The use of digital pathology and image analysis in clinical trials

Robert Pell[1] (ID), Karin Oien[2], Max Robinson[3], Helen Pitman[4], Nasir Rajpoot[5], Jens Rittscher[1], David Snead[6†], and Clare Verrill,[1†*]

on behalf of the UK National Cancer Research Institute (NCRI) Cellular-Molecular Pathology (CM-Path) quality assurance working group[‡]

[1]*Nuffield Department of Surgical Sciences, University of Oxford, and Oxford NIHR Biomedical Research Centre, Oxford, UK*
[2]*Institute of Cancer Sciences – Pathology, University of Glasgow, Glasgow, UK*
[3]*Centre for Oral Health Research, Newcastle University, Newcastle upon Tyne, UK*
[4]*Strategy and Initiatives, National Cancer Research Institute, London, UK*
[5]*Department of Computer Science, University of Warwick, Warwick, UK*
[6]*Department of Pathology, University Hospitals Coventry and Warwickshire, Coventry, UK*

*Correspondence: Clare Verrill, Nuffield Department of Surgical Sciences, University of Oxford, Level 6, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK. E-mail: clare.verrill@nds.ox.ac.uk

[†]Joint senior authors. Both authors contributed equally to the manuscript.
[‡]The members of the NCRI CM-Path Quality Assurance Panel are listed in the Acknowledgements section.

Abstract

Digital pathology and image analysis potentially provide greater accuracy, reproducibility and standardisation of pathology–based trial entry criteria and endpoints, alongside extracting new insights from both existing and novel features. Image analysis has great potential to identify, extract and quantify features in greater detail in comparison to pathologist assessment, which may produce improved prediction models or perform tasks beyond manual capability. In this article, we provide an overview of the utility of such technologies in clinical trials and provide a discussion of the potential applications, current challenges, limitations and remaining unanswered questions that require addressing prior to routine adoption in such studies. We reiterate the value of central review of pathology in clinical trials, and discuss inherent logistical, cost and performance advantages of using a digital approach. The current and emerging regulatory landscape is outlined. The role of digital platforms and remote learning to improve the training and performance of clinical trial pathologists is discussed. The impact of image analysis on quantitative tissue morphometrics in key areas such as standardisation of immunohistochemical stain interpretation, assessment of tumour cellularity prior to molecular analytical applications and the assessment of novel histological features is described. The standardisation of digital image production, establishment of criteria for digital pathology use in pre–clinical and clinical studies, establishment of performance criteria for image analysis algorithms and liaison with regulatory bodies to facilitate incorporation of image analysis applications into clinical practice are key issues to be addressed to improve digital pathology incorporation into clinical trials.

Keywords: computerised image analysis; digital image analysis; digital microscopy

Introduction

Cellular pathology facilitates diagnosis and clinical trial treatment stratification. Genomic approaches divide traditional entities into smaller subcategories requiring large multicentre, often multinational, interventional studies. Such studies require high diagnostic standards and high reporting uniformity.

Digital pathology refers to the use of computer workstations to view digital whole slide images (WSIs) obtained from high resolution scanning of glass microscope slides [1]; uses include teaching, research or primary diagnostic reporting. Digital pathology devices or image analysis algorithms used in diagnostic reporting usually require regulatory approval such as FDA (USA) or CE IVD (Europe).

Digital pathology and image analysis could ensure greater accuracy, reproducibility and standardisation of study inclusion criteria and outcomes. Automated image analysis can help to extract measurements and features that are known to be relevant. Quantification of immunohistochemical staining (IHC) is one example where automated methods are already being incorporated with some success into clinical practice [2]. In general, image analysis can provide a more reproducible quantification of morphology of individual cells or relevant tissue components such as glands. Increasingly, deep learning based approaches replace traditional image analysis algorithms. By training complex computation models directly from data it is often possible to build algorithms that surpass the capabilities of traditional image analysis methods. Examples include the scoring of PD-L1 [3], quantification of immune infiltrates to predict outcomes in testicular tumours [4], detecting sentinel lymph node metastases [5] and superior prediction of colorectal cancer outcome compared to standard morphological assessment [6]. The interpretation of human *versus* deep learning studies in pathology can be affected by methodological considerations. For example, in the Bejnordi *et al* study, pathologists were required to review 129 consecutive cases in a limited time frame which does not reproduce typical pathology reporting (where cases are mixed and difficult cases are not subjected to a time limit) [7].

Machine learning holds the promise of providing us with capabilities that not only mimic but also enhance the visual analysis pathologists perform. Beck *et al* [8] used a range of features provided by a commercial image analysis platform to identify novel stromal features associated with survival in breast cancer. Deep learning-based artificial intelligence (AI) has identified novel markers of distant metastasis in colorectal cancer [9] and novel indications of morpho-molecular associations [10]. Correlating morphological patterns with genetic stratifiers using traditional machine learning or deep learning demonstrates substantial promise [11,12]. Ultimately, it may be possible to overcome limitations of existing categorical grading systems with such tools.

Image analysis is a complex task that may involve steps such as pre-processing, accurate delineation of objects of interest, or the measurement of certain shape or texture features. Pre-analytical considerations such as standardisation of immunohistochemical staining for reproducible results must be addressed, even for relatively simple tasks such as quantifying areas of tumour demonstrating positive staining for a cytokeratin marker using a colour threshold. Overview articles [13–15] provide an introduction to different image analysis techniques.

It is possible to utilise both commercially available dedicated image analysis programmes designed for digital pathology applications [Visiopharm, Definiens Tissue Studio, Indica Labs HALO (Corrales, New Mexico, USA)], or alternatively an open source software programme such as QuPath Open Source Software for Quantitative Pathology or ImageJ (Image Process and Analysis in Java) [16–18]. Deep learning methods are increasingly incorporated into commercial and open-source image analysis applications. Radiological imaging applications used to assess lung and liver nodules already have FDA clearance for clinical use [19]; additionally, equivalent performance to human experts has been demonstrated in the interpretation of optical coherence tomography scans [20].

Image analysis has the potential to identify, extract and quantify features in greater detail in comparison to pathologist assessment, which may produce improved prediction models [6] or perform tasks beyond manual capability such as generating tumour-infiltrating lymphocyte (TIL) maps [21] and glandular maps [22]. There is now much interest in using these technologies in clinical trials [23,24].

An NCRI CM-Path Quality Assurance in Clinical Trials Workshop was held on 21 March 2017 with representation from key members of industry, regulation and pathology. Regulation, training, oversight, laboratory processes and scoring/reporting were divided up into subgroup teams, which were presented for discussion by the QA panel. These topics are covered in separate articles [25,26]. The scoring/reporting discussions included the use and validation of digital pathology and image analysis technologies in clinical trials. Currently there are limited published examples of their use in clinical trial practice, and methodological standardisation with minimal standards for publication requires further development.

In this article, we provide an overview of the utility of these technologies in clinical trials and discuss potential applications, current challenges, limitations and remaining unanswered questions that require addressing prior to routine study adoption (Table 1).

Central review of pathology in clinical trials

The value of central review

The negative impact of inter-observer variation in pathologists assessing lymphoma prompted the adoption of central case review in the 1960s [27]. Widespread central pathology case review occurs in clinical

Table 1. Key issues that require consensus for the adoption of digital pathology and image analysis

- Standardisation of operating procedures and training in digital image production
- Adoption of digital pathology infrastructure for trials
- Resource allocation for digital central review
- Data governance and research approval
- Criteria for standards in digital pathology in pre-clinical, clinical and interventional studies
- Evidence base and performance criteria for image analysis algorithms
- In the UK, development of liaison between the Medicines and Healthcare products Regulatory Agency (MHRA), the British Standards Institution (BSI), the National Institute for Health and Care Excellence (NICE) and other regulatory bodies
- Clinical trial adoption of this technology is in its infancy, and although the potential applications are exciting to all stakeholders involved, a structured and integrative approach to effectively and safely incorporate developments into practice is required.

trials; and is particularly valuable where rare and morphologically challenging diagnostic entities occur. Here, reporting pathologist experience can substantially influence reporting [28].

Currently, most central reviews occur after implementing patient management decisions for quality control prior to publication, rather than in 'real time' for trial entry. Central review requires additional slides to be produced from tissue blocks, which risks exhausting tissue required for direct patient care. In the event of review slide loss, it may not be possible to produce facsimile slides from limited remaining tissue.

Shortage of skilled trials pathologists is becoming a key issue in the conduct of clinical trials within the UK [29] and digital pathology has the potential to ameliorate this issue by linking distant sites and expanding access to expert pathologists. Rapid dissemination of identical images to multiple centres allows simultaneous case review, reducing turnaround times and ensuring consensus opinion before therapeutic allocation. Case reclassification at the end of a study could indicate suboptimal patient treatment and negate the significance of investigational findings. Duplicate tissue sections have allowed simultaneous peripheral and central review reporting in a multinational interventional trial in nephroblastoma [30] to address reporting variance before therapeutic allocation. In the study conducted by Vujanić et al, 9 of 248 simultaneously reported cases underwent diagnostic change resulting in alternative post-operative therapy. Even minor error in diagnostic accuracy can affect the statistical significance of trial outcomes. Poor biomarker validation and positive publication bias can lead to expensive negative randomised control trials [31].

The central review of 552 radical prostatectomy specimens submitted for EORTC trial 22 911 [32] showed low concordance between local assessment and central review for certain key parameters (evaluation of extra-prostatic extension and surgical margin status). Central review can facilitate identification of discordant diagnostic parameters, but this may occur after recruitment of numerous patients.

Digital pathology and central pathology review

Digital pathology can permit simultaneous case review and allow rapid access to international experts in entities of interest. Reduction or removal of the need to physically transfer slides and tissue blocks, avoiding damage or loss, is a defining logistical advantage of the digital approach [33]. Although multinational commercial platform providers can overcome technical limitations provided by trials carried out in separate countries, differing requirements between nations regarding appropriate data governance could impede disseminated review of digital material. The European General Data Protection Regulations (GDPR) which came into force on 25 May 2018, replacing the Data Protection Act (1998), have strengthened and unified data protection for individuals within the European Union (EU), whilst addressing the export of personal data outside the EU [34].

The requirement to physically transfer slides may be retained by centralised genomic analysis and biobanking requirements, or by participating units lacking appropriate infrastructure to transfer slides for digitisation. Slide hosting repositories that can accommodate large slide images (typically 0.5-4Gb), satisfy data protection legislation, meet trial protocol requirements and ensure that patient-identifiable image-associated metadata have been removed, provide a considerable resource burden. A variety of proprietary file formats and platforms pose interoperability challenges. Clinical trial protocol specification of image formats is a requirement in the absence of widespread vendor-agnostic file formats.

Cost implications

Pathologists who have not transferred to digital pathology need to be trained in line with RCPath guidelines [35], and costing of appropriate pathologists and their training should be factored into the trial business plan. Additionally, construction of standardised operating procedures (SOPs) covering scanning equipment, database construction, anonymisation of images (where appropriate) and their transfer is required: potentially

with input from a pathology working group. Appropriate training for digital image reporting and slide scanning is required, which should include appropriate information governance. SOPs for equipment should define acceptable scanning platforms, required file formats and required scanning resolution. If scanning and reporting for the trial can be done by laboratories operating under ISO 15189: 2012 or working to good clinical practice (GCP) standards, this is preferable. Should scanners with appropriate regulatory approval (CE IVD or FDA) be available, these should be used in preference to those without such quality marks.

Digital images can form image libraries allowing streamlined research as new innovations become available. H&E stained slides and notably immunofluorescence-labelled slides degrade over time and require additional preparation (potentially including further slide production from tissue blocks).

Maintenance and sustainability of data storage are thus key requirements. Ensuring strict adherence to data governance frameworks and undertaking documentation that exceeds current regulatory requirements will hopefully ensure the availability of valuable material to future research projects.

Training

Pathologist training for participation in clinical trials often involves face-to-face training. Digital pathology can facilitate flexible distance learning incorporating a wider audience than traditional face-to-face learning. Digital training programmes potentially reduce the requirement for expert pathologists leading the study to deliver face-to-face guidance. The requirement of training may be to improve standardisation of a feature routinely evaluated in diagnostic work, or alternatively to evaluate a novel feature not routinely reported: this may require a more comprehensive educational program.

Investigators can be trained by remote use of annotated digital material, either as part of an interactive web seminar or as an online training programme undertaken at the convenience of the participating pathologist. Sample images for reporting can be incorporated as an additional quality assurance measure. This is particularly valuable where pathologists are required to report a novel parameter in relatively few patients recruited from a single centre. The advantage of using direct visual feedback guided software for a novel parameter has been demonstrated in the assessment of percentage of TILs in breast cancer, which predicts response to neoadjuvant therapy [36]. This is not a standard parameter for reporting in routine practice, and participating pathologists require training to report to an agreed standard. Improved concordance between pathologists reporting percentage of TILs was demonstrated after the use of training software utilising digital images of breast tumour tissue microarrays (TMAs) [37].

Digital pathology training software allows direct demonstration of compliance with clinical trial regulatory requirements and training logs, which can additionally be used by pathologists for revalidation and appraisal purposes. Clinical trials training applications can be reformatted to drive quality improvement in routine practice.

Image analysis

Analysis of IHC

There is substantial interest in the use of image analysis programmes to assist standardisation of reporting and introduction of novel parameters where accurate assessment by pathologists is unfeasible. Standardisation of ER and PR staining interpretation in breast cancer has received considerable interest as inter-observer variation can directly impact patient care. Although expert pathologist concordance in stain interpretation was greater in comparison to machine assessment in a large study of pooled TMA cores [38], equivalent concordance was seen for HER-2 assessment. Interestingly, a small number of extreme discrepancies were identified by digital image analysis: digital image analysis screening could be initially used to highlight isolated reporting discrepancies.

Automated image analysis also allows simultaneous scoring of ER and PR restricted to tumour-rich areas only, as reported in [39] and additional studies demonstrate non-inferiority to manual HER-2 scoring in breast cancer [40].

FDA and CE IVD-cleared algorithms exist for the assessment of a small number of pathological features, such as ER, PR, HER-2 and Ki-67 expression in breast cancer. Where image analysis algorithms are used in clinical trials, an appropriate evidence base for the algorithms must be available. A separately published validation dataset describing the performance of the algorithm in measuring the variable assessed would be regarded as a minimum requirement.

Pre-analytical features

Pre-analytical staining variation and background staining are key problems in image analysis development.

Considerable variation in immunohistochemical staining and standard H&E staining can exist between laboratories (although image analysis can control for variation in staining intensity) [41]. When constructing image analysis applications, demonstrating reproducibility in comparison to standard histopathological assessment is an important requirement. Algorithm – pathologist correlation is the most commonly used method: the use of multiple pathologists is always preferable to reflect routine practice. Clear descriptions must be provided of quality control measures and validation steps for every trial where image analysis is used. This should include a careful description of algorithm validation. Measures of reproducibility should be provided in publications such as pathologist-algorithm correlation and measures of inter-pathologist variability.

Regions of interest (ROI) selection methodology should be transparently described (if applicable) and whether analysis was on ROI, hot spots, WSI or was based on a pre-selected sample (e.g. TMA spot) in order for it to be reliable and reproducible. Selection of ROI or hot spots can be completely automated, completely manual, or a combination of both. The approaches are subject to alternative potential errors, which can impact on the study design. For example, acquiring data or applying scoring systems optimised for visual interpretation and performing an AI protocol without appropriate validation is inappropriate.

Assessing tumour percentage content

There is considerable inter-observer variability between pathologist assessment of percentage tumour cells within lung [42] and colorectal [43] cancer biopsies. Inaccurate estimations of tumour cellularity have the potential to result in inaccurate reporting of key actionable variants in genes such as *EGFR*, *RAS* and *BRAF*. Utilisation of the TissueMark™ (Philips Pathology, The Philips Centre, Guildford, Surrey, UK) platform to automatically annotate tumour boundaries with assessment of percentage tumour cells indicated superior performance compared to manual assessment [44]. Standardisation of tumour sampling allows more refined molecular categorisation of tumours and facilitates stratified therapeutic approaches.

Assessing novel features

The use of digital pathology to provide novel information from histological samples is an exciting area of development. Rapid quantification and high reproducibility can be achieved by using digital image analysis,

such as with localisation and quantification of immune cell infiltrates. The quantity and localisation of T cells is the underlying basis for the Immunoscore™ (Laboratory of Integrative Cancer Immunology INSERM, Paris, France) technique in colorectal adenocarcinoma: patients with a low CD3+ and CD8+ T cell density in the tumour centre and invasive margin are at an increased risk for disease relapse [45,46]. This marker could be used to stratify adjuvant therapy in prospective randomised control trials in Stage II disease, where the absolute benefit of adjuvant cytotoxic therapy is small [47]. Novel digital morphometric signatures previously not established in the pathology community can also be mined and linked to clinical outcome, as shown recently in the case of breast cancer [8] and stage II colorectal adenocarcinoma [9].

Machine learning methods

Machine learning methods can facilitate accurate quantitative assessment of digital images at a performance level potentially exceeding that of human observers. The identification of cancer cells, stromal tissue or inflammatory cells [48] typically requires accurate ground-truth training datasets and large numbers of cases to provide optimal automated assessment. Large openly available pooled digital datasets and the establishment of Grand Challenges for computational biologists provide a means to benchmark and evaluate image analysis algorithms. The CAMELYON challenge to identify lymph node metastases demonstrates the advantage of this approach [5,7]. Large-scale interventional studies provide a relatively accessible source of suitable cases for retrospective analysis, with clear management and outcome data. This allows assessment of multiple image features by multiple image analysis programmes with subsequent univariate and multivariate analysis to identify the most relevant novel predictive and prognostic parameters or correlation with molecular markers.

Retrospective analysis of 768 pre-treatment biopsies taken as part of a large RCT investigating neoadjuvant therapeutic approaches in breast cancer (neotAnGo) identified median lymphocyte density as being independently predictive of complete pathological response to therapy on multivariate analysis [36]. This study employed machine learning methods to classify cells as carcinoma, stroma or lymphocytes, with clear description of how quality metrics involved in analysis were determined, such as the use of automated ROI. Notably, the study was able to quickly evaluate initially promising parameters and demonstrate lack of a predictive utility in an adequately powered study. The

adherence to good scientific practice by providing detailed description of techniques used and making the source code and images utilised available allowed transparency and facilitated reproducibility. Establishing such clear and transparent standards in image research practice is to be commended: maintaining algorithms and source images open to scientific scrutiny as opposed to keeping methodology closed to safeguard commercial interests can be anticipated as a key issue for the developing Digital Image Analysis community.

The discussion of applying advanced data analysis technologies to clinical trial data is by no means limited to image analysis tasks. Scott Gottlieb, Commissioner of Food and Drugs, remarks [49] that 'AI holds enormous promise for the future of medicine, and we're actively developing a new regulatory framework to promote innovation in this space and support the use of AI-based technologies'. He clearly recognises the need for including advanced digital health tools into the drug development pipeline. The Precertification Pilot Programme (Pre-Cert) provides new regulatory guidelines for permitting the use of digital technologies that include machine learning.

Centralised image analysis

Central laboratory image analysis should be considered in study design due to advantages in standardisation of results. Although image analysis provides inherent reproducibility, inter-platform variation between individual centres is a potential source of bias. Construction of SOPs with designated acceptable platforms, image analysis algorithms and validation steps ensuring inter-site reproducibility would address this issue. Where companion diagnostic techniques reliant on specific staining platforms and accompanying image analysis platforms are required, adherence to manufacturer guidance would require clear documentation. Provision of unstained slides for immunohistochemical analysis and concurrent image analysis may be required as an additional quality assurance measure to account for inter-laboratory staining variation. Ensuring both reproducible staining quality and image analysis between multiple laboratories could prevent centres participating in studies. Although centralised image analysis overcomes these issues, logistical constraints inherent to traditional studies undertaking centralised review remain. If many sites are conducting image analysis, repeatability analysis should be undertaken because of potential inter-laboratory reproducibility concerns [50].

Regulation and guidelines

Slide scanners and image analysis algorithms when intended for medical use (including diagnosis) are classed as medical devices [51]. The regulatory requirements for clinical performance studies of in vitro diagnostics (IVDs)/use of IVDs in clinical trials of medicinal products remains a subject of much debate and are fast evolving. The US FDA is testing a new Pre-Cert model [52] with the intention of demonstrating by premarket review and excellence appraisal the same quality of information as a traditional approach to ensure safety and effectiveness standards are met. This novel model for the pre-market review of digital health tools as medical devices includes implementing a new approach to the review of AI tools.

In the UK, the Medicines and Healthcare products Regulatory Agency (MHRA) regulates medicines, medical devices and blood components for transfusion in the UK. To ensure GCP compliance, the MHRA carries out inspections of trial sites (including laboratories) mostly based on a risk assessment score [53]. How oversight of clinical trials utilising these technologies as medical devices will be regulated or inspected is a subject of debate. Constantly evolving AI applications pose additional challenges, but there are regulatory cleared AI algorithms in clinical practice.

To the best of our knowledge, there are no guidelines covering the use of digital pathology or image analysis in clinical trials. The low yield of clinically actionable biomarkers from a large volume of research studies with considerable resource outlay led to the construction of the REMARK recommendations for biomarker studies in 2005 [54,55]. Studies that fail to meet these consensus requirements for a reputable biomarker study are increasingly excluded from systematic reviews of the evidence base for diagnostic approaches. A recent systematic review evaluating prognostic biomarker use in oesophageal adenocarcinoma demonstrated the effect of applying REMARK guidelines as inclusion criteria [56]. Only 36 out of 214 eligible studies (17%) were included.

Constructing similar robust recommendations for digital pathology applications would minimise the extensive waste of resources encountered in prior biomarker studies. Failure of reproducibility driven by a lack of reporting of experimental detail has been described as a major factor in the inefficient development and adoption of clinically relevant biomarkers [57]. Diligent reporting of the processes by which digital pathology applications were validated is essential to avoid repeating similar research practice failures. Direct adoption of existing biomarker guidelines could

be considered: parameters assessed by digital pathology certainly meet the criteria of a biomarker as a 'defined characteristic that is measured as an indicator of normal biological processes, pathogenic processes or responses to an exposure or intervention, including therapeutic interventions' [58].

However, the application of guidelines intended for use in biospecimen-derived biomarkers has been considered unsatisfactory when applied to imaging techniques, leading to the construction of separate guidelines for 'imaging biomarkers' [59]. Synthesis of the most appropriate approaches from both biospecimen-derived and imaging biomarkers would be logical for digital pathology applications.

The Health Research Authority (HRA) have recently updated their guidance on approval of new medical devices including software applications. The use of diagnostic platforms would require dedicated medical device clinical evaluation studies to obtain HRA approval, which would entail using systems in parallel with standard practice using light microscopy. Establishing the equivalence/non-inferiority of a digital workstream for standard diagnostic practice is being undertaken in several UK centres with supporting guidance from the Royal College of Pathologists. The adoption of clinical studies using digital platforms would require the use of flagged research ethics committees to provide expertise in digital pathology. Lack of appropriate specialist availability to participate in the ethical approval process could hinder digital pathology development. Once standard diagnostic practice in clinical trials can be facilitated by digital platforms, approval of digital image analysis applications would be feasible. It is anticipated from joint HRA/MHRA guidance on software development that IVD performance evaluations would be undertaken in line with existing biochemical biomarker assays. Both industry and academic sponsors may find provision for clinical trial insurance to indemnify against diagnostic error induced by digital image analysis platforms challenging. It is uncertain whether actuarial assessment can be based on digital image use in a healthcare system where digital image analysis is more widely adopted, such as the USA. Combined assessment approaches of Pathologist plus machine may provide complex indemnity and regulatory challenges: appropriate mentoring from radiologists and industry specialists involved in digital radiology would be recommended to overcome these issues.

Conclusions

In this paper, we describe the use and potential future applications of digital pathology and image analysis technologies in clinical trials. The technologies can play a role in central review, training and image analysis and can be used to improve assessment of standard pathological features or extract novel insights. The current landscape sees the technologies most commonly being used where feasibility is already demonstrated such as central review or quality/efficiency benefits are obvious such as quantification of immune infiltrates in immune-oncology trials.

In order to realise the potential of these technologies and dramatically improve the quality of the pathology input to clinical trials, the digital pathology community together with regulators and industry must establish practice standards for clinical trial use. Linking with international centres of excellence and involvement of other specialists such as software engineers, information network specialists is vital.

Table 1 lists the key issues that require consensus for the adoption of digital pathology and image analysis in clinical trials, and CM-Path proposes to address these issues in a future multidisciplinary workshop.

Acknowledgements

Barts Cancer Institute, Barts and the London School of Medicine and Dentistry, London, UK. Maggie Cheang, Institute of Cancer Research Clinical Trials and Statistics Unit, The Institute of Cancer Research, 15 Cotswold Road, Surrey SM2 5NG, UK. Manuel Rodriguez-Justo, Department of Gastroenterology, University College Hospitals London, London, UK. Gabrielle Rees, Department of Cellular Pathology, John Radcliffe Hospital, Oxford, UK. Manuel Salto-Tellez, Northern Ireland Molecular Pathology Laboratory, Centre for Cancer Research and Cell Biology, Queen's University Belfast, Belfast, UK. Nicholas P West, Pathology and Tumour Biology, Leeds Institute of Cancer and Pathology, University of Leeds, Leeds, UK. Ilaria Mirabile, Experimental Cancer Medicine Centres (ECMCs) Network, London, UK. Emily Howlett, Precision Medicine Team, Cancer Research UK, London, UK. Laura Stevenson, Institute of Cancer Research Clinical Trials and Statistics Unit, The Institute of Cancer Research, 15 Cotswold Road, Surrey SM2 5NG, UK. Maria da Silva, Sectra UK. Newton ACS Wong, Department of Cellular Pathology, Southmead Hospital, Bristol, UK. Sidonie Hartridge-Lambert, Bristol-Myers Squibb, London, UK. Joseph M Beecham, Nanostring, Washington, USA. Stephanie Traub, Centre for Drug Development, Cancer Research UK, London, UK. Sidath Katugampola, Centre for Drug Development, Cancer Research UK, London, UK. Sarah Blagden, Department of Oncology, University of Oxford, Oxford, UK. James Morden, Institute of Cancer Research Clinical Trials and Statistics Unit, The Institute of Cancer Research, 15 Cotswold Road, Surrey SM2 5NG, UK.

Ethics approval was not required as no study subjects were used, and no patient data was accessed. This article is the opinion expressed by all authors regarding the development of Digital Pathology Technology for Clinical Trial Use.

Author contributions statement

RP, KO, MR, DS and CV conceived the outline of the paper and areas for discussion; additional oversight/commentary was provided by the CM-Path QA Panel and HP, JR, NR. All authors were involved in writing the paper and had final approval of the submitted and published versions. This manuscript has been read and approved by all the authors, the requirements for authorship have been met and each author believes that the manuscript represents honest work.

References

1. Williams BJ, Bottoms D, Treanor D. Future-proofing pathology: the case for clinical adoption of digital pathology. *J Clin Pathol* 2017; **70**: 1010–1018.

2. Koelzer VH, Sirinukunwattana K, Rittscher J, *et al*. Precision immunoprofiling by image analysis and artificial intelligence. *Virchows Arch* 2018. https://doi.org/10.1007/s00428-018-2485-z.

3. Kapil A, Meier A, Zuraw A, *et al*. Deep semi supervised generative learning for automated PD-L1 tumor cell scoring on NSCLC tissue needle biopsies. *Sci Rep* 2018; **8**: 17343.

4. Linder N, Taylor JC, Colling R, *et al*. Deep learning for detecting tumour-infiltrating lymphocytes in testicular germ cell tumours. *J Clin Pathol* 2019; **72**: 157–164.

5. Ehteshami Bejnordi B, Veta M, van Diest PJ, *et al*. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* 2017; **318**: 2199–2210.

6. Bychkov D, Linder N, Turkki R, *et al*. Deep learning based tissue analysis predicts outcome in colorectal cancer. *Sci Rep* 2018; **8**: 3395.

7. Golden JA. Deep learning algorithms for detection of lymph node metastases from breast cancer. Helping artificial intelligence be seen. *JAMA* 2017; **318**: 2184–2186.

8. Beck AH, Sangoi AR, Leung S, *et al*. Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. *Sci Transl Med* 2011; **3**: 108ra113.

9. Sirinukunwattana K, Snead D, Epstein D, *et al*. Novel digital signatures of tissue phenotypes for predicting distant metastasis in colorectal cancer. *Sci Rep* 2018; **8**: 1–13.

10. Coudray N, Ocampo PS, Sakellaropoulos T, *et al*. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat Med* 2018; **24**: 1559–1567.

11. Schaumberg AJ, Rubin MA, Fuchs TJ. H&E-stained whole slide image deep learning predicts SPOP mutation state in prostate cancer. *BioRxiv* 2018: 064279. https://doi.org/10.1101/064279.

12. Mobadersany P, Yousefi S, Amgad M, *et al*. Predicting cancer outcomes from histology and genomics using convolutional networks. *Proc Natl Acad Sci U S A* 2018; **115**: E2970–E2979.

13. Nketia TA, Sailem H, Rohde G, *et al*. Analysis of live cell images: methods, tools and opportunities. *Methods* 2017; **115**: 65–79.

14. Hamilton PW, Bankhead P, Wang Y, *et al*. Digital pathology and image analysis in tissue biomarker research. *Methods* 2014; **70**: 59–73.

15. Robertson S, Azizpour H, Smith K, *et al*. Digital image analysis in breast pathology – from image processing techniques to artificial intelligence. *Transl Res* 2018; **194**: 19–35.

16. Bankhead P, Loughrey MB, Fernández JA, *et al*. QuPath: open source software for digital pathology image analysis. *Sci Rep* 2017; **7**: 16878.

17. Schindelin J, Rueden CT, Hiner MC, *et al*. The ImageJ ecosystem: an open platform for biomedical image analysis. *Mol Reprod Dev* 2015; **82**: 518–529.

18. Schneider CA, Rasband WS, Eliceiri KW. NIH image to ImageJ: 25 years of image analysis. *Nat Methods* 2012; **9**: 671–675.

19. Arterys: The World's First Medical Imaging Cloud Platform. [Accessed 27 December 2018]. Available from: https://www.arterys.com/

20. Wise J. AI system interprets eye scans as accurately as top specialists. *BMJ* 2018; **362**: K3484.

21. Saltz J, Gupta R, Hou L, *et al*. Spatial organization and molecular correlation of tumor-infiltrating lymphocytes using deep learning on pathology images. *Cell Rep* 2018; **23**: 181–193.

22. Graham S, Chen H, Gamper J, *et al*. MILD-net: minimal information loss dilated network for gland instance segmentation in colon histology images. *Med Image Anal* 2019; **52**: 199–211.

23. Barisoni L, Hodgin JB. Digital pathology in nephrology clinical trials, research and pathology practice. *Curr Opin Nephrol Hypertens* 2017; **26**: 450–459.

24. Butterfield LH, Disis ML, Fox BA, *et al*. SITC 2018 workshop report: Immuno-oncology biomarkers: state of the art. *J Immunother Cancer* 2018; **6**: 138.

25. Robinson M, James J, Thomas G, *et al*. Quality assurance guidance for scoring and reporting for pathologists and laboratories undertaking clinical trial work. *J Pathol Clin Res* 2019; **5**: 91–99

26. Rees G, Salto-Tellez M, Lee JL, *et al*. Training and accreditation standards for pathologists undertaking clinical trial work. *J Pathol Clin Res* 2019; **5**: 100–107

27. Wolf BC, Gilchrist KW, Mann RB, *et al*. Evaluation of pathology review of malignant lymphomas and Hodgkin's disease in cooperative clinical trials. The Eastern Cooperative Oncology Group experience. *Cancer* 1988; **62**: 1301–1305.

28. Boyett JM, Yates AJ, Burger PC, *et al*. The influence of central review on outcome associations in childhood malignant gliomas: results from the CCG-945 experience. *Neuro Oncol* 2003; **5**: 197–207.

29. Bainbridge S, Cake R, Meredith M, *et al*. Testing times to come? An evaluation of digital pathology capacity across the UK. Cancer Research UK. Published November 2016. [Accessed 2 December 2019]. Available from: https://www.cancerresearchuk.org./sites/default/files/testing_times_to_come_nov_16_cruk.pdf

30. Vujanić GM, Sandstedt B, Kelsey A, *et al*. Central pathology review in multicenter trials and studies: lessons from the nephroblastoma trials. *Cancer* 2009; **115**: 1977–1983.

31. Kleikers PW, Hooijmans C, Gob E, *et al*. A combined pre-clinical meta-analysis and randomized confirmatory trial approach to improve data validity for therapeutic target validation. *Sci Rep* 2015; **5**: 13428.

32. van der Kwast TH, Collette L, Van Poppel H, *et al*. Impact of pathology review of stage and margin status of radical prostatectomy specimens (EORTC trial 22911). *Virchows Arch* 2006; **449**: 428–434.

33. Mroz P, Parwani AV, Kulesza P. Central pathology review for phase III clinical trials: the enabling effect of virtual microscopy. *Arch Pathol Lab Med* 2013; **137**: 492–495.

34. The EU General Data Protection Regulation. [Accessed 2 January 2019]. Available from: https://www.eugdpr.org/

35. Cross S, Furness P, Igali L, *et al*. Best Practice Recommendations for Implementing Digital Pathology. Royal College of Pathologists Guidelines; 2018. [Accessed 2 December 2019]. Available from: https://www.rcpath.org/uploads/assets/uploaded/d6b14330-a8b9-4f5e-bbe443f0d56de24a.pdf

36. Raza Ali H, Dariush A, Provenzano E, *et al*. Computational pathology of pre-treatment biopsies identifies lymphocyte density as a predictor of response to neoadjuvant chemotherapy in breast cancer. *Breast Cancer Res* 2016; **18**: 21.

37. Denkert C, Wienert S, Poterie A, *et al*. Standardized evaluation of tumor-infiltrating lymphocytes in breast cancer: results of the ring studies of the international immuno-oncology biomarker working group. *Mod Pathol* 2016; **29**: 1155–1164.

38. Howat WJ, Blows FM, Provenzano E, *et al*. Performance of automated scoring of ER, PR, HER2, CK5/6 and EGFR in breast cancer tissue microarrays in the Breast Cancer Association Consortium. *J Pathol Clin Res* 2015; **1**: 18–32.

39. Trahearn N, Tsang YW, Cree IA, *et al*. Simultaneous automatic scoring and co-registration of hormone receptors in tumour areas in whole slide images of breast cancer tissue slides. *Cytometry A* 2017; **91**: 585–594.

40. Qaiser T, Mukherjee A, Reddy PBC, *et al*. HER2 challenge contest: a detailed assessment of automated HER2 scoring algorithms in whole slide images of breast cancer tissues. *Histopathology* 2018; **72**: 227–238.

41. Khan AM, Rajpoot N, Treanor D, *et al*. A nonlinear mapping approach to stain normalization in digital histopathology images using image-specific color deconvolution. *IEEE Trans Biomed Eng* 2014; **61**: 1729–1738.

42. Smits AJJ, Kummer JA, de Bruin PC, *et al*. The estimation of tumor cell percentage for molecular testing by pathologists is not accurate. *Mod Pathol* 2014; **27**: 168–174.

43. Viray H, Li K, Long TA, *et al*. A prospective, multi-institutional diagnostic trial to determine pathologist accuracy in estimation of percentage of malignant cells. *Arch Pathol Lab Med* 2013; **137**: 1545–1549.

44. Hamilton PW, Wang Y, Boyd C, *et al*. Automated tumor analysis for molecular profiling in lung cancer. *Oncotarget* 2015; **6**: 27938.

45. Pages F, Berger A, Camus M, *et al*. Effector memory T cells, early metastasis, and survival in colorectal cancer. *N Engl J Med* 2005; **353**: 2654–2666.

46. Pagès F, Kirilovsky A, Mlecnik B, *et al*. In situ cytotoxic and memory T cells predict outcome in patients with early-stage colorectal cancer. *J Clin Oncol* 2009; **27**: 5944–5951.

47. Nordlinger B, Beretta GD, Mosconi S, *et al*. Early colon cancer: ESMO clinical practice guidelines for diagnosis,treatment and follow-up. *Ann Oncol* 2013; **24**: vi64–vi72.

48. Sirinukunwattana K, Raza SEA, Tsang Y, *et al*. Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *IEEE Trans Med Imaging* 2016; **35**: 1196–1206.

49. Gottlieb S. Transforming FDA's Approach to Digital Health. [Accessed 22 January 2019]. Available from: https://www.fda.gov/NewsEvents/Speeches/ucm605697.htm.

50. Rimm DL, Leung SCY, McShane LM, *et al*. An international multicenter study to evaluate reproducibility of automated scoring for assessment of Ki67 in breast cancer. *Mod Pathol* 2019; **32**: 59–69.

51. UK Government website. Medical Devices: Software Applications (apps). [Accessed 22 January 2019]. Available from: https://www.gov.uk/government/publications/medical-devices-software-applications-apps

52. U.S. FDA Website. Precertification (Pre-Cert) Pilot Program: Milestones and Next Steps. [Accessed 22 December 2019] Available from: https://www.fda.gov/MedicalDevices/DigitalHealth/Digital HealthPreCertProgram/ucm584020.htm

53. UK Government Website. Good Clinical Practice for Clinical Trials. [Accessed 22 January 2019]. Available from: https://www.gov. uk/guidance/good-clinical-practice-for-clinical-trials

54. McShane LM, Altman DG, Sauerbrei W, *et al*. REporting recommendations for tumor MARKer prognostic studies (REMARK). *Nat Clin Pract Urol* 2005; **2**: 416–422.

55. McShane LM, Sauerbrei W, Taube SE. Reporting recommendations for tumor marker prognostic studies (REMARK): explanation and elaboration. *PLoS Med* 2012; **9**: e1001216.

56. McCormick Matthews LH, Noble F, Tod J, *et al*. Systematic review and meta-analysis of immunohistochemical prognostic biomarkers in resected oesophageal adenocarcinoma. *The Br J Cancer* 2015; **113**: 107–118.

57. Bustin SA, Huggett JF. Reproducibility of biomedical research – the importance of editorial vigilance. *Biomol Detect Quantif* 2017; **11**: 1–3.

58. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther* 2001; **69**: 89–95.

59. O'Connor JP, Aboagye EO, Adams JE, *et al*. Imaging biomarker roadmap for cancer studies. *Nat Rev Clin Oncol* 2017; **14**: 169–186.