Clinical Radiology 74 (2019) 733.e11–733.e18

Contents lists available at ScienceDirect

Clinical Radiology

journal homepage: www.clinicalradiologyonline.net

Automated data extraction and report analysis in computer-aided radiology audit: practice implications from post-mortem paediatric imaging

S.C. Shelmerdine [a,b,*], M. Singh [a], W. Norman [a], R. Jones [a], N.J. Sebire [b,c], O.J. Arthurs [a,b]

[a] Department of Radiology, Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK
[b] UCL Great Ormond Street Institute of Child Health, London, UK
[c] Department of Histopathology, Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK

ARTICLE INFORMATION

Article history:
Received 18 January 2019
Accepted 24 April 2019

AIM: To determine local departmental adherence to the paediatric post-mortem magnetic resonance imaging (MRI) protocols, using a customised automated computational approach.

MATERIALS AND METHODS: A retrospective review of 460 whole-body post-mortem MRI examinations performed at Great Ormond Street Hospital for Children over a 5.5-year period was assessed for adherence to a full or abbreviated imaging sequence protocol. A simple computer program was developed to batch process DICOM (digital imaging and communications in medicine) files, extracting imaging sequence details, followed by natural language processing (NLP) of authorised reports to automate information extraction of diagnostic image quality.

RESULTS: The program was able to extract study parameters from the entire dataset (approximately 80 GB of data) in a few hours, and retrieve information on diagnostic image quality using NLP with an overall diagnostic accuracy for data extraction of 96.7% (445/460, 95% confidence interval [CI]: 94.7–98%). The full imaging protocol was adhered to in 305/460 (66.3%) cases, and an abbreviated protocol in 140/460 (30.4%) cases. Overall, 423/460 (91.9%) of studies were of diagnostic quality. These included 298/305 (97.7%) of the full protocol, 111/140 (79.3%) of the abbreviated protocol. In only five cases were the examinations non-diagnostic for all body systems, all of whom weighed <100 g (24.7–72 g) and imaged using the abbreviated protocol.

CONCLUSION: The present study demonstrated a successful application of an automated approach for data collection for audit and quality assessment purposes using paediatric post-mortem imaging as a specific example. Re-audit of these data following change implementation will be straightforward now that the automated workflow is clearly established.

* Guarantor and correspondent: S. Shelmerdine, Department of Radiology, Great Ormond Street Hospital for Children NHS Foundation Trust, London WC1N 3JH, UK. Tel.:/fax: +44(0)20 7405 9200.
E-mail address: susie_c_s@yahoo.co.uk (S.C. Shelmerdine).

https://doi.org/10.1016/j.crad.2019.04.021
0009-9260/

Introduction

The persistent decline in consent rates for paediatric autopsy has facilitated development of non-invasive alternatives, based on imaging.[1–4] Post-mortem magnetic resonance imaging (PMMR) provides high diagnostic accuracy for perinatal and infant deaths (similar to conventional autopsy) with high concordance rates in detecting major pathological lesions.[5] PMMR performs better than post-mortem computed tomography (PMCT),[4] and is also acceptable to healthcare professionals and parents.[6,7] Consequently, use of paediatric PMMR has grown rapidly. Established working groups are embedded within several imaging societies,[8–10] and it is endorsed by the Royal College of Pathologists, with inclusion in paediatric autopsy guidelines.[11–13]

Despite these advancements, an agreed standardised national or international paediatric PMMR protocol has not been clearly defined according to age, gestation, or body weight resulting in the use of at least 15 different imaging protocols worldwide.[14] This inconsistency makes it difficult to guarantee uniformity of image quality and technique, and hinders comparison between different patient groups in multicentre studies. As one of the largest paediatric post-mortem imaging centres worldwide, the PMMR protocols used at Great Ormond Street Hospital for Children were published in 2015[14]; however, the full PMMR protocol, although designed to be comprehensive, can be time-consuming and for both clinical and timetabling reasons may be curtailed or abandoned when potentially non-diagnostic.

The purpose of this study was to assess adherence to the PMMR protocols, and understand the reasons for any variation. In order to do this efficiently, a custom computer program was designed to extract the relevant information from the DICOM (digital imaging in communications in medicine) metadata. Basic natural language processing (NLP) was also applied to analyse the study reports.[15] The aim of this computational approach was to increase the speed, accuracy, and consistency of data collection, to extract insights that may inform modifications to future protocols, and refine PMMR guidelines. Furthermore, the code used in the study is provided as an example of how automated data collection and NLP might be applied to in other imaging contexts.

Materials and methods

Study cohort

A retrospective review of the radiology information system (RIS) at Great Ormond Street Hospital for Children was conducted for all PMMR studies performed over a 5.5-year period (January 2013–July 2018). All studies were included for analysis without exclusion criteria. Written informed consent was obtained from all parents for clinical pre-autopsy PMMR, which included parental consent for use of data for audit, research, and education as part of the post-mortem imaging protocol. Ethical approval was not required for this study as it was part of a retrospective audit of imaging data, approved by the local research and development office.

Demographic data for each patient were also collected including the age at time of death, time between death and imaging (i.e. post-mortem interval), post-mortem weight (in grams), and gender. For perinatal deaths, additional information included the gestational age, maceration score at clinical autopsy (0–3; 0 representing none and 3 representing late/established maceration), and mode of death (e.g., termination of pregnancy, stillbirth, and miscarriage) from the clinical notes or autopsy report.

Imaging protocol: current practice

All PMMR imaging was performed using a 1.5 T MR machine (Avanto, Siemens Medical Solutions, Erlangen, Germany), by one of two experienced MR radiographers. The local PMMR protocols, which were taken as standard, were previously published and are included in Table 1.[14]

In brief, radiographers perform either a “full protocol” or “abbreviated protocol”. The full protocol involves three-dimensional isovolumetric T1-weighted, T2-weighted, and diffusion-weighted imaging (DWI) of the brain, spine, and torso. In addition, a susceptibility-weighted imaging (SWI) sequence of the brain and a three-dimensional high-resolution T2-weighted constructive interference steady-state (CISS) sequence covering the thorax was performed. Where a fetus is small and at the limits of image resolution, an abbreviated version of this protocol can be performed. This involves only two key sequences: three-dimensional isovolumetric T1-and T2-weighted sequences of the whole body in one acquisition (as opposed to imaging body parts separately). The cut-off for this size limitation is frequently a subjective measure, decided upon by the radiographer at time of performing the study.

The protocol does not specify the type of coil to be used, allowing operator choice. Ideally, this should be a phased-array coil with multiple elements within close proximity to the region of interest. Ordinarily, a head coil is used for neuroimaging and phase-array matrix body coil for body imaging, although these may be adjusted according to the size of the fetus or child (e.g., in smaller fetuses, the head coil alone may be sufficient to cover the head and body).

Referrals are generated for PMMR imaging via the lead pathologist responsible for the clinical case. At present, there are no restrictions for referral indication, although cases $<$200 g are not usually recommend for imaging (unless there is no other imaging alternative) given the increased likelihood of non-diagnostic imaging.[16]

Data collection and analysis

The local RIS was queried using a DICOM viewer (OsiriX, Pixmeo SARL, Switzerland). Examinations were reviewed for number and name of MR sequences, operator name, and type of coil utilised. This information was encoded in the metadata of the image files (i.e., DICOM headers) as specific

Table 1
Sequence parameters for full post-mortem MRI protocol in infant and perinatal deaths (adapted with permission from Norman et al,[14]) are given below.

Sequence	FOV (mm)	Section thickness (mm)	Matrix	Voxel size (mm)	TR (ms)	TE (ms)	Averages (NEX/NSA)	No. sections and gap	Approximate length of sequence (min)
Brain imaging									
3D FLASH T1W (sag)									
Perinatal	256	1	256/256	1×1×1	11	4.9	3	60 per slab	5.44
Child	256	1	224/256	1×1×1	11	4.9	1	160 per slab	4.20
2D DESTIR T2W (axial and coronal)									
Perinatal	100	2	172/256	0.4×0.4×2	5460	16 and 115	6	18 (1 mm)	13.46
Child	200	4	216/320	0.7×0.6×4	6180	14 and 115	1	22 (1 mm)	3.19
2D GRE T1 HEME (axial)									
Perinatal	100	4	120/256	0.5×0.4×4	800	26	4	18 (0 mm)	6.26
Child	200	5	144/256	1×0.8×5	800	26	2	18 (0 mm)	3.52
DWI (b-values 0, 500, 1000)									
Perinatal	230	5	128/128	1.8×1.8×5	2,700	96	3	19 (0 mm)	1.06
Child	230	5	128/128	1.8×1.8×5	2,700	96	3	19 (0 mm)	1.06
Spine imaging									
2D T2W TSE (sag)									
Perinatal	150	1.5	128/256	0.6×0.6×1.5	9.1	4.5	8	12 per slab	4.24
Child	300	3	272/320	1.1×0.9×3	3,050	109	3	11 per slab	5.43
3D FLASH T1W (sag)									
Perinatal	150	1.25	128/256	0.6×0.6×1.3	11	5.3	10	16 per slab	3.19
Child	350	1.40	144/256	1.4×1.4×1.4	11	4.9	6	32 per slab	5.06
Body imaging (neck to pelvis)									
3D T2W TSE (cor)[a]									
Perinatal	200	0.8	160/256	0.8×0.8×0.8	3,500	275	2	72 per slab	6.20
Child	360	1.4	226/256	1.4×1.4×1.4	3,500	173	1	96 per slab	3.42
3D T1W VIBE (cor)[a]									
Perinatal	200	0.8	160/256	0.8×0.8×0.8	5.9	2.4	8	72 per slab	5.52
Child	360	1.4	224/256	1.4×1.4×1.4	5.9	2.4	5	72 per slab	6.33
3D CISS T2W (axial) (thoracic coverage for cardiac assessment)									
Perinatal	150	0.6	192/256	0.6×0.6×0.6	5.6	2.5	10	Cover heart and lungs	29.26
Child	150	0.6	192/256	0.6×0.6×0.6	5.6	2.5	10		29.26
2D T2W tirm (axial) (TI = 150)									
Perinatal	180	5	160/256	0.7×0.7×5	5,080	109	5	Cover body and pelvis	6.58
Child	300	5	168/256	1.2×1.2×5	8,390	108	4		4.47
DWI	As for head with greater number of sections to cover chest, abdomen and pelvis	1.06							

FOV, field of view, TR, repetition time; TE, echo time; NEX, no. of excitations; NSA, number of signal averages; FLASH, fast low-angle shot; 3D, three-dimensional; 2D, two-dimensional; W, weighted; DESTIR, dual-echo short-tau inversion recovery; GRE, gradient echo; HEME, T2 weighted gradient recalled echo sequence; DWI, diffusion-weighted imaging; TSE, turbo spin-echo; CISS, constructive interference steady-state; tirm, turbo inversion magnitude sequence; TI, inversion time.

[a] The two sequences denote the imaging performed in our abbreviated PMMR protocol, with the only difference being that the coverage for both is from the head to pelvis (not neck to pelvis as stated below for full protocol).

data elements. A small computer program was designed for automated data extraction using the free, open-source "Pydicom" package[17] (https://pypi.org/project/pydicom/; see Electronic Supplementary Material, Appendix S1). Pydicom allows manipulation of DICOM data elements using the Python programming language (Python Software Foundation, https://www.python.org/). All examinations were batch processed using the program, and the resulting data were tabulated using the "pandas" data analysis library.[18]

We performed NLP on the examination reports to automate extraction of some measure of the diagnostic outcome, given that a comment regarding diagnostic image quality is required per body system using the standardised reporting template for PMMR studies. The Natural Language Toolkit (NLTK[19]) and "spaCy" were used, which are both free, open-source python packages, to create a rule-based binary classifier (i.e., diagnostic or non-diagnostic; see Electronic Supplementary Material, Appendix S2). Feature extraction involved identification of word boundaries ("tokenisation") and formation of a list of words used in each report. This list was subsequently "normalised" by converting all words to lower case. Finally, the resulting word list was searched for specific terms that suggested non-diagnostic examinations, using regular expression pattern matching. The terms used were "non-diagnostic", "uninterpretable", "quality", and "resolution".

All reports and image sequences were checked manually by one of the authors (S.C.S.) for having the same sequences as stated in the DICOM headers, and whether the reports were classified correctly as being either of diagnostic or non-diagnostic quality for each of five body systems

(neurological, thoracic, cardiac, abdominal, and musculoskeletal system). Where at least one body system was deemed to be non-diagnostic, then the study as a whole was labelled as "suboptimal" in quality. Fig 1 outlines the workflow for both extraction of imaging parameters and NLP of diagnostic image quality. Fig 2 demonstrates an example of what a radiologist would classify and report as "diagnostic quality" versus "non-diagnostic" quality for two different cases in different body areas.

Prior to data analysis, the predefined local adherence rate was set at 100% for performing all PMMR sequences as stated in local protocols. Demographic differences between cases who received the full or abbreviated protocol were compared. All data were exported to a spreadsheet (Excel, Microsoft, Redmond, Washington, USA) for collation and further analysis.

Results

Demographics

Over the 5.5-year study period, 460 PMMR examinations performed from 460 individual cases were reviewed. Of these, 402 (87.4%) were perinatal deaths (fetal and early neonatal deaths up to 7 days old), 35 (7.6%) were neonatal and infant deaths (7 days–1 year old), and the remaining 23 (5%) were aged >1 year.

There were 270 males (58.7%), median age at death was 0 days (mean: 110 days, range: 0 days–15 years), imaged at a median post-mortem interval of 8 days (mean: 9 days, range: 0–35 days) and overall median post-mortem weight of 680 g (mean: 2.8 kg, range: 13 g to 87 kg). For perinatal deaths, the median gestational age was 24 weeks (mean: 27 weeks, range: 13–42 weeks) with median maceration score of 1 (mean: 1, range: 0–3).

Data extraction

The program was able to extract study parameters from the entire dataset (approximately 80 GB of data) in <3 hours. Study reports were extracted and analysed separately before being collated.

Protocol adherence

The full PMMR protocol was adhered to in 305/460 (66.3%) cases, and the abbreviated PMMR protocol in 140/460 (30.4%) cases. The median post-mortem weight of the cases that underwent a full protocol was 2,051 g (average 3,314 g; 165–87,000 g), and for those having the abbreviated protocol the median weight was 225 g (average 264 g; 12.6–1,050 g).

Fifteen cases (15/460, 3.3%) did not have the standard abbreviated or full protocol for PMMR examination. Of these 7/15 (46.7%) cases had an incomplete full protocol (i.e., some but not all of the sequences were performed, commonly the diffusion weighted sequences). There were no clinical or radiological reporting system notes to state why this was the case or why the study was abandoned before all sequences were performed. In the other 8/15 (53.3%) cases, a customised protocol was conducted either due to the parental wishes or pathologist request. The imaging was mainly targeted to answer a specific clinical question pertaining to one or more body parts. Of these, three cases included imaging of only the head, one case of only the neck, two cases of only the thorax, and two cases where there was imaging of the thorax and abdomen, but

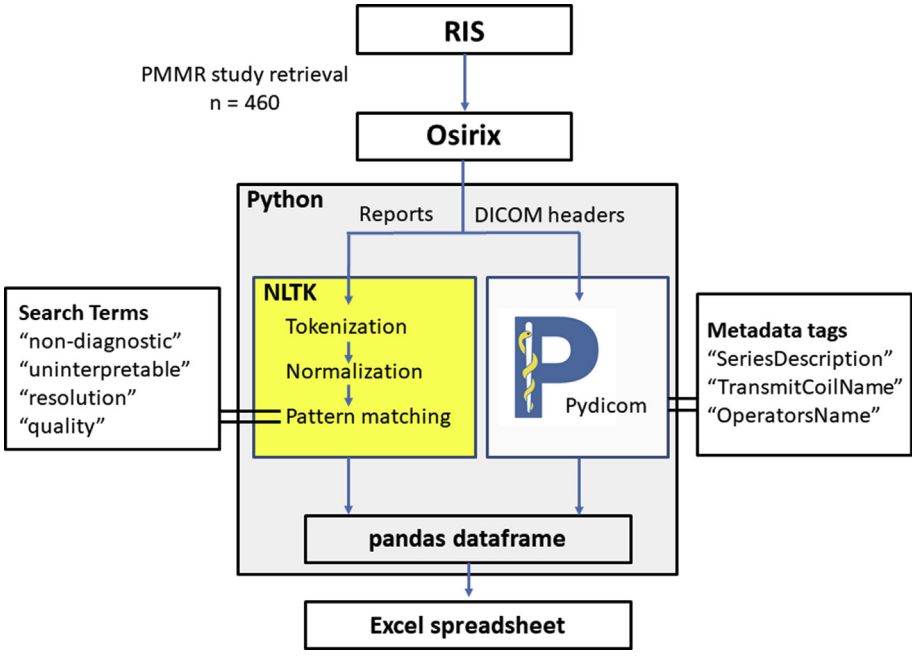

Figure 1 Workflow diagram for automated data collection utilised in the present study. RIS, radiology information system; NLTK, natural language toolkit; DICOM, digital imaging & communication in medicine.

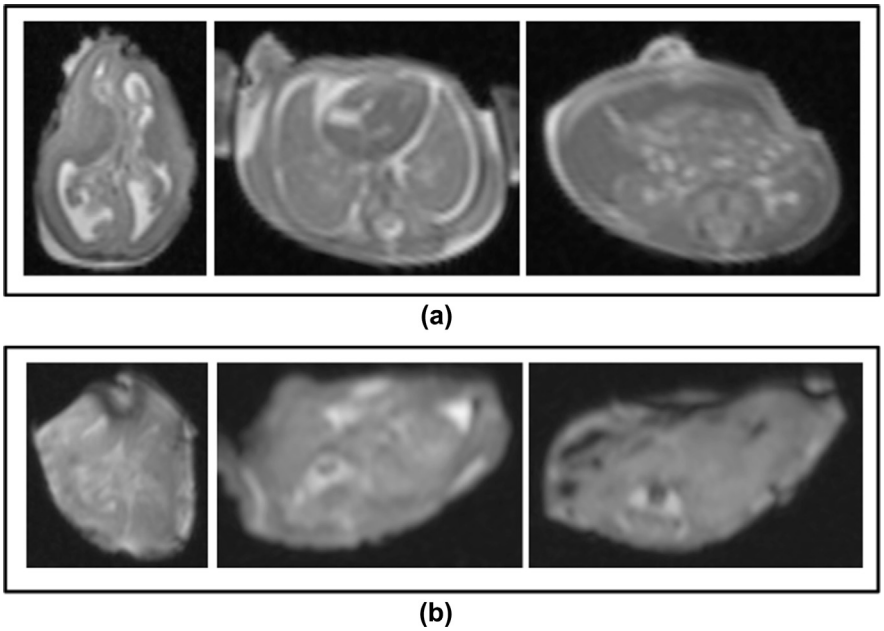

Figure 2 Diagnostic and non-diagnostic quality PMMR imaging in two different fetuses of 15 weeks gestational age, obtained 4 days after death. (a) The top row shows diagnostic quality axial T2-weighted images of the brain (top left), thorax (top middle) and abdomen, at the level of the renal hila (top right). (b) The bottom row demonstrates a "non-diagnostic" quality study for the same corresponding body parts, respectively.

not the head (in one case the child already had a recent ante-mortem MRI study of their brain, in the other case the child had a normal post-mortem CT of their head, and the referring clinical team did not deem further MRI necessary).

Diagnostic imaging quality

Overall, 423/460 (91.9%) of all studies were of diagnostic quality for all body systems imaged. Of the full protocol, 298/305 (97.7%) were diagnostic (i.e., suboptimal diagnostic rate of 2.3%) and 111/140 (79.3%) of the abbreviated protocol, which were diagnostic (i.e. suboptimal diagnostic rate of 20.7%). In only five cases were the PMMR examinations entirely non-diagnostic for all body parts examined. In all cases these were fetuses weighed <100 g (24.7–72 g) and had undergone an abbreviated protocol.

Of the seven suboptimal studies adhering to the full protocol, only one body part was deemed to be of non-diagnostic quality. Of the 29 suboptimal PMMR studies in the abbreviated protocol cohort, 5/29 were non-diagnostic for all body parts imaged. Of the remaining 24 cases, 14 were non-diagnostic for one body system, six for two body systems, one for three body systems, and two for four body systems. The breakdown of which body systems were non-diagnostic is shown in Table 2.

There were 61/460 (13.2%) PMMR examinations performed in cases weighing <200 g (four full, 56 abbreviated, and one incomplete full protocol). Of these cases, 37/61 (60.7%) were deemed as diagnostic in all body systems. These included all cases where a full protocol and the single case where the incomplete full protocol was adhered to.

There were no cases of <150 g body weight that underwent the full protocol. The full protocol was adhered to in 89.2% (248/278) cases weighing ≥450 g, with 98.8% (245/248) diagnostic image quality for all body systems. Between 150–449 g, the full protocol was adhered to in 28.4% (56/197), with 94.6% (53/56) diagnostic image quality for all body systems. Fig 3 is a graph depicting the results of the study for cases weighing ≤1,000 g in body weight.

Classification model performance

The customised NLP model had the following performance metrics compared with manual review of reports and images (labelled as "diagnostic" and "non-diagnostic/suboptimal" quality): sensitivity 99.3% (419/422, 95%

Table 2
Suboptimal post-mortem magnetic resonance (PMMR) imaging studies, divided by protocol adherence, showing which body system was deemed as non-diagnostic in each subgroup.

PMMR protocol	post-mortem magnetic resonance imaging Total no. suboptimal studies	Non-diagnostic body systems					Total non-diagnostic body systems
		Brain	Cardiac	Thoracic	Abdomen	Musculoskeletal	
Full	7	4	3	0	0	0	7
Abbreviated	29	17	24	10	9	7	67
Total studies	36	21	27	10	9	7	74

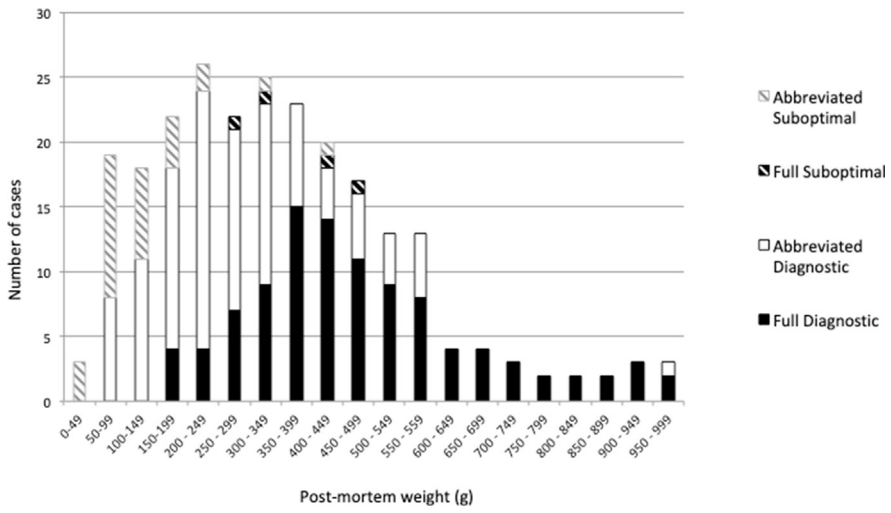

Figure 3 Bar chart demonstrating the numbers of diagnostic studies versus studies of suboptimal image quality (i.e., at least one of the body parts imaged being non-diagnostic) for fetuses at varying body weights up to 1,000 g. Both the full and abbreviated PMMR protocol figures are given. White bars denote abbreviated protocol, solid black bars denote diagnostic quality images. Those with grey stripes and black stripe patterns denote suboptimal quality imaging for the abbreviated and full protocols, respectively.

confidence interval CI: 97.9–99.8%), specificity 68.4% (26/38, 52.5–80.9%), positive predictive value 97.2% (419/431, 95.4–98.4%), negative predictive value 89.7% (26/29, 73.6–96.4%), with overall diagnostic accuracy 96.7% (94.7–98%). Given the imbalance between the numbers of diagnostic and non-diagnostic studies, a Matthews correlation coefficient of 0.78 was computed to better define the accuracy of the model.

Discussion

This study has two main findings for discussion. The first is regarding PMMR protocol adherence and the second concerns the methodology, i.e., using a computational approach to extract key data in order to perform a semi-automated audit of radiological data.

Regarding paediatric PMMR imaging, the present study shows that 66.3% adherence to the full protocol was achieved overall, and radiographers were preferentially using a limited "abbreviated" protocol in all cases weighing <150 g. Although there are no standards regarding the cut-off size for using the abbreviated protocol, this appears to be a reasonable weight limit and in line with a previous study showing that more than half of all cases imaged with PMMR will be non-diagnostic where the body weight measures <122 g.[16]

We achieved a diagnostic image quality rate of almost 100% with imaging above 450 g body weight, suggesting that in order to maximise the "clinical usefulness" of the PMMR imaging services, cases above this weight threshold should be preferentially accepted. Nevertheless, diagnostic image quality was achieved in approximately half of cases weighing <200 g, although it is recognised that there may be a selection bias as dependent on the referral pattern and parental consent for PMMR.

The decision to use the full or abbreviated protocol was subjective, usually reached in discussion between mortuary staff, radiographers, and radiologists (although some imaging performed outside clinical hours may not have had this benefit). No data are available on studies that may have been abandoned or not performed due to small body size. Nevertheless, these data reflect the clinical activity in a busy tertiary referral centre, and thus, may be used as a reference point for other centres engaged in similar activity.

This study reiterates the challenges of imaging small fetuses using PMMR. The field strength of 1.5 T is often inadequate <200 g body weight, and therefore, another imaging technique (e.g., micro-focus CT[20,21]) or higher magnetic field strength is needed.[22,23] Diagnostic imaging at 3 T PMMR has been shown to be better, particularly <20 weeks gestation, although these effects were relatively minor (non-diagnostic rates of 54% at 1.5% and 30% at 3 T [22]), and micro-CT imaging may be the better overall imaging technique for small fetal cases in this setting.[20,24,25] The present audit now highlights the limitations of current PMMR use, and raises local issues including deciding whether an abbreviated protocol is necessary or whether it should only be employed <150 g body weight, or whether to insist on a full protocol for low gestation/body weight.

The second major discussion point is the computational methodology. Manual data collection for large study cohorts is both laborious and error-prone. The presence of structured metadata in DICOM headers offers a potentially rich source of information for quality assessment of radiological practice (e.g., patient demographics, radiation doses, technique-specific parameters, etc.). Basic knowledge of computer programming can facilitate this process of "data mining", using a freely available software package (pydicom) that enables extraction of data according to DICOM tags. Python is a relatively simple and versatile cross-

platform programming language that is rapidly gaining in popularity (including specific medical imaging applications e.g. radiomics analysis with "PyRadiomics"). The present in-house program not only considerably accelerated the process of data collection, but also ensured accurate and consistent recording of the information of interest. Moreover, this approach is easily reproducible as the explicit methodology is outlined in the source code of the program, and can be repeated without any further input.

Although the local radiology post-mortem reports are written according to a suggested template (with some standardisation of report wording) they are still written as free-form text. NLP is a technique that involves computational analysis of text, which is an approach that has found numerous applications in radiology.[15] A limited NLP workflow was used, using specific keywords to identify non-diagnostic cases using search terms that captured the common words used to describe such investigations. This "rule-based" approach incorporates knowledge of standardised reporting templates as well as clinical details to generate classification models. All reports were checked manually before definitive classification as diagnostic or non-diagnostic. That said, NLP is capable of far more advanced semantic analysis (potentially incorporating radiology-specific lexicons, e.g., RadLex[26]), to extract greater meaning from reports, which will ultimately allow automatic classification without verification. More sophisticated approaches using machine learning have been applied recently to automated analysis of various study reports (CT head, lumbar spine MRI), with impressive results, although this requires much greater technical expertise).[27,28]

Although the program was written specifically for the purpose of this particular study, the automated methodology is clearly generalisable and may be equally applicable to other studies and audits where specific terminologies on patient presenting factors, outcomes, imaging sequences, and radiological findings may need to be retrieved. Although there are isolated reports of a similar approach,[29,30] and the authors are unaware of previous studies that have used this combination of automated DICOM metadata extraction and report analysis to establish patterns of clinical practice. By making this program publicly available, similar audits may now be facilitated in other radiology contexts.

Strengths of the present study include a large series of similar examinations, which lend themselves easily to automated audit, particularly as template reporting is used. The clinical activity in a busy tertiary centre is likely to reflect pragmatic practice in other departments, depending on their referral pattern. Clearly this type of approach is easily transferrable to other centres, or multi-site data, and will help to feed into ongoing work from international taskforces (e.g., European Society for Paediatric Radiology [ESPR] post-mortem imaging taskforce[8,10]) to create standardised imaging protocols and reporting templates. Highlighting inconsistent or incorrectly recorded metadata (e.g., clinical indication, operator, or coil types will help improve data recording for future studies).

The success of the present (and other) automated approaches relies on accurate information recording at the time of data acquisition. Constructing a simple NLP workflow has highlighted the need for consistent recording of diagnostic status of studies. Clearly, the low specificity of the present classification model (0.68) indicates the need for further refinement of the model rules. More extensive labelling of the reports for findings of interest might increase the utility of this NLP approach for more granular assessment. Implementing machine learning-based NLP is a natural extension of this work, but will require more data to train a statistical model, as well as greater technical expertise. The simplicity of the present rule-based approach has the benefit of a broader appeal to practising radiologists. This proof of principle study necessitated the manual checking of reports from the NLP workflow, in order to be able to assess the performance of the algorithm; however, there are many potential applications of this technique, and hopefully, it will be better used in future audit cycles.

In conclusion, the present study demonstrated a successful application of an automated approach to data collection for audit and quality assessment/improvement, using post-mortem perinatal imaging as a specific example. Re-audit of these data following implementation change will be straightforward now that the automated workflow is clearly established.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

S.C.S. is supported by a RCUK/UKRI Innovation Fellowship and Medical Research Council (MRC) Clinical Research Training Fellowship (grant ref.: MR/R00218/1). This award is jointly funded by the Royal College of Radiologists (RCR). O.J.A. is funded by a National Institute for Health Research (NIHR) Career Development Fellowship (NIHR-CDF-2017-10-037), and N.J.S. is funded by an NIHR Senior Investigator award. The authors receive funding from the Great Ormond Street Children's Charity and the Great Ormond Street Hospital NIHR Biomedical Research Centre. This article presents independent research funded by the MRC, RCR, NIHR, and the views expressed are those of the author(s) and not necessarily those of the NHS, MRC, RCR, the NIHR, or the Department of Health.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.crad.2019.04.021.

References

1. Arthurs OJ, Bevan C, Sebire NJ. Less invasive investigation of perinatal death. *BMJ* 2015;**351**:h3598.

2. Arthurs OJ, Taylor AM, Sebire NJ. Indications, advantages and limitations of perinatal postmortem imaging in clinical practice. *Pediatr Radiol* 2015;**45**(4):491–500.
3. Arthurs OJ, Hutchinson JC, Sebire NJ. Current issues in postmortem imaging of perinatal and forensic childhood deaths. *Forensic Sci Med Pathol* 2017;**13**(1):58–66.
4. Arthurs OJ, Guy A, Thayyil S, *et al*. Comparison of diagnostic performance for perinatal and paediatric post-mortem imaging: CT versus MRI. *Eur Radiol* 2016;**26**(7):2327–36.
5. Thayyil S, Sebire NJ, Chitty LS, *et al*. Post-mortem MRI versus conventional autopsy in fetuses and children: a prospective validation study. *Lancet* 2013;**382**(9888):223–33.
6. Lewis C, Hill M, Arthurs OJ, *et al*. Factors affecting uptake of postmortem examination in the prenatal, perinatal and paediatric setting. *BJOG* 2018;**125**(2):172–81.
7. Lewis C, Hill M, Arthurs OJ, *et al*. Health professionals' and coroners' views on less invasive perinatal and paediatric autopsy: a qualitative study. *Arch Dis Child* 2018 Jun;**103**(6):572–8.
8. Arthurs OJ, van Rijn RR, Sebire NJ. Current status of paediatric postmortem imaging: an ESPR questionnaire-based survey. *Pediatr Radiol* 2014;**44**(3):244–51.
9. Arthurs OJ, van Rijn RR, Taylor AM, *et al*. Paediatric and perinatal postmortem imaging: the need for a subspecialty approach. *Pediatr Radiol* 2015;**45**(4):483–90.
10. Arthurs OJ, van Rijn RR, Whitby EH, *et al*. ESPR postmortem imaging task force: where we begin. *Pediatr Radiol* 2016;**46**(9):1363–9.
11. The Royal College of Pathologists. *Guidelines on autopsy practice: neonatal death*. 2018. Available at: www.rcpath.org. [Accessed 16 May 2018].
12. The Royal College of Pathologists. *Guidelines on autopsy practice: third trimester antepartum and intrapartum stillbirth*. 15 June 2017. 2017Available at: www.rcpath.org. [Accessed 16 May 2018].
13. The Royal College of Pathologists. *Guidelines on autopsy practice: fetal autopsy (2nd trimester fetal loss and termination of pregnancy for congenital anomaly)*. 2017. Available at: www.rcpath.org. [Accessed 16 May 2018].
14. Norman W, Jawad N, Jones R, Taylor AM, Arthurs OJ. Perinatal and paediatric post-mortem magnetic resonance imaging (PMMR): sequences and technique. *Br J Radiol* 2016 Jun;**89**(1062):20151028.
15. Pons E, Braun LM, Hunink MG, *et al*. Natural language processing in radiology: a systematic review. *Radiology* 2016;**279**(2):329–43.
16. Jawad N, Sebire NJ, Wade A, Taylor AM, Chitty LS, Arthurs OJ. Body weight lower limits of fetal postmortem MRI at 1.5T. *Ultrasound Obstet Gynecol* 2016 Jul;**48**(1):92–7.
17. Mason D. SU-E-T-33: pydicom: an open source DICOM library. *Med Phys* 2011;**38**(6Part10):3493.
18. McKinney W. Data structures for statistical computing in Python. In S. van der Walt, J. Millman (editors) *Proceedings of the 9th Python in science conference* (SciPy 2010) 28 june–3 july, Austin, TexasAvailable at: http://conference.scipy.org/proceedings/scipy2010/. [Accessed 15 October 2018].
19. Bird S, Klein E, Loper E. *Natural language processing with Python*. 2nd edn. Sebastopol, CA: O'Reilly Media; 2016.
20. Hutchinson JC, Kang X, Shelmerdine SC, *et al*. Postmortem microfocus computed tomography for early gestation fetuses: a validation study against conventional autopsy. *Br J Radiol* 2018;**218**(4). 445.e1-.e12.
21. Hutchinson JC, Shelmerdine SC, Simcock IC, *et al*. Early clinical applications for imaging at microscopic detail: microfocus computed tomography (micro-CT). *Br J Radiol* 2017;**90**(1075):20170113.
22. Kang X, Cannie MM, Arthurs OJ, *et al*. Post-mortem whole-body magnetic resonance imaging of human fetuses: a comparison of 3-T vs. 1.5-T MR imaging with classical autopsy. *Eur Radiol* 2017;**27**(8):3542–53.
23. Thayyil S, Cleary JO, Sebire NJ, *et al*. Post-mortem examination of human fetuses: a comparison of whole-body high-field MRI at 9.4 T with conventional MRI and invasive autopsy. *Lancet* 2009;**374**(9688):467–75.
24. Hutchinson JC, Barrett H, Ramsey AT, *et al*. Virtual pathological examination of the human fetal kidney using micro-CT. *Ultrasound Obstet Gynecol* 2016;**48**(5):663–5.
25. Hutchinson JC, Arthurs OJ, Ashworth MT, *et al*. Clinical utility of postmortem microcomputed tomography of the fetal heart: diagnostic imaging vs macroscopic dissection. *Ultrasound Obstet Gynecol* 2016;**47**(1):58–64.
26. Radiological Society of North America R. *RSNA informatics RadLex playbook*. February 2018. Available at: http://playbook.radlex.org/playbook/SearchRadlexAction. [Accessed 3 January 2019].
27. Zech J, Pain M, Titano J, *et al*. Natural language-based machine learning models for the annotation of clinical radiology reports. *Radiology* 2018;**287**(2):570–80.
28. Tan WK, Hassanpour S, Heagerty PJ, *et al*. Comparison of natural language processing rules-based and machine-learning systems to identify lumbar spine imaging findings related to low back pain. *Acad Radiol* 2018;**25**(11):1422–32.
29. England JR, Colletti PM. Automated reporting of DXA studies using a custom-built computer program. *Clin Nucl Med* 2018;**43**(6):474–5.
30. Cutright D, Gopalakrishnan M, Roy A, *et al*. DVH Analytics: a DVH database for clinicians and researchers. *J Appl Clin Med Phys* 2018;**19**(5):413–27.