Population-based screening methods in biliary atresia: a systematic review and meta-analysis

Adam Arshad, Julian Gardiner, Carmen Ho, Philippa Rees, Karan Chadda, Alastair Baker, Alastair G Sutcliffe

ABSTRACT

Objective The aim of this study was to investigate tested methods of population-based biliary atresia (BA) screening.

Design We searched 11 databases between 1 January 1975 and 12 September 2022. Data extraction was independently done by two investigators.

Main outcome measures Our primary outcomes were: sensitivity and specificity of screening method in BA detection, age at Kasai, BA associated morbidity and mortality, cost-effectiveness of screening.

Results Six methods of BA screening were evaluated: stool colour charts (SCCs), conjugated bilirubin measurements, stool colour saturations (SCSs), measurements of urinary sulfated bile acids (USBAs), assessments of blood spot bile acids and blood carnitine measurements.

In a meta-analysis, USBA was the most sensitive and specific, with a pooled sensitivity and specificity of 100.0% (95% CI 2.5% to 100.0%) and 99.5% (95% CI 98.9% to 99.8%) (based on one study). This was followed by conjugated bilirubin measurements: 100.0% (95% CI 0.0% to 100.0%) and 99.3% (95% CI 91.9% to 99.9%), SCS: 100.0% (95% CI 0.00% to 100.0%) and 92.4% (95% CI 83.4% to 96.7%), and SCC: 87.9% (95% CI 80.4% to 92.8%) and 99.9% (95% CI 99.9% to 99.9%). SCC reduced the age of Kasai to ~60 days, compared with 36 days for conjugated bilirubin. Both SCC and conjugated bilirubin improved overall and transplant-free survival. The use of SCC was considerably more cost-effective than conjugated bilirubin measurements.

Conclusion Conjugated bilirubin measurements and SCC are the most researched and demonstrated improved sensitivity and specificity in detecting BA. However, their use is expensive. Further research into conjugated bilirubin measurements, as well as alternative methods of population-based BA screening, is required.

PROSPERO registration number CRD42021235133.

INTRODUCTION

Biliary atresia (BA) is the leading cause of liver cirrhosis in the paediatric population. The aetiology of the condition is poorly understood but results in inflammation, narrowing and destruction of the large bile ducts in the first months of life. BA is resultantly the the most common reason for paediatric liver transplantation (LTX). Epidemiological studies indicate BA occurs in approximately 1:15 000 live births in Western Europe and North America, with the highest incidence in Eastern Asia (1:6000–1:9000 births).1–3

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ There is no consensus on the most effective method of population-based screening for biliary atresia (BA).
⇒ There is no systematic or meta-analysis on this subject area.

WHAT THIS STUDY ADDS

⇒ While demonstrating the best sensitivity and specificity, conjugated bilirubin measurements are an expensive method of population-based BA screening.
⇒ A stool colour chart may reduce the age of Kasai, but their applicability to western nations, given the baseline age at Kasai without a screening intervention, is questioned.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ BA needs an effective population-based screening programme.
⇒ Further research into the practicality of conjugated bilirubin measurements, as well as alternative methods of population-based BA screening, is required.

BA presents in the first weeks of life, with neonates demonstrating jaundice and pale (acholic) stools. In contrast to physiological or breast milk associated unconjugated jaundice (which occurs in two-thirds of neonates), the jaundice in BA is prolonged and a pathological obstructive jaundice with a conjugated (direct) hyperbilirubinaemia. If undetected, neonates rapidly develop cirrhosis and subsequent liver failure.

The clinical course of BA can be improved with a Kasai portoenterostomy (Kasai), an operation that re-establishes bile flow by removing atretic bile ducts and creating a liver-intestinal anastomosis. A Kasai conducted by 30 days after birth significantly reduces the risk of subsequent LTXs. Significantly, delays in BA detection and treatment, with poor native liver survival, have been reported across paediatric hepatology centres worldwide.4–5

Overall, the nature and clinical course of BA creates a need for effective newborn screening. Additionally, BA is clearly defined, and early recognition is associated with improved clinical and potential cost savings, further supporting the need for a screening programme. To date, population-based screening programmes have been implemented in Taiwan, Brazil, Canada and Germany, but there is...
debate as to the most effective modality for this, and there has been no systematic review or meta-analysis exploring this study area.⁶–⁹ The aim of this systematic review and meta-analysis is to explore the effectiveness of tested screening methods for BA, including their sensitivity and specificity, benefits in subsequent age at Kasai, associated patient morbidity and cost savings.

METHODS

This systematic review was registered with International Prospective Register of Systematic Reviews (CRD: CRD42021235133). We included observational studies reporting outcomes of a BA screening method. Excluded study designs included opinions, reviews and non-peer-reviewed letters. The systematic review was conducted in accordance with the PRISMA statement and MOOSE guidelines (online supplemental appendix tables 1 and 2). Non-English language studies were excluded.

Eleven databases were searched to identify appropriate published white and grey literature from 1 January 1975 to 12 September 2022 (online supplemental appendix table 3). Our search strategy was created in MEDLINE OVID and consisted of 35 keywords and Medical Subject Headings (online supplemental appendix table 4). These terms were adapted for other databases. Our primary outcomes were sensitivity and specificity in BA detection, age at Kasai, BA associated morbidity and mortality, and cost-effectiveness of the screening method.

References were exported to Endnote VX9.1. The final list of articles was then exported to Rayan QCRI.¹⁰ Two reviewers screened titles and abstracts independently and were blinded. The full texts of articles deemed relevant were retrieved and assessed. Disagreements were arbitrated by a third reviewer (AGS). Quality assessments were conducted using the Newcastle-Ottawa Tool.

Data extraction and synthesis

Two reviewers (AA and KC) independently extracted appropriate data using a piloted extraction tool. Studies were grouped according to the method of screening and the outcome.

Meta-analysis

We conducted a meta-analysis of the sensitivity and specificity of population-based BA screening methods. The sensitivity and specificity of the methods were extracted from included papers, with 95% CIs calculated using the exact binomial method of Clopper and Pearson.¹¹ A further subsidiary analysis of negative predictive value (NPV) and positive predictive value (PPV) was conducted. This was performed using the following formulae:

$$PPV: \frac{1 - prevalence}{sensitivity} \times \frac{1 - prevalence}{significance} \times \frac{1 - prevalence}{(1 - specificity) \times \frac{1 - prevalence}{(1 - specificity)} \times \frac{1 - prevalence}{(1 - specificity)} \times \frac{1 - prevalence}{(1 - specificity)}$$

$$NPV: \frac{1 - prevalence}{sensitivity} \times \frac{1 - prevalence}{significance} \times \frac{1 - prevalence}{(1 - specificity) \times \frac{1 - prevalence}{(1 - specificity)} \times \frac{1 - prevalence}{(1 - specificity)} \times \frac{1 - prevalence}{(1 - specificity)}$$

RESULTS

The searches identified 9377 titles. Of these, 27 full texts were included (figure 1). This represents 2756 infants with BA and 4019847 infants without BA. Seventeen papers were suitable for a meta-analysis on the sensitivity and specificity in detecting BA.

Included studies were retrospective cohort studies (n=4),¹²–¹⁵ prospective cohort studies (n=12),²⁶–³² cross-sectional studies (n=7),¹⁹ ²⁶–³² case–control studies (n=2)³³ ³⁴ and cost–benefit analyses (n=2).³⁵ ³⁶ Two studies were conducted in the USA,²² ²⁷ two in Canada,² ²⁵ seven in China,¹³ ¹⁵ ²⁰ ²⁸ ³¹ ³² ³³ six in Japan,¹⁶ ¹⁷ ¹⁹ ²⁶ ³⁰ ³³ six in Taiwan,¹ ²⁶ ³⁰ ³¹ ³³ and two in the UK.²⁴ ³⁴ Two studies were cost-effectiveness analyses and so had no designated nation.³⁵ ³⁶ The studies had been published between 1994 and 2022 and are summarised in online supplemental appendix table 5. The majority of trialled screening methods were conducted in the first few weeks of life, with five studies assessing infants >2 months.¹⁶ ²⁶ ³⁰–³² Most studies (n=25) were deemed ‘good’ on quality assessment and two were fair (online supplemental appendix table 6A–C).

Sensitivity and specificity in detecting BA

Seventeen studies included raw data allowing us to calculate the sensitivity and specificity of their screening tool in detecting BA: five used stool colour charts (SCCs),¹⁶ ¹⁷ ¹⁸ ²⁰ ²³; 4 measured bilirubin in blood samples,¹³ ¹⁵ ²² ²⁴ ²⁷; 1 assessed urinary sulfated bile acids (USBAs),¹⁶; 2 quantified stool colour saturation (SCS)²⁶ ³¹; 4 measured bile acids in blood spots,²⁹ ³⁰ ³¹ ³²; and 1 assessed blood carnitine levels (figures 2 and 3 and online supplemental appendix tables 7 and 8).

On meta-analysis, USA was the most sensitive and specific screening approach to BA detection, with a pooled sensitivity and specificity of 100.0% (95% CI 2.5% to 100.0%) and 99.5% (95% CI 98.9% to 99.9%). This was followed by conjugated bilirubin measurements: 100.0% (95% CI 0.0% to 100.0%) and 99.3% (95% CI 91.9% to 99.9%), respectively; SCS: 100.0% (95% CI 0.0% to 100.0%) and 92.4% (95% CI 83.4% to 96.7%); SCC: 87.9% (95% CI 80.4% to 92.8%) and 99.9%
conjugated bilirubin improved both overall and transplant-related mortality: four for SCC and one for conjugated bilirubin measurements.13 21 22 25 33 Screening with either SCC or conjugated bilirubin improved both overall and transplant-free survival among patients with BA. One study demonstrated a reduction in the average hospitalisations and length of hospital stay after SCC (table 3).

**Associated costs**

Two cost effectiveness analyses and two prospective cohort studies assessed the cost-effectiveness of population-based screening for BA.7 25 35 36 Four of these studies assessed the cost-effectiveness of SCC and one looked at both SCC and conjugated bilirubin measurements. There was variation in the factors built into the cost-effectiveness model for each of the studies, and across all four studies, no model considered costs associated with complications following LTx (online supplemental appendix table 9).

Mogul et al demonstrated that SCC was associated with an overall 20-year cost saving of US$8 586 162.36 A similar result was presented by Masucci et al, with screening using a home-based SCC costing $C192 000 more than no universal screening but leading to 8 life-years gained (incremental cost-effectiveness ratio (ICER) $24 065 per life-year gained). Furthermore, screening using conjugated bilirubin testing cost $2 369 199 more than SCC and led to 5 more life-years gained (ICER: $473 840 per life-year gained).33

**Table 1:** PPV and NPV for differing methods of biliary atresia screening

<table>
<thead>
<tr>
<th>Method</th>
<th>Prevalence in population tested (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 in 100</td>
</tr>
<tr>
<td>Conjugated</td>
<td>59.1</td>
</tr>
<tr>
<td>USBA</td>
<td>66.9</td>
</tr>
<tr>
<td>SCC</td>
<td>90.0</td>
</tr>
<tr>
<td>SCS</td>
<td>9.3</td>
</tr>
<tr>
<td>Bile acid blood spot</td>
<td>4.7%</td>
</tr>
</tbody>
</table>

PPV and NPV values calculated based on pooled sensitivity and specificity estimates.
NPV, negative predictive value; PPV, positive predictive value; SCC, stool colour chart; SCS, stool colour saturation; USBA, urinary sulfated bile acid.

**DISCUSSION**

To our knowledge, this is the first systematic review and meta-analysis assessing the current methods of population-based screening for BA and concluding on all relevant studies within this important subject area. Our analysis of the literature has identified six researched methods of BA screening: (1) the use of SCC, (2) blood measurements of conjugated bilirubin, (3) measurements of USBAs, (4) analyses of SCSs, (5) measuring bile acids in blood spots and (6) blood carnitine measurements. Included studies mostly had a low risk of bias (as assessed by the Newcastle-Ottawa Scale). Two methods appeared most evidenced and superior in improving outcomes from BA: SCC and conjugated bilirubin measurements, with an overall pooled

---

**Figure 2** Meta-analysis of sensitivity of biliary atresia screening methods.

**Figure 3** Meta-analysis of specificity of biliary atresia screening methods.
sensitivity and specificity (in meta-analysis) for BA detection of 87.9% and 99.9% and 100% and 99.3%, respectively. These two methods of screening have been the most extensively researched, and both methods demonstrate a reduced subsequent age at Kasai after application, with the range of reduction being 8.4–25 days with SCC, and 19 days for conjugated bilirubin measurements. Across both strategies, this translated to reduced morbidity and mortality, with improved overall survival. Finally, from studies in America and Canada, respectively, the use of SCC provided greater cost-effectiveness compared with a conjugated bilirubin or a non-screening strategy. While the use of USBA and SCS demonstrated promising overall sensitivity and specificity results, there was a paucity of data on these methods (one and two papers, respectively) to draw any significant conclusions.

SCC has been considered a minimally invasive, efficient method of BA screening. At birth, parents are provided with a stool chart, depicting shades of coloured stools ranging from normal to acholic. Should it be considered that an infant is producing acholic stools, a referral is made for appropriate BA investigations. This referral can be made by the parents or following a healthcare professional’s review. This method of screening has been primarily investigated in East Asian nations (China, Japan and Taiwan). Its use is associated with high sensitivity and specificity in the detection of BA (88.7% and 99.9%) and a range of reduction in the age of Kasai procedure of 8.4–25.0 days. Among nations/regions using SCC-based screening, the subsequent average age at Kasai was reduced to 48.0–59.7 days. Given these aforementioned strengths, SCC is now used for population-based BA screening in Taiwan, Brazil, Canada, Germany and Switzerland, respectively.7 9 18 37 38

Despite these promising results, the benefits of introducing this screening strategy in the USA, UK and selected European countries are unclear. It should be noted that the high predictive value of the SCC method in China, Taiwan and Japan is partly due to the higher incidence of BA in these countries compared with the USA and Europe. Our data demonstrate that the average age of Kasai among East Asian countries, after using SCVs, ranges from 48.0 days to 59.7 days. Based on the limited literature available, significant reductions in the age of Kasai were observed when the preintervention age was >70 days.13 17 Non-significant reductions were reported when this age was <60 days. However, the average ages of Kasai procedure in the USA, UK and France are 63, 54 and 60, respectively, independent of a BA screening strategy.4 39 40 This limits the applicability of SCC to these nations, given the baseline age of Kasai, a concern expressed by the UK national screening committee.41 Overall, while the use of SCC may benefit BA identification in Eastern nations, its appropriateness for European and North American healthcare systems is unclear.

Studies analysing direct, conjugated bilirubin blood measurements are demonstrating promising results, with a pooled sensitivity and specificity of 100% and 99.2%, respectively, in the detection of BA within our meta-analysis. A seminal paper by Harpavat et al22 used a two-stage screening strategy of 123385 newborns across

### Table 2 Intervention for population-based screening for biliary atresia and resulting change in the age of Kasai

<table>
<thead>
<tr>
<th>Title</th>
<th>Country</th>
<th>Intervention</th>
<th>Average age at Kasai</th>
<th>Kasai at less then 60 days</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Preintervention vs postintervention</td>
<td>P value</td>
</tr>
<tr>
<td>Gu et al21</td>
<td>Japan</td>
<td>SCC</td>
<td>70.3 vs 59.7</td>
<td>0.003</td>
</tr>
<tr>
<td>Tseng et al22</td>
<td>Taiwan</td>
<td>SCC</td>
<td>51 vs 48</td>
<td>0.051</td>
</tr>
<tr>
<td>Chen et al18</td>
<td>Taiwan</td>
<td>SCC</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Lee et al12</td>
<td>Taiwan</td>
<td>SCC</td>
<td>59.9 vs 48.2</td>
<td>0.064</td>
</tr>
<tr>
<td>Lien et al20</td>
<td>Taiwan</td>
<td>SCC</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Zheng et al11</td>
<td>China</td>
<td>SCC</td>
<td>81 vs 56</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>Hsiao et al23</td>
<td>Taiwan</td>
<td>SCC</td>
<td>N/A vs 54.1</td>
<td>–</td>
</tr>
<tr>
<td>Gu et al27</td>
<td>Japan</td>
<td>SCC</td>
<td>68.1 vs 59.7</td>
<td>0.003</td>
</tr>
<tr>
<td>Harpavat et al22</td>
<td>USA</td>
<td>Conjugated bilirubin measurements</td>
<td>56 vs 36</td>
<td>0.004</td>
</tr>
</tbody>
</table>

Values provided are in days.

Bold p values are significant at p<0.05.

*N/A, not available; SCC, stool colour chart.

### Table 3 Influence of screening intervention on associated morbidity, hospital admissions and mortality

<table>
<thead>
<tr>
<th>Title</th>
<th>Country</th>
<th>Intervention</th>
<th>Overall mortality (preintervention vs postintervention)</th>
<th>Transplant-free survival (preintervention vs postintervention)</th>
<th>Average number of hospitalisations (preintervention vs postintervention)</th>
<th>Length of stay in hospital per visit (preintervention vs postintervention)</th>
<th>Liver transplantation (preintervention vs postintervention)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lee et al21</td>
<td>Taiwan</td>
<td>SCC</td>
<td>47.8% vs 21.2% (p&lt;0.001)</td>
<td>31.6% vs 56.4% (p&lt;0.001)</td>
<td>6.4 vs 5.0 (p&lt;0.001)</td>
<td>86.6% vs 81.9% (p=0.438)</td>
<td>28.6% vs 28.2% (p=0.934)</td>
</tr>
<tr>
<td>Zheng et al12</td>
<td>China</td>
<td>SCC</td>
<td>20.6% vs 10.5% (p&lt;0.05)</td>
<td>44.4% vs 52.6% (p&lt;0.05)</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Schreiber et al25</td>
<td>Canada</td>
<td>SCC</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>38.2% vs 40.4% (p&lt;0.05)</td>
</tr>
<tr>
<td>Gu and Matsui22</td>
<td>Japan</td>
<td>SCC</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>55 vs 52*</td>
</tr>
<tr>
<td>Harpavat et al22</td>
<td>USA</td>
<td>Blood</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

*P value not provided.

SCC, stool colour chart.
Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

REFERENCES


41. Committee UNS. Newborn screening for biliary atresia, external review against programme appraisal criteria for the UK national screening committee (UK NSC). 2017.