
Last revised 07-03-2008

The JOSPT as a journal that primarily focuses on clinically based research, attempts to assist readers in understanding the quality of research evidence in its published manuscripts by using a structured abstract that includes information on the “study design” and the classification of the “level of evidence.” While no single piece of information or classification can signify the quality and clinical relevance of published studies, a consistent approach to communicating this information is a step towards assisting our readers in evaluating new knowledge.

Outline of the structured abstract for research reports published in JOSPT

Study design
Objectives
Background
Methods and Measures
Results
Conclusions
Level of Evidence

Suggestions for Naming Study Designs

We recommend that authors attempt to use the following terminology when naming their research designs. Use of consistent terminology will make it easier for readers to identify the nature of the study and will allow us to track the type of studies published more easily. We recognize that this list is not all-inclusive and that more appropriate descriptors might be suitable for some studies. In those cases, investigators are encouraged to pick the most appropriate descriptors for their study. These suggestions are provided as a means of encouraging consistency where it would be both useful and informative and their use is expected where they do apply.

Quantitative Clinical Study Categories include (Therapy, Prevention, Etiology, Harm, Prognosis, Diagnosis, Differential Diagnosis, Symptom Prevalence, Economic Analysis, and Decision Analysis). For each study category there is a spectrum of study designs. These include:

2. Randomized Clinical Trial: Patients are enrolled at a relevant baseline and allocated to different intervention arms based on a random process (preferably with concealment); outcomes are ascertained prospectively.
3. Prospective Cohort: A longitudinal study where subgroups of patients are enrolled and research questions defined at a relevant baseline point (prior to when outcomes occur); patients are followed forward in time for outcomes ascertainment. For therapy or
prevention studies at least 2 groups are defined at baseline; in prognostic studies (often conducted on a single group) potential predictors are collected at baseline.

4. Retrospective Cohort: A longitudinal study where a single group or multiple groups of patients are involved in a prospective data collection but the research questions (and variables) were defined retrospectively; treatment groups or prognostic factors may have been defined after data collection was initiated e.g. database research or chart review.

5. Case-Control: A longitudinal study where patients who have the outcome of interest (cases) and control patients without the same outcome are identified/enrolled and data are collected retrospectively (recall or pre-existing data) to determine if they had the exposure (prognostic factor) or intervention of interest.

6. Cross-over Study: A single group or multiple groups are subjected to multiple treatment regimens, sequenced randomly. Typically, a “washout period” of time separates treatments to reduce residual effects of the previous treatment. If the study involves randomization, it would be a randomized crossover trial.

7. N-of-1 Randomized Clinical Trial: A single patient is enrolled at a relevant baseline and allocated to cross-over different intervention arms based on a random (preferably) concealed process; outcomes are ascertained prospectively.

8. Cross-over Case Study: A single patient is enrolled at a relevant baseline and 2 or more intervention options are compared; outcomes are ascertained prospectively.

9. Case Series: Data are collected on a single group of patients (no comparison group); limited to study of therapy or prevention. A prognostic study that includes a single group is considered a cohort study (described above).

10. Case Report: Data are collected on a single subject without using a design allowing systematic comparison against baseline or an alternative intervention.

Other Quantitative Study Categories Include

1. *Clinical measurement: Reliability, validity, responsiveness, clinimetric, psychometric, utility study for examples.

2. *Descriptive: Includes surveys and other descriptive data collection methods.

3. Consensus Statements: Systematic processes used to define or develop consensus on clinical topics (includes Clinical Practice Guidelines developed by consensus).

4. *Controlled Laboratory Study: Study performed with subjects in a laboratory setting that involves measurements related to biomechanics, electromyography, or physiology for examples. This category should be used if the study does not have a specific clinical category/purpose. Where lab based measures are used to evaluate therapy, harm, etc, they should be classified as a clinical study (above).

5. *Non-human Bench Research: Studies conducted on human tissue or animals.
*For these study categories, there are a range of possible study designs. These include:
  a. Longitudinal: data were collected at multiple time points/occasions
  b. Cross-sectional: data were collected on a single time point/occasion on 2 or more distinct groups
  c. Pre-test post-test: data were collected on a single occasion but with 2 or more time points typically prior and immediately after an intervention

Qualitative Category studies: For this study category, there are a range of possible study designs. These include:
  1. Meta-syntheses: A synthesis of the better quality qualitative studies.
  2. Ethnography: Study of a particular culture or group of people to identify their daily life patterns, meanings, and beliefs.
  3. Phenomenology: Design focuses on the lived experience of the person or a group of people.
  4. Grounded Theory: Research that seeks to understand and identify theoretical processes; themes used to develop an understanding and theoretical explanation.
  5. Case Study: An in-depth study of an individual lived experience and perspective

Recommendations
Include informative descriptors signifying the category/type of research and the specific research design
  a. The “study design” heading of the abstract should confer clear information about the design of the study using the list above, and whether the study was longitudinal or cross-sectional and whether data collection was prospective or retrospective. For some study designs this is implied, for example, randomized clinical trials (RCTs) and cohort studies are both longitudinal and prospective by definition.
  b. When applicable, the level of evidence heading of the abstract should convey information about study type based on those proposed by the Oxford Centre for Evidence-based Medicine, which consists of: therapy, prevention, etiology, harm, prognosis, diagnosis, differential diagnosis, symptom prevalence, economic, and decision analysis. Under this heading, the study type is followed by the level of evidence also as defined by the Oxford Centre for Evidence-based Medicine (for example: Therapy, level 1).
Examples
Examples of terminology for study designs where study categories based on levels of evidence apply:
- Systematic review, therapy;
- Randomized clinical trial, therapy;
- Prospective cohort, therapy;
- Retrospective cohort, therapy;
- Case-control, therapy;
- N-of-1, therapy;
- Case series, therapy;
- Case report, therapy.
** Note that similar categories could be defined for prognosis, diagnosis, etc.

Examples of terminology for study designs where the study categories based on level of evidence do not apply. In such instances, state the appropriate subcategory and design under the study design heading of the abstract and do not include the level of evidence heading:
- Clinical measurement, longitudinal;
- Clinical measurement, cross-sectional;
- Ethnographic (qualitative implied);
- Practice survey, cross-sectional;
- Bench research, pre-test-post-test
- Bench research, cross-sectional
- Non-human bench research, longitudinal
- Consensus statement

**These are example and other combinations obviously apply. Authors should expect to apply the terminology that describes their study design and have it examined during the review process.

JOSPT Policy for Naming Levels of Evidence
Use the levels of evidence published by the Oxford Center for Evidence-based Medicine, reproduced below with permission, to name the level of evidence for all studies that can be appropriately classified using the system. The original table and related notes are available at http://www.cebm.net/index.aspx?o=1025. We recognize that not all forms of studies can be classified. For example, this taxonomy provides no clear description of how to rate practice guidelines, clinical measurement studies, qualitative research, and other types of literature appearing in our Journal. Where the level of evidence system does not apply, the Study Design section of the structured abstract will still apply, and the Level of Evidence component of the abstract will not be required.
## Oxford Centre for Evidence-based Medicine Levels of Evidence (May 2001)

<table>
<thead>
<tr>
<th>Level</th>
<th>Therapy/Prevention, Aetiology/Harm</th>
<th>Prognosis</th>
<th>Diagnosis</th>
<th>Differential diagnosis/symptom prevalence study</th>
<th>Economic and decision analyses</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a</td>
<td>SR (with homogeneity*) of RCTs</td>
<td>SR (with homogeneity*) of inception cohort studies; CDRT† validated in different populations</td>
<td>SR (with homogeneity*) of Level 1 diagnostic studies; CDRT† with 1b studies from different clinical centres</td>
<td>SR (with homogeneity*) of prospective cohort studies</td>
<td>SR (with homogeneity*) of Level 1 economic studies</td>
</tr>
<tr>
<td>1b</td>
<td>Individual RCT (with narrow Confidence Interval‡)</td>
<td>Individual inception cohort study with ≥ 80% follow-up; CDRT† validated in a single population</td>
<td>Validating** cohort study with good††† reference standards; or CDRT† tested within one clinical centre</td>
<td>Prospective cohort study with good follow-up****</td>
<td>Analysis based on clinically sensible costs or alternatives; systematic review(s) of the evidence; and including multi-way sensitivity analyses</td>
</tr>
<tr>
<td>1c</td>
<td>All or none§</td>
<td>All or none case-series</td>
<td>Absolute SpPins and SnNouts††</td>
<td>All or none case-series</td>
<td>Absolute better-value or worse-value analyses † † † †</td>
</tr>
<tr>
<td>2a</td>
<td>SR (with homogeneity*) of cohort studies</td>
<td>SR (with homogeneity*) of either retrospective cohort studies or untreated control groups in RCTs</td>
<td>SR (with homogeneity*) of Level &gt;2 diagnostic studies</td>
<td>SR (with homogeneity*) of 2b and better studies</td>
<td>SR (with homogeneity*) of Level &gt;2 economic studies</td>
</tr>
<tr>
<td>2b</td>
<td>Individual cohort study (including low quality RCT; e.g., &lt;80% follow-up)</td>
<td>Retrospective cohort study or follow-up of untreated control patients in an RCT; Derivation of CDRT or validated on split-sample§§ only</td>
<td>Exploratory** cohort study with good††† reference standards; CDRT† after derivation, or validated only on split-sample§§§ or databases</td>
<td>Retrospective cohort study, or poor follow-up</td>
<td>Analysis based on clinically sensible costs or alternatives; limited review(s) of the evidence, or single studies; and including multi-way sensitivity analyses</td>
</tr>
<tr>
<td>2c</td>
<td>&quot;Outcomes&quot; Research; Ecological studies</td>
<td>&quot;Outcomes&quot; Research</td>
<td>Ecological studies</td>
<td>Audit or outcomes research</td>
<td></td>
</tr>
<tr>
<td>3a</td>
<td>SR (with homogeneity*) of case-control studies</td>
<td>SR (with homogeneity*) of 3b and better studies</td>
<td>SR (with homogeneity*) of 3b and better studies</td>
<td>SR (with homogeneity*) of 3b and better studies</td>
<td></td>
</tr>
<tr>
<td>3b</td>
<td>Individual Case-Control Study</td>
<td>Non-consecutive study; or without consistently applied reference standards</td>
<td>Non-consecutive cohort study, or very limited population</td>
<td>Analysis based on limited alternatives or costs, poor quality estimates of data, but including sensitivity analyses incorporating clinically sensible variations.</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Case-series (and poor quality cohort and case-control studies§§)</td>
<td>Case-series (and poor quality prognostic cohort studies***</td>
<td>Case-control study, poor or non-independent reference standard</td>
<td>Case-series or superseded reference standards</td>
<td>Analysis with no sensitivity analysis</td>
</tr>
<tr>
<td>5</td>
<td>Expert opinion without explicit critical appraisal, or based on physiology, bench research or &quot;first principles&quot;</td>
<td>Expert opinion without explicit critical appraisal, or based on physiology, bench research or &quot;first principles&quot;</td>
<td>Expert opinion without explicit critical appraisal, or based on physiology, bench research or &quot;first principles&quot;</td>
<td>Expert opinion without explicit critical appraisal, or based on economic theory or &quot;first principles&quot;</td>
<td></td>
</tr>
</tbody>
</table>

Produced by Bob Phillips, Chris Ball, Dave Sackett, Doug Badenoch, Sharon Straus, Brian Haynes, Martin Dawes since November 1998.
Notes
Users can add a minus-sign "-" to denote the level of that fails to provide a conclusive answer because of:

- EITHER a single result with a wide Confidence Interval (such that, for example, an ARR in an RCT is not statistically significant but whose confidence intervals fail to exclude clinically important benefit or harm)
- OR a Systematic Review with troublesome (and statistically significant) heterogeneity.
- Such evidence is inconclusive, and therefore can only generate Grade D recommendations.

| * | By homogeneity we mean a systematic review that is free of worrisome variations (heterogeneity) in the directions and degrees of results between individual studies. Not all systematic reviews with statistically significant heterogeneity need be worrisome, and not all worrisome heterogeneity need be statistically significant. As noted above, studies displaying worrisome heterogeneity should be tagged with a "-" at the end of their designated level. |
| † | Clinical Decision Rule. (These are algorithms or scoring systems which lead to a prognostic estimation or a diagnostic category.) |
| ‡ | See note #2 for advice on how to understand, rate and use trials or other studies with wide confidence intervals. |
| § | Met when all patients died before the Rx became available, but some now survive on it; or when some patients died before the Rx became available, but none now die on it. |
| §§ | By poor quality cohort study we mean one that failed to clearly define comparison groups and/or failed to measure exposures and outcomes in the same (preferably blinded), objective way in both exposed and non-exposed individuals and/or failed to identify or appropriately control known confounders and/or failed to carry out a sufficiently long and complete follow-up of patients. By poor quality case-control study we mean one that failed to clearly define comparison groups and/or failed to measure exposures and outcomes in the same (preferably blinded), objective way in both cases and controls and/or failed to identify or appropriately control known confounders. |
| §§§ | Split-sample validation is achieved by collecting all the information in a single tranche, then artificially dividing this into "derivation" and "validation" samples. |
| †† | An "Absolute SpPin" is a diagnostic finding whose Specificity is so high that a Positive result rules-in the diagnosis. An "Absolute SnNout" is a diagnostic finding whose Sensitivity is so high that a Negative result rules-out the diagnosis. |
| ‡‡ | Good, better, bad and worse refer to the comparisons between treatments in terms of their clinical risks and benefits. |
| ††† | Better-value treatments are clearly as good but cheaper, or better at the same or reduced cost. Worse-value treatments are as good and more expensive, or worse and the equally or more expensive. |
| ** | Validating studies test the quality of a specific diagnostic test, based on prior evidence. An exploratory study collects information and trawls the data (e.g. using a regression analysis) to find which factors are 'significant'. |
| *** | By poor quality prognostic cohort study we mean one in which sampling was biased in favour of patients who already had the target outcome, or the measurement of outcomes was accomplished in <80% of study patients, or outcomes were determined in an unblinded, non-objective way, or there was no correction for confounding factors. |
| **** | Good follow-up in a differential diagnosis study is >80%, with adequate time for alternative diagnoses to emerge (eg 1-6 months acute, 1 - 5 years chronic) |

Grades of Recommendation

| A | consistent level 1 studies |
| B | consistent level 2 or 3 studies or extrapolations from level 1 studies |
| C | level 4 studies or extrapolations from level 2 or 3 studies |
| D | level 5 evidence or troublingly inconsistent or inconclusive studies of any level |

"Extrapolations" are where data is used in a situation which has potentially clinically important differences than the original study situation.