (1997). Neurological findings in neonates with low Apgar in Zimbabwe. European Journal of Obstetrics & Gynecology and Reproductive Biology, 73(2), 115-119.

Woodgate, P., Jardine, L.A. (2015). Neonatal jaundice: phototherapy. BMJ clinical evidence, 2015, 0319.

World Health Organisation. (2012). Guidelines on basic newborn resuscitation. World Health Organization: Geneva. Retrieved from https://www.who.int/publications/i/item/9789241503693 Accessed 12th Jan 2023.

World Health Organization (WHO). (2022a). WHO recommendations on maternal and newborn care for a positive postnatal experience. World Health Organization: Geneva.

World Health Organisation. (2022b). Early essential newborn care: clinical practice pocket guide, 2nd edition. World Health Organization: Geneva. Retrieved from https://www.who.int/publications/i/item/9789290619659 Accessed 12th Jan 2023.

Appendices

Appendix 1: HCP systematic review example search strategy

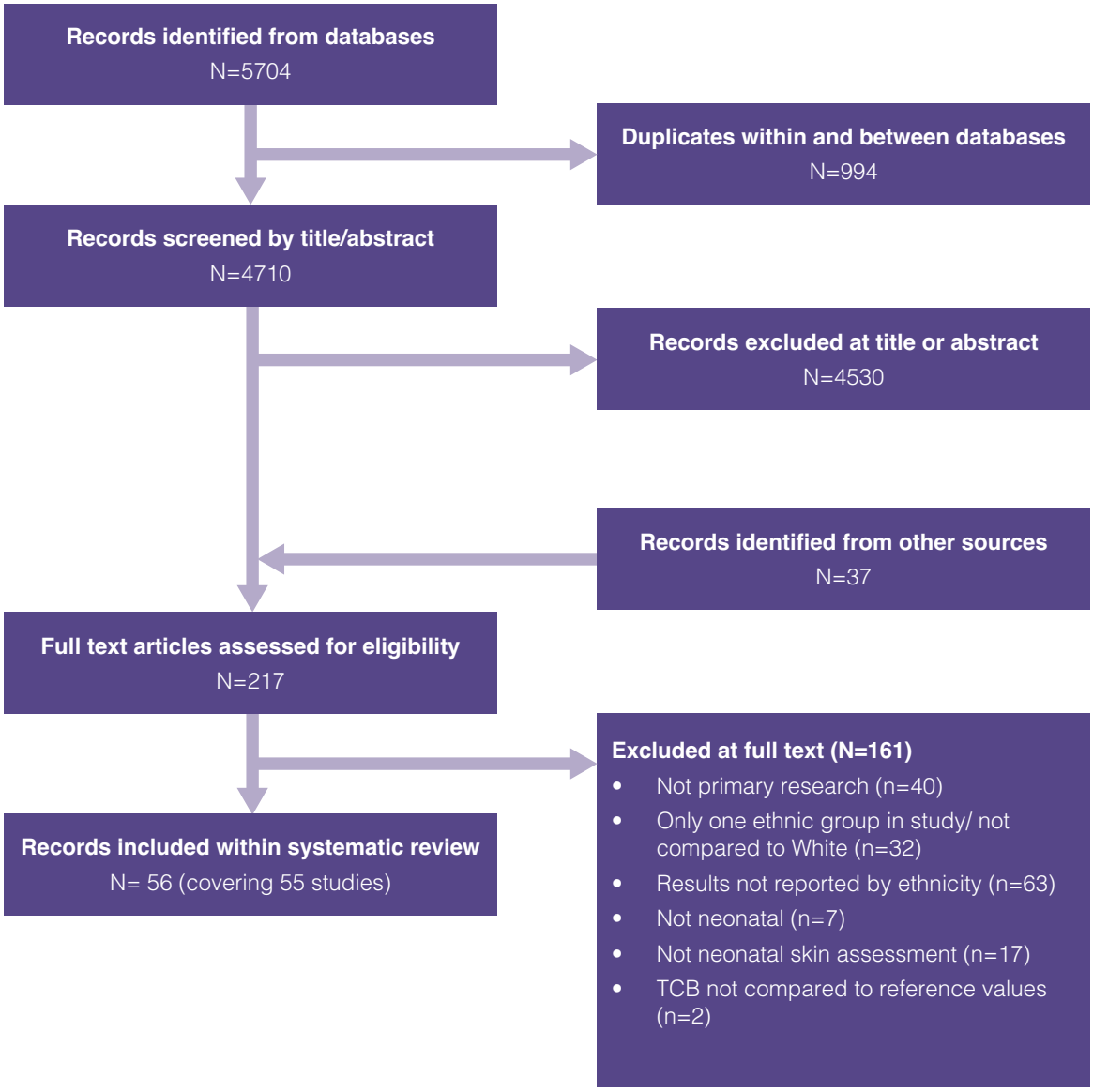
CINAHL

((MH “Jaundice”) OR (MH “Jaundice, Neonatal”) OR (MH “Hyperbilirubinemia, Neonatal”) OR (MH “Hyperbilirubinemia”) OR (MH “Anoxia”) OR (MH “Apgar Score”) OR (MH “Cyanosis”) OR Hypoxia OR hypoxemia OR oximet* OR “oxygen saturation” OR jaundice OR APGAR OR cyanosis)

AND (neonatal OR newborn OR neonate OR “new born” OR new-born OR baby OR neonatol* OR Pediatric* OR paediatric* OR preterm OR premature OR babies)

AND ((MH “Ethnic Groups+”) OR (MH “Minority Groups”) OR (MH “Cultural Diversity”) OR (MH “Asians+”) OR (MH “Black Persons”) OR (MH “Hispanic Americans”) OR (MH “Indigenous Peoples”) OR (MH “Immigrants+”) OR (MH “Transients and Migrants”) OR (MH “Emigration and Immigration”) OR Black OR Asian OR “ethnic minority” OR “minority ethnic” OR bame OR “Black or minority ethnic” OR bme OR “mixed race” OR “mixed ethnic*” OR “mixed heritage” OR skin pigmentation OR Hispanic OR Pakistan* OR Somali* OR India* OR Bangladesh* OR Chinese OR China OR Africa* OR Caribbean OR Arab OR Ethnic* OR ethno* OR race OR Racial* OR racism OR colour OR color) {LIMITED TO TITLE OR ABSTRACT OR SUBJECT}

Appendix 2: HCP systematic review PRISMA flow chart showing study identification (Page et al., 2021)



Appendix 3: Characteristics of included studies with the healthcare professional's systematic review

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Apgar score study characteristics								
Chubb et al., (2022)	UK	To review training, to assess Black, Asian, and minority ethnic babies	67 midwives and students on placement	NA	NR	Study design: Descriptive Inclusion criteria: 67 /85 midwives and students at Yeovil District NHS Foundation Trust who completed training Data Collection: Anonymous pre and post training surveys Analysis: descriptive statistics and thematic analysis of qualitative data	Not able to assess long term impact of training. Not all staff trained, but high turnout at 79%	The training was well evaluated but more training and research is required to improve safety of Black, Asian and ethnic minority families
Li et al., (2013)	USA	To evaluate if Apgar remains pertinent in practice after more than 50 years and to assess the Apgar in predicting infant survival	25,168,052 singletons 768,305 twins	24–25wks 0.18%; 26–27wks 0.25%; 28–29wks 0.34%; 30–31wks 0.56%; 32–33wks 1.13%; 34–36wks 7.38%; 37–41wks 86.95%; 41–44wks 3.22%	Non-Hispanic White 18,095,334 (66.4%) Non-Hispanic Black 4,540,838 (16.6%) Hispanic 3,300,185 (12.1%) Other 1,334,801 (4.9%) According to maternal race	Study design: Retrospective linked data Exclusion criteria: Triplets or higher (70,387); BW <500g or NK (84,177); GA <24 wks or >44 wks (490,214), 5-min Apgar 0, >10 or NR (8,637,941). Data Collection: From linked live birth and infant death datasets. Analysis: SAS v9.2. Chi-Square, Kaplan-Meier curve, Cox proportional hazard-model adjusted for maternal education, marital status, time AN care started and smoking during pregnancy. Non-Hispanic White referent	Infants born in very severe condition may have been reported as a stillbirth. Proportion of each Apgar score for each ethnicity NR	The Apgar score predicts neonatal and post-neonatal adverse outcomes in term and preterm infants. It is applicable to twins and in various race/ethnic groups

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Mihoko Doyle, et al., (2003)	USA	To assess the predictive validity of Apgar scores on infant mortality looking at race/ethnic-specific variation.	6,544,004 neonates	White: <1500g 0.6%; ≥1550g & GA≤36wks 6.7%; ≥1550g & GA>36wks 92.7% Black: <1500g 2.0%; ≥1550g & GA≤36wks 15.5%; ≥1550g & GA>36wks 82.5% MA: <1500g 0.7%; ≥1550g & GA≤36wks 8.8%; ≥1550g & GA>36wks 90.5% Apgar (0–3): White 1.7%, Black 3.1%, MA 1.8%; Apgar (4–6): White 5.9%, Black 7.3%, MA 5.8%; Apgar (7–10): White 92.4%, Black 89.6%, MA 92.4%	5,100,942 White, 1,252,870 Blacks, 190,192 MA NR how attributed ethnicity/race	Study design: Retrospective linked data Exclusion criteria: Multiple births; BW <500g; <22wks GA Data Collection: 1989–1991 NCHS Linked Birth/Infant Death files for the USA 1-minute Apgar scores. Classification as BW<1500g, or >1500g with GA≤36 wks or GA>36wks Analysis: Logistic regression models, with race/ethnic models estimated	Apgar score involves some subjectivity. Relatively small number of Mexican Americans included in study	Apgar is a strong predictor of infant survival including within race/ethnic groups
Serunian & Broman, (1975)	USA	To investigate relationship of 1-minute Apgar scores to 8-month Bayley mental and motor scores	350/391 children had Apgar and Bayley scores	1-min Apgar: White 0-3 (3.5%); 4-6 (16.3%); 7-10 (80.2%) Black 0-3 (9.0%); 4-6 (14.6%); 7-10 (76.4%) Mixed (Black & Portuguese) 0-3 (6.7%); 4-6 (9.3%); 7-10 (84.0%)	86 White, 89 Black, 75 mixed (Black & Portuguese) Racial designation based on maternal report	Study design: Prospective cohort Inclusion criteria: Participants selected from the Providence Child Development Study. Selected according to race/ethnicity Data Collection: 1 minute Apgar score. Bayley mental and motor development form at 8-months, as well as Infant Behaviour Profile and presence or absence of physical abnormalities. Analysis: Chi square tests	NR	Children with abnormal development at 8 months had significantly lower 1 minute Apgar scores

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Shankaran et al., (2004)	USA	To evaluate neurodevelop-mental sequelae among extremely high-risk infants (low birth weight, preterm and low Apgar)	304 infants survived to hospital discharge of whom 246 were seen at 18-22 months	GA: 23.6 ± 0.7 wks BW (mean): 639.3±63.6g 5-min Apgar <3: 24.4% Maternal age: 26.7±6.9 yrs Infant gender: Male 44.7%	Initial cohort Black: 182/304 (59.9%) Followed up 146/246 (59.3%) NR how attributed ethnicity/race	Study design: Prospective cohort Inclusion criteria: 304 infants surviving to hospital discharge from 12 NNUs participating in previous study Data Collection: Follow up at 18-22 months corrected age. Amiel-Tison neurologic examination, Bayley Mental Developmental Index and Psychomotor Developmental Index; medical and social history including parental education, occupation, household composition, income level Analysis: Mantel-Haenszel odds ratios	Not all surviving infants followed-up. Loss to follow-up may result in serious ascertainment bias, although clinical and demographic characteristics of those not evaluated were similar A centre-based study therefore potential for referral bias	High-risk infants (low birth weight, preterm and low Apgar) are at high risk of morbidity and mortality

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Wolf et al., (1997)	Zimbabwe	To document neurological condition of African neonates with a low Apgar score	165 babies in Zimbabwe compared to asphyxiated infants in the Netherlands (n=94) and Grenada, Caribbean (n=11)	In Zimbabwe: Primiparous 51.5%; Infant gender girl 43.6%; preterm 15.7%; congenital malformations 7%, SGA 12% BW: Zimbabwe 2846±702g; Netherlands 2879±978g. GA: Zimbabwe 38.5±2.9 wks; Netherlands 38.4±3.6 wks. Age mother: Zimbabwe 24±6.2 yrs; Netherlands 25±4.7 yrs.	Only country of origin given: Zimbabwe; Netherlands or Grenada - Caribbean.	Study design: Descriptive studies x3 Inclusion criteria: Babies with an Apgar score of ≤5 at 5 min admitted to NNU between 1 July 1991 and 30 June 1992 Data Collection: Comprehensive neurological examination adapted from Prechtl when sufficiently stable to tolerate the examination, but not before the third day after term delivery and when reached corrected term age if born preterm. The two reference groups defined asphyxia as an Apgar ≤ 6 at 3 min. Examinations performed by the same investigator. Analysis: Chi square tests, Pearson's r and ANOVA.	Control groups not strictly comparable due to different Apgar cut off points in the different countries. Unable to ascertain impact of ethnicity vs different care practices between countries	Neonatal morbidity was higher in Zimbabwe than in the comparison groups from the Netherlands and Caribbean

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Cyanosis study characteristics								
Dawson et al., (2015)	Australia	To explore whether the pinkness of an infant's tongue provided a useful indication that supplemental oxygen was required when pulse oximetry is unavailable	68 neonates (271 paired assessments after removing	GA: 38±2 wks Mean BW: 3214±545g Female n=36 (53%) Type of anaesthesia: general (n=2); epidural/spinal (n=66)	Results given by Caucasian/ non-Caucasian, proportions in each group not given.	Study design: Prospective observational study Inclusion criteria: Convenience sample of infants delivered by CS between Aug and Nov 2012. Excluded if GA<28 wks, parents did not speak English or pulse oximeter alarm messages. Data Collection: Simultaneous SpO2 and visual assessment of tongue colour when pulse oximeter applied and at 2,3,4,5,6,7 and 10 mins after birth. 38 midwives & 7 paediatric trainees carried out assessments Analysis: Sensitivity, specificity, PPV, NPV and positive and negative likelihood ratios of tongue not pink to detect SpO2 < 70%. AUC calculated.	Pulse oximeters not accurate when arterial O2 <70%, so impacted by neonate low O2 sats in first few mins of life. Only studied infants born by CS	Tongue colour was a specific but insensitive sign that SpO2 was <70%. When the tongue was pink, it was likely the infant had SpO2 >70% and supplemental oxygen not required

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Goldman et al., (1973)	US	To determine the relationship of clinical assessment of skin and mucous membrane colour of neonates to arterial oxygen saturation	93 neonates (82 instances of assessment)	Birthweight: <1500 n=20; 1500-2500 n = 52; >2500g n=21 29 were ill - respiratory distress syndrome (n=21), pneumonia (n=1), massive aspiration (n=2), congenital heart disease (n=3), seizures after birth anoxia (n=2). The rest were 'well'	Skin pigmentation assessed by 3-point scale - fair/ medium / dark. Repeated if the infant was re-studied as pigmentation can darken in neonatal period. Dark skinned infants (as judged by at least 2 of the 3 observers) n=27	Study design: Prospective cohort Inclusion criteria: Infants PN age <2 wks. Excluded if axillary temp <36°C, bruising or ecchymoses to trunk, face or hands, or if Hb <13 or > 23 gm/dl. Data Collection: Simultaneous clinical assessment and determination of arterial pO2 and pH. Assessments by 2 physicians & a nurse in brightly lit room. Cyanosis judged present or absent in 6 areas: lips, ears, trunk, nailbed, hands, region around the mouth. Blood taken after observation. Analysis: Chi square with Yates correction.	NR	Trunk and ears were the least sensitive areas and hands, nailbeds and around the mouth the most sensitive areas to detect cyanosis

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Vesoulis et al., (2022)	USA	To determine whether oxygen saturation is overestimated using pulse oximetry for Black patients.	294 infants	Mean GA: 25.8 ± 2.1 wks (Black 25.6 ± 1.9 wks vs White 25.9 ± 2.1 wks p=0.09) Mean BW: 845 ± 265 grams, (Black 805 ± 260 g vs White 875 ± 268 g p=0.02.) Black Female: 64 (52%) White female: 74 (44%) Received AN steroids: Black 70%; White 74% (p=0.49)	124 Black (42%), 170 White (58%) Classified according to parental identification on birth certificates.	Study design: Retrospective cohort study Inclusion criteria: Preterm infants with GA <32 wks, BW <1500 g and admitted to NNU between 2012 and 2019 with valid vital sign data and at least one arterial blood gas. Data Collection: Pulse oximetry with simultaneous arterial blood gas. During the study period alarm limits were set between 88 and 96% until infant reached 35 wks postmenstrual age, then alarm set to 88%. SpO2 data 30s before to 30s after arterial blood gas averaged. Analysis: R statistical package. Univariate comparison using nonparametric methods (Fisher's Exact or Mann–Whitney U test). Pearson correlation coefficient, linear regression and non-linear regression	Syringe examined for air prior to analysis. SpO2 sensor is rotated every 12 hours, but position not routinely charted. Positions other than right upper extremity are post-ductal with potential mismatch between pre- & post-ductal measurements. Binary race classification used, further differentiation of skin tone /melanin content would be beneficial in future studies.	Modest but consistent difference in SpO2 error between Black and White infants, with increased incidence of occult hypoxemia in Black infants
Jaundice study characteristics								

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Afanetti et al., (2014)	France	To evaluate accuracy of TCB in preterm and term infants of different ethnic backgrounds, using a second generation bilirubinometer.	86 neonates,	44 preterm (median [IQR] GA 30wks [2]; BW 1335g [535]; M/F ratio 0.7) 41 term/ near term (median [IQR] GA 38 wks [4.2]; BW 3030g [1088]; M/F ratio 1.3) Caucasians (median [IQR] GA 34wks [5]; BW 1660g [1020]; M/F ratio 1.3); non-Caucasians (median [IQR] GA 31.5wks [8]; BW 1760 [2010]; M/F ratio 0.6)	49 Caucasian and 36 non-Caucasian. 91% of non-Caucasians were African phenotype	Study design: Retrospective Exclusion criteria: PT within preceding 12 h, haemolytic disease, cholestasis, severe perinatal asphyxia and anomalies of skin perfusion, such as oedema or compromised hemodynamic parameters. Data Collection: Drager JM103 (sternum) vs SBr with maximal 30 min between measurements. Device Calibrated daily. Analysis: Statview and Medcalc software. Students t-test, simple linear regression, Pearson's correlation coefficient (r), Bland Altman. Multiple forward stepwise linear regression adjusted for skin tone, GA, PN age	Performed in retrospective data so can't explain unexpected differences in sex ratio and BW. Very low GA preterm neonates under-represented, with only two born <28 wks. Sample size calculation to answer research question is lacking.	TCB using JM-103 correlates with SBr regardless of gestation or skin tone
Ahmed et al., (2010)	UK	To look at agreement between BiliCheck (TCB) and SBr in babies <35 wks gestation with or without PT	57 infants (183 paired results)	All babies 26-34 wks gestation	50 were Caucasian, four were Indian subcontinent and three were mixed ethnicity. 14 paired measurements in non-Caucasian infants.	Study design: Prospective observational study Inclusion criteria: Convenience sample of all babies <35 wks GA admitted to NNU from July 2007 to June 2008. Infants requiring exchange transfusion excluded Data Collection: BiliCheck (forehead) vs SBr. TCB within 15–30 min Analysis: Linear regression analysis, ROC curve constructed. Effect of skin pigmentation assessed using Mann-Whitney rank sum test	NR	BiliCheck is safe alternative to SBr in preterm neonates, skin pigmentation did not affect performance

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Aune et al., (2020)	Norway	To evaluate a novel method of estimating bilirubin levels from colour-calibrated smartphone images.	302 infants. recruited	GA (wks) 39.4 ± 1.3 PN age (hours) 76 ± 43	59 (20%) non-Caucasian defined as: Middle Eastern, Asian, African and other or unknown	Study design: Cross-sectional prospective study Inclusion criteria: Infants GA ≥37 wks, BW ≥2500g, PN age 1 to 15 days with and without suspected jaundice recruited from two maternity wards or breastfeeding clinics. Excluded signs of diseases other than jaundice, receiving advanced medical treatment, receiving PT Data Collection: Smartphone app images (sternum) vs Dräger JM-103 orJM-105 and SBr max. 60 mins after images Analysis: IBM SPSS. Pearson's r, Bland-Altman plots, ROC curves with a cut off of SBr level of ≥250 µmol/l.	Data on light intensity not collected while obtaining the images nor was data on visual assessment of jaundice	Smartphone images correlation with SBr significantly lower than TCB correlation with SBr, but the smartphone based tool to estimate bilirubin from digital images identified severe jaundice

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Bhutani et al., (2000)	USA	To determine whether TCB is clinically equivalent to SBr in a diverse multi-ethnic sample of term and near-term infants	490 (1788 samples)	PN age: 12-98 hours BW: 2000-5665g. Mean 3404 ±518g GA: 35-42 wks. Mean 38.9 ±1.5wks	59.1% White 29.5% Black 3.46% Hispanic 4.48% Asian 3.46% other From parents' self declaration on admission	Study design: Prospective cohort study Inclusion criteria: Infants GA≥36 wks & ≥2000g or ≥35 wks & ≥2500g, from 2 well baby nurseries March 1998-Oct 1998. Excluded undergoing PT, sepsis, heart or circulatory disease, respiratory distress or clinical evidence of haemoglobinopathy. Data Collection: BiliCheck (forehead) vs SBr within 30 mins. Measured in morning ambient light. Device calibrated before each measurement. Analysis: SPSS. Data stratified for racial identity and GA. Pearson linear regression, Bland Altman, ANOVA to determine effect of GA, birthweight, race. Accuracy, NPV, PPV and likelihood ratio for BiliCheck >75th centile to detect SBr >95th centile	Proportion of infants with a SBr ≥256µmol/l was too low to determine BiliCheck accuracy in such infants.	Accuracy of predischarge BiliCheck measurement in term and near-term infants of diverse race and ethnicity demonstrated and was predictive of subsequent hyperbilirubinemia

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Bourchier et al., (1987)	New Zealand	To report on TCB use over 22-month period	2277 infants (2787 TCB measure-ments paired with SBr). Race only recorded in 729 cases		412 Caucasian; 317 Māori. Collected from maternal notes	Study design: Prospective Inclusion criteria: Term infants on PN ward with suspected jaundice. Excluded infants undergoing PT Data Collection: Minolta/ Air Shields JM- 101 (forehead & mid-sternum). SBr taken if TCB ≥17, or if requested by HCP. TCB by nursing supervisor or charge nurses Analysis: STATPRO. Students-t test and linear regression analyses	Correlation coefficients lower than for other published data may be due to paucity of low TCB and SBr values in the sample as recruited those with suspected jaundice	JM-101 was found to be a useful screening tool for who would been to have an SBr
Brits et al., (2018)	South Africa	To determine the prevalence of neonatal jaundice and secondly to explore its risk factors in healthy term neonates	96 neonates	Birthweight ranged from 2.1-4.39kg, mean 3.15kg.	Black 74 (77.1%) Mixed race 16 (16.7%) White 5 (5.2%) Asian 1 (1.0%) Maternal classification of race.	Study design: Cross sectional study Inclusion criteria: Convenience sampling, present when student researchers visited maternity ward between 01 Aug- 31 Dec 2016. All infants GA≥37 wks, healthy (no medication except HIV prevention). PN age > 6h, mother ≥18 years Data Collection: Visual assessment vs BiliCheck. Analysis: Chi square or Fisher exact.	Convenience sample used, which meant large proportion of infants born by Caesarean. Gender of infant not collected.	Jaundice was harder to diagnose in darker pigmented babies, therefore non-invasive bilirubin meter recommended

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Brucker & MacMullen, (1987)	USA	To determine if a bilirubinometer could be used in conjunction with home visits.	20 infants	GA: mean 39.4 wks (38-42 wks) BW mean 3,473g, (2,926 to 4,489g) PN age mean 63 h (18-78 h)	White, Black, Hispanic. Number in each group NR	Study design: Descriptive study Inclusion criteria: Convenience sampling of all infants discharged at ≤48 h, living in 25-mile radius who were visited for a home assessment by a nurse Data Collection: Undefined TCB device (forehead) vs SBr, simultaneously from heel prick Analysis: Pearson r-correlation coefficient	Very small sample. Comparative values for various skin tones are not provided although indicated ethnicity impacts TCB readings	A non-invasive bilirubinometer can be useful as a screening device, with no false negatives identified. however different action levels may be required for different ethnicities.
Campbell et al., (2011)	Canada	To compare the accuracy of SBr with BiliCheck in an ethnically diverse population of term and near-term infants, when used by various HCPs before discharge	430 newborns	GA: 38.8±1.4 wks BW 3289± 458g 236 Males (55%)	Asian 146 (34%) Caucasian 140 (33%) Latino 43 (10%) Indian 36 (8%) Black 34 (8%) Middle eastern 17 (4%) Unknown 14 (3%) Maternal ethnicity	Study design: Prospective cohort study Inclusion criteria: Term and near-term jaundiced neonates (35-37 wks) born from July 1, 2005, to March 1, 2007, Excluded if parental language barrier, admitted to NNU, congenital anomaly or received PT. Data Collection: BiliCheck (forehead) 30 mins after SBr prior to discharge home. Pertinent demographic data collected from chart review Analysis: Stata/SE 8.2. Paired t tests, Pearson's correlation and Lin's concordance coefficients, Bland-Altman.	Maternal ethnicity was surrogate measure for infant skin pigmentation. Study population representative of ethnic diversity of wider population, but small sample size for some ethnic subgroups. Laboratory SBr values not obtained using gold standard HPLC	TCB correlated but not predictive of SBr. Correlation between TCB and SBr did not vary by maternal ethnicity

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Ebbesen et al., (2002)	Denmark	To investigate the imprecision and inaccuracy of BiliCheck in both NNU and healthy term and near-term infants and to investigate optimal TCB measurement body site	488 (261 in NNU and 227 healthy term and near-term)	GA: mean (range) NNU 34.6wks (25–43); Healthy 38.6wks (35–43) BW: NNU 2521g (680–4645); healthy 3362g (2170–5000) NNU group only: Asphyxia, 6%; hypoglycaemia, 19%	Non-Northern European: NNU: 23/261 (9%) Helathy:16/227 (7%)	Study design: Prospective study Inclusion criteria: All infants PN age >24h, with clinical reason for SBr on a weekday from 1 Feb 1999 and 29 Feb 2000. Excluded if receiving PT, had received PT within previous 6hs or had skin infections, purpura or bruising Data Collection: BiliCheck (forehead, sternum, knee, & foot) vs SBr (heel prick). TCB performed by laboratory technician. Analysis: Multiple linear regression analyses adjusted for GA, gender, PN age and ethnic origin, as well as degree of illness in NNU group. Asymptotic t-test to investigate differences in correlation coefficients	Small sample size specially for those from ethnic minority backgrounds with lack of clarity on ethnic origin as only identified as non-North European	BiliCheck is suitable to screen healthy and NNU newborn infants. Ethnic origin was not found to influence BiliCheck measured on the forehead

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Ebbesen et al., (2012)	Denmark	To investigate two TCB devices, BiliCheck and Minolta JM-103 in preterm infants and establish cut-off values	133 neonates	GA 28+0 to 34+6 wks BW 1998g (1110–2764) Female 56/Male 77 Apgar score ≤7 at 5 min, 8% Infection: Suspected 22%, strongly suspected 4%	8 from non-Caucasian descent, (6 African, 2 Middle Eastern). Ethnicity was used as a proxy for skin tone.	Study design: Prospective observational study Inclusion criteria: Neonates PN age>24 h and <14 days, in NNU having SBr measured for clinical reasons. Excluded if received exchange transfusion or had Rh haemolytic disease, hepatic disease, or skin disease Data Collection: BiliCheck and JM-103 (forehead) vs SBr (heel prick) within 15 mins of TCB. Analysis: Pearson's correlation coefficient, ANOVA, Passing Bablok non-parametric regression analysis or Bland– Altman. Multivariate analyses adjusted for GA, gender, PN age, ethnicity and severe illness	Small proportion of non-White neonates.	In preterm neonates TCB measures underestimated SBr more at higher bilirubin levels The JM-103 gave lower TCB values than the BiliCheck. With JM-103 non-Caucasian neonates had significantly higher TCB.

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Goldman et al., (1982)	USA	To assess whether new guidelines from the marketing company enable the TCB device to identify infants requiring an SBr	125 infants (344 paired readings)	Group 1 (>38 wks) Group 1 (33 to 38 wks) and Group 3 (<33 wks)	Grp1:38 White, 46 Black. Grp2:9 White, 20 Black, Group 3: 5 White, 7 Black.	Study design: Cohort study Inclusion criteria: All infants having an SBr ordered by their primary physician at Jackson Memorial Hospital in June and July 1981. Exclusion criteria - infants receiving PT. Data Collection: Minolta/ Airshields Inc JM 101 (forehead) vs SBr. TCB within one hour of SBr, TCB readings obtained in duplicate, if readings did not agree, a third reading was obtained. Analysis: Correlation coefficient, linear regression equations, action levels	At a reading of 20,TCB devices varied by ±1, made harder as only integers present. Lack of accurate ethnic background. White included Hispanic and non-Hispanic. Black included Haitian, American and Hispanic Black infants. This may have made correlations lower than previous studies	JM 101 is a jaundice meter designed to identify which infants require an SBr rather than to estimate SBr level
Hannemann et al., (1979)	USA	To evaluate skin spectral reflectance measurement with special attention to natural skin pigmentation effects	103 neonates	Full-term, infants. 1-3 days of age	58 White and 45 Black	Study design: Prospective observational study Inclusion criteria: Full-term, infants. 1-3 days old. Data Collection: Skin spectral reflectance (posterior mid thorax) measure vs SBr – concurrent. Observations done between 10.00-13.00 and 14.00-17.00 to avoid disturbances to the nursery routine Analysis: Linear regression analysis. Regression analysis.	NR	Relationship between skin reflectance and SBr in full term infants is close to acceptable limits for clinical use and not obscured by skin pigmentation

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Hannemann et al., (1982)	USA	To evaluate the Minolta TCB device meter in White and Black infants	161 neonates	White: Term & no PT n=35; Term with photo-therapy n=7; 34-37wsk & no PT n=22; 34-37wks with PT n=23; <34 wks with PT n=25 Black ≥34wks & no PT – 42; <34wks with PT – 7	112 White; 49 Black	Study design: Prospective Inclusion criteria: Any infant for whom a SBr was ordered, including term and preterm, undergoing PT and not over a two-month period Data Collection: Minolta (JM 101) (forehead) vs SBr using American Optical bilirubinometer or by a diazo (Jendrassik) method. Mean of 3 TCB measurements taken Analysis: Linear regression analysis	Laboratory determinations performed by clinical laboratory rather than research laboratory. Laboratory SBr can lack accuracy and precision. Small sample size within ethnicity subgroups	TCB cannot be advocated for clinical use until more data available in a range of skin pigmentations
Hegyi et al., (1981)	USA	To evaluate the accuracy and precision of a TCB device in a sample of Black and White term infants	60 term healthy infants	PN age: White with no PT Day 1 n=8; Day 2 n=14; Day 3 n=23; Day 4 n=20; Day 5 n=11; Day 6 n=7; Day7-13 n=19. NR for other groups	43 White with no PT 9 White undergoing PT 8 Black	Study design: Prospective Inclusion criteria: Infants admitted to Monmouth MC newborn nursery Data Collection: Minolta Camera Co device (JM101) (9 sites: forehead, sternum, right upper abdomen, knee, sole, elbow, palm, upper back and lower back) vs SBr. TCB at the time of SBr. Analysis: Coefficient of variation to demonstrate reproducibility of the TCB index and Pearson correlation coefficient. TCB evaluated with respect to multiple measurements, body site, race, PT, PN age and SBr level	Only a small sample of Black included.	TCB is a valuable screening tool in healthy term infants, correlating in both Black and White infants

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Holland & Blick, (2009)	USA	To evaluate the use of a transcutaneous spectrophotom-eter that allows for non-invasive measurement of bilirubin levels.	343 term neonates	PN age ranged from 25 to 104 hrs (mean 38 hrs)	Race was recorded for 83.8%, consisting of 27.1% Caucasian, 54.7% Hispanic, 16.0% African American and 2.2% 'other'	Study design: Prospective cohort study Inclusion criteria: GA > 36 wks, not receiving PT, between 1 and 5 days old and admitted to a well-baby nursery in one of 3 affiliated hospitals Data Collection: BiliCheck (forehead and sternum) vs SBr within 10 mins Analysis: Linear regression. Averages weighted by number of participants in each site. Correlation coefficient r converted to z scores before comparison across sites. Bland- Altman difference plots	SBr analysed using different chemistry analyser in each laboratory reducing accuracy/reliability. No sample size calculations presented	TCB from the sternum can be useful for screening, but TCB from the forehead was affected by race with White infants having greater positive bias

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Jegathesan et al., (2021)	Canada	To determine agreement between TCB and SBr in preterm infants	296 preterm infants (with 856 TCB levels)	GA median 31.0 wks (IQR 28.0-33.0) BW: Mean 1558.8g ± 612.8 45.3% female PN age at measurement median 105h (IQR 68-151	Caucasian (n=110, 37.2%) SE Asian (n=27, 9.1%) S Asian (n=26, 8.8%) African or Caribbean (n=20, 6.7%) Hispanic (n=12, 4.1%) Middle Eastern (n=6, 2.0%) Inuit or First Nationals (n=5, 1.7%) Unknown (n=90, 30.4%) Maternal self-identification of ethnicity	Study design: Prospective cohort study Inclusion criteria: Preterm infant GA 24+0 to 35+6, PN age ≤10 days, admitted to one of 3 eligible NNU from Sept 2016 to June 2018. Excluded if hydrops, congenital malformation, infection, purpura or skin conditions. Data Collection: JM-105 (forehead and sternum) vs SBr from venous or capillary sites within 15 mins. TCB by trained nurses. TCB device calibrated daily. PT commenced as clinically indicated with PT lights turned off during SBr and TCB collection. Analysis: Bland-Altman plots, Lin's concordance correlation coefficient, sensitivity, specificity, PPV and NPV. ROC curves with respective AUC	Fewer TCB measurements were available prior to starting PT than after starting especially for infants born 24-28 wks who started on PT at a median of 26 hours of age.	TCB may offer a non-invasive approach to screening in preterm infants but underestimates after initiation of PT. Measures did not differ by ethnicity

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Jones et al., (2017)	USA	To compare JM-105 and BiliCheck with SBr in inpatient and outpatient settings. To determine a standard threshold for TCB to undertake SBr	176 infants	GA: 35wks (1.4%); 36wks (4.5%); 37wks (12.5%); 38wks (14.2%); 39wks (38.6%); 40wks (19.9%); 41+wks (7.4%); Unknown (1.1%) Male 74 (42%)	African American 27(15.3%) Asian 5 (2.8%) Hispanic/Latino 42 (23.9) Pacific Islander 1 (0.6) White 73 (42.5) Multiple races 28 (15.9)	Study design: Prospective cohort study Inclusion criteria: Infants with GA ≥35 wks admitted to newborn nursery or attending paediatric outpatient clinic, no history of PT, needing newborn metabolic screen or SBr ordered by HCP. Excluded if PN age >30 days, other than English- or Spanish-speaking parents, admitted to NNU, incarceration or developmental impairment of the mother Data Collection: JM-105 and BiliCheck devices (forehead and sternum) vs SBr from heel prick. Measures within 30 mins. Analysis: STATA. Pearson correlation and ANOVA	Not many infants with high SBr levels, so only 4 infants me criteria for PT. Only 2 infants recruited from the outpatient clinic (outpatients usually older)	TCB measurements on the sternum had the best overall correlation. JM-105 accuracy differed among races
Karen et al., (2009)	Switzerland	To assess the agreement between the Bilimed® and SBr in preterm and term infants of different skin tones	99 term (11 measures) and 51 preterm (68 measures)	Term infants: Median GA 39.1wks; BW 3300g; PN age 4 days Preterm (34-36 wks): Median GA 36wks; BW 2362.5g; PN age 4 days Preterm (28-33 wks): Median GA 30.3wks; BW 1360g; PN age 5 days	90 Caucasian 60 non-Caucasian (36 Hispanic or Middle-East, 9 African and 15 of Asian)	Study design: prospective cohort study Inclusion criteria: admitted to maternity ward or NNU not receiving PT. Data Collection: Bilimed (sternum) vs SBr with capillary sample. TCB immediately before or within 15 mins of SBr. Average of 3 TCB measures Analysis: Excel and StatView used. Pearson's correlation and Bland-Altman analysis, simple linear regression, multiple linear regression to identify confounders to predict SBr	All non-Caucasians reported collectively despite multiple ethnicities and races, potentially why limits of agreement were much wider in the non-Caucasian infants	Bilimed has no advantages and no better agreement in Caucasian or non-Caucasian infants

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Karon et al., (2008)	USA	To identify variables impacting TCB and SBr relationship. To define sensitivity and specificity of BiliCheck to predict SBr >75th or >95th percentile in term and near-term infants	177 neonates.	Median GA 39.0 wks	146 Caucasian, 19 Asian, 9 Hispanic and 3 African American infants. Mother's ethnicity	Study design: Prospective Inclusion criteria: Infants admitted to well-infant nursery between Aug 2006 to July 2007 if SBr ordered by HCP. Data Collection: BiliCheck (forehead) vs SBr (capillary or venous puncture) obtained by nurses within 30 mins. SBr analysed by diazo and Vitros methods. BiliCheck device calibrated before each measurement. Analysis: Median bias (TCB minus SBr) was calculated for both the diazo and Vitros methods	Mother's ethnicity proxy measure for infant ethnicity and most Caucasian ethnicity. A previous study found TCB underestimated SBr in infants with PN age >80 hours, but not enough older infants in this study to adequately address this	Difficult to assess TCB efficacy when SBr values are not standardized between different laboratory methods. Mother's ethnicity did not contribute to variability observed between TCB and SBr values
Kenny et al., (1984)	USA	To evaluate the Minolta/Air Shields device against SBr	53	NR	8 Black; 26 Caucasian; 12 Asian (unclear ethnicity 7) As recorded in the notes	Study design: Prospective Inclusion criteria: Infants admitted to newborn nursey. Excluded those with communicable diseases, currently under PT, receiving blood transfusions. Data Collection: Minolta/ AirShields (forehead, sternum and upper back) vs SBr, within 3 hours. TCB calibration checked with each test. Patient data e.g. race, diet, weight, GA, PN age, PT or transfusions, drugs derived from patient records Analysis: SPSS and MiniTab. Statistical tests NR	Small sample size, particularly in the ethnicity subgroups	Jaundice meters likely to gain increasing acceptance and replace subjective visual evaluation of jaundice routinely used in clinical settings, but some uncertainties as affected by race, degree of jaundice, skin trauma & angle of measurement

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Keren et al., (2009); Keren et al., (2009)	USA	To compare the predictive accuracy of alternative risk-assessment strategies used to screen for the risk of significant neonatal hyperbilirubinemia.	823 term and near-term (522 with paired assessments)	Sample with paired measurements (n=522) GA <38wks 10% Mean BW 3.3kg (±0.6) Male: 49%	For sample with paired measurements: Race: Black 274 (52%); S Asian 12 (2%); E Asian 48 (9%); White 170 (33%); Native Hawaiian/ Pacific Islander 1 (0.25); Other 15 (3%); NR 2 (0.4%) Ethnicity: Non-Hispanic 498 (95%); Hispanic: 20 (4%); NR: 4 (0.8%)	Study design: Prospective cohort study Inclusion criteria: Infants eligible for enrolment if care managed exclusively in the well infant nursery in an urban tertiary care hospital, GA ≥36wks & BW ≥2000g or GA≥35wks & BW ≥2500g. Excluded infants admitted to NNU or receiving >48h of IV antibiotics. Stratified sampling to oversample non-Black infants Data Collection: Visual assessment using the Kramer scale vs BiliCheck or SBr via heel prick if taken (TCB >75th percentile on hour specific nomogram or TCB ≥12mg/dl). Values paired if within 8 hours. TCB daily until discharge and at a home visit if discharged <72h. Nursery nurses with 2 years of nursery experience undertook visual assessment, who were not involved in the infant's care. Analysis: SAS and Stata. c-statistic (equivalent to AUC), Spearman correlation coefficient. Predictive ability to identify infants who developed significant hyperbilirubinemia, NPV of complete absence of jaundice	Despite attempts to oversample non-Black infants, over half of mothers were Black. Black race associated with lower risk of hyperbilirubinemia, therefore rates of hyperbilirubinemia may be low in this study Some infants started on PT before meeting outcome criterion, which prevented determining whether they would have developed the outcome of interest. However, sensitivity analyses demonstrated no major differences when these infants were excluded. TCB, SBr and visual assessments were not performed at the same time, but nomograms suggest increase of bilirubin in an 8 hour period is generally <1mg/dl. Nurses were not trained in jaundice assessment, nor were inter-rater reliability or impact of years qualified considered.	Cephalocaudal progression of jaundice should not be used to estimate SBr. However complete absence of jaundice can predict that an infant won't develop significant hyperbilirubinemia. Presence of jaundice may be more difficult to see in infants with darker skin pigmentation

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Lee et al., (2019)	USA & Bangladesh	To determine the validity of the novel icterometer (Bili-ruler) to detect clinically significant thresholds of hyperbilirubinemia and determine interrater reliability of icterometer scoring	790 newborns	USA (n=390): GA<37wks 19 (4.9%); BW<2500g 6 (1.5%); Male 189 (48.5%); TCB>15mg/dl 8 (2.1%) Bangladesh (n=400): GA and BW NR in >95% of sample. Male 263 (65.8%); TCB>15mg/dl 68 (18.4%)	USA (n=390): 53 Asian; 63 Black or African; 37 Hispanic or Latino; 228 Non-Hispanic White & 9 other or unknown Bangladesh (n=400): Asian: 400	Study design: Prospective Inclusion criteria: Newborns from a USA well-newborn nursery and a Bangladesh labour ward/paediatric inpatient ward if PN age <28days. Excluded if received PT, exchange transfusion, or if the hospital HCP deemed infant too ill (e.g. <2000 g, very preterm, serious illness). Data Collection: Icterometer (Bili-ruler) (forehead, nose, abdomen, palms and soles) vs TCB within 2 hours using a Drager JM-105 or SBr if measured within 2 h. Bili-ruler measurements performed with natural light, typically near a window and without fluorescent lighting. Two independent Bili-ruler measurements performed on each body part by a study physician or research assistant with average of the 2 scores used in analysis. Analysis: Stata 15.1. Simple descriptive statistics to calculate means, medians and ranges. Spearman correlation coefficients, Bland Altman, exploratory subgroup analyses by ethnicity conducted.	Bangladesh largely recruited readmissions to hospital, thus higher hyperbilirubinemia rates may not reflect general population. Also, few infants in Bangladesh had GA or BW data and had low rates of AN care and facility delivery Small number of dark skinned infants recruited in Boston and largely term so unknown efficacy of Biliruler in preterm. Insufficient data to validate in those with PN age >7 days. Inadequate SBr data to stratify by ethnicity.	Biliruler is a low cost, non-invasive tool with high diagnostic accuracy for neonatal jaundice screening.

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Leite et al., (2007)	Brazil	To compare TCB and SBr and analyse in by BW, race, GA and PN age, with or without PT	200 newborns	37.5% were receiving when SBr taken. GA Mean 35.7 (±3.73) wks BW mean 2,330 (±930) g PN age 3.25 (±1.73) days	White (66.5%), Mixed race (23.5%) Black (10%)	Study design: Prospective Inclusion criteria: From one of two public tertiary institutions Data Collection: BiliCheck vs SBr with maximum interval of 30 min. Analysis: SAS (version 8.02) and SPSS (version 10.0.7). Pearson's correlation coefficient, Bland-Altman graph constructed. ANCOVA with BW, PT, skin tone, PN age GA as co-variables. ROC curves constructed	Small number of Black neonates with no sampling strategy used to calculate sample sizes to address ethic differences. Definition of 'Black' unclear (e.g., Black African or any ethnicities such as Asian)	BiliCheck can be used up to plasma levels of 14mg/dl. Race did not significantly interfere with TCB measurements

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Linder et al., (1994)	Israel	To evaluate a simple method for reducing the influence of skin tone variability on TCB estimation.	389 infants	Term infants 123 appeared jaundiced between 3-5 days and 161 measures were obtained)	Infants were defined into fair, intermediate or dark skinned at the 4 hour TCB reading. Fair skinned n=41 (initial TCB <7.5) Intermediate skinned n=47 (initial TCB 7.5-8.9) Dark skinned n=35 (Initial TCB > 9)	Study design: Prospective cohort study Inclusion criteria: Apparently healthy infants, GA>37 wks, BW 2500-4500g, 5 minute Apgar ≥8. Excluded if maternal blood group O or Rh negative, or if other haemolytic antigens present, infants treated with PT, exchange transfusion or acute illness e.g. sepsis or respiratory distress syndrome. Convenience sample recruiting those not born at weekends or 00.00-06.00 Data Collection: Minolta AirShields JM101 (mid-sternum) 4 hours PN age subtracted from any subsequent TCB to circumvent influence of skin tone. Mean of 3 measures used. If infant subsequently appeared jaundiced JM-101 (mid-sternum) vs simultaneous SBr heel prick Analysis: linear regression analysis, student t tests and chi square tests. Specificity, sensitivity, PPV and NPV determined with plasma value of 12.9mg/dl taken as the upper acceptable limit.	Skin tone may darken in some ethnic groups over the first month of life limiting the effectiveness of correcting with 4 hour TCB reading.	Improved sensitivity, specificity and predictive values gained by subtracting TCB reading within 4 hours of birth

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Madlon-Kay, (1997)	USA	To determine how well parents, nurses, physicians, can detect presence and the severity of jaundice with an icterometer	171 newborns	GA mean 39 wks	Mothers race or ethnic group: White – 50%; Black – 24%; Asian – 13%; Hispanic – 9%; other – 4%	Design: Prospective Inclusion criteria: Convenience sample of infants 2 days or older, with parents who could read and understand instructions Data Collection: Visual assessment by nurse, physician and parent and Ingram icterometer (nose) vs SBr in 89 infants. Good lighting ensured Analysis: Pearson correlation, T-tests and ANOVA.	Bilirubin levels in the study were relatively low given brief hospital stay of most infants Small number ethnic minority neonates, with comparative values not clearly reported	The icterometer is a useful tool to assess jaundice, but more research needed to assess accuracy with high bilirubin levels.
Madlon-Kay, (2001)	USA	To compare 3 methods of clinical assessment of jaundice in newborns by home health nurses.	164 newborns	PA age mean 6.4 (±2.5) days	Race or ethnic group: 60% White; 18% Black; 6% Asian; 7% Hispanic; 9% other. Determined by the nurse	Study design: Prospective clinical trial. Inclusion criteria: newborn patients (≤2 wks old) in the PN department. Mothers excluded if not proficient in English or lived >10 miles from hospital. Newborns excluded if in admitted to NUU or received PT. Data Collection: Visual assessment by one of 12 home health nurses. If jaundiced detected, SBr estimated. Caudal progression noted by horizontal line of where jaundice ended and Ingram icterometer (nose) obtained. Analysis: SPSS v10.0.5. Pearson product moment correlations. Accuracy of estimates by subtracting nurse estimate from SBr. Subgroups assessed using 1-way ANOVA	Inability to unify methods used by the home health nurses to estimate bilirubin levels. Small number ethnic minority neonates, with comparative values not clearly reported	The method of evaluation that each nurse was most accustomed to was the most accurate in determining jaundice severity.

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Maisels et al., (2004)	USA	To evaluate the Minolta/Hill-Rom Air-Shields Transcutaneous Jaundice Meter model JM-103.	849 newborns	None reported	59.2% White (n=503); 29.8% Black (n=253); Other (n=93) including 4.5% E Asian, 3.8% Middle Eastern, 1.6% Indian/ Pakistani, & 1.1% Hispanic.	Study design: Prospective study Inclusion criteria: Convenience sample of neonates GA ≥35 wks from one of 3 hospitals with clinical indication of jaundice between Feb 1st 2001 and Dec 31st 2002. Data Collection: Minolta/ Hill-Rom Air-Shields JM-103 (mid-sternum {with a sub-sample also measured at the forehead}) vs SBr from heel prick within 1 hour. At two hospitals TCBx3 taken and averaged, At the other hospital only 1 reading taken. Mean of 3 TCB measurements provided highest degree of correlation. In a subset of 146 infants BiliCheck also assessed (included 125 (86%) White, 4 (3%) Black, 14 (10%) Middle Eastern, 2 E Asian or Indian/Pakistani). Analysis: Linear regression, Bland-Altman	Few infants had high SBr levels. No greater discrepancy between TCB and SBr with rising SBr in 'non-Black' population, but discrepancy increased in Black neonates. JM-103 generally overestimated SBr so dangerous clinical errors unlikely SBr measured using different methods at the 3 study hospitals, none using HPLC. SBr only obtained on jaundiced infants, but SBr distribution similar to general population so results could be applied to routine TCB. 40 infants tested as outpatients, all other TCBs obtained by research nurses or technicians. In clinical practice, with large number of HCPs, accuracy may be poorer	Correlation in Black infants not as high, but due to tendency to overestimate in Black neonates clinical error is unlikely

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Maya-Enero et al., (2021)	Spain	To determine the reliability of TCB in a multi-ethnic population based on skin tone according to our neonatal skin tone scale.	1359 patients	GA: Total sample 39.1wks; light 39.0 wks; medium-light 39.2 wks; medium-dark 39.2 wks; dark 39.3 wks. Average BW: Total sample 3175g; light 3127g; medium-light 3180g; medium-dark 3213g; dark 3308g Male: Total sample 54.5%; Light 179/337; medium-light 403/750; medium-dark 146/249; dark 13/23	Mother ethnicity: Caucasian 695 (51.1%); Asian 436 (32.1%); Hispanic 165 (12.1%); African American 63 (4.6%) (reflected ethnic diversity of population) Light skin (n=337): Caucasian 84.9%; Asian 7.7%; Hispanic 5.9%; African American 1.5% Light-medium (n=750): Caucasian 51.2%; Asian 30.5%; Hispanic 15.7%; African American 2.5% Medium-dark (n=249): Asian 71.9%; Hispanic 10.0% Caucasian 10.0%; African American 8.0% Dark (n=23): African American 82.6%; Asian 8.7%; Hispanic 8.7% Assigned to colour group at 24h using Neomar's skin tone scale	Study design: Prospective observational study Inclusion criteria: Infants undergoing routine metabolic screen at 48-72 hours. Exclusion criteria presence of sternal skin lesions and parents not consenting. No PT before assessment, Did not exclude those with ABO or Rh incompatibility Data Collection: Dräger Jaundice Meter JM-105™ (mid-sternum) vs SBr by heel prick. Analysis: One-way ANOVA, chi-squared tests, Bland-Altman plots, Loess local regression smoothing applied, Pearson correlation coefficient (r). For each colour, simple linear regression and, multiple linear regression models conducted through forward stepwise approach.	Not may dark toned infants recruited, which may make results for that group not comparable. Inter-observer reliability not assessed for colour assignment between colour groups 1–4. Most SBr and TCB levels were low because healthy newborns enrolled. Data collected at a single centre, so may not be generalizable.	TCB is a reliable assessment of SBr regardless of skin tone, but tends to overestimate in darker skinned neonates

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Merritt & Coulter, (1994)	USA	To evaluate the ability of Gosset icterometer to identify the need for SBr measurement in premature infants	90 premature infants (536 observations and 296 SBr)	GA: mean 31.7 wks (range 24 to 37 wks) BW mean 1676g (range 610 to 3610)	All but four of the babies were White.	Study design: Prospective Inclusion criteria: preterm infants GA <37wks Data Collection: Icterometer (nose) vs SBr within 30 mins of icterometer reading. 2 observers performed independent icterometer measurements. Lighting conditions also assessed, by making observations under 'warm' fluorescent light and 'cold' bluer fluorescent light Analysis: Statpak software. T tests, paired when appropriate. Linear regression analysis, ANOVA	NR	Icterometer provides cost effective screening for hyperbilirubinemia in preterm neonates

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Norman et al., (2022)	Sweden	To evaluate a new transcutaneous bilirubinometer (JAISY)	141 neonates (930 TCBs performed)	GA: 35 wks 2.8%; 36 wks 6.4%; 37 wks 5.0%; 38 wks 23.4%; 39 wks 26.2%; 40 wks 21.3%; 41 wks 12.8%; Missing 2.1% BW 3414g (±514) PN age; 1 day 7.1%; 2 days 13.5%; 3 days 16.3%; 4 days 32.6%; 5 days 17.0%; 6 days 9.2%; Missing 4.3% Female 49.6%; male 50.4% Mode of delivery Vaginal; 66.7%; instrumental 6.4%; CS 20.5%; Missing 6.4%	Skin tone: White 110 (78%); Darker (brown or Black) 28 (20%); Missing 3 (2%)	Study design: Prospective Inclusion criteria: GA≥ 35 wks, PN age < 10 days, not received PT or exchange transfusion. Excluded infants with major malformations, ongoing neonatal morbidity or NNU admission. Data Collection: JAISY (probe connected to a cell phone) (forehead and chest) vs Dräger JM-105 (forehead and chest). TCB and JAISY performed by specialist neonatal nurse or midwife familiar with JM-105. Mean of 3 readings taken for both devices. In 14/141 infants, a second test was performed. JM-105 calibrated at the factory prior to use. Analysis: Excel (Google drive) and Matlab R2020a. Pearson's correlation coefficient, Bland–Altman plots, coefficient of variation (CV) for repeated measurements and Chi-square test used to test for differences between CVs. Stratified analyses by GA, PN age, skin tone and investigator.	Selection of near-term and term infants only, preterm infants were excluded. The JAISY device was not evaluated in infants undergoing PT or with bilirubin values > 310 µmol/l JAISY compared to TCB not SBr	JAISY provided accurate and reproducible information on low to moderate intensity jaundice in newborns

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Palmer et al., (1982)	Australia	To evaluate TCB in Australia in term and preterm infants, in fair skinned and parents with varying skin pigmentation	100	GA: 55 infants <37 wks gestation 29 infants received PT	26 infants had parents of varying skin pigmentation: Italian (n=9); Filipino (n=5); Greek (n=4); Vietnamese (n=2); one each Egyptian, Indian, Indonesian, Japanese, Malaysian, Spanish	Study design: Prospective cohort study Inclusion criteria: Jaundiced infants. Data Collection: Minolta Camera co. Japan (JM-101) (mid sternum and mid abdomen) vs SBr from one of 7 predetermined sites. TCB undertaken by medical staff and nurses Analysis: Regression analysis; ANOVA	NR	The TCB gave reproducible results in term and pre term and in fair and in those with varying degrees of pigmentation
Raimondi et al., (2012)	Italy	To compare the performance of three widespread TCB devices on a multiracial population of term and late pre-term neonates.	289 neonates (343 meas-urements obtained)	GA (range) 35-41 wks BW mean 3060g range 1800-4350g PN age 4-424h Male 119: Female 170	Ethnicity: Caucasians n=253; non-Caucasians (West African) n=36	Study design: Prospective Inclusion criteria: GA >35 wks, PN age 4-75h, in well baby nursery with SBr ordered by HCP from Jan -Dec 2009. Excluded infants with Rh or ABO isoimmunisation, major congenital malformations, haemoglobinopathies or evidence of liver disease. Data Collection: BiliCheck), Bilimed and JM-103 (forehead) vs SBr (heel prick) within 20 mins of TCB. TCB performed in the ambient morning light, protected from direct sunlight and avoided areas with hair, bruises, nevi or other skin anomalies. TCB devices used in a random order. Analysis: MedCalc software v.11. Linear regression analysis, Pearson correlation coefficients, Bland-Altman plots and ROC	No mention of how they dealt with repeated measures	BiliCheck, JM-103 but not Bilimed are reliable screening tools in multiracial neonatal population

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Robertson et al., (2002)	USA	To compare a new TCB device, BiliCheck (using multiple wavelength analysis of reflectance data), with the JM-102 (two- wavelength bilirubinometer) to estimate SBr	101 infants	GA: mean 37.7±2.2 wks BW mean 3179±723g PN age 50±18h	Race reported as: 70 Caucasian, 21 African American, 6 Hispanic, 4 Asian Skin pigmentation on forehead graded using Skin Tone Chart (SpectRX). Colour defined as light (skin score 1 to 4, n=80) or dark (skin score 5 to 8, n=21)	Study design: Prospective Inclusion criteria: Newborn term infants requiring an SBr for clinical purposes from Jan 9 - Dec 18, 2000. Excluded if receiving PT. Data Collection: BiliCheck and JM-102 (forehead) vs SBr (heel prick) within 15mins. Order of TCB devices randomized. Procedures in all infants undertaken by one HCP. All infants only exposed to ambient fluorescent lighting in the nursery. JM-102 correlated to SBr within the institution. Analysis: Bland-Altman.	Approx 50% of the sample was above the 75th percentile on the nomogram (as infants were recruited when the clinician wanted a SBr due to visible jaundice)	More variability with JM-102 than BiliCheck. BiliCheck accuracy was not affected by skin tone, but the accuracy of the JM-102 was affected by skin tone

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Rubaltelli et al., (2001)	Europe: UK, France, Switzerland, Italy, Germany	To determine whether BiliCheck correlates with SBr and whether race, GA, PN age or BW impact TCB	210 infants (35 at each site)	GA ≤36 wks 19.8%; >36 wks 80.2% BW: < 2500g 16.3%; 2500-3499g 53.6%; ≥3500g 30.1% PN age: <48h 16.3%; 48-71h 24.4%; 72-95h 34.9%; ≥96h 24.4%	140 White; 31 Asian; 14 Hispanic; 9 African; 16 infants of 'other' races.	Study design: Prospective cohort study Inclusion criteria: Newborn infant GA ≥30 wks, PN age <28 days undergoing SBr as part of normal care in six institutions from 4 European countries. Excluded those with known skin disorders, receiving PT or exchange transfusions. Data Collection: BiliCheck (forehead & sternum) vs SBr (heel prick or venous) within 30 mins. Repeated on 111 infants to determine BiliCheck precision. Device calibrated according to manufacturer's instructions. Analysis: SAS software. Pearson correlation co-efficient, linear regression. Sensitivity and specificity estimated for a range of values, ROC curves. ANCOVA to determine impact of GA, PN age, race and BW	Different laboratory methods used at each study site to determine SBr. The transatlantic transport of the samples to undergo HPLC-B analysis, required extended storage, which may have affected the samples. Higher bilirubin levels noted in the HPLC-B may have been due to evaporation of the sample. Few African newborns in the sample	BiliCheck is a reliable substitute for SBr. BiliCheck accuracy is independent of race

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Rubio et al., (2017)	France	To evaluate TCB diagnostic accuracy for very preterm neonates	167 preterm infants (481 measure-ments)	GA mean 27.6±1.6wks (range 24-29.9). 24–26 wks 17%; 26–28 wks 33%; 28–30 wks 50% BW mean 985±248g (range 470-1740). Small for GA (<10th percentile) 10% Males 58% Born by CS 68% All required PT at least once: 52% required PT in first 3 days; 16% from day 4-7; 2% during the second week.	Skin tone Pale 127 (76%) Intermediate 37 (22%) Dark 3 (2%) NR how skin tone was determined	Study design: Prospective multi-centre study Inclusion criteria: Preterm infants GA <30 wks, born March 2013 to Aug 2014 in 3 hospitals. One or more SBr taken in first 15 days. Excluded those with conjugated hyperbilirubinemia. Those undergoing PT or who had already received PT were not excluded. Data Collection: BiliCheck (shielded zones of skin if PT in last 12 hour) vs SBr. TCB carried out immediately by NNU nurse. Collected GA, gender, BW, skin tone & condition, mode of birth, antibiotics, PN age, Hb. Analysis: STATA 13.0. One paired SBr-TCB measurement per child with and without PT randomly selected. Sensitivity, specificity, PPV, NPV, positive and negative likelihood ratios. Pearson linear regression analysis, Bland-Altman. Multivariate logistic regression.	SBr analysed by French standard (diazo method) rather than HPLC Information such as skin maturity, body temperature, albuminemia and capillary pH not taken into account, so they may be confounders 46% of PPV/NPV measures were during or within 12 hours of PT so caution required when interpreting as NPV may be lower	TCB could be useful in preterm neonates

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Samiee-Zafarghandy et al., (2014)	Canada	To assess the diagnostic accuracy of the JM 103 as a screening tool for neonatal jaundice and explore differential effects based on skin tone	503 infants recruited (52 missing skin tone information, leaving 451 infants, with 598 pairs of measure-ments)	GA (wks) (mean±SD); Light 39.7±1.3; Medium 38.9±1.5; Dark 39.0±1.6 Preterm (<37 wks); Light 2.0%; Medium 8.0%; Dark 8.1% BW (g) (mean±SD); Light 3382±494; Medium 3414±545; dark 3237±522 PN age (h) (mean±SD); Light 51.3±20.5; Medium 61.1±29.8; Dark 58.9±25.6 Female 46.1%; Light 54.9%; Medium 44.5%; Dark 47.3%	Skin tone: 51 light, 326 medium and 74 dark. Assigned using reference colour swatches by one of eight HCPs. 2 HCPs assigned 82.7%	Study design: Prospective Inclusion criteria: all healthy newborns GA ≥35 wks, admitted to the mother-baby unit 23 June - 14 Sept 2011. Excluded if admitted to NNU, had SBr during or within 12h of discontinuing PT or if the corresponding TCB measurement not performed. Data Collection: Konica Dräger Minolta/Air Shields JM 103 (mid sternum) vs SBr (capillary samples) within 30 mins. TCB taken by nurse 3 times and averaged. SBr performed a) at discharge; b) if visibly jaundiced; c) by HCP order; d) during outpatient appointment within 7-10 days of life Analysis: SAS software. Spearman's correlation, Bland-Altman. Random effects model to create ROC curve as repeated measures on one infant	Small number of dark and light skin tone infants with SBr >230 µmol/l Cut-offs proposed in study were for all infants with GA ≥35 wks, so might not apply to higher risk infants Inter-rater reliability for skin tone assessment; not assessed. Colours chosen to determine skin pigmentation lighter than those used by Wainer et al., which limits generalisability	JM-103 is useful screening tool regardless of infant skin tone

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Schmidt et al., (2009)	USA	To determine the accuracy and precision of TCB versus SBr in preterm neonates ≤34 wks gestation.	90 (G1 24-28wks n=30; G2 29-31 wks n=29; G3 32-34wks n=31)	GA (median): G1 26wks; G2 30wks; G3 33.5wks BW [median (range)]: G1 940 (370–1530)g; G2 1481 (890–2030)g; G3 2033 (980–2989) g Gender M:F: G1 21:9; G2 15:14; G3 15:16 CS: G1 52%; G2 55%; G3: 65% Apgar 1 min [median (range)]: G1 4 (0-9): G2 6 (1-9): G3 8 (6-9). Apgar 5 min: G1 7 (0-9): G2 8 (4-9): G3 9 (7-9)	G1: Hispanic 66%; African American 17%; Caucasian/other 17% G2: Hispanic 70%; African American 20%; Caucasian/other 10% G3: Hispanic 75%; African American 19%; Caucasian/other 6%	Study design: Prospective Inclusion criteria: Preterm neonates GA ≤34 wks, in NNU June 2007 to June 2008 if SBr ordered as part of routine care. Excluded if hydrops fetalis, severe haemolytic disease, receiving or had received PT or an exchange transfusion or were considered non-viable. Data Collection: JM-103 (sternum) vs SBr within 45 mins. TCB by one or two HCPs (two of the authors or 13 experienced NNPs). Maximum of two SBr vs TCB comparisons for each infant. JM-103 used according to manufacturer recommendations with daily quality control Analysis: Kruskal–Wallis one-way ANOVA, Chi-square or Fisher's exact tests, Bland–Altman plot and predictive indices using target SBr values of >6, >8 and >10 mg/dl. Linear regression with independent variables: BW, GA, PN age, Apgar 1 min, Apgar 5 min, umbilical artery pH	Neonatal skin thickness in preterm neonates not compared to term, so results may not be applicable to older preterm neonates when skin thickening occurs. Although the more immature neonates required a higher level of intensive care, results not stratified for severity of illness or poor peripheral perfusion within groups. Attempted forehead measurements initially, but discontinued due to clinical concern regarding the pressure applied with the device	TCB and SBr correlate significantly in preterm neonates. Routine monitoring of preterm neonates with TCB may improve clinical monitoring for hyperbilirubinemia

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Slusher et al., (2004)	Nigeria	To determine whether TCB correlates SBr in indigenous, darkly pigmented African newborns with varying degrees of skin pigmentation, some of whom had developed kernicterus.	127 infants in analysis. Site 1 (n=29); Site 2 (n=98)	BW mean ± SD (range): Total 2.72±0.62 (1.28–4.16); . Site 1 2.62±0.65 (1.40–3.70); Site 2 2.75±0.61 (1.28–4.16) Male: Total 60%; Site 1 59%; Site 2 60%.	Pigmentation: Site 1: Light 3 (18%); Medium 12 (70%); Dark 2 (12%) Site 2: Light 57 (61%); Medium 28 (30%); Dark 9 (9%) Skin pigmentation determined by the researcher who enrolled patients; 1 HCP at each site performed all skin pigmentation assessments	Study design: Prospective study Inclusion criteria: Jaundiced infants PN age ≤ 2 wks admitted to one of two hospitals from May 2000 to January 2002, regardless of place of birth, GA, or health status. Data Collection: BiliCheck (forehead) vs SBr (heel prick site 1, venepuncture site 2), done simultaneously. TCB and SBr before PT. One HCP at each site performed TCB. Analysis: Unpaired t tests, Linear regression analysis, Bland-Altman. Differences in skin pigmentation between infants the 2 sites compared by Chi square analysis	The high intercept (in mg/dl) may reflect dietary fat-soluble carotenoids (from maternal consumption of red palm oil) which pass through breastmilk to infant. Laboratory facilities and SBr collection methods varied by site, which may lead to significant SBr variation Skin pigmentation assessed by 1 HCP at each site. Inter-rater variability not measured and may account for differences in pigmentation distribution	TCB is useful and reliable for estimating SBr in pigmented neonates

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Starowicz et al., (2020)	Australia	To evaluate the accuracy of the Kejian 8000 (KJ-8000) TCB device in infants of differing ethnicity and GA	201 infants (416 paired SBr and TCB samples analysed)	GA median (range) 36 (24–41) wks. ≥37wks 45%; 32–36wks 36%; <32wks 19% BW median (range) 2668 (760–4770)g PN age median (range) 109.3 (12.6–943.1)h Male n=107 (53.2%)	Ethnicity (% infants): Caucasian 77%; Non-Caucasian 23% {Aboriginal/ Torres Strait 5%; Islander 7%; Polynesian 5%; Asian 5%; Indian subcontinent 1}	Study design: Prospective study Inclusion criteria: Term and preterm infants who had a SBr level measured from Nov 2015 to July 2017. PN age >12h. Excluded if receiving PT and for 12h after PT. Data Collection: Kejian-8000 (forehead) vs SBr (blood gas or laboratory) within 30 mins. TCB measured 3 times and average taken. HCOs trained to use the device. Analysis: Microsoft Excel software. Linear regression, Pearson correlation coefficient and square of the correlation coefficient (r2). Bland–Altman plot	Term infants' data difficult to capture within outpatient due to TCB often beyond the 30 mins required. Multiple readings collected from preterm inpatients, so fewer term paired samples. KJ-8000 not recommended in preterm infants, so large contribution of this subgroup dilutes overall correlation and agreement for the total population	Correlation of KJ-8000 with SBr was not as good as reported for other bilirubinometers, but may have some value for screening in term infants

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Szabo et al., (2004a)	Switzerland	To clinically assess jaundice by Kramer method and 2 TCB devices vs SBr in healthy preterm infants with GA 34-36.9 wks. To analyse whether the three non-invasive methods are affected by GA, BW, skin tone or ambient light and establish cut-off values to predict SBr >190µmol/l with >95% confidence.	69 healthy infants (107 measure-ments)	GA 34-36.9wks (median 35.7wks). BW median birthweight 2530g (range 2020–3630 g).	White (n=60); 'non-White' (n=9) {Asian n=5; Indian n=1; African origin n=3}	Study design: Prospective study. Inclusion criteria: Healthy preterm newborns GA 34-36.9 wks, PN age ≤6days, BW ≥2000 g recruited from maternity ward or NNU 1 March 2002 to 28 Feb 2003, Excluded jaundice above zone 3 according to Kramer, within 48h), positive direct Coombs test, birthweight <10th percentile for GA, any sign or symptom of illness such as infection, respiratory distress or feeding intolerance. Excluded if PT already started Data Collection: Visual assessment with Kramer and BiliCheck (forehead and sternum) and Minolta Airshields JM-102 (sternum - mean of 2 measurements) vs SBr. Observed every 8 hours. SBr taken when Kramer zone 3 reached or with metabolic screening test at PN age 72-96h. Avoided skin areas with bruising, excessively hairy or hypermelanotic. Lighting conditions recorded: artificial light/ cloudy daylight /sunshine. The PI performed TCB using both devices. Analysis: SPSS software. ROC curve, linear regression. Multiple comparisons per infant included and considered independent, as time gap of >24h and different lighting, body weight and nurse involved. Mann- Whitney U-test assessed skin tone, Kruskal-Wallis test ambient lighting	Strength was that all TCB device measurements were undertaken by one PI Very small sample of 'non-White' infants	JM 102 showed best performance by ROC, followed by BiliCheck and visual assessment least effective. Skin tone effects BiliCheck and visual assessment

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Szabo et al., (2004b)	Switzerland	To compare the clinical assessment of jaundice using Kramer and two TCB devices with SBr and to investigate the effect of skin tone and ambient light on the three non-invasive methods	140 infants	GA median 39wks (range 37+0 to 41+6) BW median 3320 g (range 2050–4400 g). 7% of infants small GA (BW <10th percentile)	92 White; 48 'non-White' (18 Asian, 30 Indian or African)	Study design: Prospective study Inclusion criteria: Full-term healthy newborns, BW≥2000g, PN age ≤6 days, recruited from maternity ward from July 1st, 2002 to June 30th, 2003. Exclusion: haemolysis, jaundice within the first 36h, PT. Data Collection: Visual assessment with Kramer and BiliCheck (forehead & sternum) and Minolta Airshields JM-102 (2 measurements from sternum higher one included) vs SBr (heel prick) within 10 mins. Kramer assessed by nurse in charge and by PI every 12 hours. PI performed TCB measurements avoiding skin areas with bruising, excessive hair or hypermelanotic. Ambient light, daylight or fluorescent recorded. SBr taken when Kramer reached zone 3 or with routine metabolic screening at 72h. Analysis: SPSS software. Linear regression, intraclass correlation coefficient, Bland-Altman, sensitivity, specificity, AUC in ROC curve	NR	BiliCheck more variable in 'non-White' than JM-102, with BiliCheck underestimating in 'non-White' infants

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Taylor et al., (2015)	USA	To characterize discrepancies between TCB and SBr among newborns receiving care at multiple nursery sites across the USA	4994 newborns (8319 TCB measure-ments; 769 newborns with at least 1 paired TCB/ SBr)	GA mean ± SD 39.1±1.4wks BW mean ± SD 3333±500g Vaginal delivery 74.5% Feeding: Exclusively breast 46.4%; mixed 41.5%; formula only 12.1%	Race/ethnicity of newborns collected if known, otherwise race/ethnicity of mother recorded White 64.2%); 'non-White 35.8% (American Indian/ Alaska Native 1.0%; African American 24.9%; Asian 7.7%; Pacific Islander/ Native Hawaiian 0.5%; Multiple races 1.6%; Other 0.2%) (Race missing in 143 cases) 24.5% Hispanic ethnicity.	Study design: Retrospective study Inclusion criteria: Survey sent to 82 nurseries to determine how jaundice assessed. Those that screened all /virtually all neonates with TCB invited to participate (27/38 eligible sites provided data). Eligibility infants GA ≥35wks, admission in 2xtwo-week periods (Jan 2012 to June 2013), no Rh isoimmunization, at least 1 TCB measurement. Not receiving PT. Data Collection: BiliCheck or JM-103 or both (measured forehead/ sternum /multiple /other). vs SBr within 2 hours. Data abstracted from medical records including results of all TCB tests performed PN age <120h Analysis: Descriptive statistics and correlation, variables statistically associated with a TCB-SBr difference in bivariate analyses included in multivariate model. Generalised Estimating Equation techniques used due to multiple measurements from the same newborn. Logistic regression	Decision to obtain SBr varied by site and HCP, so SBr-TCB differences may not be representative of all TCB measurements. 18.7% cases missing race. Only compared White vs other & African vs other Pre-requisite survey collected clinical practice, therefore may not reflect actual practice at time of TCB & SBr TCB assessed in inpatient setting, so may not be generalizable to outpatient settings. Data combined from multiple TCB device brands, may account for low correlation.	TCB provides a reasonable estimate of SBr. Differences between TCB-SBr levels are increased in African American neonates

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Taylor et al., (2017)	USA	To assess accuracy of a technology based on the analysis of digital images of newborns obtained using a smartphone application called BiliCam.	530 newborns had complete BiliCam images	PN age mean 75.2±28.9h (range 12-163h)	Race: White 55.4%; African American 20.8%; Asian American 21.2%; Pacific Islander/ Hawaiian native 2.8%; American Indian/ Alaskan native 2.8%; Other 1.5%. (Multiple races listed by some participants) Ethnicity: Hispanic /Latino 26.3% Infant ethnicity and race provided by parents at enrolment	Study design: Prospective study Inclusion criteria: Healthy newborn infants PN age <7 days, GA ≥35 wks in 7 sites across USA Oct 2014- July 2016. Excluded if received PT. At 2 sites newborns enrolled when ≤24 hours old, with follow-up visit at 3 to 5 days to obtain BiliCam images and SBr. 2 other sites BiliCam images obtained if SBr requested due to clinical jaundice; 2 sites BiliCam images obtained when SBr routinely screened of or if neonate clinically jaundiced. Data Collection: BiliCheck or Draeger JM-103 vs BiliCam (sternum 3 flash images 3 without flash with colour calibration card in situ) vs SBr Attempts to take BiliCam images within 2 hours of SBr. Analysis: Correlation; subgroup analyses for racial and ethnic groups. Bland-Altman analysis.	SBr levels measure by different methods in different laboratories, resulting in more variability in SBr measurements than is generally appreciated. Correlations between BiliCam and SBr may have been higher if only single laboratory used to determine SBr	The Smartphone app provided accurate estimates of SBr showing an inexpensive technology could be used to screen newborns
Thomson et al., (2008)	UK	To determine if alternative TCB devices to the BiliCheck are of sufficient clinical equivalence	235 newborns (all had JM-103 check and 45 also had BiliCheck)	PN age median 21h.	60% Asian from the Indian subcontinent; 21% White; 14% Black; 5% mixed race/other ethnicity	Study design: Descriptive study Inclusion criteria: GA >35wk. Data Collection Konica Minolta JM-103 with subsample also having BiliCheck with SBr taken if TCB >220 µmol/l or above local PT chart treatment line. Analysis: Correlation, Bland–Altman plots	NR	BiliCheck provided accurate SBr estimate across all ethnicities, not impacted by skin tone unlike JM-103. JM-103 deemed clinically equivalent but underestimated in White and overestimated in Black infants.

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
van der Geest et al., (2021)	Netherlands	To assess knowledge and skills of maternity care assistants regarding neonatal hyper-bilirubinemia	1465 survey responses	NA	NR	Study design Descriptive study – survey Inclusion criteria: Maternity care assistants in the Netherlands Data Collection: Nationwide survey sent to 9065 maternity care assistants disseminated via weblink on 29 Jan 2018, promoted on social media page twice, survey closed 28th Feb 2018. A subgroup of maternity care assistants undertook training in neonatal hyperbilirubinemia at 7 birth centres, with 99 completing pre and post training questionnaires. Analysis: Descriptive statistics, Wilcoxon, Mann Whitney, chi square or Fisher exact tests	Low response rate (16%) may have introduced nonresponse bias, with respondents more interested in the topic leading to overestimation of knowledge. Data was not collected among other maternity HCPs, e.g. midwives, obstetric nurses, obstetricians and paediatricians. Six knowledge questions may have not effectively discriminated high/ low knowledge Participants were asked not to look up correct answers but unable to assess this within an online survey	Background knowledge of neonatal hyperbilirubinemia among maternity care assistants was adequate and improved with training. Estimation of SBr from skin colour was inadequate, with no association between self perceived capacity to assess jaundice and actual ability

Author	Country	Aim	Sample size	Neonatal characteristics	Ethnicity details	Methodology	Limitations	Study Conclusion
Wainer et al., (2009)	Canada	To evaluate the performance of the Konica Minolta/Air-Shields JM-103 jaundice meter according to neonatal skin tone	938 infants (but only 774 SBr/TCB pairs for analysis)	GA: Mean 39.1±1.2wks. (Range 37-42) BW: Mean 3166±447g (range 2096-4765g) Male 48.7%: Female 51.3%	Maternal declared ethnicity of infant: Asian 41.3%; Middle Eastern 9.5%; Aboriginal 3.0%; Black 4.6%; White 41.7%; 13 cases NR Infants categorized by skin tone into 3 groups ('light'/ 'medium' /'dark') using two reference cosmetic colours. Categorized by study coordinator in 79% of infants. Light n=347 (44.0%); Medium n=412 (52.3%); Dark n=15 (1.9%). N=14 missing skin tone data excluded from analysis	Study design: Prospective cohort study Inclusion criteria: Healthy infants GA ≥37wks, attending a regional centre from 1 Dec, 2004 to 31 Dec, 2005 (excluding July & Aug 2005). Excluded if home address outside geographical area served by Public Health Nurses (PHN), major malformation, received prior PT, admitted to NNU for >24h Data Collection: Konica Minolta/Air-Shields JM-103 (forehead) vs SBr with routine metabolic screen at 24h, if TCB value within 20μmol/l of PT level or if clinically necessary. TCB obtained at 12, 24, 48, 72h and 7 days. TCB & SBr paired if within 60 mins. TCB devices calibrated according to manufacturer's instructions. Change in approved SBr calibrators during study accounted for in analysis. Infant weight and feeding method noted Analysis: Stata 9.0. Multivariate linear regression. Medium skin tone referent. Bradley-Blackwood analyses (based on Bland–Altman), Lin concordance correlation coefficient	Small number of infants with dark skin tone so unable to analyse separately, therefore more data required. Skin tone not re-assessed after enrolment, so impact of any change in skin tone during the early neonatal period on device performance not evaluated. Inter-observer bias minimized as >79% of infants allocated by one person. Devices regularly calibrated according to the manufacturer's recommendations, but no inter-device correlation testing was performed	TCB and SBr measurements had greater disagreement at high SBr concentrations. Too few dark skin toned infants were included to full evaluate performance in this group

USA – United States of America
UK – United Kingdom
NA – not applicable
NK – not known
NR – not reported
GA – gestational age
BW – birthweight
Wks – weeks
AN – antenatal
MA – Mexican American
Hb - haemoglobin
Yrs – years
SGA – small for gestational age
CS – Caesarean section
ROC – receiver operating characteristics
AUC – Area under the curve
PPV - positive predictive value
NPV – negative predictive value
PDA - patent ductus arteriosus
IQR – interquartile range
NNU – neonatal unit
Min – minute
h – hours
HCP – healthcare professionals
MC – medical centre
IV – intravenous
S – South
E - East
AN – antenatal
SpO2 – oxygen saturation using pulse oximetry

Appendix 4.

Critical appraisal results for each study included within the healthcare professional's systematic review

Author	Clear research question	Data collected addressed research question	Quantitative non-randomised study					Quantitative descriptive study				
			Q1	Q2	Q3	Q4	Q5	Q1	Q2	Q3	Q4	Q5
Apgar studies												
Chubb et al., (2022)	✓	✓						✓	?	?	X	✓
Li et al., (2013)	✓	✓	✓	✓	✓	✓	✓					
Mihoko Doyle et al., (2003)	✓	✓	✓	✓	✓	?	✓					
Serunian & Broman, (1975)	✓	✓	✓	✓	?	✓	✓					
Shankaran et al., (2004)	✓	✓	?	✓	✓	✓	✓					
Wolf et al., (1997)	✓	?						✓	✓	X	✓	X
Cyanosis studies												
Dawson et al., (2015)	✓	✓						✓	?	✓	?	✓
Goldman et al., (1973)	✓	✓	?	✓	✓	✓	✓					
Vesoulis et al., (2022)	✓	✓	✓	✓	X	X	✓					
Jaundice studies												
Afanetti et al., (2014)	✓	✓	✓	✓	✓	✓	✓					
Ahmed et al., (2010)	✓	✓	?	✓	✓	X	✓					
Aune et al., (2020)	✓	✓	?	✓	✓	?	✓					
Bhutani et al., (2000)	✓	✓	?	✓	✓	✓	✓					
Bourchier et al., (1987)	?	?	?	X	?	?	X					
Brits et al., (2018)	✓	✓	✓	✓	✓	✓	✓					
Brucker & MacMullen, (1987)	✓	✓	?	✓	✓	✓	✓					
Campbell et al., (2011)	✓	✓	✓	✓	✓	✓	✓					

Author	Clear research question	Data collected addressed research question	Quantitative non-randomised study					Quantitative descriptive study				
			Q1	Q2	Q3	Q4	Q5	Q1	Q2	Q3	Q4	Q5
Ebbesen et al., (2002)	✓	✓	✓	✓	✓	✓	✓					
Ebbesen et al., (2012)	✓	✓	✓	✓	✓	✓	✓					
Goldman et al., (1982)	✓	✓	✓	✓	✓	✓	✓					
Hannemann et al., (1979)	✓	✓	X	✓	✓	?	✓					
Hannemann et al., (1982)	✓	✓	✓	✓	✓	✓	✓					
Hegyi et al., (1981)	✓	✓	?	✓	✓	✓	✓					
Holland & Blick, (2009)	✓	✓	?	✓	✓	✓	✓					
Jegathesan et al., (2021)	✓	✓	?	✓	✓	✓	✓					
Jones et al., (2017)	✓	✓	✓	✓	✓	✓	✓					
Karen et al., (2009)	✓	✓	✓	✓	✓	✓	✓					
Karon et al., (2008)	✓	✓	?	✓	✓	✓	✓					
Kenny et al., (1984)	✓	✓	?	✓	✓	✓	✓					
Keren et al., (2009); Keren et al., (2009)	✓	✓	✓	✓	✓	✓	✓					
Lee et al., (2019)	✓	✓	✓	✓	✓	✓	✓					
Leite et al., (2007)	✓	✓	?	✓	✓	✓	✓					
Linder et al., (1994)	✓	✓	?	✓	✓	X	✓					
Maldon-Kay, (1997)	✓	✓						?	?	✓	✓	✓
Madlon-Kay, (2001)	✓	✓	?	✓	✓	✓	✓					
Maisels et al., (2004)	✓	✓	✓	✓	✓	✓	✓					
Maya-Enero et al., (2021)	✓	✓	?	✓	✓	?	✓					
Merritt & Coulter, (1994)	✓	✓	?	✓	✓	✓	✓					
Norman et al., (2022)	✓	✓	?	✓	✓	✓	?					
Palmer et al., (1982)	✓	✓	?	✓	X	✓	✓					
Raimondi et al., (2012)	✓	✓	?	✓	✓	X	✓					

Author	Clear research question	Data collected addressed research question	Quantitative non-randomised study					Quantitative descriptive study				
			Q1	Q2	Q3	Q4	Q5	Q1	Q2	Q3	Q4	Q5
Robertson et al., (2002)	✓	✓	?	✓	✓	X	✓					
Rubaltelli et al., (2001)	✓	✓	?	✓	✓	✓	✓					
Rubio et al., (2017)	✓	✓	?	✓	?	✓	✓					
Samiee-Zafarghandy et al., (2014)	✓	✓	✓	✓	✓	✓	✓					
Schmidt et al., (2009)	✓	✓	?	✓	?	✓	✓					
Slusher et al., (2004)	✓	✓	?	✓	✓	✓	✓					
Starowicz et al., (2020)	✓	✓	X	✓	✓	✓	✓					
Szabo et al., (2004a)	✓	✓	?	✓	✓	✓	✓					
Szabo et al., (2004b)	✓	✓	?	✓	✓	✓	✓					
Taylor et al., (2015)	✓	✓	?	?	X	✓	?					
Taylor et al., (2017)	✓	✓	?	✓	✓	X	✓					
Thomson et al., (2008)	✓	✓						?	✓	✓	?	✓
van der Geest et al., (2021)	✓	✓						✓	✓	✓	?	✓
Wainer et al., (2009)	✓	✓	?	✓	X	X	✓					

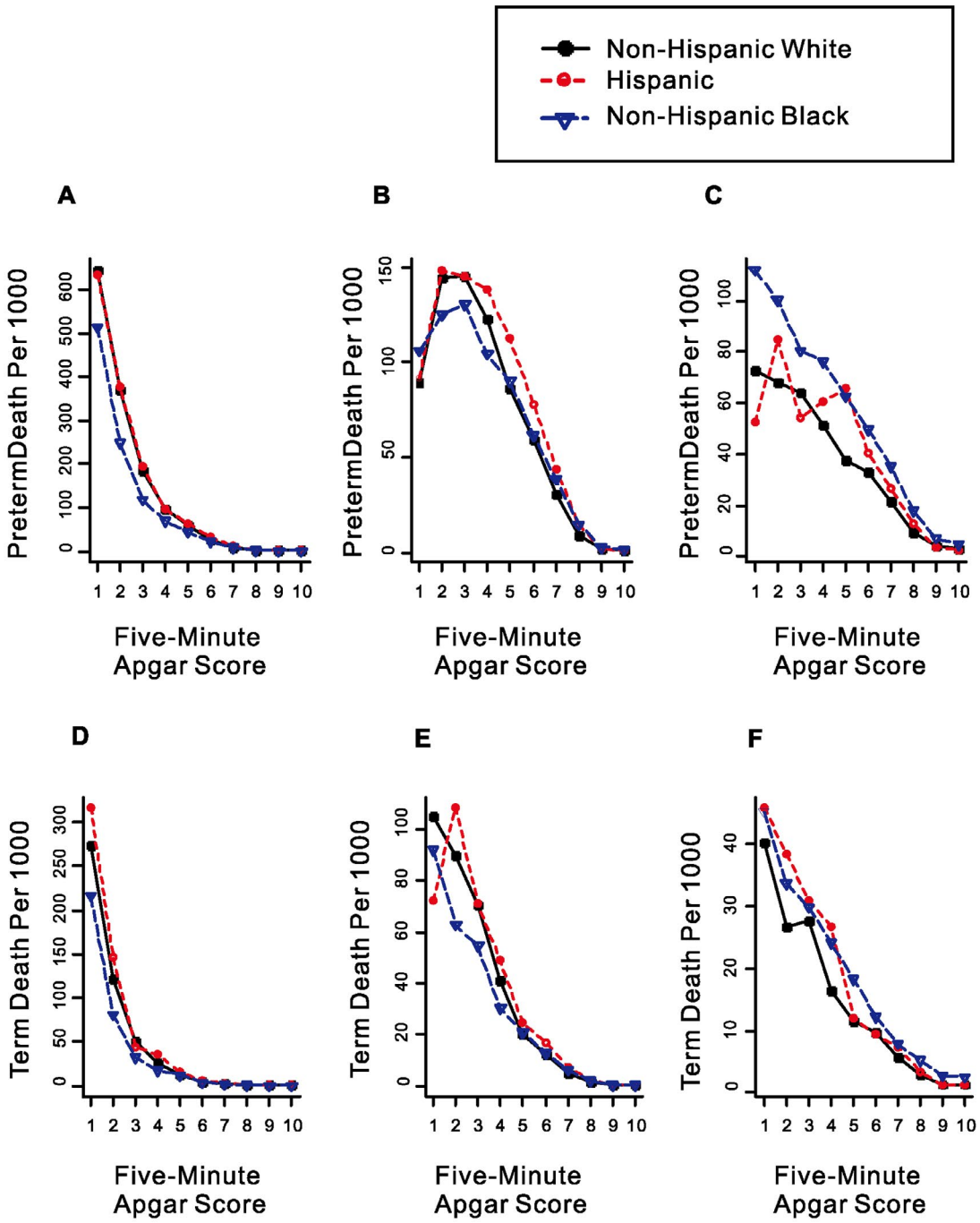
✓ - adequately addressed
X – not adequately addressed
? unclear risk of bias
N/A – not applicable

Quantitative non-randomised study questions:
Q1 - Are the participants representative of the target population?
Q2 - Are measurements appropriate regarding both the outcome and intervention (or exposure)?
Q3 - Are there complete outcome data?
Q4 - Are the confounders accounted for in the design and analysis?
Q5 - During the study period, is the intervention administered (or exposure occurred) as intended?

Quantitative descriptive study questions:
Q1 - Is the sampling strategy relevant to address the research question?
Q2 - Is the sample representative of the target population?
Q3 - Are the measurements appropriate?
Q4 - Is the risk of nonresponse bias low?
Q5 - Is the statistical analysis appropriate to answer the research question?

Appendix 5: Mortality rates for Black, White and Hispanic neonates according to 5 minute Apgar score

From Li et al., 2013.



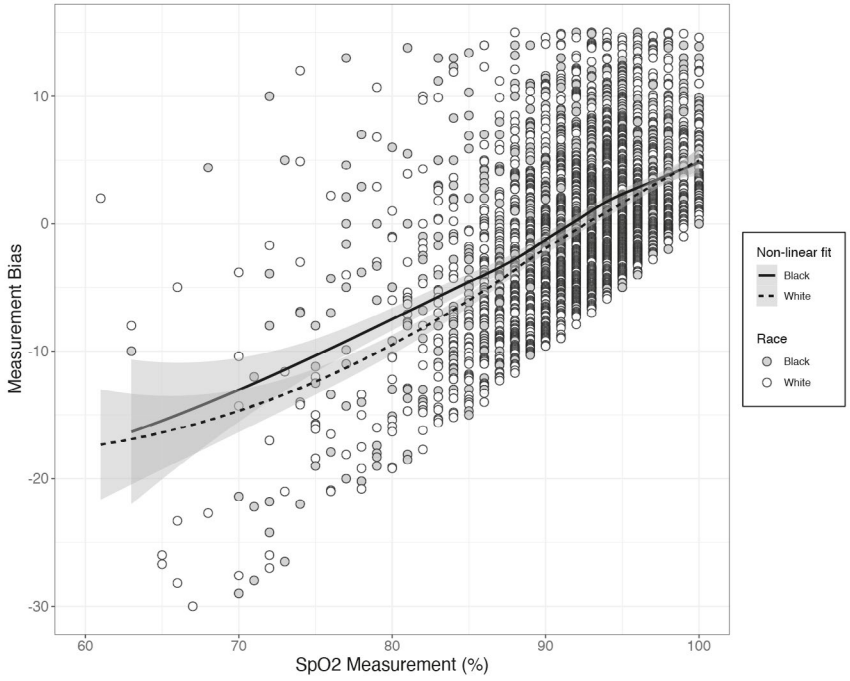
Mortality rates for White, Black and Hispanic infants by five-minute Apgar score.

Presented according to gestational age. Preterm (24–36 weeks) presented in Figures A–C and term (37–41 weeks) presented in Figures D–F.

Time of Infant Death ≤1 day presented in Figures A and D. Time of Infant Death between 2-27 days presented in Figures B and E. Time of Infant Death ≥28 days presented in Figures C and F.

Appendix 6: Mean bias (pulse oximetry saturation minus arterial oxygen saturation) clustered by pulse oximetry saturation and race

From Vesoulis et al., 2022



Black neonates shown as grey dots and White neonates as white dots.

Non-linear regression lines shown in solid line for Black neonates and dashed for White neonates.

Shaded area shows 95% confidence interval.

At SpO2 saturations <95% there was a consistent difference between White and Black infants, with Black infants having higher pulse oximetry saturations at each arterial oxygen saturation level.

Appendix 7: Full Bland Altman results

Presented as Mean bias ± imprecision (95% CI/ limits of agreement)

	First author	Gestation at birth	Sample size	Overall	White	Black/ African American/ West African	Asian/ Asian American	Hispanic	Mixed ethnicity/ "non-White"	"Medium"	"Dark"	Under-/over-estimate/variability significance
BilliCheck	Slusher et al., (2004)	Term and preterm	127	Bias of 8.55 ±130.0 µmol/l	18.8 ±92.3 µmol/l (-73.5 to 111.1) Y					13.7 ±133.4 µmol/l (-119.7 to 147.1)	-3.4 ±198.4µmol/l (-201.8 to 194.9)	Variability significantly increased with pigmentation
	Szabo et al., (2004a)	Preterm (34-36+6 wks)	69	Sternum 10 ±31µmol/l Forehead -8 ±33µmol/l					Underestimated SBr by 33 µmol/l			Significantly under-estimated in 'non-White'
	Szabo et al., (2004b)	Term	140		-16 ± 52µmol/l (-68 to 36)				-25 ±68µmol/l (-93 to 43)			Under-estimated in 'non-White' (p<0.001) Variability increased in 'non-White' (p=0.02)
	Thomson et al., (2008)	Term and late preterm (<35 wks)	45	-10 (95% CI -19.0 to -1.6 µmol/l			-10± 13.5 µmol/l (95% CI -23.5 to 3.5)					Mean difference NS (p=0.94)
	Karon et al., (2008)	Term and Preterm (≥32 wks)	177									Variability NS (p=0.2207)
	Holland & Blick, (2009)	"Term" (≥36 wks)	343		Forehead 29.1 ±10.3µmol/l (18.8 to 39.3) Sternum 25.7 ±6.9µmol/l (18.8 to 32.5)	Forehead 6.8 ±17.1µmol/l (-10.3 to 23.9) Sternum 35.2µmol/l (22.3 to 47.9)		Forehead 6.8 ±8.6 µmol/l (-1.7 to 15.4) Sternum 32.5 µmol/l (27.4 to 39.3)				Variability significantly different (p<0.01) - forehead. NS - sternum
	Ahmed et al., (2010)	Preterm (<35 wks)	57									Mean difference not significant (SBr-TCB). Mann Whitney rank sums 'non-White' 1207, White 15,629, p=NS
	Raimondi et al., (2012) ∫	Term and late preterm (≥35 wks)	289	3.4 µmol/l (-49.6 to 58.1)		0.0 ±65.0 µmol/l (-65.0 to 65.0)						NR

	First author	Gestation at birth	Sample size	Overall	White	Black/ African American/ West African	Asian/ Asian American	Hispanic	Mixed ethnicity/ "non-White"	"Medium"	"Dark"	Under-/over-estimate/variability significance
JM-102	Szabo et al., (2004b)	Term	140		± 2 SD equivalent to 56µmol/l				± 2 SD equivalent to 56µmol/l			Variability NS
JM-103	Maisels et al., (2004) ∫	Term and late preterm (≥35 wks)	849	-4.5 ± 10.9 µmol/	Approx 6.8 ±29.1µmol/l (-22.2 to 35.9)	Approx 26.5 ±55.6 µmol/l (-29.1 to 82.1)	Approx -3.4 ±37.6 µmol/ (-41.0 to 34.2) §					NR
	Thomson et al., (2008)	Term and late preterm (<35 wks)	235	-11 ±5.9µmol/l (-16.9 to -5.2)	-38 ±8.3µmol/l (-46.3 to -29.7)	28 ±10.8µmol/l (17.2 to 38.8)	-1 ±6.9µmol/l (-7.9 to 5.9)l					Underestimated in White and overestimated in Black (p<0.001)
	Wainer et al., (2009) ∫	Term	938 (744 cases in analysis)	(-58.32 to 28.20 µmol/l)	(-60.00 to 13.68 µmol/l) Y					(-51.12 to 31.97 µmol/l)		NR
	Raimondi et al., (2012) ∫	Term and late preterm (≥35 wks)	289	-8.6 ± 61.6µmol/l (-70.1 to 53.0)		15.4 µmol/l (-56.4 to 88.9)						NR
	Afanetti et al., (2014) ∫	Term and preterm	86		-1.0 ±56.7 µmol/l (-57.7 to 55.7)				9.3 ±60.6 µmol/l (-51.3 to 69.9)			NR
	Samiee-Za-farghandy et al., (2014) ∫	Term and late preterm (≥35 wks)	451	-13.3 ± 52.8 µmol/l (-66.9 to 39.4)	-15.0 ±34.0µmol/l (-49 to 19) Y					-18.7 ±50.4 µmol/l (-69.1 to 31.7)	11.6 ±44.6 µmol/l (-33.0 to 56.2)	TCB underestimated SBr in light and medium skin. Overestimated dark skin. NS but deemed clinically significant

	First author	Gestation at birth	Sample size	Overall	White	Black/ African American/ West African	Asian/ Asian American	Hispanic	Mixed ethnicity/ "non-White"	"Medium"	"Dark"	Under-/over-estimate/variability significance
JM-105	Jones et al., (2017)	Term and late preterm (≥35 wks)	176	Sternum: -3.6 ±19.7µmol/l Forehead: 3.8 ± 20.4µmol/l	Sternum: -9.8 ±3.5µmol/l (-13.3 to -6.3) Forehead: NS	Sternum: 15.22 ±8.7µmol/l (6.5 to 23.8) Forehead: 24.7 ±8.1µmol/l (16.6 to 32.8)		Sternum): NS Forehead: 9.2 ± 4.6 µmol/l (4.6 to 13.7)				On Sternum: Overall p<0.001 underestimated White, overestimated Black. Forehead: White NS (p=0.18), overestimated Black and Hispanic
	Maya-Enero et al., (2021) ∫	Term and preterm (≥31 wks)	1359	20.2 ± 51.3µmol/ (-31.1 to 71.7)	12.0 ±53.4µmol/ (-41.4 to 65.3) Y					Medium-light: 18.5 ±49.6µmol/ (-31.1to 68.1) Medium-dark: 32.4 ±54.6µmol/ (-22.2 to 87.0)	31.9 ±55.5µmol/ (-23.6 to 87.4)	Bias increased between light and medium-light, medium-dark and dark (all p<0.001). Significant between medium-light and medium-dark (p<0.001) and dark (p=0.023)
	Jegathe-san et al., (2021)	Preterm	296	Whole sample: -24.5 ±78.8µmol/l (-103.3 to 54.3) Prior to phototherapy 1.6 ±75.0µmol/l (-73.4 to 76.5)	-30.6µmol/l (95% limits of agreement -106.2 to 45.1)	-12.0 ±64µmol/l (-76.0 to 52.0)	Southeast Asian: -25.2 ±75.1 µmol/l (-100.3 to 49.9) South Asian: -9.8 ±84.3 µmol/l (-94.1 to 74.5)					Mean bias NS
Kejian KJ-8000	Starowicz et al., (2020)	Term and preterm	201	–5.9 ±95.5 µmol/L (–101.4, 89.6)	- 10.6 ±86.9µmol/l (-97.4 to 76.3)				8.7 ±114.7µmol/l (-105.9 to 123.4)			Overestimated and less precise in darker skinned infants

	First author	Gestation at birth	Sample size	Overall	White	Black/ African American/ West African	Asian/ Asian American	Hispanic	Mixed ethnicity/ "non-White"	"Medium"	"Dark"	Under-/over-estimate/variability significance
Bilimed	Karen et al., (2009) ∫	Term and preterm (≥28 wks)	150		-16 ± 121μmol/l (-137 to 105)				-10 ± 174μmol/l (-184 to 164)			Limits of agreement wider in 'non-White'
	Raimondi et al., (2012) ∫	Term and late preterm (≥35 wks)	289	13.7μmol/l (-61.6 to 90.6)		12.0μmol/l (-80.4 to 102.6)						NR
Mixed*	Taylor et al., (2015)	Term and late preterm (≥35 wks)	769	80.4 ± 60.8μmol/l	10.9 ± 58.1μmol/l	26.7 ±49.9μmol/l		11.5 ±57.1μmol/l				TCB-SBr difference. Significant in African American ($p<0.001$); White ($p<0.001$) and Hispanic ethnicity ($p=0.02$)

Data presented as mg/dl were converted to μmol/l

∫ - Some studies presented TCB-SBr and others SBr-TCB, for ease of interpretation results were converted to all show TCB-SBr

§ - included Asian, Indian Pakistani, Middle Eastern and Hispanic

Y - Study comparison was "light/ fair skin pigmentation" rather than White per se

Mixed* - *Combination of BiliCheck and JM-103

wks - weeks

Appendix 8: Sensitivity, specificity, positive predictive value, negative predictive value and blood tests avoided

	First author	Gestation at birth	Sample size	Sensitivity/specificity/Positive predictive value/Negative predictive value					
				Overall	White	Black/ African American	Mixed ethnicity/"non-White"	Medium	Dark
JM 101	Goldman et al., (1982)	Term and preterm	125		For > 38 weeks: FP=2.1%; FN=3.2% For 33-38 weeks: FP=2.4%; FN 0% For < 33 weeks: FP=0%; FN=0% ◊	For > 38 weeks: FP=7.4%; FN=6.5% For 33-38 weeks: FP=5.9%; FN 0% For < 33 weeks: FP=0%; FN=0%			
	Linder et al., (1994)	Term	389	Ability to predict SBr>220µmol/l 73%/96%/ 76%/96%	Ability to predict SBr>220µmol/l 67%/96%/ 50%/98% Y			Ability to predict SBr>220µmol/l 92%/98%/ 92%/98%	Ability to predict SBr>220µmol/l 80%/100%/ 100%/ 98%
JM-103	Samiee-Zafarghandy et al., (2014)	Term and late preterm (≥35 wks)	451		SBR cut off 110µmol/l TCB cut off 85 µmol/l 96.7%/80.9%/ 87.9%/ 94.4% Optimal SBR cut off 170µmol/l Optimal TCB cut off 145 µmol/l 100%/ 95%/ 84.6%/ 100% SBR cut off 230µmol/l NA/NA/ NA/NA Y			SBR cut off 110 µmol/l Optimal TCB cut off 90 µmol/l 95.2%/81.6%/ 92.3%/ 87.9% SBR cut off 170µmol/l Optimal TCB cut off 130 µmol/l 95.9%/ 75.1%/ 69.5%/ 96.9% SBR cut off 230µmol/l Optimal TCB cut off 155 µmol/l 96.2%/75.5%/ 43.2%/ 99.0%	SBR cut off 110 µmol/l Optimal TCB cut off 110 µmol/l 95.3%/80.6%/ 87.2%/ 92.6% SBR cut off 170µmol/l Optimal TCB cut off 170 µmol/l 100%/ 89.3%/ 75.0%/ 100% SBR cut off 230µmol/l Optimal TCB cut off 210 µmol/l 100%/ 91.0%/ 53.8%/ 100%

	First author	Gestation at birth	Sample size	Sensitivity/specificity/Positive predictive value/Negative predictive value					
				Overall	White	Black/ African American	Mixed ethnicity/"non-White"	Medium	Dark
JM-103	Wainer et al., (2009)	Term only	938 (744 in analysis)	SBr>150µmol/l TCB 70 µmol/l 100%/24.9%/27.4%/ 100% BTA 19.4% TCB 80 µmol/l 99.4%/34.3%/30.0%/ 99.5% BTA 26.9% TCB 190 µmol/l 38.6%/99.7%/97.1%/ 85.1% BTA 91.2% TCB 200 µmol/l 31.6%/100%/100%/ 83.8% BTA 93.0% SBR>200µmol/l TCB 130 µmol/l 100%/80.8%/32.7%/ 100% BTA 73.9% TCB 140 µmol/l 98.5%/85.7%/39.2%/ 99.8% BTA 78.6% TCB 220 µmol/l 54.5%/99.7%/94.7%/ 95.9% BTA 95.1% TCB 230 µmol/l 45.5%/100%/100%/ 95.2% BTA 96.1% SBr>250µmol/l TCB 160 µmol/l 100%/90.1%/31.1%/ 100% BTA 86.3% TCB 170 µmol/l 97.0%/91.8%/34.4%/ 99.9% BTA 88.0% TCB 240 µmol/l 60.6%/99.7%/90.9%/ 98.3% BTA 97.2% TCB 250 µmol/l 57.6%/100%/100%/ 98.1% BTA 97.5%	SBR>150µmol/l Y TCB 100 µmol/l 100%/72.0%/51.3%/ 100% BTA 55.6% TCB 110 µmol/l 97.5%/81.0%/60.2%/ 99.1% BTA 63.1% TCB 150 µmol/l 53.2%/99.3%/95.5%/ 87.8% BTA 87.3% TCB 160 µmol/l 45.6%/100%/100%/ 86.2% BTA 89.6% SBR>200µmol/l Y TCB 130 µmol/l 100%/85.4%/33.8%/ 100% BTA 79.5% TCB 140 µmol/l 95.8%/90.4%/42.6%/ 99.7% BTA 84.4% TCB 200 µmol/l 62.5%/99.7%/93.8%/ 97.3% BTA 95.4% TCB 210 µmol/l 54.2%/100%/100%/ 96.7% BTA 96.3% SBr>250µmol/l Y TCB 160 µmol/l 100%/92.8%/33.3%/ 100% BTA 89.6% TCB 170 µmol/l 91.7%/94.0%/35.5%/ 99.7% BTA 91.1% TCB 230 µmol/l 50.0%/99.7%/85.7%/ 98.2% BTA 98.0% TCB 240 µmol/l 41.7%/100%/100%/ 98.0% BTA 98.6%			SBR>150µmol/l TCB 70 µmol/l 100%/ 17.1%/25.5%/ 100% BTA 13.3% TCB 80 µmol/l 98.9%/ 24.9%/27.2%/ 98.8% BTA 19.7% TCB 170 µmol/l 62.6%/ 99.1%/95.0%/ 90.3% BTA 85.4% TCB 180 µmol/l 54.9%/ 100%/100%/ 88.7% BTA 87.9% SBR>200µmol/l TCB 140 µmol/l 100%/ 82.2%/38.9%/ 100% BTA 73.8% TCB 150 µmol/l 95.2%/ 87.6%/46.5%/ 99.4% BTA 79.1% TCB 220 µmol/l 61.9%/ 99.5%/92.9%/ 95.8% BTA 93.2% TCB 230 µmol/l 54.8%/ 100%/100%/ 95.1% BTA 94.4% SBr>250µmol/l TCB 190 µmol/l 100%/ 94.1%/47.7%/ 100% BTA 89.3% TCB 200 µmol/l 95.2%/ 95.4%/52.6%/ 99.7% BTA 90.8% TCB 240 µmol/l 71.4%/ 99.7%/93.8%/ 98.5% BTA 96.1% TCB 250 µmol/l 66.7%/ 100%/100%/ 98.2% BTA 96.6%	

FP – false positive
FN – false negative
BTA blood tests avoided
◊ - included Hispanic and non-Hispanic Whites
ϕ –compared to TCB not SBr

§ - Medium dark and dark skin tone combined
Y - Study comparison was “light/ fair skin pigmentation” rather than White per se

	First author	Gestation at birth	Sample size	Sensitivity/specificity/Positive predictive value/Negative predictive value					
				Overall	White	Black/ African American	Mixed ethnicity/"non-White"	Medium	Dark
JM-105	Maya-Enero et al., (2014)	Term and preterm (≥31 wks)	1359		≥ 38 wks BTA 21.05% 35-37 wks BTA 6.9% 31-34 wks BTA 5.26% Y			≥ 38 wks BTA 18.1% 35-37 wks BTA 11.1% 31-34 wks BTA 0%	≥ 38 wks BTA 18.27% 35-37 wks BTA 27.78% 31-34 wks BTA 8.33% §
KJ-8000	Starowicz et al., (2020)	Term and preterm	201	Ability to predict phototherapy (TCB within 50μmol/l of treatment line) 82.7%/52.7%/20.0%/ 95.9% BTA 48.3%	Ability to predict phototherapy (TCB within 50μmol/l of treatment line) 81.4%/58.6%/ 23.6%/ 95.2% BTA 53.2%		Ability to predict phototherapy (TCB within 50μmol/l of treatment line) 88.9%/35.2%/ 11.9%/ 97.0% BTA 33.0%		
Visual	Keren et al., (2008,2009) ϕ	Term and late preterm (≥35 wks)	812 (522 had paired TCB)	Absence of visual jaundice: Sensitivity = 95%, NPV 99% to rule out significant hyperbilirubinemia	NPV for absence of visual jaundice to identify bilirubin value in low-risk zone 83%	NPV for absence of visual jaundice to identify bilirubin value in low-risk zone 84%			

Appendix 9: Full results for other statistical test undertaken or results reported to assess the impact of skin tone on assessment of jaundice

	First author	Gestation at birth	Sample size	Type of statistical test	Full results	
BiliCheck	Rubaltelli et al., (2001)	Term and preterm (>30 wks)	210	Analysis of covariance: TCB was independent of race, birthweight, gestational age and PN age of infant when SBr was included in the model. However, only 4.3% of their sample was African descent	Results of Analysis of Covariance Analysis for contributors to BiliCheck Forehead error Source df Type III Mean Square F P value Race 1 1.878 1.878 0.61 0.4357 GA 1 7.249 7.249 2.35 0.1265 BW 2 11.605 5.802 1.88 0.1546 PN age 3 14.125 4.708 1.53 0.2082 SBr 1 217.012 217.012 70.50 0.0001	
	Ebbesen et al., (2002)	Term and preterm	488	Multiple linear regression - Ethnic origin was not found to influence TCB i.e. BiliCheck corrects sufficiently for this (p=0.49 in NNU jaundice infants and 0.20 in 'healthy' jaundiced infants).	Results of the multiple linear regression analysis [TCB vs SBr] in healthy term and near-term infants Variable Estimate (95%CI) p-Value Intercept 25.37 (12.47, 38.27) <0.0001 SBr (µmol/l) 0.79 (0.72, 0.85) <0.0001 GA (wks) (≥39 = REF) 36–38 -0.78 (-8.07, 6.50) 0.83 ≤35 11.35 (-3.20, 25.90) 0.13 Gender (Female = REF) Male -1.93 (-8.78, 4.93) 0.58 PN day (Days 1-4 = REF) ≥5 -3.37 (-11.92, 5.18) 0.44 Ethnic origin (Northern European = REF Non-North. Europ. -8.87 (-22.32, 4.59) 0.20	Results of the multiple linear regression analysis [TCB vs SBr] in infants in NNU: Variable Estimate (95% CI) p-Value Intercept 34.90 (21.71, 48.09) <0.0001 SBr (µmol/l) 0.81 (0.76, 0.87) <0.0001 GA (wks) (≥39 = REF) 36–38 1.96 (-7.71, 11.63) 0.69 ≤35 6.77 (-1.75, 15.30) 0.12 Gender (Female = REF) Male -9.76 (-16.26, -3.27) 0.003 PN day (Days 1-4 = REF) ≥5 -6.85 (-13.59, -0.12) 0.046 Ethnic origin (Northern European = REF) Non-North.Europ. -3.97 (-15.34, 7.39) 0.49 Degree of disease (Mild = REF) Severe 4.67 (-2.76, 12.09) 0.22

	First author	Gestation at birth	Sample size	Type of statistical test	Full results
BiliCheck	Robertson et al., (2002)	Term (mean GA 37.7 ± 2.2 wks so includes late preterm	101	Linear regression. Skin tone had no impact on BiliCheck	Regression coefficients showing effect of skin tone in predicting SBr from TCB reading: Value SE p value BC meter reading 0.937 0.043 0.000 Skin tone 0.0019 0.134 0.890
	Leite et al., (2007)	Term and preterm	200	ANCOVA: Race was not a significant variable	Univariate analysis Skin tone (p=0.1560) White (n=133) BiliCheck 8.82±3.70 mg/dl SBr 7.97±3.94 mg/dl Black (n=20) BiliCheck 8.32±3.53 mg/dl SBr 8.16±4.42 mg/dl Mixed race (n=47) BiliCheck 9.03±4.54 mg/dl SBr 8.45±4.76 mg/dl
	Ebbesen et al., (2012)	Preterm	133	Multiple regression	Multiple regression analysis results for BiliCheck using TCB as dependent variable (n=133) Coefficient (95% CI) p value SBr (μmol/L) 0.71 (0.63–0.79) <0.001 GA (d) -0.34 (-0.69–0.02) 0.06 PN age (h) -0.05 (-0.14–0.03) 0.19 Sex (Male = REF) Female -6.02 (-13.87–1.83) 0.14 Ethnicity (White = REF) 'Non-White' 10.02 (-6.51–26.54) 0.24 Severe illness (No = REF) Yes 4.01 (-9.52–17.54) 0.56
	Rubio et al., (2017)	Preterm only	167	Multivariate multilevel logistic regression	The only factor significantly influencing discordance between TCB and SBr was postnatal age: the difference decreased 3 points per day in the first 2 weeks of life ($p < 0.01$). No impact was found from GA or skin tone
JM-101	Hegyi et al., (1981)	Term	60	Regression	Skin tone significantly influenced the intercept of the regression line.
	Hannemann et al., (1982)	Term and late preterm	161	Regression: Intercept much higher for Black ≥34 weeks than White ≥34 weeks	Significance not given
	Kenny et al., (1984)	Term	53	Regression	When estimates were adjusted by initial measurements of both TCB and SBr, we found the relative error in estimating SBr relative to its measurement was 7.4% for thirteen White infants, 8.3% for four Black infants and 20.9% for five Asian infants. (A relative error in estimating SBr > 10.8% was significant in their laboratory - therefore Asian significant)

	First author	Gestation at birth	Sample size	Type of statistical test	Full results
JM-102	Robertson et al., (2002)	Term (mean GA 37.7 ± 2.2 wks so includes late preterm	101	Linear regression	Regression coefficients showing effect of skin tone in predicting SBr from TCB reading Value SE p value JM meter reading 0.7074 0.069 0.000 Skin tone -0.771 0.240 0.002 (i.e. impact of skin tone with JM-102)
	Szabo et al., (2004a)	Preterm (34-36+6 wks)	69	Mann Whitney	Not impacted by skin tone
JM-103	Maisels et al., (2004)	Term and late preterm (≥35 wks)	849	Proportion	TCB differed from SBr by > 3mg/dl (51 µmol/L), in 2.0% of White and 17.4% of Black infants and 3.2% of other infants. In all Black infants the JM-103 value was higher than the SBr value
	Schmidt et al., (2009)	Preterm	90	Chi square	NS relationship between birth weight, GA, mode of birth, gender, ethnicity, 5 min Apgar score, umbilical artery pH, ventilatory support or PN age at SBr and the absolute TCB/SBr difference.
	Wainer et al., (2009)	Term only	938 (744 in analy-sis)	Multivariate linear Regression: tendency to overestimate in dark skinned infants	Multivariate linear regression analysis of skin tone on SBr vs. TCB (Adj. R2 = 0.86) Coefficient (95% CI) P value Intercept 17.8 (14.04–21.57) TCB 0.93 (0.90–0.96) <0.001 Skin tone (Medium skin tone = REF) Light skin tone 12.20 (9.27–15.12) <0.001 Dark skin tone -31.20 (-41.57–-20.83) <0.001
	Ebbesen et al., (2012)	Preterm	133	Multiple regression	Results of the multiple regression analysis for JM-103 using TCB as dependent variable (n=133) Coefficient (95% CI) p value SBr (µmol/L) 0.73 (0.66–0.81) <0.001 GA (d) 0.20 (-0.12–0.53) 0.22 PN age (h) -0.12 (-0.20 to –0.05) 0.001 Sex (Male = REF) Female -0.19 (-7.00–7.38) 0.96 Ethnicity (White = REF) 'Non-White' 29.60 (14.48–44.73) <0.001 Severe illness (No = REF) Yes -4.29 (-16.67–8.10) 0.50
	Afanetti et al., (2014)	Term and preterm	86	Multiple linear regression	Skin tone (p=0.73) and gestational age (p=0.60) were not predictors of SBr when the model included TBC (p<0.001))

	First author	Gestation at birth	Sample size	Type of statistical test	Full results	
JM-105	Maya-Enero et al., (2021)	Term and preterm (≥31 wks)	1359	Multiple linear regression: significant for skin tone in the model with and without GA	Results for multiple linear regression as a function of SBr: β (95%CI) p value TCB 0.88 (0.87; 0.90) <0.001 Skin tone (Light = REF) Medium-clear −0.22 (−0.40; −0.04) 0.019 medium-dark −0.81 (−1.04; −0.58) <0.001 dark −0.80 (−1.32; −0.28) 0.003	
Bilimed	Karen et al., (2009)	Term and preterm (≥28 wks)	150	Multiple stepwise linear regression	Birthweight was a confounding factor to predict SBr (p<0.001), but skin tone was not (p=0.38) nor GA (p=0.96)	
Mixed*	Taylor et al., (2015)	Term and late preterm (≥35 weeks)	769	Multivariate regression analysis: White and Hispanic not associated with TCB-SBr difference after adjusting for other factors; African American infants significantly higher difference. Chance of TCB-SBr difference ≥2mg/dl: higher in African American and lower in Hispanic	Association between patient and nursery characteristics for TCB-SBr difference. Co-efficient p value Hour of age 0.05 <0.001 TCB (JM-103) -0.87 <0.001 Anatomic site (Forehead = REF) Chest -0.24 0.24 Chest and forehead 0.53 0.04 African American 0.67 0.002 White -0.30 0.11 Hispanic -0.002 0.99	TCB reading ≥2mg/dl higher than SBr African American OR 2.19 (95% CI 1.09-4.39) TCB ≥3mg/dl higher than SBr Hispanic ethnicity OR 0.47 (95% CI 0.30-0.71) TCB ≤2mg/dl lower than SBr African American OR 0.13 (95% CI 0.03-0.56) White OR 3.41 (95%CI 1.36-8.57))
Undefined	Brucker & MacMullen, (1987)	Term	20	Unclear which test used. Sample size only 20	SBr and TCB readings were not significant according to type of feeding, GA, maternal age or ethnic origin	
Visual -Kramer	Szabo et al., (2004a)	Preterm (34-36+6 wks)	69	Mann Whitney	Skin tone significantly affected clinical estimation by nurses - estimating 0.5 dermal zones higher in infants with White skin to infants with 'non-White' skin	

	First author	Gestation at birth	Sample size	Type of statistical test	Full results
Visual -Kramer	Keren et al., (2008, 2009) ϕ	Term and late preterm (≥35 wks)	812 (522 had paired TCB)	Logistic regression	Grade 4 (or higher) (on Kramer scale) jaundice was associated with the development of significant hyperbilirubinemia in non-Black infants (OR 6.9 95% CI 2.2-22.0). Black infants did not demonstrate this association - 7/353 Black infants with grade 0-3 jaundice developed hyperbilirubinemia, compared to 0 out of the 10 with grade 4 or above jaundice.
Visual	Brits et al., (2018) ϕ	Term	96	Not statistically tested	Only 10% (4/41) of Black babies diagnosed with jaundice appeared clinically jaundiced Compared to 5/12 non-Black infants
Visual (Icterometer)	Merritt & Coulter, (1994)	Preterm	90	Not statistically tested	Only 4 'non-infants' were included. Their icterometer measurements plotted within the 99th percentile range for White infants. But there was a tendency for SBr level to fall below the mean at the given icterometer value (i.e. icterometer slightly overestimated SBr), but too small numbers to form any firm conclusions of the utility of the icterometer in this group of neonates
	Madlon-Kay, (1997)	Mainly term	171	T-test and ANOVA	The accuracy of the HCP estimates of bilirubin levels affected by physician/nurse level of training or experience or by neonatal race or ethnic group
	Madlon-Kay, (2001)	NR	164	ANOVA	Accuracy of nurses estimates of bilirubin levels unaffected by race or ethnic group.
	Lee et al., (2019) ϕ	Term and preterm	790	Mean TCB was similar at each icterometer score for different ethnicities (being within 2mg/dl across all ethnicities at the same Biliruler score with an adequate sample size of > 10)	Mean TCB (± standard deviation) in mg/dl for the different ethnicities at each Biliruler score: Biliruler score 1.5 Black/African American 6.1 (±3.5) (n= 6); Hispanic/Latino 3.5 (±2.3) (n=4); Asian 4.5 (±2.7) (n=9); Non-Hispanic White 4.2 (±2.7) (n=31) Biliruler score 2 Black/African American 5.9 (±2.2) (n=14); Hispanic/Latino 6.2 (±2.5) (n=19); Asian 4.8 (±3.0) (n=111); Non-Hispanic White 6.3 (±2.5) (n=80) Biliruler score 2.5 Black/African American 9.0 (±3.0) (n=18); Hispanic/Latino 7.5 (±1.8) (n=10); Asian 6.9 (±3.0) (n=63); Non-Hispanic White 8.1 (±2.1) (n=51) Biliruler score 3 Black/African American 9.9 (±3.2) (n=14); Hispanic/Latino 8.9 (±2.3) (n=2); Asian 8.2 (±3.5) (n=74); Non-Hispanic White 9.7 (±2.5) (n=29) Biliruler score 3.5 Black/African American 11.3 (± 3.5) (n=7); Hispanic/Latino 13.0 (n=1); Asian 11.9 (± 3.2) (n=32); Non-Hispanic White 10.1 (± 2.8) (n= 18)

TCB – transcutaneous bilirubin
SBr – serum bilirubin
NS – not significant

GA – gestational age
PN – post natal
SE – standard error

NNU – neonatal unit
REF - referent category
95% CI – 95% confidence interval

Mixed* - BiliCheck or JM-103
ϕ –compared to TCB not SBr
NR – not reported

Appendix 10: Example search strategy from the systematic review of the experiences of parents or carers

Example search strategy within CINAHL database

((MH mothers+) OR (MH fathers+) OR 'maternity service user' OR (MH parents+) OR (MH Maternal attitude) OR (MH Paternal Attitude) OR (MH parental attitude))

AND

(neonatal OR newborn OR neonate OR 'new born' OR new-born OR baby OR preterm OR premature OR babies)

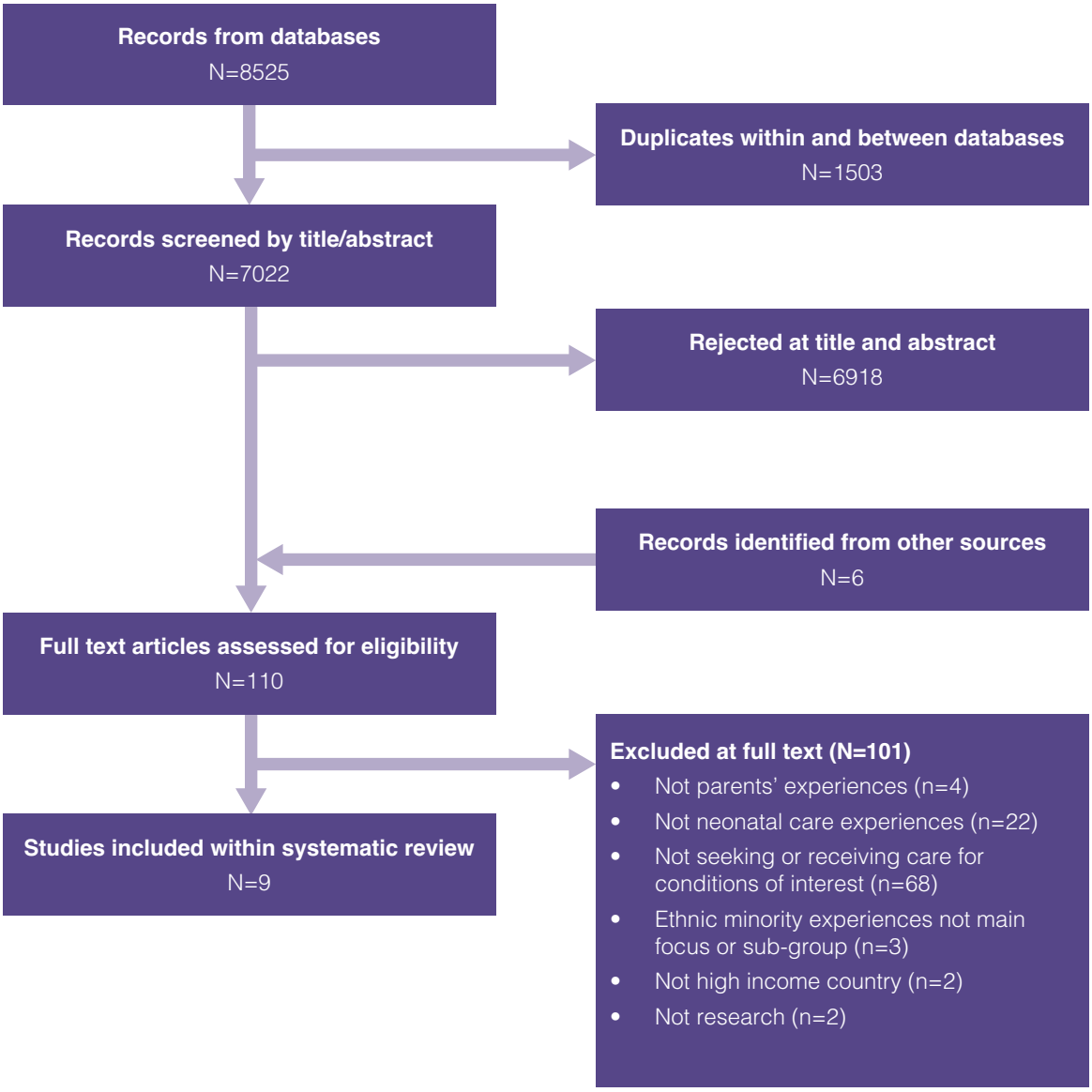
AND

((MH 'Ethnic Groups+') OR (MH 'Minority Groups') OR (MH 'Cultural Diversity') OR (MH 'Asians+') OR (MH 'Black Persons') OR (MH 'Hispanic Americans') OR (MH 'Indigenous Peoples') OR (MH 'Immigrants+') OR (MH 'Transients and Migrants') OR (MH 'Emigration and Immigration') OR Black OR Asian OR 'ethnic minority' OR 'minority ethnic' OR bame OR 'Black or minority ethnic' OR bme OR 'mixed race' OR 'mixed ethnic*' OR 'mixed heritage' OR skin pigmentation OR Hispanic OR Pakistan* OR Somali* OR India* OR Bangladesh* OR Chinese OR China OR Africa* OR Caribbean OR Arab OR Ethnic* OR ethno* OR race OR Racial* OR colour OR color) LIMITED TO TITLE OR ABSTRACT OR SUBJECT

AND

((MH 'Jaundice') OR (MH 'Jaundice, Neonatal') OR (MH 'Hyperbilirubinemia, Neonatal') OR (MH 'Hyperbilirubinemia') OR (MH 'Anoxia') OR (MH 'Apgar Score') OR (MH 'Cyanosis') OR Hypoxia OR hypoxemia OR oximet* OR 'oxygen saturation' OR jaundice OR APGAR OR cyanosis OR (MH 'Anemia+') OR anemia OR anaemia OR anemic OR anaemic OR (MH 'Healthcare Disparities') OR (MH 'Health Services Accessibility+') OR (MH 'Health Services+') OR (experience OR perspective OR view OR attitude OR 'use' OR access* OR utili* OR block* OR hurdl* OR barrier* OR hindr* OR hinder* OR obstacle* OR exclu* OR discrimin* OR disparit* OR disproportion* OR inequal* OR unequal* OR inadequat* OR insuffic* OR stratif* OR limit* OR lack* OR unreliab* OR poor* OR poverty* OR depriv* OR disadvantag* OR insecur* OR insensit* OR status* OR entitl* OR benefit* OR deliver)) LIMITED TO TITLE OR ABSTRACT OR SUBJECT

Appendix 11: PRISMA flow chart showing study identification from the systematic review of the experiences of parents or carers (Page et al., 2021)



Appendix 12: Critical appraisal results for studied enclosed in the systematic review of parents' experiences.

		Birthrights, 2022	Casey et al., 1992	Fivexmore, 2022	Hannon et al., 2001	Hurst, 2001	Lee et al., 2009	Palau et al., 2019	Tarnow-Mordi & Pickering, 1983	Witt et al., 2022
	Are there clear research questions?	Y	Y	Y	Y	Y	Y	Y	N	Y
	Do the collected data allow to address the research questions?	Y	Y	Y	Y	Y	Y	Y	N	Y
Qualitative studies	QU1 – Is the qualitative approach appropriate to answer the research question?				Y	Y	Y			Y
	QU2 – Are the qualitative data collection methods adequate to address the research question?				Y	Y	N			?
	QU3 – Are the findings adequately derived from the data?				Y	Y	Y			Y
	QU4 – Is the interpretation of results sufficiently substantiated by data?				Y	Y	Y			Y
	QU5 – Is there coherence between qualitative data sources, collection, analysis and interpretation?				Y	Y	Y			Y
Quantitative descriptive studies	QU1 – Is the sampling strategy relevant to address the research question?		Y	Y				Y	?	
	QU2 – Is the sample representative of the target population?		?	?				N	N	
	QU3 – Are the measurements appropriate?		N	Y				Y	N/A	
	QU4 – Is the risk of nonresponse bias low?		?	N				Y	?	
	QU5 Is the statistical analysis appropriate to answer the research question?		Y	Y				Y	N	
Mixed methods studies	QU1 – Is there an adequate rationale for using a mixed methods design to address the research question?	Y								
	QU2 – Are the different components of the study effectively integrated to answer the research question?	Y								
	QU3 – Are the outputs of the integration of qualitative and quantitative components adequately interpreted?	Y								
	QU4 – Are divergences and inconsistencies between quantitative and qualitative results adequately addressed?	Y								
	QU5 – Do the different components of the study adhere to the quality criteria of each tradition of the methods involved?	?								

Appendix 13: Policy review search terms

BAPM
There was no facility to search on the BAPM website therefore the 'Resources' tab was searched manually for relevant documents.

Institute of Health Visitors
"Neonatal" "Neonate" "cyanosis" "hypoxia" "Asian" "Black" "Colour" "Color" "Jaundice" "pulse oximetry" "oxygen saturation"

Neonatal Nurses Association
No search terms. The 'Learn' tab was screened including the links to additional resources which led to "Neonatal nursing – knowledge to support leaning in practice" resources:
https://www.herts.ac.uk/study/schools-of-study/health-and-social-work/course-subject-areas/nursing/childrens-nursing/neonatal-nursing

NICE
"Neonatal" "Neonate" "cyanosis" "hypoxia" "Asian" "Black" "Colour" "Color" "Jaundice" "pulse oximetry" "oxygen saturation"

Office for Health Improvement & Disparities
Filtered by 'guidance and regulation'
"Black Neonate" "Asian Neonate"
"Neonatal" "Neonate" "cyanosis" "hypoxia" "Asian" "Black" "Colour" "Color" "Jaundice" "pulse oximetry" "oxygen saturation"

Resuscitation council
"Neonatal" "Neonate" "cyanosis" "hypoxia" "Asian" "Black" "Colour" "Color"

Royal College of Midwives
"Apgar" "Neonatal" "cyanosis" "hypoxia" "Asian" "Black" "Colour" "Color"

Royal College of Nursing
"Neonatal" "Neonate" "cyanosis" "hypoxia" "Asian" "Black" "Colour" "Color" "Jaundice" "pulse oximetry" "oxygen saturation"

WHO
"Neonatal" "Neonate" "cyanosis" "hypoxia" "Asian" "Black" "Colour" "Color" "Jaundice" "pulse oximetry" "oxygen saturation"

nhsrho.org

@nhs_rho