Clinical Radiology

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

--Manuscript Draft--

Manuscript Number:	CRAD-D-19-00063R2
Full Title:	Automated data extraction and report analysis in computer-aided radiology audit; practice implications from post mortem paediatric imaging.
Article Type:	Original Paper
Corresponding Author:	Susan C Shelmerdine, MBBS BSc MRCSEng FRCR Great Ormond Street Hospital For Children NHS Foundation Trust London, UNITED KINGDOM
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Great Ormond Street Hospital For Children NHS Foundation Trust
Corresponding Author's Secondary Institution:	
First Author:	Susan C Shelmerdine, MBBS BSc MRCSEng FRCR
First Author Secondary Information:	
Order of Authors:	Susan C Shelmerdine, MBBS BSc MRCSEng FRCR
	Mrinal Singh
	Wendy Norman
	Rod Jones
	Neil J Sebire, FRCPath
	Owen J Arthurs, FRCR, PhD
Order of Authors Secondary Information:	
Abstract:	Aim To determine local departmental adherence to our paediatric post-mortem MRI imaging protocols, using a customised automated computational approach. Materials and Methods A retrospective review of 460 whole body post-mortem MRI performed at our institution over a 5 ½ year period was assessed for adherence to a full or abbreviated imaging sequence protocol. We developed a simple computer program to batch process DICOM files, extracting imaging sequence details, followed by natural language processing (NLP) of authorised reports to automate information extraction of diagnostic image quality. Results Our program was able to extract study parameters from the entire dataset (approximately 80GB 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% CI: 94.7 – 98.0%). 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 5 cases were the examinations non-diagnostic for all body systems, all of whom weighed <100g (24.7 – 72g) and imaged using the abbreviated protocol. Conclusion We have 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 this data following change implementation

	will be straightforward now that we have clearly established the automated workflow.

Full Title:

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

Short Title:

Computer aided radiology audit: paediatric PMMR

Authors & Affiliations:

Susan C Shelmerdine [a,b], susie_c_s@yahoo.co.uk

Mrinal Singh [a], mrinal.singh@nhs.net

Wendy Norman [a], wendy.norman@gosh.nhs.uk

Rod Jones [a], rod.jones@gosh.nhs.uk

Neil J Sebire [b,c], neil.sebire@gosh.nhs.uk

Owen J Arthurs [a,b] owen.arthurs@gosh.nhs.uk

[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

Corresponding author:

Dr. Susan Shelmerdine, Paediatric Radiology Research Fellow, Department of Radiology

Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK WC1N 3JH

E mail: susie_c_s@yahoo.co.uk

Tel/Fax +44(0)20 7405 9200

Keywords: Post-mortem, radiology, MRI, children, paediatric, autopsy

Type of manuscript: Original research

Word count: 3233

Figure count: 3

Table count: 2

Funding:

SCS 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 OJA is funded by a National Institute for Health Research (NIHR) Career

Development Fellowship (NIHR-CDF-2017-10-037), and NJS 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.

Role of Funding Source:

The funding sources stated had no involvement or influence on the conception, data collection or analysis of this study.

Conflicts of Interest:

The authors have no conflicts of interest to declare.

Author Contributions

1 Guarantor of integrity of the entire study - OJA
2 Study concepts and design – OJA, NJS, SCS, MS
3 Literature research – SCS, MS
4 Clinical studies – SCS, MS, RJ, WN
5 Experimental studies / data analysis – SCS, MS, RJ, WN
6 Statistical analysis – SCS, MS
7 Manuscript preparation – OJA, NJS, SCS, MS
8 Manuscript editing – OJA, NJS, SCS, MS

2 April 2019

Deputy Editor, *Clinical Radiology*

Dear Dr Julie Cox,

Thank you to you and your reviewers for your invaluable feedback regarding our manuscript entitled "Automated data extraction and report analysis in computer-aided radiology audit; practice implications from post mortem paediatric imaging"; Manuscript ID: **CRAD-D-19-00063**.

In this second revision, we have made the following amendments to our article as suggested below (responses in **bold font**). Since the only changes relate to figures and not to text in the main manuscript, the manuscript document has not been tampered with and the previously submitted 'revised clean' version has been carried forward in this submission.

Advisory Editorial Comments:

1. The MR images will be limited by the acquisition matrix (and size of the specimens) - normal production process will confirm they are adequate but despite one of the reviewers' comments I think this is likely and would not suggest further revision of these images.
   Thank you for your understanding – this has been left untouched.

2. Fig 1 is difficult to read as the text is not clear no matter what resolution the image is set to: please revise.
   Figure 1 has been reworked to make the font larger and the text boxes bigger to allow for better readability. This has been uploaded and the previous figure 1 removed.

I can confirm that co-authors have read and agree to the changes in the manuscript above.

Abstract

Aim

To determine local departmental adherence to our paediatric post-mortem MRI imaging protocols, using a customised automated computational approach.

Materials and Methods

A retrospective review of 460 whole body post-mortem MRI performed at our institution over a 5 ½ year period was assessed for adherence to a full or abbreviated imaging sequence protocol. We developed a simple computer program to batch process DICOM files, extracting imaging sequence details, followed by natural language processing (NLP) of authorised reports to automate information extraction of diagnostic image quality.

Results

Our program was able to extract study parameters from the entire dataset (approximately 80GB 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% CI: 94.7 – 98.0%). 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 5 cases were the examinations non-diagnostic for all body systems, all of whom weighed <100g (24.7 – 72g) and imaged using the abbreviated protocol.

Conclusion

We have 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 this data following change implementation will be straightforward now that we have clearly established the automated workflow.

Manuscript

Introduction

The persistent decline in consent rates for paediatric autopsy has facilitated development of non-invasive alternatives, based on imaging [1-4]. Post-mortem MRI (PMMR) provides high diagnostic accuracy rates 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, we published our PMMR protocols in 2015[14]. However, our full PMMR protocol, whilst 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 our own adherence to our PMMR protocols, and understand the reasons for any variation. In order to do this efficiently, we designed a custom computer program to extract the relevant information from Digital Imaging in Communications in Medicine (DICOM) metadata. We also applied basic natural language processing (NLP) to analyse the study reports [15]. With this computational approach we hope 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 we provide the code used in our study 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 our institution was conducted for all PMMR studies performed over a 5½ 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 our 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 our local research and development (R&D) office.

Demographic data for each patient was 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 to 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 on a 1.5T MR scanner (Avanto, Siemens Medical Solutions, Erlangen, Germany), by one of two experienced MR radiographers. Our local PMMR protocols, which we took as our standard, has been previously published and are included in **Table 1**[14].

In brief, our radiographers perform either a 'full protocol' or 'abbreviated protocol'. The full protocol involves three-dimensional isovolumetric T1, 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 is 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 sequence 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.

Our 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 we have no restrictions for referral indication, although we usually do not recommend imaging in cases less than 200g (unless there is no other imaging alternative) given the increased likelihood of non-diagnostic imaging [16].

Data Collection and Analysis

We queried our local RIS 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 data elements. We designed a small computer program for automated data extraction using the free, open-source “Pydicom” package[17] (https://pypi.org/project/pydicom/) (see **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 our program, and the resulting data was tabulated using the “pandas” data analysis library[18].

We performed natural language processing (NLP) on the examination reports to partially automate extraction of some measure of diagnostic outcome, given that a comment regarding diagnostic image quality is required per body system using our standardised reporting template for PMMR studies. We used Natural Language Toolkit (NLTK [19]) and “spaCy” - both free, open-source python packages—to create a rule-based binary classifier (i.e. diagnostic or non-diagnostic) (see **Supplementary Material,**

Appendix S2). Feature extraction involved identification of word boundaries (“tokenization”) and formation of a list of words used in each report. This list was subsequently “normalized” by converting all words to lower case. Finally, we searched the resulting word list 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 manually checked by one of the authors (SCS) for having the same sequences as stated in the DICOM headers, and ~~also~~ whether the reports were correctly classified as being either 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. **Figure 1** outlines our workflow for both extraction of imaging parameters and NLP of diagnostic image quality. **Figure 2** demonstrates an example of what a radiologist would classify and report as a ‘diagnostic quality’ versus ‘non-diagnostic’ quality study for two different cases in different body areas.

Prior to data analysis, our 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 Corporation, USA) for collation and further analysis.

Results

Demographics

Over the 5 ½ year study period we reviewed 460 PMMR examinations performed from 460 individual cases. 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 to 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 680g (mean: 2.8kg, range: 13g – 87kg). 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

Our program was able to extract study parameters from the entire dataset (approximately 80GB of data) in less than three 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 2051g (average 3314g; 165g – 87,000g), and for those having the abbreviated protocol the median weight was 225g (average 264g; 12.6 – 1050g).

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, 3 cases included imaging of only the head, 1 case of only the neck, 2 cases of only the thorax and 2 cases where there was imaging of the thorax and abdomen, but not the head (in one case the child already had a recent antemortem 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. 298/ 305 (97.7%) of the full protocol 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 5 cases were the PMMR examinations entirely non-diagnostic for all body parts examined. In all cases these were fetuses weighing <100g (24.7 – 72g) and had undergone an abbreviated protocol.

Of the 7 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, 6 for two body systems, 1 for three body system and 2 for four body systems. The breakdown of which body systems were non-diagnostic are shown in **Table 2**.

There were 61/460 (13.2%) PMMR examinations performed in cases weighing <200g (4 full, 56 abbreviated, 1 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.

We did not scan any cases with the full protocol below 150g body weight. The full protocol was adhered to in 89.2% (248/278) cases weighing 450g or more, with 98.8% (245/248) diagnostic image quality for all body systems. Between 150 – 449g, the full protocol was adhered to in 28.4% (56/197), with 94.6% (53/56) diagnostic image quality for all body systems. See **Figure 3** for a graph depicting the results of our study for cases weighing up to 1000g in body weight.

Classification Model Performance

Our 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% 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 – 98.0%). Given the imbalance between the numbers of diagnostic and non-diagnostic studies, we computed a Matthews correlation coefficient of 0.78 to better define accuracy of the model.

Discussion

This study has two main findings for discussion. The first is regarding PMMR protocol adherence and the second concerns our 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, our study shows that we achieved 66.3% adherence with the full protocol overall, and our radiographers were preferentially using a limited ‘abbreviated’ protocol in all cases weighing <150g. Whilst we do not have any standards regarding the cut-off size for using the abbreviated protocol, this appears to be a reasonable weight limit and in line with our previous study showing that more than half of all cases imaged with PMMR will be non-diagnostic where the body weight measures less than 122g[16].

We achieved an almost 100% diagnostic image quality rate with imaging above 450g body weight suggesting that in order to maximise the ‘clinical usefulness’ of our post-mortem MRI imaging services, we should preferentially accept cases above this weight threshold. Nevertheless, we did achieve diagnostic image quality in approximately half of cases weighing <200g, although we recognise that there may be a selection bias as we are dependent upon our referral pattern and parental consent for post mortem imaging.

We also recognise that 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). We did not have data available on studies that may have been abandoned or not performed due to small body size. Nevertheless, this data reflects 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 re-iterates the challenges of imaging small fetuses at PMMR. Field strength of 1.5T is often inadequate below 200g body weight and therefore another imaging technique (e.g. micro-focus computed tomography (micro-CT)[20,21]) or higher magnetic field strength is needed [22,23]. Diagnostic imaging at 3T PMMR has been shown to be better particularly below 20 weeks gestation, although these effects were relatively minor (non-diagnostic rates of 54% at 1.5% and 30% at 3T [22]), and micro-CT imaging may be the better overall imaging modality for small fetal cases in this setting [20,24,25]. Our 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 below 150g body weight, or whether to insist on a full protocol for low gestation / body weight.

The second major discussion point is our 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 radiologic practice (e.g. patient demographics, radiation doses, modality specific parameters, etc). We have shown that a 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”). Our 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 our 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. Natural language processing (NLP) is a technique that computational analysis of text - an approach that has found numerous applications in radiology [15]. We used a limited NLP workflow 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 manually checked 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 that we anticipate 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 MR), with impressive results, although this requires much greater technical expertise)[27,28].

Whilst our 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 we 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 our study include a large series of similar examinations which lend themselves easily to automated audit, particularly as we use template reporting. Our 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 on-going 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 our (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 our 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 our 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 we are only beginning to understand the potential applications of this technique and hope to better use it in future audit cycles.

We conclude that we have 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 this data following change implementation will be straightforward now that we have clearly established the automated workflow.

Figure legends

Figure 1

Workflow diagram for automated data collection utilised in our methodology. RIS = Radiology Information System; NLTK = Natural Language ToolKit, DICOM = Digital Imaging & Communication in Medicine

Figure 2

Diagnostic and non-diagnostic quality post-mortem MRI 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.

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 1000g. Both the full and abbreviated post-mortem MRI imaging 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.

Table 1. Sequence parameters for full post-mortem MRI protocol in infant and perinatal deaths (adapted with permission from **BLINDED**) are given below. The two sequences followed by '*' 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).

Sequence	FOV (mm)	Slice thickness (mm)	Matrix	Voxel size (mm)	TR (ms)	TE (ms)	Averages (NEX/NSA)	Number slices and gap	Approximate length of sequence (min)
BRAIN IMAGING									
3D FLASH T1-w (sag)									
Perinatal	256	1	256/256	1.0 x 1.0 x 1.0	11	4.9	3	60 per slab	5.44
Child	256	1	224/256	1.0 x 1.0 x 1.0	11	4.9	1	160 per slab	4.20
2D DESTIR T2-w (axial and coronal)									
Perinatal	100	2	172/256	0.4 x 0.4 x 2.0	5460	16 and 115	6	18 (1mm)	13.46
Child	200	4	216/320	0.7 x 0.6 x 4.0	6180	14 and 115	1	22 (1mm)	3.19
2D GRE T1 HEME (axial)									
Perinatal	100	4	120/256	0.5 x 0.4 x 4.0	800	26	4	18 (0mm)	6.26
Child	200	5	144/256	1.0 x 0.8 x 5.0	800	26	2	18 (0mm)	3.52
DWI (b-values 0, 500, 1000)									
Perinatal	230	5	128/128	1.8 x 1.8 x 5.0	2700	96	3	19 (0mm)	1.06
Child	230	5	128/128	1.8 x 1.8 x 5.0	2700	96	3	19 (0mm)	1.06
SPINE IMAGING									
2D T2-w TSE (sag)									
Perinatal	150	1.5	128/256	0.6 x 0.6 x 1.5	9.1	4.5	8	12 per slab	4.24
Child	300	3	272/320	1.1 x 0.9 x 3.0	3050	109	3	11 per slab	5.43
3D FLASH T1-w (sag)									
Perinatal	150	1.25	128/256	0.6 x 0.6 x 1.3	11	5.3	10	16 per slab	3.19
Child	350	1.40	144/256	1.4 x 1.4 x 1.4	11	4.9	6	32 per slab	5.06
BODY IMAGING (NECK TO PELVIS)									
3D T2-w TSE (cor)*									
Perinatal	200	0.8	160/256	0.8 x 0.8 x 0.8	3500	275	2	72 per slab	6.20
Child	360	1.4	226/256	1.4 x 1.4 x 1.4	3500	173	1	96 per slab	3.42
3D T1-w VIBE (cor)*									
Perinatal	200	0.8	160/256	0.8 x 0.8 x 0.8	5.9	2.4	8	72 per slab	5.52
Child	360	1.4	224/256	1.4 x 1.4 x 1.4	5.9	2.4	5	72 per slab	6.33
3D CISS T2-w (axial) (thoracic coverage for cardiac assessment)									
Perinatal	150	0.6	192/256	0.6 x 0.6 x 0.6	5.6	2.5	10	Cover heart and lungs	29.26
Child	150	0.6	192/256	0.6 x 0.6 x 0.6	5.6	2.5	10		29.26
2D T2-w tirm (axial) (Ti = 150)									
Perinatal	180	5	160/256	0.7 x 0.7 x 5.0	5080	109	5	Cover body and pelvis	6.58
Child	300	5	168/256	1.2 x 1.2 x 5.0	8390	108	4		4.47
DWI	As for head with greater number of slices to cover chest, abdomen and pelvis								1.06

Table 2. Suboptimal PMMR studies, divided by protocol adherence, showing which body system was deemed as non-diagnostic in each subgroup.

PMMR Protocol	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

References

1. Arthurs OJ, Bevan C, Sebire NJ. Less invasive investigation of perinatal death. *BMJ (Clinical research ed)* 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 (London, England)* 2013; **382**(9888): 223-33.

6. Lewis C, Hill M, Arthurs OJ, Hutchinson C, Chitty LS, Sebire NJ. 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, Hutchinson JC, Chitty LS, Sebire N. Health professionals' and coroners' views on less invasive perinatal and paediatric autopsy: a qualitative study. *Archives of disease in childhood* 2018.

8. Arthurs OJ, van Rijn RR, Sebire NJ. Current status of paediatric post-mortem imaging: an ESPR questionnaire-based survey. *Pediatr Radiol* 2014; **44**(3): 244-51.

9. Arthurs OJ, van Rijn RR, Taylor AM, Sebire NJ. 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. Pathologists RCo. Guidelines on autopsy practice: Neonatal Death. 2018. 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 2017. 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. www.rcpath.org (accessed 16 May 2018).

14. BLINDED XXXX

15. Pons E, Braun LM, Hunink MG, Kors JA. Natural Language Processing in Radiology: A Systematic Review. *Radiology* 2016; **279**(2): 329-43.

16. BLINDED XXXX

17. Mason D. SU-E-T-33: Pydicom: An Open Source DICOM Library. *Medical Physics* 2011; **38**(6Part10): 3493-.

18. McKinney W. Data Structures for Statistical Computing in Python. *Proceedings of the 9th Python in Science Conference* 2010: 51 - 6.

19. Bird S, Klein E, Loper E. Natural Language Processing with Python. 2nd Edition ed. Sebastopol, California 95472: 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, Sebire NJ, Arthurs OJ. 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 (London, England)* 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.

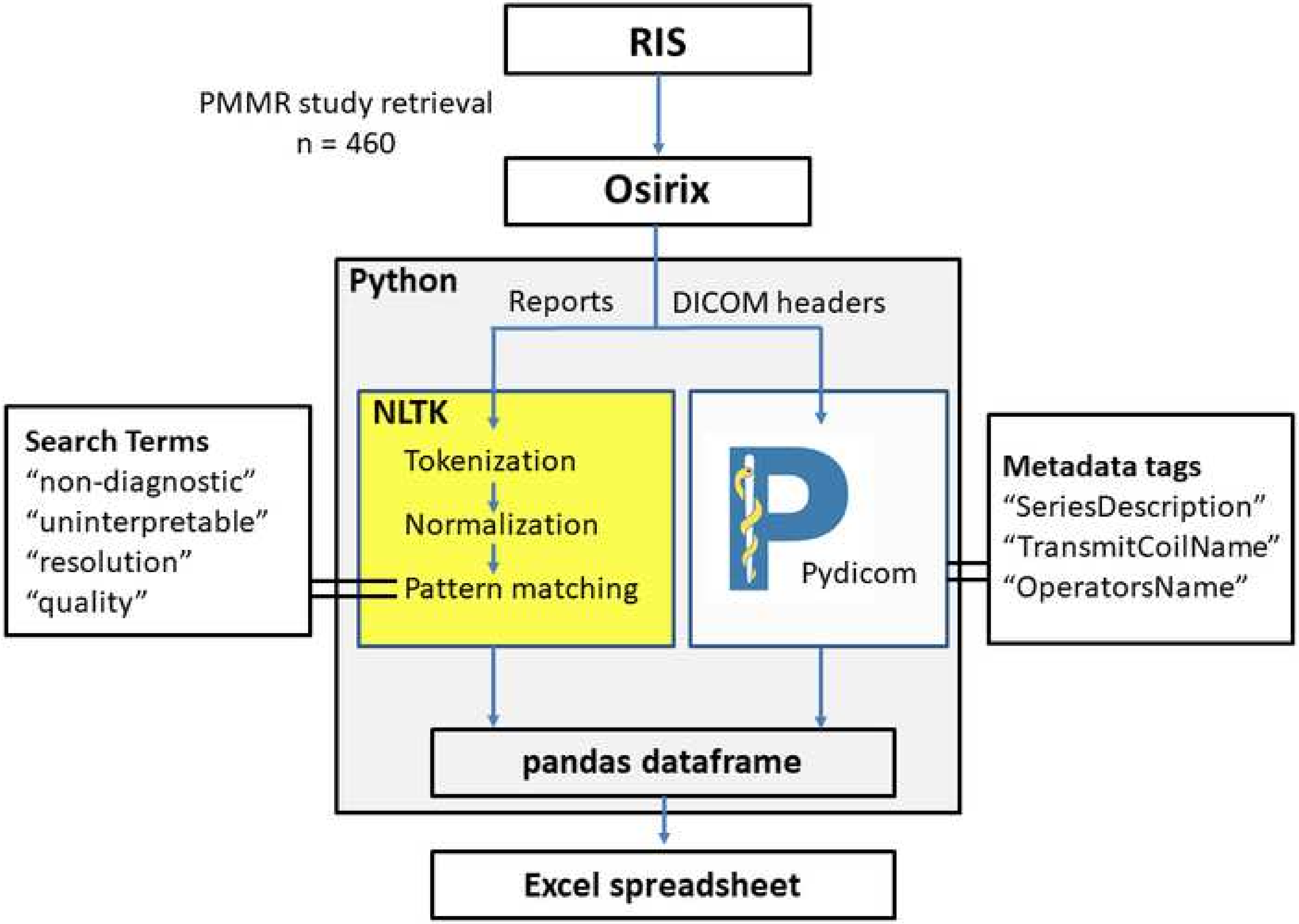
26. Radiological Society of North America R. RSNA Informatics RadLex Playbook. February 2018 2018. http://playbook.radlex.org/playbook/SearchRadlexAction (accessed 3rd January 2019 2019).

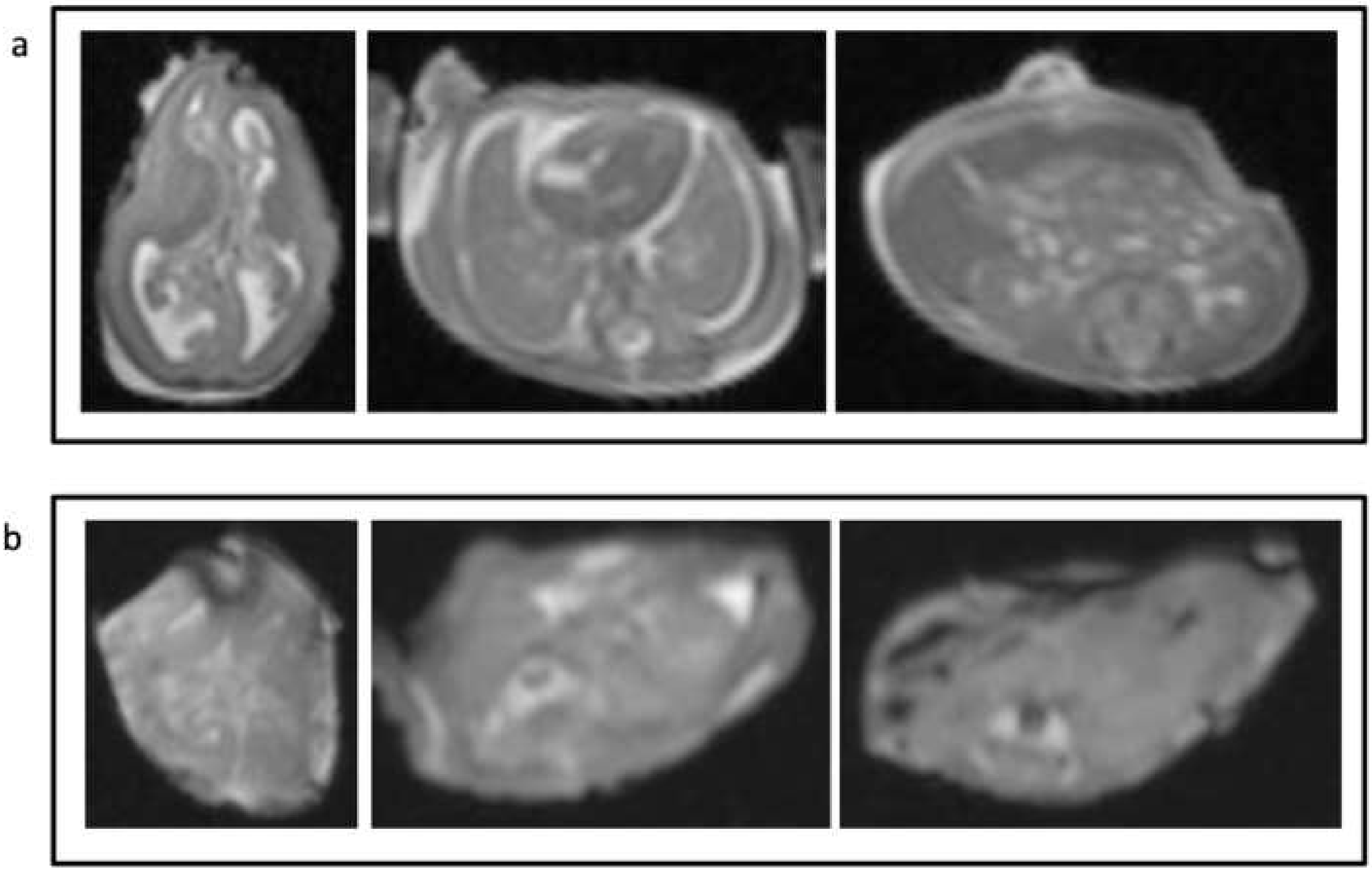
27. Zech J, Pain M, Titano J, et al. Natural Language-based Machine Learning Models for the Annotation of Clinical Radiology Reports. *Radiology* 2018; **287**(2): 570-80.

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. *Clinical nuclear medicine* 2018; **43**(6): 474-5.

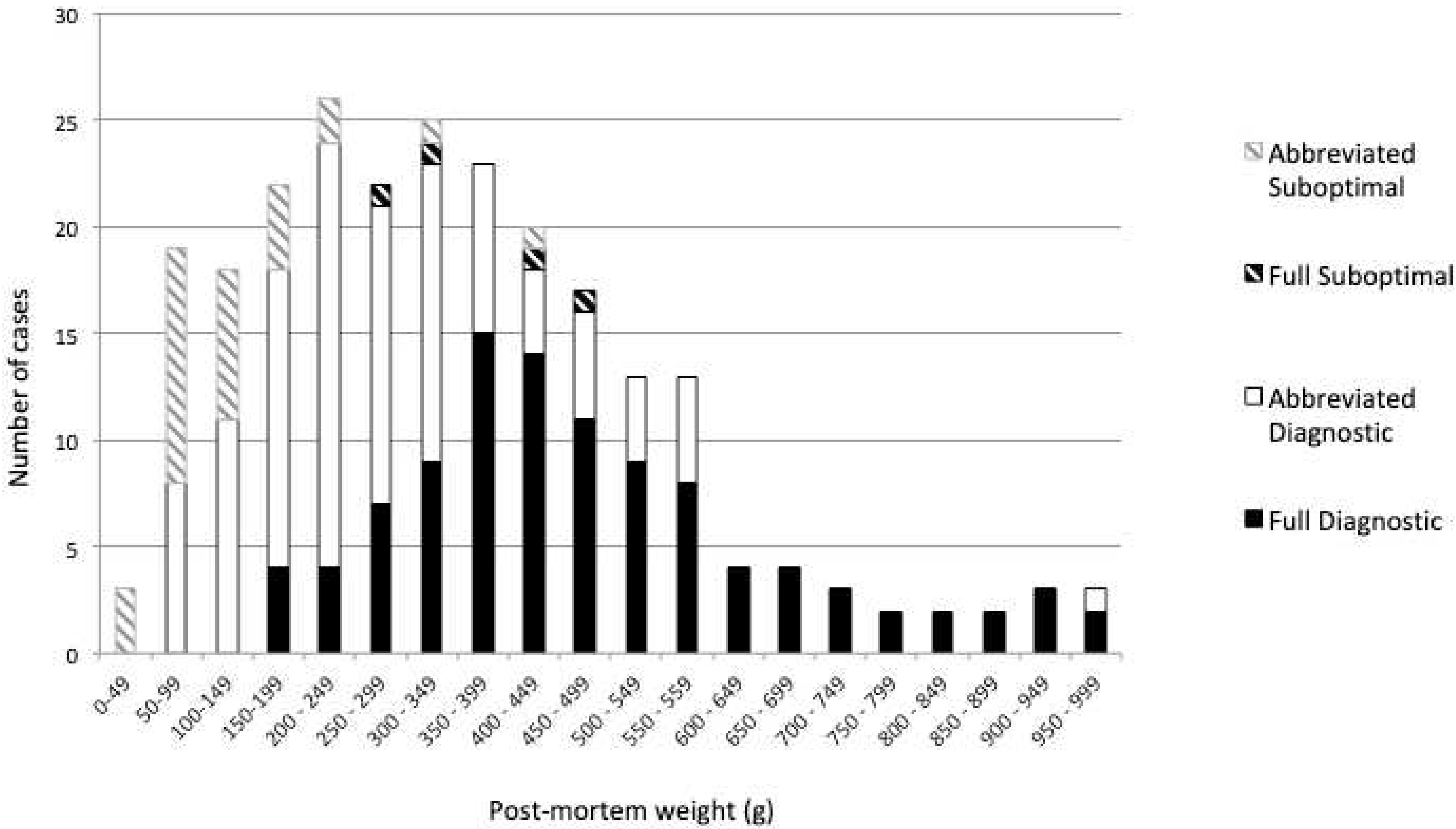
30. Cutright D, Gopalakrishnan M, Roy A, Panchal A, Mittal BB. DVH Analytics: A DVH database for clinicians and researchers. *Journal of applied clinical medical physics* 2018; **19**(5): 413-27.


RIS
PMMR study retrieval
n = 460
Osirix
Python
Reports
DICOM headers
NLTK
Tokenization
Normalization
Pattern matching
Pydicom
Search Terms
"non-diagnostic"
"uninterpretable"
"resolution"
"quality"
Metadata tags
"SeriesDescription"
"TransmitCoilName"
"OperatorsName"
pandas dataframe
Excel spreadsheet


a
b

Revised Figure 3 600dpi

Click here to access/download;Revised Figure;Figure 3.tiff



Electronic Supplementary Material

Appendix S1. Pydicom Code

```
import pandas as pd
import numpy as np
from collections import OrderedDict
import glob
import pydicom

def sequence_extractor(source, sequences):
  rows_list = []
  for file in glob.glob(source):
      ds = pydicom.dcmread(file, force=True, specific_tags=['PatientID',
'SeriesDescription', 'TransmitCoilName', 'OperatorsName'])
      coil = getattr(ds, 'TransmitCoilName', None)
      opname = getattr(ds, 'OperatorsName', None)
      newrow = OrderedDict([
          ('id', ds.PatientID),
          ('seq', ds.SeriesDescription),
          ('coil', coil),
          ('opname', opname)
          ])
      rows_list.append(newrow)

  df = pd.DataFrame.from_dict(rows_list)
  df2 = df.groupby(['id','seq']).size().unstack('seq')
  seq_pmmr = df2[sequences]
  seq_other = df2.drop(sequences, axis=1)

#Specify path to DICOM files
source = '/Path/to/folder/*/*/*/*.dcm'

#Specify precise list of sequence names (as recorded in metadata)
sequences = [
    'fl3D_t1_sag',
    't2_destir_tra',
    't2_destir_cor',
    't2_fl2d_tra_haem',
    'ep2d_dwi_tra',
    'ep2d_dwi_tra_ADC',
    't2_tse_rst_sag',
    'fl3D_t1_sag spine',
    'fl3D_t1_sag_spine',
    't2_tse3d_vfl_ns_cor',
    'VIBE fs cor',
    't2_tirm_tra_dark-fl_pat2',
    't2_ci3d_iso Heart'
]

#Run function
sequence_extractor(source, sequences)
```

Appendix S2. Natural Language Programming Code

```
import spacy
import pandas as pd
import numpy as np
```

```
import nltk
from nltk.tokenize.toktok import ToktokTokenizer
import re
import unicodedata
from spacy import displacy
from spacy.matcher import Matcher
from spacy.matcher import PhraseMatcher
import os
import glob
from pathlib import Path

def pmmr_nlp(source, terms):
    nlp = spacy.load('en', disable = ['ner'])
    tokenizer = ToktokTokenizer()
    stopword_list = nltk.corpus.stopwords.words('english')
    stopword_list.remove('no')
    stopword_list.remove('not')
    stopword_list.remove('both')

    def remove_stopwords(text):
        tokens = tokenizer.tokenize(text)
        tokens = [token.strip() for token in tokens]
        filtered_tokens = [token for token in tokens if token not in
stopword_list]
        filtered_text = ' '.join(filtered_tokens)
        return filtered_text

    def remove_special_characters(text, remove_digits=False):
        pattern = r'[^a-zA-z0-9\s]' if not remove_digits else r'[^a-zA-
z\s]'
        text = re.sub(pattern, '', text)
        return text

    def normalize(report, remove_digits = False):
            #make lowercase
            report = report.lower()
            #remove extra newlines
            report = re.sub(r'[\r|\n|\r\n]+', ' ',report)
            #remove extra whitespace
            report = re.sub(' +', ' ', report)
            #remove special characters
            special_char_pattern = re.compile(r'([{.(-)!}])')
            report = special_char_pattern.sub(" \\1 ", report)
            report = remove_special_characters(report,
remove_digits=remove_digits)
            #remove stopwords
            report = remove_stopwords(report)
            return report

    nlp.vocab.strings.add('DIAGNOSTIC-YIELD')
    diag = nlp.vocab.strings['DIAGNOSTIC-YIELD']

    def add_ent(matcher, doc, i, matches):
        # Get the current match and create tuple of entity label, start
and end.
        # Append entity to the doc's entity
        match_id, start, end = matches[i]
        doc.ents += ((diag, start, end),)
```

```
    pm = PhraseMatcher(nlp.vocab)
    terminology_list = terms
    patterns = [nlp(text) for text in terminology_list]
    pm.add('TerminologyList', add_ent, *patterns)

    dict = []
    for file in sorted(glob.glob(source)):
        report=open(file).read()
        doc = nlp(normalize(report))
        pm_matches = pm(doc)
        ent_diag = len([ent.label_ for ent in doc.ents if
ent.label_=='DIAGNOSTIC-YIELD'])
        fn = Path(file).stem
        data = {"filename": fn, "diag": ent_diag}
        dict.append(data)
    output = pd.DataFrame(dict)
    return(output)

#Specify path to folder containing all reports as txt files
source = '/path/to/reports/*.txt'

#Specify search terms in list
terms = ['non diagnostic', 'not diagnostic', 'nondiagnostic']

#Run function
pmmr_nlp(source, terms)
```

Declaration of interests

- ✓ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Highlights

1. Automated data extraction allows rapid DICOM metadata and report keyword compilation.
2. Our PMMR protocol gave diagnostic image quality in 98.8% cases weighing >450g
3. PMMR in fetuses weighing <200g, were more likely to be non-diagnostic.

Please wait...

If this message is not eventually replaced by the proper contents of the document, your PDF viewer may not be able to display this type of document.

You can upgrade to the latest version of Adobe Reader for Windows®, Mac, or Linux® by visiting http://www.adobe.com/go/reader_download.

For more assistance with Adobe Reader visit http://www.adobe.com/go/acrreader.

Please wait...

If this message is not eventually replaced by the proper contents of the document, your PDF viewer may not be able to display this type of document.

You can upgrade to the latest version of Adobe Reader for Windows®, Mac, or Linux® by visiting http://www.adobe.com/go/reader_download.

For more assistance with Adobe Reader visit http://www.adobe.com/go/acrreader.

Please wait...

If this message is not eventually replaced by the proper contents of the document, your PDF viewer may not be able to display this type of document.

You can upgrade to the latest version of Adobe Reader for Windows®, Mac, or Linux® by visiting http://www.adobe.com/go/reader_download.

For more assistance with Adobe Reader visit http://www.adobe.com/go/acrreader.

Please wait...

If this message is not eventually replaced by the proper contents of the document, your PDF viewer may not be able to display this type of document.

You can upgrade to the latest version of Adobe Reader for Windows®, Mac, or Linux® by visiting http://www.adobe.com/go/reader_download.

For more assistance with Adobe Reader visit http://www.adobe.com/go/acrreader.