**Use of test accuracy study design labels in (developing) NICE’s Diagnostic Guidance**

**- Data extraction guidance and criteria**

Objective:

The overall purpose of this project is to investigate the use of DTA classification terminology (study design labels) in of NICE diagnostic guidelines and evidence reports.

The primary objectives are:

* To describe the range of study design terms and labels in NICE evidence reports of diagnostic accuracy reviews.
* To investigate whether different weight is given to different study designs in the final guidance.

Moreover, this study is a part of another parallel project, which aims to standardize classification terms for better reporting, quality/validity assessment and transparency of DTA studies. Hence, a secondary objective is to collect information on the different types of designs and corresponding labels (incl. non-labels), to inform this development of labels (the reason for identifying domains, please see below).

Search and selection:

A list of all guidelines either published, under development or in consultation is identified from NICE diagnostic webpage:

https://www.nice.org.uk/guidance/published?type=dg, https://www.nice.org.uk/guidance/indevelopment?type=dg, https://www.nice.org.uk/guidance/inconsultation?type=dg,

Eligibility criteria:

All NICE diagnostic assessment reports (DARs) and respective published guidelines (incl. relevant abbendums and erratums) that includes a diagnostic question\* and a clear test accuracy review\*\*, are included. If a report includes other reviews besides the diagnostic, these are also included, however, data extraction is only performed for the diagnostic accuracy question and reviews.

Reports of only monitoring, predictive and prognostic questions are excluded, since these could have other design, features and labels not relevant to the description of primary diagnostic accuracy studies).

\* a instant classification of condition positive and negative (including results of an 2x2 table from an index test and reference standard)

\*\* a systematic review assessing *test accuracy* (including identification of an index and reference standard, excluding analytical accuracy/analytical performance) reporting a diagnostic outcome measure (e.g. SN, SP, AUC, PPV, NPV, LR)

The DAR and related documents are found under the tab “evidence”, in each link.

Definitions used

As no standardized terminology for DTA studies exists, we developed a set of definitions for this current study.

**Study design**: The methodological research strategy/plan of a study question (hypothesis)/objective, incl. any specific features (an individual methodological study characteristic).

**Study design label:**

A (standard/common) classification term used to describe or classify/distinguish studies by design or feature. E.g. Two-gate/case-control study.

**Design description/non-label**:

A non-standardized (descriptive) term or sentence used to describe or classify/distinguish studies by design or design feature. E.g. “Two studies assessed two of the index tests within the same population” for describing a comparative study.

To ensure extraction of only design/features that are relevant (i.e. neither too broad/vague or too specific/detailed) for classification of DTA studies in general, we developed a set of criteria for identification of labels and non-labels of qualified design/features. For a label/non-label to be extracted, it has to be:

1. A study design/feature that falls under at least one of the main domains of: study type (research phase/stage), participant eligibility, sampling method/data collection, tests (type; number), participant flow (incl. test order) or analysis (adopted from STARD).
2. A study design/feature should be clearly recognized without the need for topic-specific knowledge of a test and/or target condition.
3. Moreover, we do not consider specific numbers (e.g. of sample size), individual countries, or details of methods such as threshold decisions, details of test conduct or analysis plan as design features relevant for general classification of DTA.

If features that remains unclear appears, albeit above criteria, these will be discussed among reviewers.

Example 1: GD32 (DAR pg. 54, paragraph 2 in results):

“*All included studies were conducted in hospital-based colposcopy clinics and used a prospective cohort design. All patients underwent colposcopy with an adjunctive colposcopy technology, except for participants included in two DYSIS two-arm studies that included a separate parallel control group examined with colposcopy alone.(74, 80) Six studies were conducted in more than one centre (42, 57, 74, 80, 88, 94)*”

**Labels:** prospective cohort design (domain: participant flow and sampling/data collection)

**Descriptions:** hospital-based colposcopy clinics (domain: setting); Six studies were conducted in more than one centre (domain: location)

“***All patients underwent colposcopy*** *with an adjunctive colposcopy technology,* ***except f****or participants included in* ***two DYSIS two-arm studies that included a separate parallel control group******examined with colposcopy alone***”. This describes patient flow, however, this is (too) specific for the current review/test/condition questions (to be a standard design feature for DTA studies) and for this reason, not extracted/included as use of “non-label /description”.

Overview of extraction and sections in DG and DARs:

We will extract the use of labels in the following two contexts:

1. The use of labels and descriptions/non-labels from sections that are expected or identified to make use of labels, for describing the available evidence/studies.
2. If different weight is given to study designs (identified by use of labels and non-labels) from sections that are expected or identified to distinguish between study designs for final recommendations
3. The use of labels are extracted from section of:

* From the systematic review (DAR), labels used in:
  + Abstract (all)
  + Scientific summary:
    - Methods/results
    - Discussion/conclusion
  + Methods:
    - Defining eligibility criteria
  + Results:
    - Description/characterization of included studies
    - The proportion of studies excluded by design (from flow diagram)
    - Exclusion table (in appendix)
    - Quality assessment
  + Discussion and conclusion
* Published guideline
  + Recommendation
  + Evidence section
  + Committee discussion
  + Recommendations for future research

1. If labels are used to give different weight to (differentiate) between studies, were extracted from section of:

* Systematic review/DAR
  + Scientific summary: discussion and conclusion (i.e. final recommendation)
  + Discussion and conclusion
* Published guideline:
  + Recommendation
  + Evidence section
  + Committee discussion
  + Recommendations for future research
* For the scope of this study, we focus only on design labels used in *accuracy review/evidence*, from the DAR and guidelines.
* The extraction is developed so that sections can be identified in the table of content of each report. However, if it is not possible to identify a section (incl. the information related to the accuracy review/studies separately from the review of clinical effectiveness), this is indicate by “Section not identified (SNI)” (incl. the executive summary. Please be aware, not to confuse it with “None”, “Not Specified (NS)” and ”NA”).
* If the same label/non-label occurs several times within one section, this should only be extracted once (as we are not interested in the frequency of each label/non-label within a section
* Please insert new line for every excerpt in the same cell.