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

Adam Arshad ,[1] Julian Gardiner,[1] Carmen Ho,[1] Philippa Rees ,[1] Karan Chadda,[1] Alastair Baker,[2] Alastair G Sutcliffe [3]

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/archdischild-2022-324946).

[1]Population, Policy & Practice Department, UCL GOS Institute of Child Health, London, UK
[2]Department of Paediatric Hepatology, Kings College Hospital, London, UK
[3]Population, Policy & Practice Department, UCL GOS Institute of Child Health, Institute of Child Health, London, UK

Correspondence to
Dr Adam Arshad, University College London Hospitals NHS Foundation Trust, London, UK; adam.arshad95@outlook.com

Received 7 October 2022
Accepted 30 January 2023
Published Online First 16 February 2023

To cite: Arshad A, Gardiner J, Ho C, *et al*. *Arch Dis Child* 2023;**108**:468–473.

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 demonstrate 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.

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.

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]

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 LTx. 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 V.X9.1. 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:

$$\text{PPV}: \quad \text{sensitivity} \times \text{prevalence}/\text{sensitivity} \times \text{prevalence} + (1 - \text{specificity}) \times (1 - \text{prevalence})$$

$$\text{NPV}: \quad \text{specificity} \times (1 - \text{prevalence}) / (1 - \text{sensitivity}) \times \text{prevalence} + \text{specificity} \times (1 - \text{prevalence})$$

RESULTS

The searches identified 9377 titles. Of these, 27 full texts were included (figure 1). This represents 2756 infants with BA and 4 019 847 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),[6 7 16–25] cross-sectional studies (n=7),[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,[7 25] seven in China,[13 15 20 28 29 31 32] six in Japan,[16 17 19 26 30 33] six in Taiwan[6 12 14 18 21 23] 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).

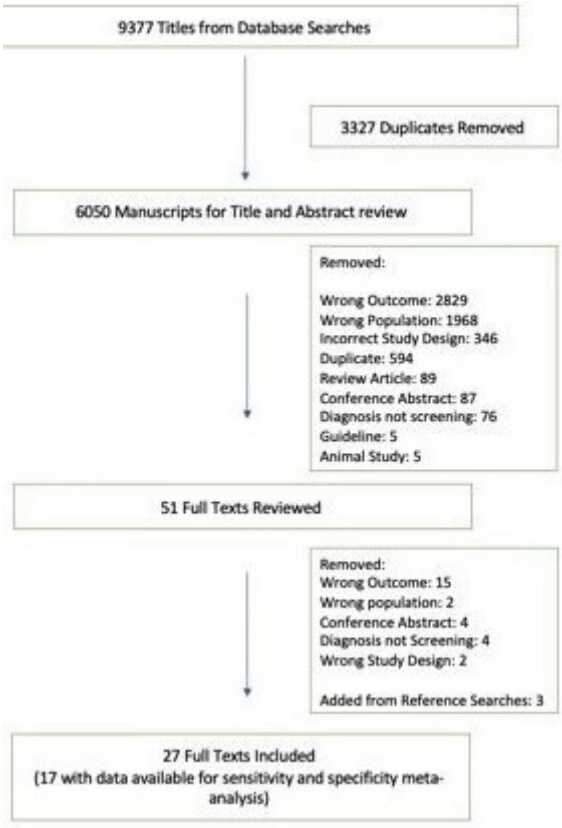

Figure 1 Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram demonstrating included and excluded studies and reason for exclusion in the systematic review of population-based screening for biliary atresia.

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: 5 used stool colour charts (SCCs)[14 17 18 20 23]; 4 measured bilirubin in blood samples[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[19 28 29 34]; and 1 assessed blood carnitine levels[29] (figures 2 and 3 and online supplemental appendix tables 7 and 8).

On meta-analysis, USBA 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 99.8%). 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 0.00% 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%

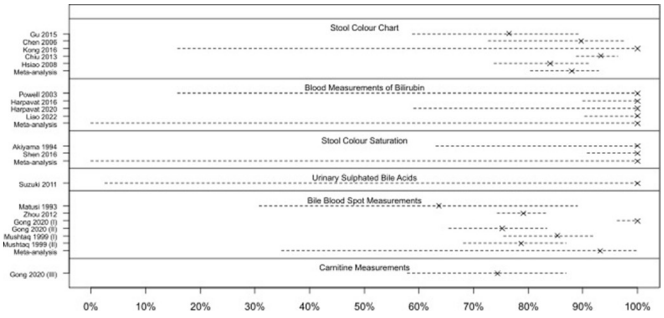

Figure 2 Meta-analysis of sensitivity of biliary atresia screening methods.

(95% CI 99.9% to 99.9%); bile acid blood spot measurements: 93.2% (95% CI 34.8% to 99.7%) and 95.5% (95% CI 65.8% to 99.5%); and blood carnitine measurements: 74.4% (95% CI 57.9% to 87.0%) and 94.0% (95% CI 93.5% to 94.5%).

Using a population prevalence of BA of 1 in 15 000, results showed that all methods of BA screening provided an NPV of 100%. SCC demonstrated the highest PPV of 5.6%. This was followed by USBA (1.3%) and conjugated bilirubin measurement (0.9%) (table 1).

Resulting age at Kasai

Data for the effect on screening on the resulting age of Kasai were available for eight studies: seven using SCC and one using conjugated bilirubin measurements (table 2).[6 12 13 17 18 21–23 33] All studies using SCC demonstrated a reduction in the age of Kasai, with the range of reduction being 8.4–25.0 days, and the average age at Kasai of 48.0–59.7 days after SCC.

One study explored the effect of conjugated bilirubin screening on the age of Kasai. Harpavat *et al* conducted a two-stage BA screening programme. In stage 1, all newborns were tested within the first 60 hours of life. In stage 2, patients with a positive result were retested and considered positive if the bilirubin was greater than the stage 1 result or 1 mg/dL. This intervention demonstrated a significant reduction in Kasai age, with a between-group difference (intervention vs non-screened cohort) of 19 days and improvements in Kasai under 30-days (preintervention and postintervention, 12.5% vs 57.9%, p=0.003).[22]

Associated morbidity, hospital admission and mortality

Five studies explored BA-associated morbidity, hospital admission and 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).

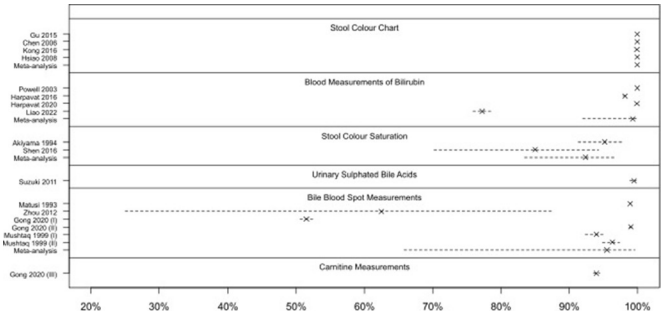

Figure 3 Meta-analysis of specificity of biliary atresia screening methods.

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).[35]

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

PPV				
Method	Prevalence in population tested (%)			
	1 in 100	1 in 1000	1 in 10 000	1 in 15 000
Conjugated	59.1	12.5	1.4	0.9
USBA	66.9	16.7	2.0	1.3
SCC	90.0	47.0	8.1	5.6
SCS	9.3	1.0	0.1	0.1
Bile acid blood spot	4.7%	0.5%	0.0%	0.0%
NPV				
Method	Prevalence in population tested (%)			
	1 in 100	1 in 1000	1 in 10 000	1 in 15 000
Conjugated	100.0	100.0	100.0	100.0
USBA	100.0	100.0	100.0	100.0
SCC	99.9	100.0	100.0	100.0
SCS	100.0	100.0	100.0	100.0
Bile acid blood spot	99.8	100.0	100.0	100.0

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

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

Title	Country	Intervention	Average age at Kasai		Kasai at less then 60 days	
			Preintervention versus postintervention	P value	Preintervention versus postintervention (%)	P value
Gu *et al*[17]	Japan	SCC	70.3 vs 59.7	**0.003**	34.0 vs 55.9	**>0.05**
Tseng *et al*[12]	Taiwan	SCC	51 vs 48	0.051	68.9 vs 73.6	0.31
Chen *et al*[18]	Taiwan	SCC	–	–	N/A vs 58.6	
Lee *et al*[21]	Taiwan	SCC	59.9 vs 48.2	0.064	68.4 vs 73.7	0.242
Lien *et al*[6]	Taiwan	SCC	–		49.4 vs 65.7	**0.020**
Zheng *et al*[13]	China	SCC	81 vs 56	**<0.05**	35.3 vs 64.5	**<0.05**
Hsiao *et al*[23]	Taiwan	SCC	N/A vs 54.1	–	47.2 vs 74.3	**0.004**
Gu *et al*[17]	Japan	SCC	68.1 vs 59.7	**0.003**	55.9 vs 40.4	0.109
Harpavat *et al*[22]	USA	Conjugated bilirubin measurements	56 vs 36	**0.004**	12.5* vs 57.9	**0.003**

Values provided are in days.
Bold p values are significant at p<0.05.
No p value provided.
*Presented data for Kasai <30 days (12.5% vs 57.9%, p=0.003).
N/A, not available; SCC, stool colour chart.

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 SCCs, 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 al*[22] used a two-stage screening strategy of 123 385 newborns across

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

Title	Country	Intervention	Overall mortality (preintervention vs postintervention)	Transplant-free survival (preintervention vs postintervention)	Average number of hospitalisations (preintervention vs postintervention)	Length of stay in hospital per visit (preintervention vs postintervention)	Liver transplantation (preintervention vs postintervention)
Lee *et al*[21]	Taiwan	SCC	47.8% vs 21.2% (p<0.001)	31.6% vs 56.4% (p<0.001)	6.4 vs 5.0 (p<0.001)	86.6% vs 81.9% (p=0.438)	28.6% vs 28.2% (p=0.934)
Zheng *et al*[13]	China	SCC	20.6% vs 10.5% (p<0.05)	44.4% vs 52.6% (p>0.05)	–	–	38.2% vs 40.4% (p>0.05)
Schreiber *et al*[25]	Canada	SCC	–	–	–	–	55 vs 52*
Gu and Matsui[33]	Japan	SCC	–	197.2 vs 81 months (p=0.017)	–	–	–
Harpavat *et al*[22]	USA	Blood	–	70.8% vs 94.7% (p=0.06)	–	–	–

*P value not provided.
SCC, stool colour chart.

14 hospitals in Southern Texas to screen for BA. This involved an initial measurement of conjugated bilirubin within the first 60 hours of life. If the result was above the reference values for conjugated bilirubin, a repeat test was conducted at 2–3 weeks of age. This method provided a sensitivity of 100% (95% CI 56% to 100%), a specificity of 99.9% (95% CI 99.9% to 99.9%) and a PPV of 5.9% (95% CI 2.6% to 12.2%). The screening strategy reduced the age of Kasai (when compared with the preintervention average) from 56 vs 36 days (p=0.004) and provided a greater post-Kasai survival rate (71% vs 95%, p=0.060).[22]

While providing positive results, further research into this modality of population-based screening is required. In the aforementioned study, the number of true positive cases was small (n=7), providing broad CIs in the reported statistics. Follow-up was for 1–4 years after screening, and if cases had been missed, it would significantly alter reported CIs. Furthermore, screening was positive for 112 infants without BA, resulting in a PPV of 5.9% and associated costs of further investigations.

Pivotal to the applicability of any future population screening method for BA is an analysis of its cost-effectiveness. Cost-effectiveness analyses have been conducted in the USA and Canada and demonstrate an overall cost–benefit from implementing an SCC screening programme (compared with conjugated bilirubin or a non-screening strategy). Specifically, the use of conjugated bilirubin measurements is deemed to have an unacceptable high cost.[35] However, existing literature must be updated with new emerging data and must also consider additional costs associated with LTx (eg, complications such as acute rejection management, cancer risk, etc). Furthermore, as discussed, the cost associated with false-positive results should be factored into any cost-effectiveness analysis. In Harpavat *et al*, over 50% of the 112 patients with false-positive results were unable to be diagnosed, despite extensive investigations (that included liver biopsies), incurring significant healthcare costs and potential iatrogenic harm to infants.[22] Finally, future studies should only assess infants in the first weeks of life. This is the cohort we must develop an effective screening tool for to achieve the target of Kasai of <30 days.

Our systematic review and meta-analysis demonstrates that the current literature is divergent as to an efficacious, practical and cost-effective method for BA screening. SCC has shown significant reductions in the age of Kasai in Eastern Asian nations, but its applicability to Western healthcare systems is unclear. Conjugated bilirubin measurements appear to provide the greatest sensitivity and specificity in BA detection but incur significant healthcare-related costs. Until further research is conducted on large sample sizes, BA remains a condition in need of an effective population-based screening programme.

Twitter Philippa Rees @PhilippaCRees

Contributors AA designed the study; conducted data extraction, data analysis and data interpretation; and further wrote the original draft and reviewed the final manuscript. JG conducted statistical analysis for the meta-analysis. CH contributed to the data analysis. PR contributed to study design and data interpretation. KC aided the data extraction of the required studies from databases for this study. AB designed the study and aided the data interpretation. AA acts as guarantor.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request.

ORCID iDs
Adam Arshad http://orcid.org/0000-0002-3870-3663
Philippa Rees http://orcid.org/0000-0002-1074-5837
Alastair G Sutcliffe http://orcid.org/0000-0001-8542-6155

REFERENCES

1 Sanchez-Valle A, Kassira N, Varela VC, *et al*. Biliary atresia: epidemiology, genetics, clinical update, and public health perspective. *Adv Pediatr* 2017;64:285–305.
2 Schreiber RA. Newborn screening for biliary atresia. *JAMA* 2020;323:1137–8.
3 Noorulla F, Dedon R, Maisels MJ. Association of early direct bilirubin levels and biliary atresia among neonates. *JAMA Netw Open* 2019;2:e1913321.
4 Serinet M-O, Wildhaber BE, Broué P, *et al*. Impact of age at kasai operation on its results in late childhood and adolescence: a rational basis for biliary atresia screening. *Pediatrics* 2009;123:1280–6.
5 Nio M, Wada M, Sasaki H, *et al*. Effects of age at kasai portoenterostomy on the surgical outcome: a review of the literature. *Surg Today* 2015;45:813–8.
6 Lien T-H, Chang M-H, Wu J-F, *et al*. Effects of the infant stool color card screening program on 5-year outcome of biliary atresia in Taiwan. *Hepatology* 2011;53:202–8.
7 Woolfson JP, Schreiber RA, Butler AE, *et al*. Province-wide biliary atresia home screening program in British Columbia: evaluation of first 2 years. *J Pediatr Gastroenterol Nutr* 2018;66:845–9.
8 Borgeat M, Korff S, Wildhaber BE. Newborn biliary atresia screening with the stool colour card: a questionnaire survey of parents. *BMJ Paediatr Open* 2018;2:e000269.
9 Bezerra JA. Biliary atresia in Brazil: where we are and where we are going. *J Pediatr (Rio J)* 2010;86:445–7.
10 Ouzzani M, Hammady H, Fedorowicz Z, *et al*. Rayyan-a web and mobile APP for systematic reviews. *Syst Rev* 2016;5:210.
11 Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* 1934;26:404–13.
12 Tseng JJ, Lai MS, Lin MC, *et al*. Stool color card screening for biliary atresia. *Pediatrics* 2011;128:e1209–15.
13 Zheng J, Ye Y, Wang B, *et al*. Biliary atresia screening in Shenzhen: implementation and achievements. *Arch Dis Child* 2020;105:720–3.
14 Chiu C-Y, Chen P-H, Chan C-F, *et al*. Biliary atresia in preterm infants in taiwan: a nationwide survey. *J Pediatr* 2013;163:100–3.
15 Liao FM, Chang KC, Wu JF, *et al*. Direct bilirubin and risk of biliary atresia. *Pediatrics* 2022;149:e2021053073.
16 Suzuki M, Muraji T, Obatake M, *et al*. Urinary sulfated bile acid analysis for the early detection of biliary atresia in infants. *Pediatr Int* 2011;53:497–500.
17 Gu Y-H, Yokoyama K, Mizuta K, *et al*. Stool color card screening for early detection of biliary atresia and long-term native liver survival: a 19-year cohort study in japan. *J Pediatr* 2015;166:897–902.
18 Chen S-M, Chang M-H, Du J-C, *et al*. Screening for biliary atresia by infant stool color card in Taiwan. *Pediatrics* 2006;117:1147–54.
19 Matsui A, Sasakia N, Arakawa Y, *et al*. Neonatal mass screening for biliary atresia: A pilot study in tochigi prefecture, Japan. *Screening* 1993;2:201–9.
20 Kong Y-Y, Zhao J-Q, Wang J, *et al*. Modified stool color card with digital images was efficient and feasible for early detection of biliary atresia-a pilot study in Beijing, China. *World J Pediatr* 2016;12:415–20.
21 Lee M, Chen SC-C, Yang H-Y, *et al*. Infant stool color card screening helps reduce the hospitalization rate and mortality of biliary atresia: a 14-year nationwide cohort study in taiwan. *Medicine (Baltimore)* 2016;95:e3166.
22 Harpavat S, Garcia-Prats JA, Anaya C, *et al*. Diagnostic yield of newborn screening for biliary atresia using direct or conjugated bilirubin measurements. *JAMA* 2020;323:1141–50.
23 Hsiao C-H, Chang M-H, Chen H-L, *et al*. Universal screening for biliary atresia using an infant stool color card in Taiwan. *Hepatology* 2008;47:1233–40.
24 Powell JE, Keffler S, Kelly DA, *et al*. Population screening for neonatal liver disease: potential for a community-based programme. *J Med Screen* 2003;10:112–6.
25 Schreiber RA, Masucci L, Kaczorowski J, *et al*. Home-Based screening for biliary atresia using infant stool colour cards: a large-scale prospective cohort study and cost-effectiveness analysis. *J Med Screen* 2014;21:126–32.
26 Akiyama T, Yamauchi Y. Use of near infrared reflectance spectroscopy in the screening for biliary atresia. *J Pediatr Surg* 1994;29:645–7.
27 Harpavat S, Ramraj R, Finegold MJ, *et al*. Newborn direct or conjugated bilirubin measurements as a potential screen for biliary atresia. *J Pediatr Gastroenterol Nutr* 2016;62:799–803.

28 Zhou K, Lin N, Xiao Y, *et al*. Elevated bile acids in newborns with biliary atresia (ba). *PLoS One* 2012;7:e49270.
29 Gong Z, Wu Y, Zheng L, *et al*. Can free carnitine or bilirubin in blood be used in neonatal screening for biliary atresia? *Eur J Pediatr Surg* 2020;30:459–64.
30 Muraji T, Harada T, Miki K, *et al*. Urinary sulfated bile acid concentrations in infants with biliary atresia and breast-feeding jaundice. *Pediatr Int* 2003;45:281–3.
31 Shen Z, Zheng S, Dong R, *et al*. Saturation of stool color in HSV color model is a promising objective parameter for screening biliary atresia. *J Pediatr Surg* 2016;51:2091–4.
32 Xiao Y, Zhou Y, Zhou K, *et al*. Targeted metabolomics reveals birth screening biomarkers for biliary atresia in dried blood spots. *J Proteome Res* 2022;21:721–6.
33 Gu YH, Matsui A. Long-Term native liver survival in infants with biliary atresia and use of a stool color card: case-control study. *Pediatr Int* 2017;59:1189–93.
34 Mushtaq I, Logan S, Morris M, *et al*. Screening of newborn infants for cholestatic hepatobiliary disease with tandem mass spectrometry. *BMJ* 1999;319:471–7.
35 Masucci L, Schreiber RA, Kaczorowski J, *et al*. Universal screening of newborns for biliary atresia: cost-effectiveness of alternative strategies. *J Med Screen* 2019;26:113–9.
36 Mogul D, Zhou M, Intihar P, *et al*. Cost-effective analysis of screening for biliary atresia with the stool color card. *J Pediatr Gastroenterol Nutr* 2015;60:91–8.
37 Wildhaber BE. Screening for biliary atresia: swiss stool color card. *Hepatology* 2011;54:367–8.
38 Madadi-Sanjani O, Blaser J, Voigt G, *et al*. Home-based color card screening for biliary atresia: the first steps for implementation of a nationwide newborn screening in germany. *Pediatr Surg Int* 2019;35:1217–22.
39 Hopkins PC, Yazigi N, Nylund CM. Incidence of biliary atresia and timing of hepatoportoenterostomy in the united states. *J Pediatr* 2017;187:253–7.
40 Davenport M, Ong E, Sharif K, *et al*. Biliary atresia in England and Wales: results of centralization and new benchmark. *Journal of Pediatric Surgery* 2011;46:1689–94.
41 Committee UNS. *Newborn screening for biliary atresia. external review against programme appraisal criteria for the UK national screening committee (UK NSC)*. 2017.

Supplementary Appendix - Table 1: PRISMA checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	4
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	4 - 6
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	5 - 6
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	5 - 6
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	5 - 6
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	5 - 6
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	5 - 6
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	5 - 6

Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	5 - 6
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	5 - 6
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	5 - 6
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	5 - 6
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	5 - 6
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	5 - 6
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	7
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	7
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	7
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	7 - 10
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	7 - 10
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	7
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	7
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	11 - 15
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	15

Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	15
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data), role of funders for the systematic review.	5

Supplementary Appendix Table 2: MOOSE Checklist

Reporting Criteria	Reported (Yes/No)	Reported on Page
Reporting of Background		
Problem definition	Yes	4
Hypothesis statement	Yes	4 - 5
Description of Study Outcome(s)	Yes	4 - 5
Type of exposure or intervention used	No	N/A
Type of study design used	Yes	5
Study population	Yes	5
Reporting of Search Strategy		
Qualifications of searchers (eg, librarians and investigators)	Yes	1
Search strategy, including time period included in the synthesis and keywords	Yes	5 - 6
Effort to include all available studies, including contact with authors	Yes	5 -6
Databases and registries searched	Yes	5
Search software used, name and version, including special features used (eg, explosion)	Yes	5 - 6
Use of hand searching (eg, reference lists of obtained articles)	Yes	5
List of citations located and those excluded, including justification	Yes	7
Method for addressing articles published in languages other than English	Yes	5
Method of handling abstracts and	Yes	5

unpublished studies		
Description of any contact with authors	Yes	5
Reporting of Methods		
Description of relevance or appropriateness of studies assembled for assessing the hypothesis to be tested	Yes	5
Rationale for the selection and coding of data (eg, sound clinical principles or convenience)	Yes	5
Documentation of how data were classified and coded (eg, multiple raters, blinding, and interrater reliability)	Yes	5
Assessment of confounding (eg, comparability of cases and controls in studies where appropriate	Yes	5
Assessment of study quality, including blinding of quality assessors. stratification or regression on possible predictors of study results Y	Yes	5
Assessment of heterogeneity	No	Cannot be conducted within out study. Bias assessments conducted
Description of statistical methods (eg, complete description of fixed or random effects models, justification of whether the chosen models account for predictors of study results, dose-response models, or cumulative meta-analysis) in sufficient	Yes	5

detail to be replicated		
Provision of appropriate tables and graphics	Yes	See Tables and Figures
Reporting of Results		
Table giving descriptive information for each study included	Yes	See Tables and Figures
Results of sensitivity testing (eg, subgroup analysis)	Yes	See Tables and Figures
Indication of statistical uncertainty of findings	Yes	See Tables and Figures
Reporting of Discussion		
Quantitative assessment of bias (eg, publication bias)	Yes	7
Justification for exclusion (eg, exclusion of non–English-language citations)	Yes	5
Assessment of quality of included studies	Yes	7
Reporting of Conclusions		
Consideration of alternative explanations for observed results	Yes	11 - 15
Generalization of the conclusions (i.e., appropriate for the data presented and within the domain of the literature review)	Yes	11 - 15
Guidelines for future research	Yes	15
Disclosure of funding source	Yes	5

Supplementary Appendix Table 3: Databases searched for systematic review of population-based screening for Biliary Atresia

Database	Date range searched	Date searched	Number of results
Cochrane Central Register of Controlled Trials	1946 - current	10.09.2022	306
EBSCO–CINAHL (Cumulative Index to Nursing and Allied Health Literature)	1971 - current	10.09.2022	206
Google Scholar	-	10.09.2022	1670
Ovid–EMBASE	1974 – September 10th 2022	10.09.2022	1836
Ovid–HMIC (Health Management Information Consortium)	1979 to September 2022	10.09.2022	0
Ovid–MEDLINE	1946 - current	10.09.2022	953
Ovid–MEDLINE E-pub ahead of print	September 10th 2022	10.09.2022	0
Ovid–MEDLINE In-Process and Other Non-Indexed Citations	1946 - current	10.09.2022	56
PubMed	1963 - current	10.09.2022	807
Scopus	-	10.09.2022	1463

Web of Knowledge (science citation index expanded and conference proceedings citation index science)	1969 - current	10.09.2022	2080

Supplementary Appendix Table 4: Medline Ovid Search Strategy to identify studies on population-based screening for Biliary Atresia

1. exp Infant/ or exp Infant, Newborn/
2. exp Child/
3. neonate.mp
4. baby.mp
5. newborn.mp
6. neonates.mp
7. neonatal.mp
8. Screen.mp.
9. Screening.mp
10. exp Mass Screening/
11. exp Neonatal Screening/
12. exp Jaundice/ or exp Jaundice, Obstructive/ or exp Jaundice, Neonatal/
13. exp Cholestasis, Extrahepatic/ or exp Cholestasis/ or exp cholestasis, Intrahepatic/
14. exp Liver/
15. exp Bilirubin/
16. biliary.mp or exp Biliary Tract Surgical Procedures/ or Biliary Tract Diseases/ or exp Biliary Atresia/ or exp Biliary Tract/ or exp Liver Cirrhosis, Biliary/
17. cohort*.tw.
18. exp Epidemiologic Methods/
19. exp Case-Control Studies/
20. (case$ and control$).tw.
21. exp Cohort Studies/
22. exp Retrospective Studies/
23. exp Cross-Sectional Studies/
24. Animals/
25. animal stud*.mp.
26. exp "Review"/
27. exp Case Reports/

28. 1 or 2 or 3 or 4 or 5 or 6 or 7
29. 8 or 9 or 10 or 11
30. 12 or 13 or 14 or 15 or 16
31. 17 or 18 or 19 or 20 or 21 or 22 or 23
32. 24 or 25 or 26 or 27
33. 28 and 29 and 30 and 31
34. 33 not 32
35. Limit 34 to (English language and yr= '1975 – current')

Supplementary Appendix Table 5: Characteristics of included studies exploring population-based screening for Biliary Atresia

Study (study design)	Country	Age at Testing	Number Included	Screening Method	Outcomes	Comments
Akiyama et al. 1994 (Cross-sectional study)	Japan	Healthy Group (Mean) – 50.1 months BA group (mean) – 30 months	200 Healthy Infants, 8 BA and 8 Neonatal Hepatitis	Infrared reflectance spectrometry of Stool Samples	Sensitivity: 100%, Specificity: 95.2%	
Suzuki et al. 2011 (Prospective cohort study)	Japan	39 weeks	1148	Measurement of urinary sulfated bile acid (USBA)	Sensitivity - 100%, Specificity - 96%, NPV - 4%, PPV - 100%	Author's state high FPR may be secondary to the use of an ordinary mail collection and delivery system with no temperature regulation and the 3- to 6-day interval between sampling and receipt.
Masucci et al. 2019 (Cost-Effectiveness Analysis)	N/A	N/A	N/A	N/A	SCC cost approximately $192,000 more than no universal screening but led to eight life-years gained (incremental cost-effectiveness ratio (ICER) $24,065 per life-year gained). Screening using conjugated bilirubin testing	

					versus the colour card cost $2,369,199 more and led to five more life-years gained (ICER - $473,840 per life year gained), and so was not cost-effective.	
Gu et al. 2015 (Prospective Cohort Study)	Japan	Unclear, appears physician stools reviewed at 1 month of age	264071	SCC	Sensitivity - 76.5% (95% CI: 62.2 - 90.7), Specificity - 99.9% (95% CI: 99.9 - 100.0) NPV - 99.9% (95% CI: 99.9 - 99.9), PPV - 12.7% (8.2% - 17.3%) Age at Kasai before SCC 70.3 days. After SCC 59.7 (p = 0.03) Improved 5-, 10- and 15-year native liver survival (87.6%, 76.9% and 48.5%) compared to studies conducted in US, UK and France	
Tseng et al. 2011 (Retrospective Cohort Study)	Taiwan	Unclear, appears physician reviewed stools at 1 month of age	2,246,924 born before SCC. 1029879 born after SCC.	SCC	Median age at first presentation decreased (47 vs. 43, p = 0.028). Late referrals decreased from 9.5% to 4.9%. The median age of Kasai operation decreased (51 vs. 48. p = 0.051). The proportions of Kasai operation within 60 days decreased (68.9% vs. 73.6%, p = 0.31)	
Chen et al. 2006 (Prospective Cohort Study)	Taiwan	Stools reviewed at 1 month of age	Type 1: 29412 Type 2: 37632	SCC (Type 1: Labeled and Type 2: Unlabeled)	For the detection of BA before 60 days: • Type 1: Sensitivity 86.7%, Specificity 99.9%. NPV - 99.9%, PPV - 41.9%. • Type 2: Sensitivity: 88.8%, Specificity: 99.9%, NPV: 99.9%, PPV: 20.0	

					17 (58.6%) infants received a Kasai operation before 60 days of age.	
Woolfson et al. 2018 (Prospective Cohort Study)	Canada (British Columbia)	Stools reviewed daily up to 1 month of age	87,583	SCC	Sensitivity: 50%, Specificity: 99%, NPV: 99% and PPV: 4% Set-up and Operational Costs in 1st Year: $80,154.63. Operational costs in 2nd year: $330,033.82. Additional cost of $50,120.81 for program launch in its inaugural first year. After program start up, ISCC cost per birth, including ongoing administrative expenses was $0.86	
Harpavat et al. 2016 (Cross-sectional study)	USA	Newborns (exact age not detailed)	BA Cohort: 61 Non-BA Cohort: 9102	Conjugated Blood Measurements	Sensitivity - 100% (95% CI: 87.7 - 100), Specificity - 98.2% (95% CI: 97.9 - 98.4)	
Matsui et al. 1993 (Prospective Cohort Study)	Japan	1 month of age	104,309	Total 3x- OH bile acids were extracted from dried blood spots	Sensitivity: 63.6%, PPV: 0.62%	
Kong et al. 2016 (Prospective Cohort Study)	China	Daily check until 4 months of age	29 799	SCC	Sensitivity: 100%, Specificity: 99.9%, PPV - 8.3% (95% CI: 2.7-19.4)	
Lee et al. 2016 (Prospective Cohort	Taiwan	Review at 2 months	513 BA cases (Comparison done by BA cases before	SCC	SCC reduced the average Kasai operation age (59.9 vs. 48.2, p = 0.064).	

Study)		of age	and after screening introduced)		SCC reduced hospitalization rate in the first 2 years of life (6.4 vs. 5.0, p < 0.001). SCC also reduced the death rate within the UK (47.8 vs. 21.2, p < 0.001) and percentage of infants having neither LTX nor death (31.6% vs. 56.4%, P < 0.001). Finally, there was no significant difference in the rate of LTx (28.6% vs. 28.2, p = 0.934).	
Zhou et al. 2012 (Cross sectional study)	China	4 days after birth	292 normal infants, 17 neonatal jaundice and 8 biliary atresia	Bile acids from dried blood spots	With a cutoff of 0.63 mmol/L, produces a sensitivity: 79.1 (74.3 - 83.2), specificity: 62.5 (25 - 87.5)	
Lien et al. 2016 (Prospective Cohort Study)	Taiwan	Unclear, appears daily from birth	191 BA Infants	SCC	3-year overall survival improved after SCC implementation (64.0% vs. 89.2% P < 0.001). The 5-year survival rates with native liver in cohorts A and B were (37.5% vs. 64.3%, P = 0.01). The 5-year overall survival rates were 89.3% vs. 55.7%, (P < 0.001).	
Harpavat et al. 2020 (Prospective Cohort Study)	USA	After Birth	123,279 infants	Conjugated Blood Measurements	Sensitivity: 100.0% (95% CI, 56.1%-100.0%), Specificity: 99.9% (95% CI, 99.9%-99.9%), PPV: 5.9% (95% CI, 2.6%-12.2%), NPV: 100.0% (95% CI, 100.0%- 100.0%) Screening reduced age at presentation (56 vs. 36 days, p = 0.004) and proportion having Kasai < 30 days (12.5% vs. 57.9%, p = 0.003). Screening reduced the age the patient referred	

					to a specialist (44 vs. 25 days, p 0.003). After Kasai, infants in screening group had significantly faster time of bilirubin normalization, but no sig. difference in transplant free survival. Screening infants more likely to have a normal conjugated bilirubin by 90 days (41.7% vs. 78.9%, p = 0.03).	
Mogul et al. 2015 (Cost effectiveness Study)	N/A	N/A	N/A	SCC	With no screening, the 20-year cost was $142,479,725 with 3702 life- years, 74 deaths and 158 liver transplants. With SCC B, the cost was $133,893,563 with 3731.7 life-years, 71 deaths and 147 liver transplants. There was a >97% probability that screening with the stool color card would be cost saving and associated with an increase in life-years gained. Among all parameters, only stool color card specificity was associated with the potential for screening to no longer be cost saving.	
Gong et al. 2020 (Cross Sectional Study)	China	3 – 14 days after birth	52, 862	Free carnitine, unconjugated bilirubin (UBIL), Bilirubin monoglucuronide (BMG), and Bilirubin diglucuronide (BDG) in dry blood spots	Direct Bilirubin: • Using 30 u/mol as cut off - Sensitivity: 100%, Specificity: 52%. • Using 140 u/mol as cut off - sensitivity: 75%, Specificity: 99%. Free Carnitine: • Using 38 u/mol as cut off: sensitivity 85%, Specificity: 85%. • Using 38 u/mol as cut off - Sensitivity: 75%, Specificity: 94%.	

Zheng et al. 2020 (Retrospective cohort study)	China	Unclear – appears daily from birth	118 BA cases	SCC	SCC reduced age at Kasai (56 vs. 81, $p < 0.05$), Length of stay in hospital (44 vs. 49, $p < 0.05$). It improved 2-year native liver survival rate (44.4% vs. 52.6%, $p < 0.05$) and survival (20.6% vs. 10.5%, $p < 0.05$).	
Chiu et al. 2013 (Retrospective Cohort Study)	Taiwan	Daily from birth	197 BA Cases	SCC	Sensitivity in detecting BA using SCC before 60 days: 92.8%. 96.3% in the preterm infants	
Muraji et al. 2003 (Cross Sectional Study)	Japan	21 – 138 days	58 infants with Breast feeding Jaundice. 16 BA infants	Urinary excretion of sulfated bile acid	Sensitivity - 100%, FPR - 1.0%	
Hsaio et al. 2008 (Prospective Cohort Study)	Taiwan	Daily from Birth	422273 Infants	SCC	2004: Sensitivity - 72.5%, 2005 - 97.1% Proportion of Kasai < 60 days: 47.2% prior to SCC, vs. 60% in 2004 and 74.3% in 2005 (once SCC introduced) 1976 - 2000 ($p = 0.004$). Delayed operation rate beyond 90 days decreased over time, from 15.3% in 1976-2000 to 10.3% in 2002-2003 and 0% in 2004 and 2005	
Powell et al. 2003 (Prospective Cohort Study)	United Kingdom	Babies under 28 days	27654	Conjugated Bilirubin	Using bilirubin cut off of 18 u/mol/l: True positives: 2, False negatives: 0 False positives: 10, True negatives: 23,107	

Schreiber et al. 2014 (Prospective Cohort Study)	Canada	Daily up to 4 weeks	6187	SCC	Liver transplants decreased from 55 (no screening) to 52 (SCC). For a Canadian population, the increase in cost for passive screening, compared with no screening, is $213,584 and the gain in life years is 9.7 ($22,000 per life-year gained).	
Mushtaq et al. 1999 (Case-Control Study)	United Kingdom	Infants < 1 year	218 infants with cholestasis	Mass Spectrometry on Blood Spots	Sensitivity/Specificity/PPV/NPV: cut off of 25 umol/l produced figures of 85.3%, 94.0%, 14.2, and 0.16, and a cut off of 35 umol/l 70.5%, 97.8%, 32.0, and 0.30, respectively	Unfortunately, there is too much overlap between bile acid concentrations in infants with cholestasis and those in control infants for this to be used as a single screening test for cholestatic hepatobiliary disease in general and biliary atresia
Shen et al. 2016 (Cross Sectional Study)	China	Neonates ranging from 18 – 94 days	40 BA cases, 40 Neonates with Pneumonia	Light Spectrometry (with phone application – POOPMD)	Sensitivity - 100%, Specificity - 34/40	
Gu et al. 2017 (Case-Control Study)	Japan	Unclear – appears daily until 1 month physician review	148 BA cases	SCC	Kasai < 60 days: 55.9% vs. 40.4% (p = 0.109), Native liver survival 197.2 months before SCC vs. 81 months after SCC, p = 0.017)	

Liao et al. 2022	China	Newborns 0 – 60 days	38 BA cases	Direct Bilirubin	Using $\geq$ 1 mg dL as cut-off: Sensitivity 100% Specificity 77.26%	
Xiao et al. 2022	China	36 – 40 weeks	21 BA cases	THCA, 2-hydroxyglutaric acid, and indoleacetic acid in dried blood spots	Sensitivity of 90.48% (95% CI: 69.62% – 98.83%) and specificity of 92% (95% CI: 84.84% – 96.48%).	

SCC: Stool Colour Chart, PPV: Positive Predictive Value; NPV; Negative Predictive Value, 95% CI: 95% Confidence Intervals

Supplementary Appendix Table 6a: Quality assessment scores for cohort studies exploring population-based screening for Biliary Atresia

Study (Year)	Selection				Comparability		Exposure/Outcome			Subtotal Assessment			Conclusion
	1	2	3	4	1a	1b	1	2	3	S	C	E/O	
Cohort Studies													
Suzuki et al. (2019)	*	*	*	*	*	*	*	*	*	Good	Good	Good	Good
Gu et al. (2015)	*	*	*	*	*	*	*	*	*	Good	Good	Good	Good
Tseng et al. (2011)	*	No	*	*	No	No	*	*	*	Good	Poor	Good	Good
Chen et al. (2006)	*	*	*	*	*	*	*	*	*	Good	Good	Good	Good
Woolfson et al. (2018)	*	*	*	*	*	*	*	*	*	Good	Good	Good	Good
Matsui et al. (1993)	*	*	No	*	*	*	*	*	Follow-up rate 80%	Good	Good	Fair	Good
Kong et al. (2016)	*	*	*	*	*	*	*	*	*	Good	Good	Good	Good
Lee et al. (2016)	*	*	*	*	*	*	*	*	No statement	Good	Good	Fair	Good

Lien et al. (2011)	*	No (different provinces)	*	*	Different Years	Different Region	*	*	*	Good	Poor	Good	Good
Harpavat et al. (2020)	*	Different Source	*	*	Different Years	Different Region	*	*	*	Good	Poor	Good	Fair
Zheng et al. (2020)	*	*	No description	*	Different Years	*	*	*	*	Good	Poor	Good	Fair
Chiu et al. (2013)	*	No	*	*	N/A	N/A	*	*	No Statement	Good	-	Good	Good
Hsaio et al. (2008)	*	*	*	*	Different Years	Different Region	*	*	*	Good	Poor	Good	Good
Powell et al. (2003	*	No unexposed group	*	*	N/A	N/A	*	*	84.70%	Good	-	Good	Good
Schreiber et al. (2014)	*	No unexposed group	*	*	N/A	N/A	*	*	40% return rate	Good	-	Good	Good
Liao et al. (2022)	*	No unexposed group	*	*	N/A	N/A	*	*	*	Good	-	Good	Good

Supplementary Appendix Table 6b: Quality assessment scores for cross sectional studies exploring population-based screening for Biliary Atresia

Study (Year)	Selection				Comparability	Exposure/Outcome		Subtotal Assessment Conclusion			Conclusions
	1	2	3	4	1a	1	2	1	2	3	
Cross sectional studies											
Akiyama et al. (1994)	*	No	*	**	No	*	*	Good	Poor	Good	Good
Harpavat et al. (2016)	*	No	*	**	No	*	*	Good	Poor	Good	Good
Zhou et al. (2012)	*	No	*	**	**	**	*	Good	Good	Good	Good
Gong et al. 2020	*	No	*	**	No	**	*	Good	Poor	Good	Good
Muraji et al. (2013)	*	No	*	**	No	**	*	Good	Poor	Good	Good

Shen et al. (2016)	*	No	*	**	**	*	*	Good	Good	Good	Good
Xiao et al. (2022)	*	No	*	**	**	*	*	Good	Good	Good	Good

Supplementary Appendix Table 6c: Quality assessment scores for case-control studies exploring population-based screening for Biliary Atresia

Study (Year)	Selection				Comparability		Exposure/Outcome			Subtotal Assessment			Conclusion
	1	2	3	4	1a	1b	1	2	3	S	C	E/O	
Case-Control Studies													
Mushtaq et al. 1999	No	*	N/A	*	*	No	*	*	*	Good	Fair	Good	Good
Gu et al. 2017	No	*	N/A	*	*	No	*	*	*	Good	Fair	Good	Good

Supplementary Appendix Table 7: Summary of study data for a meta-analysis on the sensitivity and specificity of BA screening methods

Study Number	Reference	Method	Numbers of			
			True Positives	False Negatives	False Positives	True Negatives
4	Gu 2015	Stool Colour Chart	26	8	177	263859
6	Chen 2006		26	3	65	78090
10	Kong 2016		2	0	22	22775
18	Chiu 2013		181	13	n/a	n/a
20	Hsiao 2008		63	12	279	422273
21	Powell 2003	Blood Measurements of Bilirubin	2	0	10	23107
8	Harpavat 2016		35	0	166	8936
14	Harpavat 2020		7	0	122	123140
25	Liao et al. 2022		36	0	929	3157
1	Akiyama 1994	Stool Colour Saturation	8	0	10	198
24	Shen 2016		38	0	6	34
2	Suzuki 2011	Urinary Sulphated Bile Acids	1	0	6	1141
9	Matusi 1993	Bile Blood Spot Measurements	7	4	1129	103173
12	Zhou 2012		n/a	n/a	n/a	n/a
16	Gong 2020 (I)		97	0	4894	5204
16	Gong 2020 (II)		73	24	100	9908
23	Mushtaq 1999		n/a	n/a	n/a	n/a
16	Gong 2020 (III)	Carnitine Measurements	29	10	600	9408

Gong 2020 (I) results with cut-off bilirubin > 30 µmol/l.
Gong 2020 (II) results with cut-off bilirubin > 140 µmol/l.
Gong 2020 (III) results with cut-off free carnitine > 45 µmol/l.

Supplementary Appendix Table 8: Sensitivity and specificity for studies, with 95% confidence intervals.

Reference	Method	Sensitivity		Specificity	
		Estimate	Confidence Interval	Estimate	Confidence Interval
Gu 2015	Stool Colour Chart	76.47%	(58.83%,89.25%)	99.93%	(99.92%,99.94%)
Chen 2006		89.66%	(72.65%,97.81%)	99.92%	(99.89%,99.94%)
Kong 2016		100.00%	(15.81%,100.00%)	99.90%	(99.85%,99.94%)
Chiu 2013		93.30%	(88.81%,96.38%)	n/a	n/a
Hsiao 2008		84.00%	(73.72%,91.45%)	99.93%	(99.93%,99.94%)
Meta-analysis		87.90%	(80.40%, 92.80%)	99.99%	(99.99 – 99.99%)
Powell 2003	Blood Measurements of Bilirubin	100.00%	(15.81%,100.00%)	99.96%	(99.92%,99.98%)
Harpavat 2016		100.00%	(90.00%,100.00%)	98.18%	(97.88%,98.44%)
Harpavat 2020		100.00%	(59.04%,100.00%)	99.90%	(99.88%,99.92%)
Liao 2022		100.00%	(90.26%,100.00%)	77.26%	(75.95%,78.54%)
Meta-analysis		100.00%	(00.00%. 100.00%)	99.3%	(91.90% - 99.99%)
Akiyama 1994	Stool Colour Saturation	100.00%	(63.06%,100.00%)	95.19%	(91.34%,97.67%)
Shen 2016		100.00%	(90.75%,100.00%)	85.00%	(70.16%,94.29%)
Meta-analysis		100.00%	(0.00%, 100.00%)	92.4%	(83.4% - 96.7%)
Suzuki 2011	Urinary Sulphated Bile Acids	100.00%	(2.50%,100.00%)	99.48%	(98.86%,99.81%)
Matusi 1993	Bile Blood Spot Measurements	63.64%	(30.79%,89.07%)	98.92%	(98.85%,98.98%)
Zhou 2012		79.10%[1]	(74.30%,83.20%)	62.50%[1]	(25.00%,87.50%)
Gong 2020 (I)		100.00%	(96.27%,100.00%)	51.53%	(50.55%,52.51%)
Gong 2020 (II)		75.26%	(65.46%,83.46%)	99.00%	(98.79%,99.19%)
Mushtaq 1999 (I)		85.30%[1]	(75.50%,92.00%)	94.00%[1]	(92.30%,95.30%)
Mushtaq 1999 (II)		78.70%[1]	(68.10%,86.90%)	96.30%[1]	(94.90%,97.40%)
Meta-analysis		93.20%	(34.80%, 99.70%)	95.50%	(65.80% - 99.50%)
Gong 2020 (III)	Carnitine Measurements	74.36%	(57.87%,86.96%)	94.00%	(93.52%,94.46%)

[1] Sensitivity / specificity and their confidence intervals are derived from ROC curve.

Gong 2020 (I) results with cut-off bilirubin > 30 µmol/l.

Gong 2020 (II) results with cut-off bilirubin > 140 µmol/l.

Gong 2020 (III) results with cut-off free carnitine > 45 µmol/l.

Mushtaq 1999 (I) results with cut-off bilirubin > 25 µmol/l.

Mushtaq 1999 (II) results with cut-off bilirubin > 30 µmol/l.

Supplementary Appendix Table 9: Breakdown of factors included within cost-effectiveness analyses of BA screening methods

Paper	Type of BA Screening	Model Used	Cost of screening setup	Considered LTx costs	Considered Immunosuppression Cost	Considered Liver Transplant Follow-up	Considered Liver Transplant Complications
Woolfson et al.	SCC	Simple cost of set-up first and second year	Y	N	N	N	Y
Schreiber et al.	SCC	Markov Model	Y	N	N	N	Y
Masucci et al.	SCC and Conjugated Bilirubin Measurements	Markov Model	Y	Y	Y	Y	Y
Mogul et al.	SCC	Markov Model	Y	Y	Y	Y	Y

SCC: Stool Colour Chart, LTx: Liver Transplantation, Y: Yes, N: No