

EVIDENCE – BASED MEDICINE OVERVIEW OF THE MAIN STEPS FOR DEVELOPING AND GRADING GUIDELINE RECOMMENDATIONS.

P. Abrams, S Khoury

INTRODUCTION

The International Consultation on Urological Diseases (ICUD) is a non-governmental organization registered with the World Health Organisation (WHO). In the last ten years Consultations have been organised on BPH, Prostate Cancer, Urinary Stone Disease, Nosocomial Infections, Erectile Dysfunction and Urinary Incontinence. These consultations have looked at published evidence and produced recommendations at four levels; highly recommended, recommended, optional and not recommended. This method has been useful but the ICUD believes that there should be more explicit statements of the levels of evidence that generate the subsequent grades of recommendations.

The Agency for Health Care Policy and Research (AHCPR) have used specified evidence levels to justify recommendations for the investigation and treatment of a variety of conditions. The Oxford Centre for Evidence Based Medicine have produced a widely accepted adaptation of the work of AHCPR. (June 5th 2001

<http://minerva.minervation.com/cebm/docs/levels.html>).

The ICUD has examined the Oxford guidelines and discussed with the Oxford group their applicability to the Consultations organised by ICUD. It is highly desirable that the recommendations made by the Consultations follow an accepted grading system supported by explicit levels of evidence.

The ICUD proposes that future consultations should use a modified version of the Oxford system which can be directly ‘mapped’ onto the Oxford system.

1. 1st Step: Define the specific questions or statements that the recommendations are supposed to address.

2. 2nd Step: Analyse and rate (level of evidence) the relevant papers published in the literature.

The analysis of the literature is an important step in preparing recommendations and their guarantee of quality.

2.1 What papers should be included in the analysis ?

- Papers published, or accepted for publication in the peer reviewed issues of journals.
- The committee should do its best to search for papers accepted for publication by the peer reviewed journals in the relevant field but not yet published.
- Abstracts published in peer review journals should be identified. If of sufficient interest the author(s) should be asked for full details of methodology and results. The relevant committee members can then ‘peer review’ the data, and if the data confirms the details in the abstract, then that abstract may be included, with an explanatory footnote. This is a complex issue – it may actually increase publication bias as “uninteresting” abstracts commonly do not progress to full publication.

- Papers published in non peer reviewed supplements will not be included. An exhaustive list should be obtained through:

I. the major databases covering the last ten years (e.g. Medline, Embase, Cochrane Library, Biosis, Science Citation Index)

II. the table of contents of the major journals of urology and other relevant journals, for the last three months, to take into account the possible delay in the indexation of the published papers in the databases.

It is expected that the highly experienced and expert committee members provide additional assurance that no important study would be missed using this review process.

2.2 How papers are analysed ?

Papers published in peer reviewed journals have differing quality and level of evidence.

Each committee will rate the included papers according to levels of evidence (see below).

The level (strength) of evidence provided by an individual study depends on the ability of the study design to minimise the possibility of bias and to maximise attribution.

is influenced by:

- **the type of study**

The hierarchy of study types are:

- Systematic reviews and meta-analysis of randomised controlled trials
- Randomised controlled trials
- Non-randomised cohort studies
- Case control studies
- Case series
- Expert opinion

- **how well the study was designed and carried out**

Failure to give due attention to key aspects of study methodology increase the risk of bias or confounding factors, and thus reduces the study's reliability.

The use of **standard check lists** is recommended to insure that all relevant aspects are considered and that a consistent approach is used in the methodological assessment of the evidence.

The objective of the check list is to give a quality rating for individual studies.

• **how well the study was reported**

The ICUD has adopted the CONSORT statement and its widely accepted check list. The CONSORT statement and the checklist are available at <http://www.consort-statement.org>

2.3 How papers are rated ?

Papers are rated following a « **Level of Evidence scale** ».

ICUD has modified the Oxford Center for Evidence-Based Medicine levels of evidence.

The levels of evidence scales vary between types of studies (ie therapy, diagnosis, differential diagnosis/symptom prevalence study).

the Oxford Center for Evidence-Based Medicine Website:
<http://minerva.minervation.com/cebm/docs/levels.html>

3. 3rd Step: Synthesis of the evidence

After the selection of the papers and the rating of the level of evidence of each study, the next step is to compile a summary of the individual studies and the overall direction of the evidence in an **Evidence Table**.

4. 4th Step: Considered judgment (integration of individual clinical expertise)

Having completed a rigorous and objective synthesis of the evidence base, the committee must then make a judgement as to the grade of the recommendation on the basis of this evidence. This requires the exercise of judgement based on clinical experience as well as knowledge of the evidence and the methods used to generate it. Evidence based medicine requires the integration of individual clinical expertise with best available external clinical evidence from systematic research. Without the former, practice quickly becomes tyrannised by evidence, for even excellent external evidence may be inapplicable to, or inappropriate for, an individual patient: without current best evidence, practice quickly becomes out of date. Although it is not practical to lay our “rules” for exercising judgement, guideline development groups are asked to consider the evidence in terms of quantity, quality, and consistency; applicability; generalisability; and clinical impact.

5. 5th Step: Final Grading

The grading of the recommendation is intended to strike an appropriate balance between incorporating the complexity of type and quality of the evidence and maintaining clarity for guideline users.

The recommendations for grading follow the Oxford Centre for Evidence-Based Medicine.

The levels of evidence shown below have again been modified in the light of previous consultations. There are now 4 levels of evidence instead of 5.

The grades of recommendation have not been reduced and a “no recommendation possible” grade has been added.

6. Levels of Evidence and Grades of Recommendation Therapeutic Interventions

All interventions should be judged by the body of evidence for their efficacy, tolerability, safety, clinical

effectiveness and cost effectiveness. It is accepted that at present little data exists on cost effectiveness for most interventions.

6.1 Levels of Evidence

Firstly, it should be stated that any level of evidence may be positive (the therapy works) or negative (the therapy doesn't work). A level of evidence is given to each individual study.

- **Level 1** evidence (incorporates Oxford 1a, 1b) usually involves meta-analysis of trials (RCTs) or a good quality randomised controlled trial, or 'all or none' studies in which no treatment is not an option, for example in vesicovaginal fistula.
- **Level 2** evidence (incorporates Oxford 2a, 2b and 2c) includes "low" quality RCT (e.g. < 80% follow up) or meta-analysis (with homogeneity) of good quality prospective 'cohort studies'. These may include a single group when individuals who develop the condition are compared with others from within the original cohort group. There can be parallel cohorts, where those with the condition in the first group are compared with those in the second group.
- **Level 3** evidence (incorporates Oxford 3a, 3b and 4) includes:

good quality retrospective 'case-control studies' where a group of patients who have a condition are matched appropriately (e.g. for age, sex etc) with control individuals who do not have the condition.

good quality 'case series' where a complete group of patients all, with the same condition/disease/therapeutic intervention, are described, without a comparison control group.

- **Level 4** evidence (incorporates Oxford 4) includes expert opinion where the opinion is based not on evidence but on 'first principles' (e.g. physiological or anatomical) or bench research. The Delphi process can be used to give 'expert opinion' greater authority. In the Delphi process a series of questions are posed to a panel; the answers are collected into a series of 'options'; the options are serially ranked; if a 75% agreement is reached then a Delphi consensus statement can be made.

6.2 Grades of Recommendation

The ICUD will use the four grades from the Oxford system. As with levels of evidence the grades of evidence may apply either positively (do the procedure) or negatively (don't do the procedure). Where there is disparity of evidence, for example if there were three well conducted RCT's indicating that Drug A was superior to placebo, but one RCT whose results show no difference, then there has to be an individual judgement as to the grade of recommendation given and the rationale explained.

- **Grade A** recommendation usually depends on consistent level 1 evidence and often means that the recommendation is effectively mandatory and placed within a clinical care pathway. However, there will be occasions where excellent evidence (level 1) does not lead to a Grade A recommendation, for example, if the therapy is prohibitively expensive, dangerous or unethical. Grade A recommendation can follow from Level 2 evidence. However, a Grade A recommendation needs a greater body of evidence if based on anything except Level 1 evidence
- **Grade B** recommendation usually depends on consistent level 2 and or 3 studies, or 'majority evidence' from RCT's.
- **Grade C** recommendation usually depends on level 4 studies or 'majority evidence' from level 2/3

studies or Dephi processed expert opinion.

- **Grade D** “No recommendation possible” would be used where the evidence is inadequate or conflicting and when expert opinion is delivered without a formal analytical process, such as by Dephi.

7. Levels of Evidence and Grades of Recommendation for Methods of Assessment and Investigation

From initial discussions with the Oxford group it is clear that application of levels of evidence/grades of recommendation for diagnostic techniques is much more complex than for interventions. The ICUD recommend, that, as a minimum, any test should be subjected to three questions:

1. does the test have good technical performance, for example, do three aliquots of the same urine sample give the same result when subjected to ‘stix’ testing?
2. Does the test have good diagnostic performance, ideally against a “gold standard” measure?
3. Does the test have good therapeutic performance, that is, does the use of the test alter clinical management, does the use of the test improve outcome?

For the third component (therapeutic performance) the same approach can be used as for section 6.

8. Levels of Evidence and Grades of Recommendation for Basic Science and Epidemiology Studies

The proposed ICUD system does not easily fit into these areas of science. Further research needs to be carried out, in order to develop explicit levels of evidence that can lead to recommendations as to the soundness of data in these important aspects of medicine.

CONCLUSION

The ICUD believes that its consultations should follow the ICUD system of levels of evidence and grades of recommendation, where possible. This system can be mapped to the Oxford system.

There are aspects to the ICUD system that require further research and development, particularly diagnostic performance and cost effectiveness, and also factors such as patient preference.