SUPPLEMENTAL MATERIAL 2:

Glossary of evidence types listed in the Hierarchy of Evidence for Tumor Pathology

- **Animal studies:** we use this term here to group together all *in vivo* (whole animal, not just tissues and cells) studies that do not involve humans. This could include experimental models or drug trials in animals.

  Example

  NOTE: studies with animal derived material are usually only allowed in the etiology and pathogenesis sections of the World Health Organisation Classification of Tumours but are excluded from (and so are not for consideration in) other sections.

- **Population-based descriptive studies:** citation of published information directly from national/international registers of cancer diagnoses and statistics (cancer registry data). Here we refer to population-based cancer registries that are a particular type of registry that collect information of all cancer cases occurring in a specific geographic unit. We do not include hospital registries which may not be as robust and may to be used in studies that fall into another category of evidence (such as cohort studies).

  Example

- **Case-control studies:** an observational analytical study where A group of patients/cases are sometimes pair-matches or sometime compared with a group denominated control (usually healthy people, often matched by sex and age) to explore the association between a risk factor, characteristic or treatment and the case/control status. Some case-control studies use a control group that comprises also cancer cases, different from the “case group” according to a specific characteristic

(sometimes known as case-case studies). Case-control studies may also include nested, individually matched, and case-cohort designs.

Example

- **Case report:** an observational report of a new single patient/case describing salient features with some unique insight or educational point of interest. Case reports that then include other cases in the literature are still considered as a case report (not a case series).

  Example

- **Case series:** an observational descriptive study which is essentially a collection of case reports with the aim of establishing common characteristics in a group of patients/cases; there is neither randomization not comparator group. Their goal is merely descriptive. Case series here may also include longitudinal case series.

  Example

- **Clinical laboratory test validation studies:** we are using this term to refer to a technical laboratory validation study conducted in a controlled setting with deliberately selected (also known as 'cherry-picked') or enriched cases (e.g. 30 positive and 30 negative) patients/cases to confirm a test is able to detect something of interest (e.g. an antibody stain). Usually these only analyze the number of cases correctly called. This could be thought of as similar to phase I/II clinical trials. Some may refer to these as 'verification' studies and some may distinguish between verification and validation, however for the purposes of this work we have combined these types of study.

  Example

  NOTE: We are deliberately distinguishing between these and other diagnostic test studies (see below) however, which are more comprehensive, use representative cases

that are not cherry picked, and are aimed to establish indicators such as sensitivity and specificity.

- **Cohort studies:** an observational analytical study where one or more cohorts of people (can be patients) are followed to analyze the influence of different exposures, characteristics, or treatments in the selected end-point(s) after follow-up, with analytical tools that allow for censured follow-up (individuals in the cohort who do not develop the end-point, being particularly critical for cases not followed after the end of the study). In classical epidemiology people exposed and unexposed to a particular risk factor are followed to study the association of this factor with the incidence of the disease. In cohorts of patients, characteristics of the tumor, exposures or treatments are associated with other endpoints during follow-up, typically recurrence or death. This is the classical design in prognostic studies and often also in predictive studies. These may be **prospective cohort studies**, whereby a group of patients are identified and followed over the proceeding period (example). Alternatively, these may be **retrospective cohort studies**, whereby the researchers look back at a selected cohort and try to make inferences on what has already been recorded (example). The latter may have a higher risk of bias or deficiencies in the dataset. Cohort studies that cannot be classified as retrospective or prospective (either it is uncertain, mixed, or just impossible to classify) by default are designated as retrospective.

- **Consensus studies:** a paper outlining the consensus of a group of experts, following a formal process such as a Delphi method, a structured survey with the aim of establishing consensus, or a formal meeting with voting. Simple surveys used to support a position are excluded here. Editorials, letters, white papers, or position type statements written by a small group of experts are also excluded here.

  Example

- **Diagnostic agreement/reproducibility studies:** an observational study where a group of testers (e.g., pathologists) are compared with each other (int*er*-observer) or with themselves (int*ra*-observer) at a later date to determine the level of agreement (also known as test reproducibility or reliability). These often are reported with outcomes such as kappa statistics (scores) or agreement rates/percentages. Not to be confused with consensus studies.

  Example

- **Diagnostic test accuracy studies:** a study where a group of patients/cases, already diagnosed by a gold standard or other well-established test/criteria (such as histological characteristics) for a particular disease, are exposed to a new test (such as immunohistochemistry, PCR, FISH) to evaluate that new test's accuracy - outcomes are often sensitivity, specificity, predictive values, etc. Cases must not be cherry-picked and are unselected, thus representative of the true, real-life population that the test may be used on in the future. These are usually a consecutive series cohort of eligible patients.

  Example

  NOTE: we distinguish between accuracy and precision (reproducibility) for the purposes of the hierarchy (see agreement/reproducibility studies above).

- **Mechanistic laboratory studies**: an experimental/fundamental/basic science study in a wet laboratory setting - conducted using human cells, tissues, or biological molecules. This would include in vitro studies (those on isolated tissue, organs, and/or cells) and ex vivo studies (living tissue is used that has been artificially created in the laboratory or donated by living organisms). These do no not involve the use of living whole organisms (animals). We use the term to refer to studies usually focusing on a

tumour and not a patient, and do not usually generate findings linked to clinical outcomes or applications (as compared with cohort studies).

Example

NOTE: Mechanistic studies in humans and their samples/images etc. are classified into 'mechanistic clinical studies' (below). Mechanistic studies do not include studies of clinical tests carried out in the laboratory for diagnostic purposes (these are considered diagnostic test accuracy studies instead).

- **Mechanistic clinical studies:** mechanistic studies in patients using imaging, biochemistry, molecular, or other technologies and where clinical outcomes are available and compared.

  Example

  NOTE: mechanistic studies do not include studies of clinical tests carried out in the laboratory for diagnostic purposes (these are considered diagnostic test accuracy studies instead).

- **Molecular biology database studies**: published work based upon online database entries / records from formally curated repositories of biological molecules such as nucleic acids (including genes and genomes), proteins and metabolites, their modifications or interactions. Examples include TCGA, OMIM, HPA, GenBank, HMDB. These sources may cite other forms of evidence.

  Example

  NOTE: studies with animal derived material are usually only allowed in the etiology and pathogenesis sections of the WCT but are excluded (and so are not for consideration) from other sections.

- **Observational trials:** for the purposes of this work, we are using this term to refer an interventional (single arm) study where no control group was used, or no randomization was carried out (or both) - this may be because outcomes in participants before and after intervention are being compared, and a control/placebo is not needed.

  Example

  NOTE: this category is observational trials, not observational studies in general.

- **Other cross-sectional studies:** cross sectional studies are a snapshot of the status of a group of people (or patients/cases) at one point in time, such as a study of prevalence or a census, and can explore the association between an exposure or characteristic and a disease or subgroup of patients. We are identifying here 'other' cross-sectional studies because there we have distinguished some special types of cross-sectional evidence which have been listed separately (e.g., diagnostic test accuracy studies, population-based descriptive studies).

  [Example not given as this is an 'other' category.]

  NOTE: Interview and patient/population survey studies are not included here, these were excluded altogether from the hierarchy (see Table 1).

- **Randomized-controlled trials & Studies derived from randomized-controlled trials:** an interventional study where participants (usually patients) are randomized into a treatment group and a control, standard of care or placebo group. In these studies, the exposure or treatment is assigned by the research team, either using a blinded method or not – for the purposes of this work we did not distinguish between blinded and non-blinded studies. Sometimes these studies have a primary outcome that is directly related to the subheading of interest, but sometimes the work is carried out on the clinical trial cohort only as an add on piece of work at a later date (the

cohort was re-used for new research). The latter are generally more robust than, for example, a cohort study due to the systematic nature of data collection and follow up.

Example

NOTE: Studies that do not have randomization or control groups are classified above as 'observational studies'.

- **Systematic review:** a comprehensive review with a clearly formulated question (e.g., using PICO) that uses systematic and explicit methods to identify, select and critically appraise relevant primary research, and to extract and analyze data from the studies that are included in the review. These reviews may have used meta-analysis (if possible) or a more discussion-based (narrative) analysis. For the purposes of this work, we also regard mapping/gap map, scoping, umbrella or overview reviews as systematic reviews, as well as any other systematically designed evidence synthesis work. Traditional narrative or literature reviews are not considered systematic reviews and are excluded from the hierarchy (see Table 1).

  Example

  NOTE: rapid reviews are placed in Level P2 as these have a higher risk of bias than full systematic reviews. This category also includes clinical guidelines which are developed using systematic review methodologies (example).

Types of cited evidence which do not meet the criteria of any of the above are excluded from the Hierarchy of Evidence for Tumor Pathology. See also Table 1 (note - terminology evolved during the project and may not match those given above).