PERSPECTIVE

Quality assurance guidance for scoring and reporting for pathologists and laboratories undertaking clinical trial work

Max Robinson[1]* , Jacqueline James[2], Gareth Thomas[3], Nicholas West[4] , Louise Jones[5], Jessica Lee[6], Karin Oien[7], Alex Freeman[8], Clare Craig[9], Philip Sloan[10], Philip Elliot[5], Maggie Cheang[11], Manuel Rodriguez-Justo[12], and Clare Verrill,[13]
on behalf of the UK National Cancer Research Institute (NCRI) Cellular-Molecular Pathology (CM-Path) clinical trials working group[†]

[1]*Centre for Oral Health Research, Newcastle University, Newcastle upon Tyne, UK*
[2]*School of Medicine, Dentistry and Biomedical Sciences, Centre for Cancer Research and Cell Biology, Institute for Health Sciences, Queen's University Belfast, Belfast, UK*
[3]*Faculty of Medicine Cancer Sciences Unit, Southampton University, Southampton, UK*
[4]*Pathology and Tumour Biology, Leeds Institute of Cancer and Pathology, University of Leeds, Leeds, UK*
[5]*Centre for Tumour Biology, Barts Cancer Institute, Barts and the London School of Medicine and Dentistry, London, UK*
[6]*Strategy and Initiatives, National Cancer Research Institute, London, UK*
[7]*Institute of Cancer Sciences – Pathology, University of Glasgow, Glasgow, UK*
[8]*Department of Pathology, University College London Hospitals NHS Foundation Trust, London, UK*
[9]*Genomics England, London, UK*
[10]*Department of Cellular Pathology, Newcastle upon Tyne Hospitals NHS Trust, Newcastle upon Tyne, UK*
[11]*Institute of Cancer Research Clinical Trials and Statistics Unit, The Institute of Cancer Research, Surrey, UK*
[12]*Department of Gastroenterology, University College Hospitals London, London, UK*
[13]*Nuffield Department of Surgical Sciences, University of Oxford, and Oxford NIHR Biomedical Research Centre, Oxford, UK*

*Correspondence to: M Robinson, Centre for Oral Health Research, Newcastle University, Framlington Place, Newcastle-upon-Tyne NE2 4BW, UK. E-mail: max.robinson@newcastle.ac.uk

†The members of the NCRI CM-Path Quality Assurance Panel are listed in the Acknowledgements section.

Abstract

While pathologists have always played a pivotal role in clinical trials ensuring accurate diagnosis and staging, pathology data from prognostic and predictive tests are increasingly being used to enrol, stratify and randomise patients to experimental treatments. The use of pathological parameters as primary and secondary outcome measures, either as standalone classifiers or in combination with clinical data, is also becoming more common. Moreover, reporting of estimates of residual disease, termed 'pathological complete response', have been incorporated into neoadjuvant clinical trials. Pathologists have the expertise to deliver this essential information and they also understand the requirements and limitations of laboratory testing. Quality assurance of pathology-derived data builds confidence around trial-specific findings and is necessarily focused on the reproducibility of pathological data, including 'estimates of uncertainty of measurement', emphasising the importance of pathologist education, training, calibration and demonstration of satisfactory inter-observer agreement. There are also opportunities to validate objective image analysis tools alongside conventional histological assessments. The ever-expanding portfolio of clinical trials will demand more pathologist engagement to deliver the reliable evidence-base required for new treatments. We provide guidance for quality assurance of pathology scoring and reporting in clinical trials.

Keywords: pathology; clinical trials; quality assurance; immunohistochemistry; biomarkers

Introduction

A Quality Assurance in Clinical Trials Workshop was convened by the UK National Cancer Research Institute Cellular-Molecular Pathology Initiative (NCRI CM-Path). The workshop was held on 21 March 2017 with representation from the UK Medicines and Healthcare products Regulatory Agency

(MHRA), industry and pathology. Four subgroups were formed to tackle issues around regulation: training: trial oversight: and scoring and reporting. The subject areas were researched by the subgroup members prior to the workshop and presented by the subgroup leads for general discussion by the workshop participants. This article provides an overview of the discussions around scoring and reporting of pathology parameters in clinical trials. Discussions around the benefits and challenges in using digital pathology and image analysis were wide-ranging and will therefore be presented in a separate article. The workshop formulated 'practice points' to help pathologists navigate the development, set up and delivery of clinical trials and these are placed in *italics* at the end of each section and compiled in Table 1.

Scoring and reporting of pathology parameters is an interpretive skill; consequently there is an acknowledged subjective variation in clinical practice from pathologist to pathologist, even within the same centre. For clinical trials, the challenge is to control for variations in individual pathologist practice to limit the introduction of bias and errors. Common pathology parameters used in clinical trials include: diagnosis; grading; staging; biomarker scoring; pathological complete response (pCR); resection margins and recurrence.

Diagnosis, grading and staging are routinely carried out in clinical practice following well-established guidelines (e.g. Royal College of Pathologists, College of American Pathologists, Union for International Cancer Control, American Joint Committee on Cancer). The information can be collected from participating sites using a carefully designed 'pathology case report form' (Pathology CRF). Central review of at least a proportion of specimens is desirable for quality assurance purposes, but the extent of review will depend on the aims and objectives of the trial, diagnostic confidence, reproducibility of grading schemes and the ease of application of staging systems.

Immunohistochemistry, *in situ* hybridisation or molecular profiling of tumours may be used for trial entry, patient stratification and randomisation into treatment arms. If a trial is multicentre, it must be considered whether testing will be carried out at the participating sites or co-ordinated by a central laboratory. Generally, it is preferable that biomarker testing is carried out at the participating sites as they benefit from proximity to patient care and this mitigates against delays in the patient pathway. In the case where laboratory tests are not generally available or are specialised tests developed for the trial, then a central laboratory with the requisite expertise and appropriate

Table 1. Summary of scoring and reporting practice points

- Scoring and reporting should be carried out blinded to treatment allocation and clinical outcomes.
- The effect of diagnostic drift and chronological bias should be considered throughout the trial, in particular if scoring and reporting is carried out over several years and/or if new reporting guidelines are published.
- Double reporting and/or central review should be considered to increase confidence in the pathology data.
- The formation of a 'Pathology Working Group' should be considered to oversee the pathology aspects of a study.
- Training requirements for pathologists participating in clinical trials need to be addressed and will depend on whether the parameter is an established clinical test or a novel biomarker.
- Auto-stainers should be used for immunohistochemistry in preference to manual staining.
- Staining should be carried out in an ISO15189:2012-accredited laboratory or laboratory working to GCP standards and is mandatory for primary or secondary endpoints of CTIMP clinical trials.
- Measures to assure confidence in the reproducibility of pathologist scoring should be considered, for example, by reporting levels of inter and intra-observer agreement.
- The pathologists in the trial should agree what is considered background staining and thus filtered out of pathologist scoring and this should be kept under review during the trial.
- Companion and complementary diagnostic tests should be used in preference to laboratory-devised assays. Manufacturers' instructions must be followed to produce a reliable result.
- Estimates of uncertainty, defined in ISO 15189:2012, should be made for clinical trial tests where possible.
- The requirement for pathologist calibration should be considered prior to the trial opening.
- If an external quality assurance scheme (e.g. UK NEQAS ICC & ISH) is available for the trial test then trial laboratories should participate.
- Clear definitions of reporting of pathological complete response and SOPs for standardised assessment need to be agreed and established prior to the trial opening.
- Pathologists should be actively involved in developing trial specific bio-resources.
- Pathologists' contributions to clinical trials are wide ranging and should be recognised and appropriately funded.
- Early pathologist engagement and input into trial design is essential.

level of accreditation is necessary to run the trial effectively. It is important to consider realistic turnaround times when considering central laboratory testing and to manage the expectations of the participating sites and trial co-ordinators. Trial sample logistics, laboratory capacity and likelihood of repeat tests need to be considered. Whether tests are delivered at participating sites or a central laboratory, quality assurance measures need to be considered and implemented.

It is well known that the results of laboratory tests are influenced by a myriad of variables (Table 2) [1]. Pre-analytical variables are those associated with preservation and preparation of tissue prior to testing. Analytical variables are specific to the test carried out. Post-analytical variables, particularly associated with

Table 2. Pathology variables that may influence pathology scoring and reporting of immunohistochemistry and *in situ* hybridisation [1]

Variable	Description	Examples
Pre-analytical	Tissue preservation	Time to fixation, warm and cold ischaemia times
		Type of fixative, concentration, pH, temperature and duration
		Size of sample versus volume of fixative
	Tissue processing	Dehydration and clearing schedules
	Section preparation	Section thickness, temperature and duration of section drying
		Fresh cut sections versus stored sections versus paraffin wax 'dipped' sections
	Block storage	Temperature and duration of block storage
Analytical	Laboratory test	Type of proprietary auto-stainer
		Tissue conditioning, unmasking reagents, pH, temperature and incubation time
		Blocking reagent
		Primary antibody, concentration, incubation time
		Detection system
Post-analytical	Pathologist	Experience
		Specialist reporting
		Biomarker scoring training and accreditation

immunohistochemistry, are a consequence of subjective assessment by the pathologist, which is at best semi-quantitative, but variability can be estimated by measuring the degree of inter-observer and intra-observer agreement. Molecular testing of tissue homogenates using 'grind and bind' techniques benefit from quantitative results, but each specific assay has characteristic technical challenges and they lack the morpho-molecular features of tests carried out on intact tissue sections [2].

Pathology reporting in clinical trials

Ideally, assessment of pathology-based clinical trial endpoints should be masked or blinded to randomisation and clinical outcomes. For exploratory outputs, a staged approach may be more appropriate, such as that proposed by the Society for Toxicological Pathology [3]. The first stage is an un-blinded comparison of treated and control specimens to identify consistent changes and to develop scoring criteria. This strategy facilitates identification of subtle, treatment related findings that can be consistently differentiated from

those that occur in controls [4]. An independent pathology peer review with targeted blinding may also help minimise bias [5].

Scoring and reporting should be carried out blinded to treatment allocation and clinical outcomes.

Diagnostic drift and chronological bias

Diagnostic drift and chronological bias should be considered when setting up a clinical trial. Diagnostic drift is a gradual change in nomenclature, grading of lesions, or scoring of a biomarker within a single study over time. It is a source of inconsistency that can negatively affect detection of treatment-related changes or the determination of 'no-effect' levels [3]. Chronological bias is defined as the evolutionary process of a grading system, whereby more sensitive and specific criteria for grade assignment are clarified, learned and disseminated. As pathologists gain experience, subtleties of the evolving system are applied to interpretation of tissue sections [6]. This is also a source of inconsistency that can negatively affect detection of treatment-related changes or the determination of the 'no-effect' levels [4]. Plans for monitoring diagnostic drift and chronological bias should be considered for trials projected to recruit over several years. Comparison of pathology data at planned intervals throughout the life of the study, for example at the beginning, middle and end of the study, is recommended.

The effect of diagnostic drift and chronological bias should be considered throughout the trial, in particular if scoring and reporting is carried out over several years and/or if new reporting guidelines are published.

Double reporting and central review

Double reporting to verify pathological parameters such as diagnosis, grade, stage and biomarker scores is recommended, but may not be feasible in real time. Pathology review of cases should be considered as part of the trial design. The extent of the review process will depend on the number of cases, the complexity of the scoring, and the contribution of pathology parameters to the trial endpoints. Statistical advice should be sought when designing the review strategy. Central review by a lead pathologist or a group of pathologists should be completed before final data analysis and trial publication. For example, Speight *et al* (2015) described a system for review of cases from patients with oral epithelial dysplasia that involved four pathologists with adjudication of disagreements towards a consensus diagnosis as the gold standard for clinical

trials [7]. Central pathology review may be 'blinded', meaning independent assessment without knowledge of the participating site data or 'un-blinded' where the original pathology report or pathology CRF is available and data are simply checked for any disagreements.

Double reporting and/or central review should be considered to increase confidence in the pathology data.

Pathology working groups

Pathology working groups have been used in some trials to good effect. For example, ProtecT (Prostate Testing for Cancer and Treatment), a randomised controlled trial comparing active monitoring, radiotherapy and radical prostatectomy in patients with localised prostate cancer, employed a Pathology Working Group that oversaw the pathology aspects of the trial and published impactful pathology-specific papers from the trial [8,9]. Such groups enhance the quality of a study by actively engaging pathologists in the development of trial protocols, providing training, formulating pathology standard operating procedures (SOPs), conducting central review, resolving discrepancies and disseminating the trial findings to the pathology community.

The formation of a 'Pathology Working Group' should be considered to oversee the pathology aspects of a study.

Providing quality assured pathology tests

To ensure that pathology laboratory tests are scored and reported in a consistent manner it is recommended that a systematic process is established comprising, in sequence, education, training and calibration.

Education and training

Pathologists are most likely to be asked to provide immunohistochemistry-based tests for clinical trials. Such tests may form part of the recommended datasets for the histopathological reporting of cancer, in which case the expertise is likely to be already established, for example the scoring of oestrogen and HER2 receptors in breast cancer [10,11]. Furthermore, they are likely to be supported by external quality assurance programmes (Table 3) [12]. For novel biomarkers, where expertise is not established, it is essential that participating pathologists understand the clinical context of the test and the role of the test in the trial;

Table 3. United Kingdom National External Quality Assessment Service for immunohistochemistry and *in situ* hybridisation (UK NEQAS ICC & ISH) quality assurance modules for immunocytochemistry [12]

Disease	Test	Drug
Breast cancer	Oestrogen receptors	Tamoxifen
	Progesterone receptors	Aromatase inhibitors
	HER2	Trastuzumab
Gastric cancer	HER2	Trastuzumab
Alimentary tract – Lynch syndrome	MLH1, PMS2 MSH2, MSH6	Not applicable
Alimentary tract – Gastrointestinal stromal tumour (GIST)	CD117 (c-Kit), DOG-1	Imatinib
Non-small cell lung cancer	ALK	Crizotinib
	PD-L1 (pilot scheme)	Nivolumab Pembrolizamab

screening, recruitment, stratification or randomisation. Knowledge of the performance characteristics of the test is essential in order to recognise sub-optimal tests. For example, assessment of external analyte controls (same-slide based proprietary cell line controls or characterised tissue controls) and review of expected staining of internal controls are required before scoring the tumour. Trial-specific training may be delivered by face-to-face site meetings, establishing trial-specific pathology working groups or using on-line training modules [13]. Such meetings can also be used to collect feedback from the participating pathologists on the SOPs associated with tissue collection, processing, staining and scoring. In some circumstances, particularly with more complex SOPs, it is best practise to run trial-specific protocols, prior to the trial opening, to identify any operational problems.

Training requirements for pathologists participating in clinical trials need to be addressed and will depend on whether the parameter is an established clinical test or a novel biomarker.

Immunohistochemistry

Guidance on validation of immunohistochemistry assays in a clinical service is already available [14]. This guidance can be followed when incorporating immunohistochemistry into a clinical trial. The assay must be assessed for accuracy and analytical sensitivity/specificity. Once validation is completed, the assay should be regularly monitored and performance against known external positive controls assessed [14]. In the UK, in the setting of a clinical trial, if immunohistochemistry forms part of the primary or secondary outcomes, then this must be performed to a standard to satisfy the MHRA, who inspect laboratories

undertaking laboratory work for Clinical Trial of an Investigational Medicinal Product (CTIMP) trials. Immunohistochemistry can be performed manually or on a proprietary automated stainer (auto-stainer). The latter produce highly reliable tests and should be used in preference to manual staining. Ideally, clinical trial tests should be carried out in an ISO15189: 2012-accredited laboratory or a laboratory working to Good Clinical Practice (GCP) standards. Such stringency may not be required for exploratory tests, where biomarker development is the aim.

Auto-stainers should be used for immunohistochemistry in preference to manual staining.

Staining should be carried out in an ISO15189: 2012-accredited laboratory or laboratory working to GCP standards and is mandatory for primary or secondary endpoints of CTIMP clinical trials.

Scoring of immunohistochemistry

Numerous scoring methods have been devised for the assessment of immunohistochemical staining. The majority are based on the assessment of two parameters: intensity of the staining and proportion of the tumour that is stained (Table 4) [15]. Occasionally the scoring system includes an assessment of the tumour and the immune cells; an example is PD-L1 testing in non-small cell lung cancer [16]. Scoring systems associated with recognised predictive and prognostic markers should be scored according to guidelines [10,11,17]. For exploratory tests, researchers should investigate the literature and implement validated scoring systems or devise a scoring system to adequately capture the data and to facilitate the development of clinically relevant 'cut-offs'. Tumour heterogeneity is recognised as an important variable and needs to be considered when devising scoring systems. Whilst it is recognised that pathologist scoring is semi-quantitative and subjective, measurement of inter-observer and intra-observer agreement provides evidence of the reproducibility of the scoring system and increases the degree of confidence in the data [18]. For example, a study that examined Ki-67 (MIB-1) scoring across eight laboratories and a central reference laboratory demonstrated high intra-laboratory reproducibility (intraclass correlation = 0.94; 95% CI 0.93–0.97), but only moderate inter-laboratory reproducibility (intraclass correlation = 0.71, 95% CI = 0.47–0.78). Factors contributing to inter-laboratory discordance were tumour region selection, counting method and subjective assessment of staining positivity. Formal counting methods gave more consistent results than visual

Table 4. Scoring systems for immunohistochemistry [15]

Scoring method	Intensity score (IS)	Proportion score (PS/%)	Calculation
H score	0, 1, 2, 3	0–100% (continuous)	H score = (1 × %IS1) + (2 × %IS2) + (3 × %IS3) Range 0–300
Allred score	0, 1, 2, 3	1 = <1 2 = 1–10 3 = 10–33 4 = 33–66 5 = >66	Allred score = IS + PS Range 0, 2–8
Additive quick score	0, 1, 2, 3	1 = 0–4 2 = 5–19 3 = 20–39 4 = 40–59 5 = 60–79 6 = 80–100	Additive quick score = IS + PS Range 0–9
Multiplicative quick score	0, 1, 2, 3	1 = 0–4 2 = 5–19 3 = 20–39 4 = 40–59 5 = 60–79 6 = 80–100	Multiplicative quick score = IS × PS Range 0–18

estimation [19]. There is considerable interest and investment in developing quantitative image analysis for scoring immunohistochemistry [1].

Measures to assure confidence in the reproducibility of pathologist scoring should be considered, for example by reporting levels of inter and intra-observer agreement.

Filtering

Individual pathologists have different thresholds for observing and recording incidental morphological changes, background staining or unexpected staining patterns, called filtering. Recording such data, particularly for novel biomarkers, may be important as the information contributes to an understanding of the characteristics of a particular test and informs biomarker development and refinement.

The pathologists in the trial should agree what is considered background staining, and thus filtered out of pathologist scoring, and this should be kept under review during the trial.

Companion and complementary diagnostic tests

Increasingly, the development of new drugs requires specific diagnostic tests to guide treatment, called companion and complementary diagnostic tests. Companion tests are mandatory prior to administering a targeted drug, whereas complementary tests are used to guide patient selection for treatment, but are not used as 'gate-keepers' for access to the drug. The rationale

for recommending such tests is based on the principle of precision medicine: delivering the right drug to the right patient at the right time, but is also driven by the requirements of regulatory bodies such as the US FDA and the UK MHRA. The paradigm for companion testing is breast cancer where HER2 receptor testing (e.g. HercepTest, Agilent Dako, Glostrup, Denmark) is required prior to treatment with Trastuzumab. Such diagnostic tests are manufactured to high standards and satisfy the requirements for an *in vitro* diagnostic (IVD) medical device, accredited by regulatory bodies. Furthermore, UK National External Quality Assurance Scheme (NEQAS) immunocytochemistry (ICC) and *in situ* hybridisation (ISH) have shown that laboratories using IVD tests outperform those using laboratory developed assays [20]. One of the issues around the use of IVDs is that the tests tend to be expensive and the manufacturers' instructions need to be followed precisely to produce a valid test result. Typically, the instructions specify the use of a specific proprietary auto-stainer, which may not be available in participating laboratories. The problem can be resolved by setting up central laboratory testing at a site with the designated staining platform that can deliver quality assured tests.

Companion and complementary diagnostic tests should be used in preference to laboratory-devised assays. Manufacturers' instructions must be followed to produce a reliable result.

Estimates of uncertainty

Scoring immunohistochemistry may be subject to 'estimates of uncertainty of measurement'. ISO 15189:2012 states that 'the laboratory shall at least attempt to identify all the components of uncertainty and make a reasonable estimation, and shall ensure that the form of reporting of the result does not give the wrong impression of uncertainty'. This applies to numerical counts of immunohistochemically positive cells and working examples are provided in guidelines provided by the UK Royal College of Pathologists [21]. The recommendations include consideration of the best methods to achieve clinically reliable measurements, ensuring that these are defined in SOPs and working to ensure that the measurement procedures are consistent between pathologists [21].

Estimates of uncertainty, defined in ISO 15189:2012, should be made for clinical trial tests where possible.

Calibration

It may be necessary to calibrate the laboratories and pathologists prior to opening a trial. This can be

achieved by providing a 'test set' of cases and compiling the results across the centres and providing feedback. This can be an iterative process aimed at achieving a pre-specified level of performance or more formal accreditation of the laboratory for the trial. Typically, accredited laboratories would have high levels of agreement with a reference laboratory or demonstrate high levels of agreement with other laboratories taking part in the trial. Statistical tests, such as Cohen's Kappa coefficient (κ) [22] or intraclass correlation coefficient, can be used to measure inter-laboratory or inter-observer agreement. In the case of Cohen's Kappa coefficient to assess categorical variables, the acceptable level of agreement is controversial. Some researchers consider a κ of >0.6 (moderate agreement) acceptable, whereas others stipulate a κ of >0.8 (strong agreement) [22]. A statistician should be consulted for advice on the most appropriate statistical tests to use. Ideally, the statistician should analyse the data independently of the pathologist(s) to ensure objective assessment.

The requirement for pathologist calibration should be considered prior to the trial opening.

External quality assurance

For established biomarkers, where external quality assurance programmes already exist (e.g. UK NEQAS for ICC and ISH), trial laboratories should participate in such schemes and demonstrate satisfactory performance (Table 3) [12].

If an external quality assurance scheme (e.g. UK NEQAS ICC & ISH) is available for the trial test then trial laboratories should participate.

Genomic testing

Factors such as acceptable testing platforms, nucleic acid extraction kits, modality of testing and the thresholds for positive results need to be set out in SOPs. Fresh-frozen tissue is the gold standard for techniques where nucleic acid quality requirements are high, for example whole genome sequencing. Formalin-fixed paraffin-embedded samples may be acceptable but are subject to the numerous pre-analytical variables listed in Table 2 [1,23]. If an external quality assurance scheme for example, the UK NEQAS Molecular Genetics programme exists for a particular test, then laboratories should participate [24]. Pathologists should review trial-specific patient information sheets and consent forms to ensure that laboratory tests

relating to germline mutations are disclosed and that the participant understands the consequence of a positive result and the implications for family members.

Pathological complete response

Pathological estimates of residual disease following trial-specific interventions, termed 'pathological complete response', are being utilised to expedite the reporting of clinical trials, where previously only extended clinical follow up has been used [25–28]. It is important that the participating pathologists have an understanding of the definition of pCR; for example in breast cancer trials, variable definitions have been used allowing or excluding ductal carcinoma *in situ* from the definition, which influences trial outcomes [29]. Nevertheless, central pathology review of a phase 3 neoadjuvant breast cancer trial (ARTemis) has shown good concordance of pCR with participating centre assessment, even in the absence of guidelines and a trial-specific reporting proforma [30]. For colorectal cancer, the UK Royal College of Pathologists reporting guidelines specify that the entire scar should be embedded and three deeper levels cut on each block prior to calling a pCR. These principles should be followed in trials and it is recommended that all pCR cases are centrally reviewed. It is also important to specify whether residual nodal disease should contribute to pCR, as assessment of mesorectal lymph nodes has been shown to reduce the rate of pCR [31]. Detailed SOPs are required to standardise block selection and sectioning protocols to ensure consistent assessment of patient specimens across the trial sites [26]. Similar principles can be applied to other morphological parameters such as defining and measuring resection margins in surgical trials and processing sentinel lymph node specimens.

Clear definitions of reporting of pathological complete response and SOPs for standardised assessment need to be agreed and established prior to the trial opening.

Bio-resource for translational research

Tissue collected and curated as part of a clinical trial (frozen tissue and/or formalin-fixed paraffin-embedded tissue) is associated with high quality clinical data and pathological classification. Consequently, it is an important legacy that can be used for translational research; for example, the tissue can be used to

develop and refine predictive tests. At trial inception, pathologists should actively promote tissue collection and incorporate the costs into the trial grant application or formulate applications for a companion study. The trial documentation, specifically the consent form and patient information sheet, must be worded appropriately. If samples are to be kept beyond the end of the trial, then consent needs to be enduring and ideally generic. Samples may be moved into a biobank with the correct consent provisions, ensuring that the relevant regulations are followed; for example, in England, tissue must be moved under an Human Tissue Authority (HTA) licence and meet HTA standards when the research ethics approval for the study expires.

Pathologists should be actively involved in developing trial specific bio-resources.

Conclusion

Our Quality Assurance in Clinical Trials Workshop highlighted the importance of ensuring reproducible scoring and reporting of pathology parameters in clinical trials. Key considerations include pathologist training and calibration, and measurement of inter-observer variation. There are opportunities to evaluate image analysis systems in clinical trials, alongside conventional histological assessment, to moderate 'uncertainty of measurement'. In any case, pathologists' engagement in clinical trials is essential to help deliver the reliable evidence-base required for new treatments and robust pathology processes are vital to the credibility of studies.

Pathologists' contributions to clinical trials are wide ranging and should be recognised and appropriately funded.

Early pathologist engagement and input into trial design is essential.

Acknowledgements

Author contributions statement

CV conceived and chaired the NCRI CM-Path Quality Assurance in Clinical Trials Workshop. All authors contributed to the workshop. CV and MR composed the first draft of the manuscript. All authors critically reviewed and approved the final version of the manuscript.

References

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

2. Jones JL, Oien KA, Lee JL, *et al*. Morphomolecular pathology: setting the framework for a new generation of pathologists. *Br J Cancer* 2017; **117**: 1581–1582.

3. Crissman JW, Goodman DG, Hildebrandt PK, *et al*. Best practices guideline: toxicologic histopathology. *Toxicol Pathol* 2004; **32**: 126–131.

4. US Department of Health and Human Services Food and Drug Administration Center for Drug Evaluation and Research (CDER). Considerations for Use of Histopathology and Its Associated Methodologies to Support Biomarker Qualification. Guidance for Industry. 2016. [Accessed 6 August 2018]. Available from: https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm285297.pdf.

5. Morton D, Sellers RS, Barale-Thomas E, *et al*. Recommendations for pathology peer review. *Toxicol Pathol* 2010; **38**: 1118–1127.

6. Kondylis FI, Moriarty RP, Bostwick D, *et al*. Prostate cancer grade assignment: the effect of chronological, interpretative and translation bias. *J Urol* 2003; **170**: 1189–1193.

7. Speight PM, Abram TKJ, Floriano PN, *et al*. Interobserver agreement in dysplasia grading: toward an enhanced gold standard for clinical pathology trials. *Oral Surg Oral Med Oral Pathol Oral Radiol* 2015; **120**: 474–482.

8. Hamdy FC, Donovan JL, Lane JA, *et al*. 10-year outcomes after monitoring, surgery, or radiotherapy for localized prostate cancer. *N Engl J Med* 2016; **375**: 1415–1424.

9. Oxley J, Simpkin A, Goepel J, *et al*. Gleason drift in the NIHR ProtecT study. *Histopathology* 2015; **66**: 438–446.

10. Hammond ME, Hayes DF, Dowsett M, *et al*.; American Society of Clinical Oncology; College of American Pathologists.American Society of Clinical Oncology/College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *J Clin Oncol* 2010; **28**: 2784–2795.

11. Wolff AC, Hammond ME, Hicks DG, *et al*. Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. *J Clin Oncol* 2013; **31**: 3997–4013.

12. United Kingdom National External Quality Assessment Service for Immunohistochemistry and In Situ Hybridisation (UK NEQAS ICC & ISH). [Accessed 6 September 2018]. Available from: https://ukneqas.org.uk/.

13. Dennis E, Banks P, Murata LB, *et al*. Validation of an electronic program for pathologist training in the interpretation of a complex companion diagnostic immunohistochemical assay. *Hum Pathol* 2016; **56**: 194–203.

14. Fitzgibbons PL, Bradley LA, Fatheree LA, *et al*. Principles of analytic validation of immunohistochemical assays. Guideline from the College of American Pathologists Pathology and Laboratory Quality Center. *Arch Pathol Lab Med* 2014; **138**: 1432–1443.

15. Lee M, Lee CS, Tan PH. Hormone receptor expression in breast cancer: postanalytical issues. *J Clin Pathol* 2013; **66**: 478–484.

16. Buttner R, Gosney JR, Skov BG, *et al*. Programmed death-ligand 1 immunohistochemistry testing: a review of analytical assays and clinical implementation in non-small-cell lung cancer. *J Clin Oncol* 2017; **35**: 3867–3876.

17. Dowsett M, Nielsen TO, A'Hern R, *et al.*; International Ki-67 in Breast Cancer Working Group.Assessment of Ki67 in breast cancer: recommendations from the International Ki67 in Breast Cancer working group. *J Natl Cancer Inst* 2011; **103**: 1656–1664.

18. Elliott K, McQuaid S, Salto-Tellez M, *et al.* Immunohistochemistry should undergo robust validation equivalent to that of molecular diagnostics. *J Clin Pathol* 2015; **68**: 766–770.

19. Polley MY, Leung SC, McShane LM, *et al.* An international Ki67 reproducibility study. *J Natl Cancer Inst* 2013; **105**: 1897–1906.

20. Ibrahim M, Parry S, Wilkinson D, *et al.* ALK immunohistochemistry in NSCLC: discordant staining can impact patient treatment regimen. *J Thorac Oncol* 2016; **11**: 2241–2247.

21. Helliwell T, Giles T, Fensome-Rimmer S. ISO 15189:2012 – An approach to the assessment of uncertainty of measurement for cellular pathology laboratories. The Royal College of Pathologists 2015. [Accessed 30 October 2018]. Available from: https://www.rcpath.org/asset/83AABBC1-EF3B-48F5-9DD3F4D1 6E45A80F.

22. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)* 2012; **22**: 276–282.

23. Robbe P, Popitsch N, Knight SJL, *et al.* Clinical whole genome sequencing from routine formalin fixed, paraffin-embedded specimens; pilot study for the 100,000 genomes project. *Genet Med* 2018; **20**: 1196–1205.

24. United Kingdom National External Quality Assessment Service for Molecular Genetics. [Accessed 6 September 2018]. Available from: https://www.ukneqas-molgen.org.uk/molecular-pathology. html.

25. Nagtegaal ID, West NP, van Krieken JH, *et al.* Pathology is a necessary and informative tool in oncology clinical trials. *J Pathol* 2014; **232**: 185–189.

26. Provenzano E, Bossuyt V, Viale G, *et al.*; Residual Disease Characterization Working Group of the Breast International Group-North American Breast Cancer Group Collaboration.Standardization of pathologic evaluation and reporting of postneoadjuvant specimens in clinical trials of breast cancer: recommendations from an international working group. *Mod Pathol* 2015; **28**: 1185–1201.

27. Bossuyt V, Provenzano E, Symmans WF, *et al.* Recommendations for standardized pathological characterization of residual disease for neoadjuvant clinical trials of breast cancer by the BIG-NABCG collaboration. *Ann Oncol* 2015; **26**: 1280–1291.

28. Sheri A, Smith IE, Johnston SR, *et al.* Residual proliferative cancer burden to predict long-term outcome following neoadjuvant chemotherapy. *Ann Oncol* 2015; **26**: 75–80.

29. Von Minckwitz G, Untch M, Blohmer JU, *et al.* Definition and impact of pathologic complete response on prognosis after neoadjuvant chemotherapy in various intrinsic breast cancer subtypes. *J Clin Oncol* 2012; **30**: 1796–1804.

30. Thomas JSJ, Provenzano E, Hiller L, *et al.* Central pathology review with two-stage quality assurance for pathological response after neoadjuvant chemotherapy in the ARTemis trial. *Mod Pathol* 2017; **30**: 1069–1077.

31. Swellengrebell HA, Bosch SL, Cats A, *et al.* Tumour regression grading after chemoradiotherapy for locally advanced rectal cancer: a near pathologic complete response does not translate into good clinical outcome. *Radiother Oncol* 2014; **112**: 44–51.