Are non-invasive or minimally invasive autopsy techniques for detecting cause of death in prenates, neonates and infants accurate? A systematic review of diagnostic test accuracy

Hannah O’Keefe 1,2, Rebekka Shenfine, 2 Melissa Brown, 2 Fiona Beyer, 1,2 Judith Rankin 2

ABSTRACT

Objectives  To assess the diagnostic accuracy of non-invasive or minimally invasive autopsy techniques in deaths under 1 year of age.

Design  This is a systematic review of diagnostic test accuracy. The protocol is registered on PROSPERO.

Participants  Deaths from conception to one adjusted year of age.

Search methods  MEDLINE (Ovid), EMBASE (Ovid), CINAHL (EBSCO), the Cochrane Library, Scopus and grey literature sources were searched from inception to November 2021.

Diagnostic tests  Non-invasive or minimally invasive diagnostic tests as an alternative to traditional autopsy.

Data collection and analysis  Studies were included if participants were under one adjusted year of age, with index tests conducted prior to the reference standard. Data were extracted from eligible studies using piloted forms. Risk of bias was assessed using Quality Assessment of Diagnostic Accuracy Studies-2. A narrative synthesis was conducted following the Synthesis without Meta-Analysis guidelines. Vote counting was used to assess the direction of effect.

Main outcome measures  Direction of effect was expressed as percentage of patients per study.

Findings  We included 54 direct evidence studies (68 articles/trials), encompassing 3268 cases and eight index tests. The direction of effect was positive for postmortem ultrasound and antenatal echography, although with varying levels of success. Conversely, the direction of effect was against virtual autopsy. For the remaining tests, the direction of effect was inconclusive.

A further 134 indirect evidence studies (135 articles/trials) were included, encompassing 6242 perinatal cases. The addition of these results had minimal impact on the direct results.

Conclusions  Current evidence is insufficient to make firm conclusions about the generalised use of non-invasive or minimally invasive autopsy techniques in relation to all perinatal population groups.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ This systematic review followed PRISMA-S, PRISMA-DTA and Synthesis without Meta-Analysis reporting standards.
⇒ The search was conducted by a trained information specialist and peer-reviewed by a Cochrane information specialist.
⇒ Publication bias was reduced by considering all study designs for inclusion.
⇒ The direction of effect may be overestimated due to the inclusion of low powered studies.
⇒ A narrative synthesis was conducted for each of the index tests as it was not possible to derive estimates of effect size.

CRD42021223254.

BACKGROUND

The devastating loss of a child has emotional, psychological and physical implications for parents. While healthcare practices have come a long way in reducing death under 1 year of age, approximately 5% of this population still die annually.1 Understanding why this has happened can provide some support to parents through the grieving process and contribute to identifying potential risks for future pregnancies.2 An autopsy can provide additional information or a change of diagnosis regarding cause of death in up to 76% of cases.3

Over the last 20 years, there has been a decline in the uptake of autopsy for the perinatal population.4 5 Studies have shown a decrease in uptake of 15%–20% in England and Wales between the 1990s and 2000s, with <60% of cases gaining consent for autopsy.6 7 The Alder Hey hospital retention inquiry, 1999–2001, saw children’s
organs and tissues being held after post-mortem without consent.7 8 Organ retention was one of the biggest concerns for parents when consenting to autopsy, following disclosures of unlawful organ retention (after the Alder Hey episode) and was still listed among the seven major themes describing barriers to postmortem uptake in two recent papers.7 8 9 Disfigurement of the child is of similar concern to the parents with many fearing invasive techniques will cause further trauma and feeling a need to protect the child.10 11 This is paralleled with religious beliefs where any disfigurement or delay to burial is unacceptable, for example, in Muslim or Jewish faiths, which generally require burial within 48 hours and any desecration of the body is forbidden.12 Implications for research have also been highlighted in response to decreasing uptake rates. Collating statistics for research can help to identify knowledge gaps and encourage funding bodies to drive the development of new interventions to prevent such deaths from occurring. Offering non-invasive or minimally invasive alternatives to autopsy is well supported by parents, religious groups and clinical pathologists, although with some caveats around the impact on the pathology specialism. This has been widely reflected in the qualitative evidence along with confirmation that non-invasive or minimally invasive techniques will likely increase uptake rates for the perinatal population.2 13–15

Cause of death can be distinguished as underlying, intermediate and immediate. Appropriate non-invasive or minimally invasive techniques may vary depending on which of these categories the postmortem is trying to identify.16 According to the Royal College of Pathologists (RCP) guidelines traditional autopsy is used to determine immediate cause of death.17–19 However, underlying cause of death is often used in mortality statistics. Guidelines from the RCP state that non-invasive or minimally invasive techniques should not be used alone, instead they should be used as adjuncts to traditional autopsy.17–19 In some circumstances, the family is charged for these alternative services, with costs starting at ~£500.20–22 In 2012, the UK National Health Service (NHS) recommended that 30 specialist mortuary-based imaging centres should be adopted across the country, five of which should be specialist paediatric centres, combining pathology and radiology services. They also recommended that a flat rate for autopsy services should be implemented, regardless of the techniques used, although recognising that this would likely inflate the NHS costs burdened by local authorities, but prevent costs being passed to the family.23 This vision is slowly coming to fruition with the Great Ormond Street centre and planned openings at facilities across the country, including adult imaging centres in Northumbria and Preston.20 24 However, there is little systematically compiled quantitative evidence to support these NHS recommendations for paediatric centres. To date, seven systematic reviews have attempted to address the evidence.25–31 However, it is clear that there are major limitations to the search methodologies in these reviews, potentially introducing bias, which means that the results cannot be interpreted with confidence. With the exception of the review by Wojcieszek et al, these reviews all focus on imaging techniques and do not consider other viable non-invasive or minimally invasive techniques.26–31 Similarly, the population groups are highly variable. Some consider a whole of life population, while others consider few very distinct perinatal populations. This makes it difficult to assess how these techniques perform across the subpopulations of the perinatal period (see online supplemental file 1 for a critique of the literature). This systematic review focuses on deaths from conception up to one year of age. For consistency, we adhere to UK terminology for five subpopulations and consider index tests which are underpinned by imaging, visual, verbal or laboratory techniques. While these tests are not commonly used in mortuaries at present, the majority are used routinely in clinical practice and are already well defined (see online supplemental file 2 for details of the clinical role of the index tests). Given the existing clinical application of the index tests, there is a broad spanning literature base which could be classified into two distinct groups. Studies where the primary aim was to assess non-invasive and minimally invasive techniques as an alternative to traditional autopsy can be classified as direct evidence. Conversely, studies where the primary aim was assessing these techniques for routine diagnostic use but follow-up in an autopsy population can be classified as indirect evidence.

**Aim**

To establish the direction of effect for non-invasive and minimally invasive autopsy techniques when discerning the cause of death in preivable, loss in utero, stillborn, neonatal and infant populations up to one adjusted year of age.

**Objectives**

The specific objectives were to:

- Synthesise the existing literature to assess which non-invasive and minimally invasive autopsy techniques have been studied as alternatives to traditional autopsy.
- Undertake a comparative analysis of the level of agreement and disagreement between the index test(s) and the reference standard.
- Assess the direction of effect for the index test(s) and reference standard using a vote counting approach.

**METHODS**

The methodologies used in this systematic review follow the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) reporting standards for search strategies PRISMA-S, Cochrane standards of conduct for Diagnostic Test Accuracy reviews, and the PRISMA diagnostic test accuracy reporting standards PRISMA-DTA.32–34 The protocol was published in PROSPERO in January 2021 (online supplemental file 3) and a full
protocol is available through the project website https://research.ncl.ac.uk/perinatal-post-mortem-dta/. 35

Patient and public involvement
Patients or publics were not involved in the design, or conduct, or reporting or dissemination plans of our research.

Eligibility criteria
Population
This systematic review includes loss of life from conception to one adjusted year. Population cohorts have been split according to UK terminology as follows. Preivable deaths include spontaneous and elective abortions at <24 weeks of gestation. Stillbirths include deaths at ≥24 weeks gestation. Loss in utero has been used for deaths where it was not possible to determine the gestational age or where gestational age has been given as a range that straddles the 24-week cut-off for preivable deaths and stillbirths. Neonatal death includes deaths within 28 unadjusted days from birth. Infant deaths include deaths up to one adjusted year of age. All causes of death were included, with the only exception relating to elective abortions with no indication of fetal compromise as no diagnostic benefit will be gained from autopsy in a large proportion of cases.

Index tests
The focus of this systematic review is non-invasive and minimally invasive techniques. Imaging techniques, amniocentesis, Chorionic Villus Sampling (CVS), fetal-free DNA analysis, percutaneous and endoscopic biopsy, umbilical cord and placental examinations, virtual and verbal autopsy have been included. In some instances, these techniques were carried out during the antenatal period, either prior to or after fetal demise. This systematic review only included studies where the index tests were performed antenatally or in the postmortem period. Therefore, studies where the index test was performed on living perinates in the post-partum period prior to death were excluded.

Reference standard
Full or partial traditional autopsies are the gold standard in postmortem diagnosis. Full autopsy includes examination of the body and internal organs, often with organ dissection and biopsy. Partial autopsy is performed on a specific region or regions of the body and/or specific organs or organ systems. Therefore, both techniques have been considered as reference standard within this review.

Study design
All study designs where individuals had received both the index test(s) and the reference standard were included. However, invasive procedures or removal of tissue samples during reference standard autopsy may affect the accuracy of any index test(s) conducted afterwards, potentially leading to false positive or negative results. Incisions or tissue sampling may mask details that would otherwise be obtained through imaging techniques had they been conducted initially. Therefore, studies where the index test(s) were performed secondary to the reference standard were excluded. Both prospective and retrospective studies were included to ensure coverage of retrospective analysis performed on index test(s) completed during the antenatal period. Studies reporting mixed populations were accepted if it was possible to isolate the populations of interest. Similarly, those with mixed comparators were included where the autopsy population could be distinguished from other non-autopsy-based comparators. Conference abstracts have been included where enough information was available to determine the methods used by the researchers and either the cause of death diagnosed by the index test(s) and reference standard or summary statistics with an indication of the number of index test(s) and reference standards in agreement. Qualitative evidence and systematic reviews were excluded from this review; however, the reference lists of systematic reviews were hand searched for eligible studies.

Search methods and information sources
A search strategy was designed by an information specialist (HO) in Medline(Ovid) and reviewed by a Cochrane Information Specialist using the Peer Review of Electronic Search Strategies (PRESS) checklist. 36 The search was translated into other databases as appropriate. The search strategy comprised of population, index test and reference standard concepts and a diagnostic study design filter was applied. No date or language restrictions were placed on the search and searches were conducted on 18 January 2021 and rerun on 17 November 2021 (online supplemental file 4).

Electronic searches
Bibliographic databases were searched from inception to 17 November 2021. The main databases were Medline(Ovid); Embase(Ovid); CINAHL(EBSCO); Cochrane library(Wiley) and Scopus. The reference lists of systematic reviews were hand checked for eligible studies. In addition to the main databases, Web of Science—Conference Proceedings Citation Index; ClinicalTrials.gov and WHO international clinical trials registry platform were searched for conference abstracts and trials. The Royal College of Obstetrics and Gynaecologists; Royal College of Midwives and Royal College of Paediatrics and Child Health were searched for grey literature.

Study selection
Duplicate references were removed using EndNote V.x9.01. 37 Title and abstracts and full-text screening were conducted by a single reviewer (HO) and 20% was screened independently at each stage by a second reviewer (RS or MB), after which clarification of eligibility took place before completion of screening. Where full texts were not freely available, they were requested through interlibrary loans. Articles written in non-English
language were translated to English prior to full-text screening. Discrepancies were resolved by discussion between the review team (n=37 title and abstracts; n=24 full texts).

**Data collection process**

In a change to protocol, studies were collated into groups of either direct or indirect evidence for data extraction. Pneumatically forms were then used to extract data from eligible studies. Authors were contacted where information was missing or unclear. Accuracy measures were recorded on a per-patient basis where possible, or a per-patient basis otherwise. Studies where the autopsy was clinically non-diagnostic in some cases have been included. However, the non-diagnostic cases have not contributed to direction of effect in the analysis.

**Risk of bias and applicability**

Risk of bias was assessed using an adaptation of the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2) tool, in line with author recommendations (see online supplemental table 1 for adaptations). The QUADAS-2 tool comprises four domains covering patient selection, index test, reference standard and flow and timing. Signalling questions are used in each domain to assess the extent to which the study has minimised bias in the way the patients were enrolled and the setting in which the study took place, the conduct and interpretation of results for both the index and reference standard tests, the timing between test and whether all patients received both tests.

**Synthesis of results**

A narrative synthesis was conducted following the PRISMA-DTA (online supplemental table 2) and Synthesis without Meta-Analysis (SWiM) reporting guidelines. Where possible, results have been converted from true positive (TP), true negative (TN), false positive (FP), false negative (FN) measures to numbers and percentages for consistency of reporting. Specifically, TP+TN being combined to demonstrate full agreement between tests, and FP+FN combined to show partial or partial disagreement between tests. Those with >50% full agreement between the index test and reference standard within a cohort are deemed to have a positive direction of effect towards the index test as an alternative to traditional autopsy. To this end, a vote counting approach was taken for this synthesis. Vote counting is the process of counting the number of studies where the results from the index and reference standard tests matched, and those where they disagreed. We have used the terminology of ‘favour’, ‘favourable’, ‘favouring’ and ‘favoured’ throughout the manuscript to express the direction of effect, that is to say when the index test results matched the reference standard results (denoted as favoured index test) and when they did not (denoted as favoured reference standard). This does not indicate usefulness or accuracy of the tests, instead it refers to the observed variability between tests as a means of demonstrating non-inferiority. Results with >80% agreement within a cohort were considered highly favourable, those with 51%–80% agreement were considered moderately favourable.

**RESULTS**

**Study selection**

Title and abstract screening was performed on 2117 references, 1633 were excluded as they did not meet our inclusion criteria. At full-text stage, 484 references were screened for eligibility. Of these, 281 were excluded as they did not meet the inclusion criteria (figure 1, online supplemental table 3). Articles and trials were grouped into direct (n=68) and indirect (n=135) evidence. Direct evidence was article, which focused on the index test(s) specifically for autopsy purposes. Indirect evidence was articles where the intention of the index test(s) was not for autopsy but did follow-up in an autopsy population, providing valuable insight into the suitability of these tests for autopsy purposes.

**Study groupings**

Three studies had multiple published articles which discussed the same population. The risk of double counting was minimised by carefully selecting the most comprehensive articles where different index tests had been described in the same study (online supplemental table 4).

This resulted in 50 of the 62 direct evidence papers or abstracts being synthesised, along with six trial records. However this only amounted to 54 studies in total as some articles described different tests from the same study. Of the 134 indirect evidence papers or abstracts, 133 were synthesised as well as a single trial record. None of the seven included trials identified from trial registries reported interim or final results and only four remain active. A trial in France (Unnamed Investigators) has been terminated due to lack of recruitment and the trial by Blaser in Canada has been withdrawn. The final trial by Fuchs, conducted in France, was completed in August 2020.

**Risk of bias and applicability**

Risk of bias assessment was performed for all papers or abstracts (n=196), regardless of study groupings, but not for trial registry records (n=7). We performed risk of bias assessment on all studies for transparency, particularly when considering reporting and within study biases across the grouped articles.

**Direct evidence**

Out of 62 articles, eight (13%) articles had a low risk of bias, 45 (72%) articles had some concerns, and nine (15%) articles had a high risk of bias (figure 2).

The overall ratings of the applicability of the articles to the review question found one (2%) article...
had a high level of concern, 241 20 (31%) had some concerns, 105 200 205 206 212 214 215 217 220 222 225 226 231–234 237–240 and 41 (67%) had a low level of concern. 182–194 196–199 201–204 207–211 213 216 218 219 221 223 224 227–229 235 236 242 243

Indirect evidence
For both risk of bias and applicability to the review question, the majority (n=129) had some concerns, 41 42 44–67 69–87 89–149 151–170 172–174 and five had a high level of concern 43 68 88 150 171 (figure 3).

Study characteristics
Postmortem MRI was the most widely studied technique among the direct evidence, with 24 studies and three trials. Conversely, antenatal CT, amniocentesis and placental examination were not present in any of the direct evidence studies. Similarly, little direct evidence (n=1) was seen for both virtual autopsy and verbal autopsy. The indirect evidence for antenatal ultrasound was considerable (n=96). Antenatal MRI and antenatal echography were also widely demonstrated through indirect studies, with 31 and 22 studies, respectively. All remaining techniques in the indirect evidence had relatively low total study numbers (n<10).

Among the indirect evidence, antenatal ultrasound had the highest total number of cases (n=5746), as expected from the volume of studies. For the direct evidence, antenatal ultrasound and postmortem MRI had the highest total number of cases, 1506 and 1669, respectively, despite differences in the number of studies for each technique (figure 4).

Two studies did not report the country of origin. 173 241 The remaining studies were split across 40 countries, with the UK having the highest output. A total of 26 direct and 12 indirect studies (15.84%) came from the UK. The USA was the second largest producer, with 21 direct and two indirect studies. Conversely, only one direct study was identified from China; 231 however, this country had the largest output of indirect studies (n=20), making it the third largest producer of outputs (figure 5).

Synthesis of results
None of the seven trials included in this systematic review reported interim or final results. 175–181 The results from
Figure 2  Risk of bias assessment for the 62 direct evidence papers and abstracts. Risk of bias assessment performed with QUADAS-2 evaluation tool. Inclusive rating for each domain and overall score. Priority order of scoring for each domain was as follows: high>some concern>low. QUADAS-2, Quality Assessment of Diagnostic Accuracy Studies-2.
each of the remaining studies have been described below in terms of favourability of each test in different perinatal populations. It is important to note, this demonstrates the direction of effect but not the degree of that effect (figure 4).

**Antenatal ultrasound**

**Direct evidence**

Among the direct evidence, diagnostic accuracy results for antenatal ultrasound in a previable population appeared
to be inconclusive with only a single study demonstrating a 50:50 split between agreement and disagreement.208

In a loss in utero population, results are conflicting with four studies in favour of ultrasound, one borderline favourable and two suggesting a large proportion of partial agreements making ultrasound unfavourable.215 220 228 230 233 243

Another found favour for ultrasound in cases of CNS anomalies but not for somatic tissue anomalies.229

The final study considered loss in utero, stillbirths and neonates. This study had partial agreements, resulting in autopsy being the favourable technique205 (online supplemental figure 1A).

For antenatal ultrasound, the majority of studies had some concerns around the risk of bias,205 208 215 220 228–230 233 while one was at high risk (online supplemental figure 1A).243 Struksnaes et al contributed highly to the favourability of antenatal ultrasound in a loss in utero population but had a high risk of bias due to conduct of the reference standard in domain 3 of the QUADAS-2 assessment. The autopsy was not performed by a prenatal and paediatric specialist and the results of the index were not blinded.243

Indirect evidence

The indirect evidence indicated that in a preivable population, 15 studies favoured autopsy,86 91 92 101 102 112 119 129 131 133 152 156 160 172 whereas 13 favoured antenatal ultrasound.41 45 53 55 57 66 73 88 94 98 104 107 166

In contrast, antenatal ultrasound was favourable in the majority of studies (n=35) for a loss in utero population.42 43 46 56 60 68 78–80 83 84 89 90 93 95 97 99 100 109 114 118 126 125 127 130 135 138 142 150 155 157 161 169 170 173 Only 12 studies favoured autopsy,70–72 96 115 125 126 134 149 154 163 164 and 5 studies presented a 50:50 split between agreement and disagreement.44 132 144 162 171

In a stillbirth population, the number of studies favouring the index test was higher than those favouring the reference standard. Eighteen studies favoured antenatal ultrasound,4 12 44 52–54 57 75 77 78 85 89 99 108 124 128 133 137 166 and 11 favoured autopsy.

Similarly, studies assessing a neonatal population tended to favour the index test.51 32 68 73 78 89 90 134 Seven studies favoured the reference standard Gupta et al.66 86 91 119 130 136 137 However, two studies were inconclusive.60 65
The final piece of indirect evidence considered a single infant and found in favour of autopsy\textsuperscript{83} (online supplemental figure 1B).

Five of the 96 articles assessing antenatal ultrasound had high risk of bias,\textsuperscript{13} 68 88 150 171 the remainder had some concerns (online supplemental figure 1B). Akgun\textit{et al}, Tudorache\textit{et al} and Zsansett\textit{et al} considered a loss in utero population, Faugstad\textit{et al} considered loss in utero and neonatal populations, and Indrani\textit{et al} considered previable population. With the exception of the study by Zsansett\textit{et al} which was inconclusive, these studies favoured antenatal ultrasound.\textsuperscript{43} 68 88 150 171 However, all had a high risk of bias due to domain 3, where the results of the reference standard had not been interpreted without knowledge of the index test results.

**Antenatal echography**

**Direct evidence**

Only two direct evidence studies were available for antenatal echography. Both showed similar results in favour of the index test.\textsuperscript{231} 232 Both studies showed some concerns around the risk of bias (online supplemental figure 1E).

**Indirect evidence**

Only one study was in favour of autopsy in a previable population.\textsuperscript{65} The remaining studies favoured antenatal echography.\textsuperscript{117} 122 141

All 11 studies including a loss in utero population were favourable of the index test.\textsuperscript{48} 49 50 62 70 87 102 103 106 121 153

Similarly, in a stillbirth population, the majority of studies favoured the index test,\textsuperscript{74} 101 139 151 163 164 with only two favouring autopsy.\textsuperscript{140} 174

Among the neonatal population, two studies favoured antenatal echography,\textsuperscript{121} 131 and one study was indiscriminate between the index test and reference standard.\textsuperscript{65} All studies showed some concerns around the risk of bias (online supplemental figure 1D).

**Antenatal CT**

**Direct evidence**

Two studies considered a loss in utero population,\textsuperscript{81} 154 one of these also considered neonates.\textsuperscript{154} The final study considered a stillbirth population.\textsuperscript{61} In all cases, antenatal CT was favoured over autopsy. These studies demonstrated some concerns around the risk of bias (online supplemental figure 1E).

**Indirect evidence**

In a loss in utero population, one study was in favour of antenatal MRI, whereas the other was inconclusive, with partial agreement or disagreement in 50% of the population\textsuperscript{182} 237 (online supplemental figure 1F).

The two studies for antenatal MRI were opposing, with one having high risk of bias,\textsuperscript{237} and the other low risk (online supplemental figure 1F).\textsuperscript{182} The study by Griffiths\textit{et al} had a high risk of bias as it did not employ consecutive or random sampling, as assessed in domain 1.\textsuperscript{237} This study was favourable for antenatal MRI and removal of this high risk study would alter the findings of this review.

**Indirect evidence**

The indirect evidence for antenatal MRI showed that six studies favoured the index test in a previable population.\textsuperscript{64} 67 69 113 160 166 However, in another six studies, autopsy was favourable,\textsuperscript{116} 129 135 143 146 152 and one study presented a 50:50 split between agreement and disagreement.\textsuperscript{65} All 11 studies considering a loss in utero population favoured the index test.\textsuperscript{44} 70 80 103 105 111 130 135 162 163 170

In a stillbirth population, there was a mixture of results, with seven studies favouring the index test,\textsuperscript{69} 75 77 133 150 160 166 three studies favouring the reference standard\textsuperscript{65} 85 167 and one having a split opinion.\textsuperscript{52}

Contrastingly, four studies assessing a neonatal population favoured autopsy,\textsuperscript{52} 65 145 147 with a fifth study being indiscriminate between the tests\textsuperscript{130} (online supplemental figure 1G).

All indirect evidence studies considering antenatal MRI had some concerns around the risk of bias (online supplemental figure 1G).

**Virtual autopsy, X-ray babygram**

**Direct evidence**

This single study favoured autopsy in a mixed population of loss in utero, stillbirth and infants. This was largely due to a high proportion of partial agreement.\textsuperscript{198} This study was considered to have some concerns around the risk of bias.\textsuperscript{198}

**Biopsy**

**Direct evidence**

There were three studies and one trial covering populations of loss in utero, stillbirth, neonates and infants. For loss in utero, the findings favoured autopsy with 100% disagreement between index tests and reference standard.\textsuperscript{236} In stillbirths, biopsy is highly favoured, whereas, in neonates, it is moderately favoured.\textsuperscript{213} For infants, results suggest biopsy is favoured, but more so in a younger infant ≤5 months\textsuperscript{218} (online supplemental figure 1H). Two of these studies showed some concerns around risk of bias,\textsuperscript{213} 218 and the other had a high risk of bias as not all cases were included in the analysis (online supplemental figure 1H).

**Placental examination**

**Indirect evidence**

A single study favoured placental examination for identifying blood flow restrictions in unexplained stillbirths but had some concerns around the risk of bias.\textsuperscript{124}

**Amniocentesis**

**Indirect evidence**

Amniocentesis in a loss in utero population was highly favourable in three studies, all of which were for the identification of congenital infection.\textsuperscript{47} 100 138 However, concerns were shown around the risk of bias (online supplemental figure 1I).
Verbal autopsy

Direct evidence
For stillbirths, verbal autopsy was unfavourable. However, in neonates, it appeared that verbal autopsy was favourable for diagnosis of infections, but unfavourable for congenital malformations, intrapartum complications, preterm complications and other diseases. This study for verbal autopsy showed some concerns around the risk of bias.

Postmortem CT

Direct evidence
A single study investigated postmortem CT in a preivable population and was in favour of CT. Two studies looked at a loss in utero population, in both cases, CT was generally favourable. However, one of these studies split the results by positive, negative and neutral markers and found that autopsy was favourable for neutral markers. One study assessed neonates only, showing favour for CT. One study included neonates and infants as a mixed population, again this was in favour of CT. One study had a mixed population of stillbirths and infants. This study favoured CT for identifying intracerebral and intraventricular haemorrhages, but found it unfavourable for germinal matrix haemorrhages. In contrast, the final study favoured autopsy in an infant-only population (online supplemental figure I).

Indirect evidence
Two of these studies demonstrated a high risk of bias and one was at low risk (online supplemental figure I). The studies by Guddat et al and Sandrini et al were both in favour of postmortem CT in a loss in utero and neonatal population, respectively. However, both had a high risk of bias for domain 3, which considered the performance of the reference standard. The pathologist was not a prenatal and paediatric specialist in the Guddat et al study and it was unclear whether the results of the reference standard were interpreted without knowledge of the index test results. In the Sandrini et al study, it was not clear whether the reference standard was conducted by a specialist and the results were interpreted without blinding of the index test. The study by Sandrini et al also had issues regarding domain 2, which focuses on the conduct of the index tests. The results of the index test were not interpreted without knowledge of the results of the reference standard, resulting in a high risk of bias in this domain.

Postmortem ultrasound

Direct evidence
A single study included previable and stillbirths as separate populations for some but not all outcomes. Nevertheless, in all cases, postmortem ultrasound was favourable. One study included stillbirths and infants and was highly favourable for identifying intracerebral and intraventricular haemorrhages but only moderately favourable for germinal matrix haemorrhages. Three studies included loss in utero only and favoured ultrasound. The final study included a mixed population of loss in utero and stillbirths, favouring ultrasound (online supplemental figure 1K).

Indirect evidence
Two studies demonstrated low risk of bias and three showed some concerns. The final study had a high risk of bias (online supplemental figure 1K). The study by Votino et al favoured postmortem ultrasound for loss in utero but did not include all cases in the analysis. This resulted in a high risk of bias in domain 4.

Post-mortem MRI

Direct evidence
Four studies included a previable population only. One has not reported the results sufficiently enough to extract the data. One was indiscriminate with a 50:50 split of agreement and disagreement. The final studies were in favour of postmortem MRI, however, one of these only favoured MRI in combination with external examination and blood tests, not MRI alone.

Indirect evidence
Two studies considered a previable and stillbirth population and favoured MRI. A single study included a mixed population of previable, loss in utero, stillbirth and neonates. This study found in favour of MRI when compared with full autopsy; however, favoured the reference standard when considering partial autopsy.

Post-mortem ultrasound

Direct evidence
A single study included previable and stillbirths as separate populations for some but not all outcomes. Nevertheless, in all cases, postmortem ultrasound was favourable. One study included stillbirths and infants and was highly favourable for identifying intracerebral and intraventricular haemorrhages but only moderately favourable for germinal matrix haemorrhages. Three studies included loss in utero only and favoured ultrasound. The final study included a mixed population of loss in utero and stillbirths, favouring ultrasound (online supplemental figure 1K).

Indirect evidence
Two studies demonstrated low risk of bias and three showed some concerns. The final study had a high risk of bias (online supplemental figure 1K). The study by Votino et al favoured postmortem ultrasound for loss in utero but did not include all cases in the analysis. This resulted in a high risk of bias in domain 4.
of the index test. Finally, two studies reported an infant only population. One favoured autopsy due to partial agreement in all cases. The second favoured autopsy over MRI alone, but favoured MRI in combination with external examination and blood tests (online supplemental figure 1L). The majority of studies for postmortem MRI showed some concern around the risk of bias. Two studies had a low risk, while three studies had high risk (online supplemental figure 1L). The study by Alderliesten et al was in favour of postmortem MRI for loss in utero but showed a high risk of bias in domain 4 as not all cases were included in the final analysis. Griffiths et al was also in favour of postmortem MRI for both loss in utero and stillbirths, however they did not use a consecutive or random sample, resulting in high risk of bias in domain 1. Finally, the study by Hart et al, which favoured the reference standard for infants had a high risk of bias in domain 3 as both signalling questions were refuted.

Indirect evidence

Three indirect studies considered postmortem MRI. In all cases, the study included a previable population. Contrasting results were seen, with two case studies favouring autopsy as a result of partial corroboration between tests. The final study of 28 individuals favoured postmortem MRI, claiming 100% diagnostic accuracy. All studies had some concerns around the risk of bias (online supplemental figure 1M).

DISCUSSION

Summary of evidence

This systematic review included direct evidence from 54 studies of diagnostic accuracy for non-invasive or minimally invasive autopsy techniques in the perinatal population. Postmortem MRI was the most widely investigated test with 24 studies and three trials including this technique. Other index tests were explored in fewer studies, ranging from one to nine direct evidence studies per technique. A substantial volume of indirect evidence has also been included in the review, with 134 studies. Antenatal ultrasound was covered by 96 of these indirect studies, making it the most widely investigated technique overall. Greater emphasis has been placed on the direct evidence as these studies were specifically designed to assess the index tests as an alternative to traditional autopsy. While direct studies offer the most robust evidence, it is important to recognise the value of indirect studies, particularly when direct evidence is sparse. The inferences drawn from scant direct evidence may be supported or superseded by indirect evidence. In cases where no direct evidence is available, indirect evidence can help determine whether clinical trials are warranted.

Direct evidence from six studies, supported by indirect evidence from one study, suggested that postmortem ultrasound was generally favoured across all subpopulations investigated for this technique, yet with varying degrees of success. Similarly, direct evidence of postmortem CT from eight studies, supported by a single indirect study, demonstrated favourability in preivable, stillbirth and infant populations, but traditional autopsy was preferable in loss in utero and neonatal populations. Conversely, the direct evidence of virtual autopsy (X-ray babygram) in a loss in utero, stillbirth and infant population showed favour for traditional autopsy (figure 4).

While direct evidence was available for antenatal echography (n=2), the majority was indirect evidence (n=22). This technique was favoured across all subpopulations investigated. Indirect evidence of antenatal CT, placental examination and amniocentesis were favourable for the use of these techniques, however, in limited populations. Studies have shown that time elapsed between events and the conduct of verbal autopsy has an effect on the prediction of the cause of death due to the reliability on memory of the interviewees. The same could be expected for placental or cord examination conducted at the time of birth and relayed at a later date, particularly when the birth is uneventful and there is no apparent issue with the child at the time of delivery. The single indirect evidence study for placental examination meant it was not possible to analyse differences in performance for preivable, stillbirth and loss in utero populations versus the populations of neonates and infants. While this is not a direct limitation of this systematic review, it does prompt caution around the interpretation of the results for these techniques.

It was found that all other evidence was inconclusive in the perinatal population as a whole, whereas the use of these techniques was favoured or refuted in individual subpopulations (figure 4). For example, direct evidence of biopsies was favourable in three of the four subpopulations studied; however, unfavourable in a loss in utero population making it inconclusive overall. The findings also varied depending on the reference standard used. Cohen et al studied postmortem MRI with loss in utero, preivable, stillbirth and neonatal cases; 10 (10%) of which underwent partial autopsy. There was a stark contrast in findings among the two groups: there was a tendency to favour postmortem MRI in those undergoing full autopsy; whereas, when a partial autopsy was conducted, the reference standard was the favourable test in all cases. Partial autopsy is usually conducted when maceration of particular organs or systems is severe or upon parental request. In the Cohen et al study, maceration scores did not affect which cases received full or partial autopsy, suggesting parental consent was the deciding factor. This raises questions around why partial autopsy does not show any agreement with postmortem MRI in these cases.

Strengths and Limitations

This systematic review has a number of strengths. The reporting of results followed the PRISMA-S and PRIMA-SA-DTA reporting standards (online supplemental table
about the reference standard. It would be expected that concerns, this mostly arose from domain 3 question 1 (direct evidence n=42; indirect evidence n=126) had some impact to this, there were a number of seemingly applicable studies which could not be included due to the stringent inclusion and exclusion criteria laid out in this systematic review.

The risk of bias assessment was performed with an adaptation of the QUADAS-2 tool, in accordance with the guidance (online supplemental table 1). The resulting changes to the signalling question altered the outcome of the assessment in some cases to unclear or high but made the risk of bias assessment much more applicable to the review question. For example, the majority of studies (direct evidence n=42; indirect evidence n=126) had some concerns, this mostly arose from domain 3 question 1 about the reference standard. It would be expected that a low risk of bias would be given for all studies under the original domain 3 question 1. Confluence between risk of bias assessments was 100% and minor comments were made between reviewers in <1% of the data extraction which did not result in changes to the data interpretation or findings of this systematic review.

A vote counting approach was taken for the analysis due to limitations with data reporting in the included studies. However, in cases where the studies have low power, particularly case studies, a vote counting methodology can give a false impression of the true direction of effect. In addition to this, vote counting will not take account of any results, which show some partial agreement between the reference standard and index test(s). Here, those with partial agreement have been counted as favourable towards the reference standard. This included cases where the reference standard may have revealed findings which were not identified by the index test or vice versa.

In a clinical setting index tests, such as ultrasound, CT or MRI may be beneficial for identifying certain types of congenital anomalies but not others. This has been shown to hold true in the context of autopsy. This means that comparing results between studies may introduce confounding as the populations and conditions of interest may vary considerably. This systematic review has limited this by narratively describing the results in terms of subgroups of the perinatal population; however, it was not feasible to group the outcomes by cause of death. This was partly due to inconsistent reporting of conditions or anatomy, where study authors have looked either more generally at cause of death or at specific scenarios. Furthermore, no subgroup analysis was performed by date of publication. This may have potential implications as technology has developed rapidly since the early 2000s. However, dates have been provided for all studies (online supplemental figure 1) and only 5 of the 62 direct evidence studies were published prior to 2000 (figure 2).

A simplified version of the data extraction process was conducted for indirect evidence studies. While this is a limitation in the review process, the simplified data extraction has contributed significantly to this systematic review and a comprehensive extraction process would likely not have impacted the reporting of results. It is plausible that extensive data extraction may have allowed subgroup analysis in this set of data, although that would not have been comparable to the direct evidence as it does not support subgroup analysis. The substantial volume of indirect evidence (n=134 studies) has included non-invasive or minimally invasive autopsy techniques that were not found in the body of direct evidence, adding to the strength of this systematic review as a whole.

Implications for clinical practice

The inconclusive direction of effect for antenatal ultrasound and antenatal MRI in this population bears significant weighting on current clinical practice where parents may decide to terminate a pregnancy as a result of these tests. Many of the studies included loss in utero due to termination of pregnancy as a result of the index tests. The techniques have mixed favourability in indirect studies and are inconclusive in all direct studies. These inconclusive results may be due to the specific anomalies which led to termination, as are often the focus of these articles. While this systematic review only considers the diagnostic accuracy of the tests relative to autopsy, and not to live birth, it is worth a renewed effort to evaluate the effectiveness of antenatal ultrasound and/or MRI under different conditions. As stated in the National Institute for Health and Care Excellence (NICE) guidelines for antenatal care, diagnostic results from antenatal ultrasound and MRI do have limitations and parents should be made aware of this.

The NHS guidance to open new imaging centres throughout the UK recommended that five centres be open specifically for perinatal and child autopsies. These recommendations were primarily based on the lack of uptake for traditional autopsy and scant
findings from prior studies that had not been systematically assessed. As part of the report, specialist centres were asked to complete a questionnaire. Only two centres responded and no comment was made on the results of these responses. In addition, the report openly states that the review conducted by Thayyil et al found insufficient evidence to support the use of postmortem MRI (online supplemental file 1). In this systematic review, the evidence for postmortem cross-sectional imaging (CT and MRI) is still inconclusive, 10 years after the findings by Thayyil et al, showing no indication for the use of these techniques as a replacement for traditional autopsy. The NHS report, originally conducted in 2012, should be reviewed in light of the systematic evidence produced since 2010 and an up-to-date benefit-to-cost ratio for these centres should be produced.

Implications for policy

By UK law, only neonates and infants with unexplained or suspicious deaths are required to have a postmortem. However, autopsy may be offered by NHS pathology services for deaths occurring at 12 or more weeks of pregnancy, or any neonatal and infant deaths. Current NICE and NHS policy states that guidelines from the RCP should be followed as a minimum standard. At present, these guidelines emphasise that non-invasive or minimally invasive techniques should only be used as an adjunct to traditional autopsy. However, partial autopsy may be offered when consent to full autopsy is denied.

All autopsy procedures should be conducted within 7 days, with urgent examinations being performed within 48 hours. It is also recommended that perinatal pathology services have a minimum of two whole-time equivalent prenatal and paediatric pathology specialists, led by a specialist consultant in perinatal pathology, each performing a minimum of 50 perinatal autopsies per year. This is a particular point of interest as the number of prenatal and paediatric pathologists is declining and there are fewer specialists available to fulfil the policy requirements for staffing in perinatal pathology services. This means that specialists are having to travel across the country to provide their services, increasing the waiting time for autopsy. There are very few trainees in this specialism, so it is expected that these delays will continue for the foreseeable future unless uptake rates increase and the specialism is revived. The consequences for parents are considerable, without specialists parents may not receive accurate information, which could be detrimental for risk management of future pregnancies. These alternative techniques could be performed by radiographers or general pathologists with results sent electronically to specialist pathologists where review is needed. This would reduce strain on current services and could potentially reduce NHS costs.

There are clear benefits of employing non-invasive and minimally invasive autopsy techniques, such as reduced requirements on specialists and specialist services, as well as quicker turnaround time and return to the family. However, the current evidence does not support their use as standalone procedures. Therefore, changes to these policies cannot be recommended.

Implications for research

It is notable that there is a substantial lack of standardised reporting among these quantitative studies, specifically anatomical descriptions and reporting of numerical results. To address this, agreement in reporting practices is needed among prenatal and paediatric researchers and clinicians. A strategy for this may be best deployed by one of the leading bodies, such as the Royal College of Paediatric and Child Health. While the vast majority of studies included in this review appear to be assessing the immediate cause of death, it is not definitively clear from the reporting. This has implications when considering which tests may be appropriate for identifying underlying, intermediate and immediate cause of death as the accuracy of the techniques may differ under each circumstance. This should be reported plainly throughout manuscripts, enabling comparison across the different categories of cause of death. Furthermore, very few studies reported sufficient sensitivity and specificity data, which should be a primary statistic in diagnostic accuracy reports. Without this, it is not possible to derive degree of effect measurements, summary statistics, confidence intervals or investigate heterogeneity. A key recommendation for future research is to include sensitivity and specificity analysis in all reports.

Further research is needed to understand what leads to early loss of life and to help produce interventions to prevent death or minimise risk. In light of a lack of supportive evidence for introducing these new techniques, increasing the uptake of traditional autopsy will provide the advantage of data collation for research purposes. There is a recognisable body of qualitative literature discussing the barriers to traditional autopsy; however, there are few studies considering enablers for consent. One recommendation for future research is to conduct qualitative studies with parents, guardians, religious representatives, healthcare professionals and supporting charities to better understand potential enablers in this area.

CONCLUSIONS

The current evidence is insufficient to support the routine use of non-invasive or minimally invasive techniques in autopsy practice for all perinatal subpopulations. Postmortem ultrasound and antenatal echography are the most promising with both direct and indirect evidence favouring the use of the index test. However, without sufficient sensitivity and specificity data, the degree of effect cannot be deduced. Moreover, the heterogeneity of the results cannot be formally assessed. This is the same for each of the index tests evaluated here and poses a substantial barrier to the clinical and parliamentary acceptance of these findings.


Open access

by guest. Protected by copyright.
or rejection of such techniques. The current guidelines from the RCP should remain in place and continue to be adhered to.\textsuperscript{17}\textsuperscript{19}

Acknowledgements With thanks to Sheila Wallace for careful peer review of the search strategy, to Dr Ryan Kenny, Dr Malcolm Moffat and Dr Svetlana Glinyanaya for helpful comments and revision of the manuscript.

Contributors HK: guarantor, conceptualisation, data curation, formal analysis, methodology, writing, reviewing and editing the manuscript. NS: data curation, reviewing and editing the manuscript. MB: data curation, reviewing and editing the manuscript. FB: conceptualisation, supervision, data curation, methodology, reviewing and editing the manuscript. JR: conceptualisation, supervision, data curation, methodology, reviewing and editing the manuscript. FB: conceptualisation, supervision, data curation, methodology, writing, reviewing and editing the manuscript. RS: data curation, methodology, writing, reviewing and editing the manuscript.


Disclaimer The views expressed are those of the author(s) and not necessarily of the NIHR or the Department of Health and Social Care.

Map disclaimer The inclusion of any map (including the depiction of any boundaries therein), or of any geographic or locational reference, does not imply the expression of any opinion whatsoever on the part of BMJ concerning the legal status of any country, territory, jurisdiction or area of or in its territories. Any such expression remains solely that of the relevant source and is not endorsed by BMJ. Maps are provided without any warranty of any kind, either express or implied.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data sharing is not applicable as no new datasets were generated and/or analysed for this study. No additional data available.

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 or omission arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iD

Hannah O’Keefe http://orcid.org/0000-0002-0107-711X

REFERENCES


Hong L, Xie MX. Multiple ultrasonic techniques combined in screening the fetal congenital heart diseases clinical value. *Eur Heart J Suppl* 2012;14:A35.


Ctri YK. Which heart diseases cause children to die? clinicopathological analysis of heart disease in pediatric age group, 2021


