THE EVIDENCE BASIS OF RECOMMENDATIONS FOR DECISIONS IN HEALTH POLICY

Marcial Velasco Garrido[1]
Prof. Reinhard Busse, MPH[1,2]

Berlin, Juni 2003

[1]Dept. Health Care Management
Technische Universität Berlin

[2]European Observatory on Health Care Systems
Berlin Hub

Contents

Introduction 3
Section 1. Health Technology Assessment –supporting decision makers 3
Section 2. What is the evidence in an assessment? 5
Section 3. Evidence basis for two selected interventions 14
References 36

Introduction

European health care systems face the common challenge of delivering effective care, that is of providing interventions which really work. Experiences from the recent past have shown that some interventions which were thought to be beneficial, turned to be – in the light of careful evaluation – at the best case of no benefit, at the worst case even harmful. Delivering interventions which do not work or produce more harm than good represents poor quality of the health system and a waste of resources. These concerns can be seen as the main reason for the increasing demand to base decisions about the provision of health interventions on *evidence*. However decision makers usually can not themselves use the evidence directly, either because of a lack of time resources, or because of a lack of expertise.

In this paper we will first briefly present *health technology assessment* as a genuine approach to support the decision making processes at the policy level with evidence. To make appropriate use of the information provided in health technology assessments, users need to apprehend the underlying concept of *evidence*. Therefore we will next explore the notion of *evidence*, as it is understood in the evidence-based approach to health care. Finally, we will present and discuss the conclusions and recommendations from HTA reports and/or systematic reviews, which have synthesised the available evidence on two selected health interventions, namely screening for prostate cancer and nicotine replacement therapy.

Section 1. Health Technology Assessment –supporting decision makers

The so called ”best practice” initiatives – Evidence Based Medicine (EBM), Clinical Practice Guidelines (CPGs) and Health Technology Assessment (HTA) – are committed to the work of collecting and analysing the *evidence* in a systematic way, and to make information available for different decision purposes about which interventions do really work, presenting it in the form of synthesis of the evidence or as recommendations (Perleth et al. 2001). Their common approach is the attempt to comprehensively cover the best available evidence on one topic, in order to draw the most valid conclusions about an intervention.

Of these activities, HTA is the approach with the explicit aim of supporting the process of decision making at the policy level of health care (e.g. decisions about including an intervention in the benefit catalogue, about purchasing expensive technologies, etc.). It is a multidisciplinary activity that systematically examines the different aspects (e.g. safety, effectiveness, cost-effectiveness, organisational implications, social consequences, etc.) of interventions whose aim is to improve health, or to improve the performance of the health system (Busse et al. 2002).

HTA is a policy-driven activity, hence assessments are undertaken when policy makers commission them and, ideally, with the explicit aim of informing a particular decision process. The beginning of any assessment is thus the so called *policy question*, that is the decision maker’s need for information. The policy question represents the commissioners’ scope of the problem and the context in which the assessment will be carried out. The aspects included in the policy question are listed in Table 1. Ideally, the policy question should be worded in close co-operation between the commissioners and the researchers, since close two-way communication has been demonstrated to be one of the determinants to increase use of evidence in policy-making (Innvaer et al. 2002).

Table 1. Aspects included in the policy question (Busse et al. 2002)

Who initiated the assessment?/ Who commissioned it?	• Policy makers • Health care providers • Third party payers • Patients' advocate
Why is an assessment needed right now?	• New technology • Changes in old technology • New indications for old technology • New findings • Structural / organisational changes • Safety concerns • Ethical concerns • Economic concerns
Which decision is it going to support?	• Investment decisions • Market licensure • Inclusion in / exclusion from benefit catalogue • Planning of capacities • Guidance on best-practice • Investment in further research
Who represents the primary target audience for the report?	• Political decision makers • Third party payers • Hospital managers / administrators • Clinicians • Citizens / patients

The context in which the research is carried out may lead to some financial or time constraints which determine the methods used and the extent/comprehensiveness of the assessment. The scope and level of detail of HTA vary considerably depending on who commissioned a study and why. Therefore, it is crucial to clearly explain that context, so that readers of HTA (other than those who initiated and commissioned the study) can better assess whether the report can be also relevant for their own problems. The scope of the assessment and its recommendations are determined by the policy question.

In order to give an evidence based answer to the problems outlined in the policy question the researchers undertaking the assessment will need to specify the policy question in terms of safety, efficacy, effectiveness, psychological, social, ethical, organisational, professional and economic aspects. These aspects may be able to be addressed with available scientific evidence and data, but they either have not yet been sufficiently answered or have answers that are not accessible and/ or appropriate for the use of decision-making. The task of the researchers is to retrieve the available evidence and prepare the information in a way which it is useful for policy makers.

The *research questions* drive how the rest of the assessment is going to be conducted, the aspects which will be evaluated and those which will not. The formulation of the research questions is a crucial part of the assessment, since they transpose the policy question into questions which can be answered by evaluating scientific evidence. Thus there should be a feed-back loop to the commissioner(s) of the assessment to ensure that the research questions represent a useful "translation" of the policy question(s).

Once the questions have been made clear, the HTA researchers will try to identify and collect the best available evidence which allows them to give valid answers to those questions. The researchers will summarise the evidence and, based on the answers given to the research questions, formulate recommendations for policy making, if appropriate. The product of the health technology assessment process is the so called assessment report or HTA-report, which

ideally includes a summary useful for policy makers, containing information tailored to them. The provision of a brief summary including clear recommendations is one of the main demands from decision makers on researchers in order to increase their own use of evidence (Innvaer et al. 2002).

In our previous work (Busse et al. 2002) we have recommended that HTA researchers provide two kind of reports of their synthesis work, one tailored to the commissioners (the so called *executive summary*) and one targeting the scientific community which differ in it extent and content (see Table 2).

Table 2. Differences between "Executive Summary" and "Scientific Summary Report"

Executive Summary	**Scientific Summary Report**
Addressed to local decision makers ("executives")	Addressed to the HTA and Scientific Community
Focuses on recommendations and conclusions	Stresses the context of the HTA and methodological aspects, in addition to conclusions and recommendations
Written in agencies'/institutions' official tongue(s)	Available in English
Quickly informs decisions	Allows for critical appraisal of relevance, quality, and main findings

So far, we have been talking about the issue of basing decisions in health care on the best **evidence**, but what is exactly meant by this term? In the next section we will explain in some detail the notions of *evidence* and *best evidence*, as they are applied in the HTA community. We will also present some of the approaches used to present the strength of the evidence in a transparent and standardised way.

Section 2. What is the evidence in an assessment?

Evidence based medicine has been defined as "the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients" (Sackett et al. 1996). As stated in the definition, the origin of this evidence based approach is to be seen in the clinical medicine. As stated above, the demand to base decisions on evidence has been extended to other areas of health care, such as public health interventions and health care policy making.

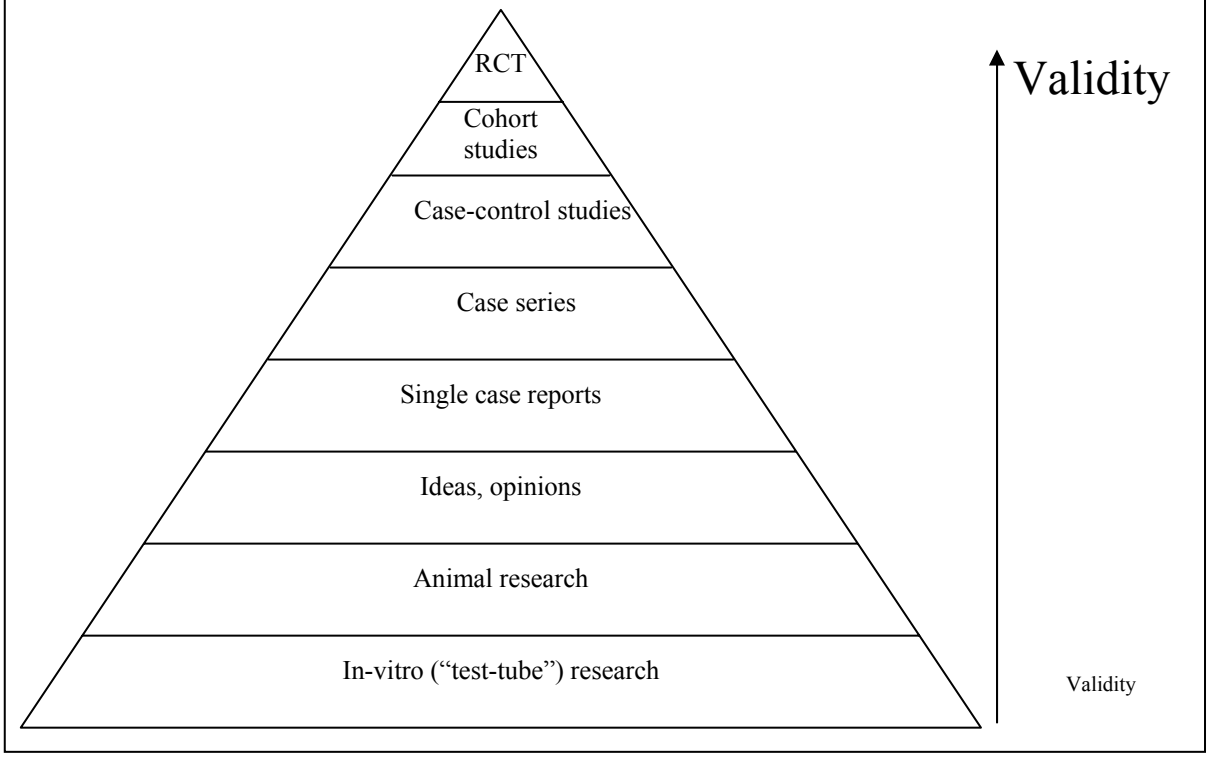
In this context, **evidence** is understood as the product of systematic observation or experiment and it is inseparable from the notion of data (McQueen and Anderson 2001). That is, evidence based medicine is to be seen as a positivist approach. The results from research, that is from empirical *observation* or from *experiments*, in which data are systematically collected and rigorously analysed following a pre-established plan, represent the evidence which, ideally, underpins the decisions. The results from research are usually published in scientific journals, thus searching for the best evidence can be seen as synonymous with searching the literature for results of studies (i.e. for publications)[1].

In addition, the idea to base decisions on the **best available** evidence implies a "hierarchy" of the evidence. Since the evidence comes from research, one can say that there is an underlying hierarchy of research designs. There are pieces of research (i.e. trials) which are considered to

[1] Although it is desirable (and sometimes necessary) to search from evidence from sources other than the published literature, this is always not possible because of resources constraints. Many systematic reviews and assessment focus mainly on published results. Most of the work done on the classification and appraisal of the evidence has been concentrated on evidence in published form, and specially concerning benefit from interventions.

be *better* and, consequently, other which are considered to be *worse*. Evidence from *good* research will be therefore considered to be *better* as evidence resulting from *worse* research. A hierarchy of research design is depicted in Figure 1, whereby the width of the levels represents the amount of published results.

Figure 1. Hierarchy of research designs to base health policy decisions upon (based mainly on internal validity)



For the evaluation of the benefit from an intervention (e.g. reducing mortality from an specific cause) evidence from *experiments* is considered to be better than evidence from non-experimental observations, and among experiments, some *study designs* (e.g. including an explicit comparison group) are considered to be better than others, thus ranking higher in the hierarchy of research design. The main underlying rationale of the hierarchy is the level of **internal validity**. The level of internal validity tells us how likely it is that an observed benefit from an intervention is in fact attributable to that intervention. Put simply, when we observe a benefit from a given intervention, there are two possible explanations:

1. The first one is that the observed benefit has really been caused by the intervention itself, e.g. a reduction in mortality is fully attributable to the intervention.
2. The second possibility is that the observed benefit *seems* to be caused by the intervention but in fact other factors are responsible for the observed cause-effect association, and the intervention itself does indeed not have any benefit (or even produces harm). These factors may be chance, errors in collecting or interpreting the data (so called *bias*) or the effect of additional variables (so called *confounding*) (Hennekens and Buring 1987).

It has been demonstrated, that the most appropriate study design to minimise these effects, i.e. with the higher level of internal validity, is the so-called "randomised controlled trial" (RCT)[2], with an adequate number of participants. Therefore this design takes the highest

[2] In such a trial, at least to groups are built, for example one to receive the intervention and another one to receive a placebo intervention. The assignment of the participants to one of the groups occurs by chance (random), in order to minimise the effects of selection bias or of confounding.

place in the pyramid of research designs. The higher in the hierarchy, the more internal validity has a research design, and the more can we be sure that an observed benefit from the intervention is in fact attributable to it. As we go down in the pyramid, it is more likely that the reality looks different as the findings from the study make us think.

The extent to which the information provided by a study has significant clinical or policy relevance has been defined as the "non-methodological quality" of the evidence (Lohr & Carey 1999). In the pyramid shown above research on animal models or in test tubes ranks lowest, because the non-methodological quality of its evidence is very poor[3]. This is due to the fact, that one can not draw any direct conclusions about the benefit of an intervention (e.g. in reducing mortality) from this kind of evidence.

The Task Force on Community Preventive Services (TFCPS) has developed a more elaborated algorithm to help classify the different types of research on health interventions, according to its **suitability** for assessing the effectiveness of interventions (Briss et al. 2000). Depending on their internal validity, single pieces of research are classified in different categories of suitability (i.e. again a kind of hierarchy of research). This approach is very interesting because it takes into account, that the RCT is not always the most appropriate (or feasible)[4] research design, and also because it provides a good and detailed systematisation of the different kinds of research which can be used to evaluate the effectiveness of an intervention. In Figure 2 the link between the algorithm and the classification of suitability is presented.

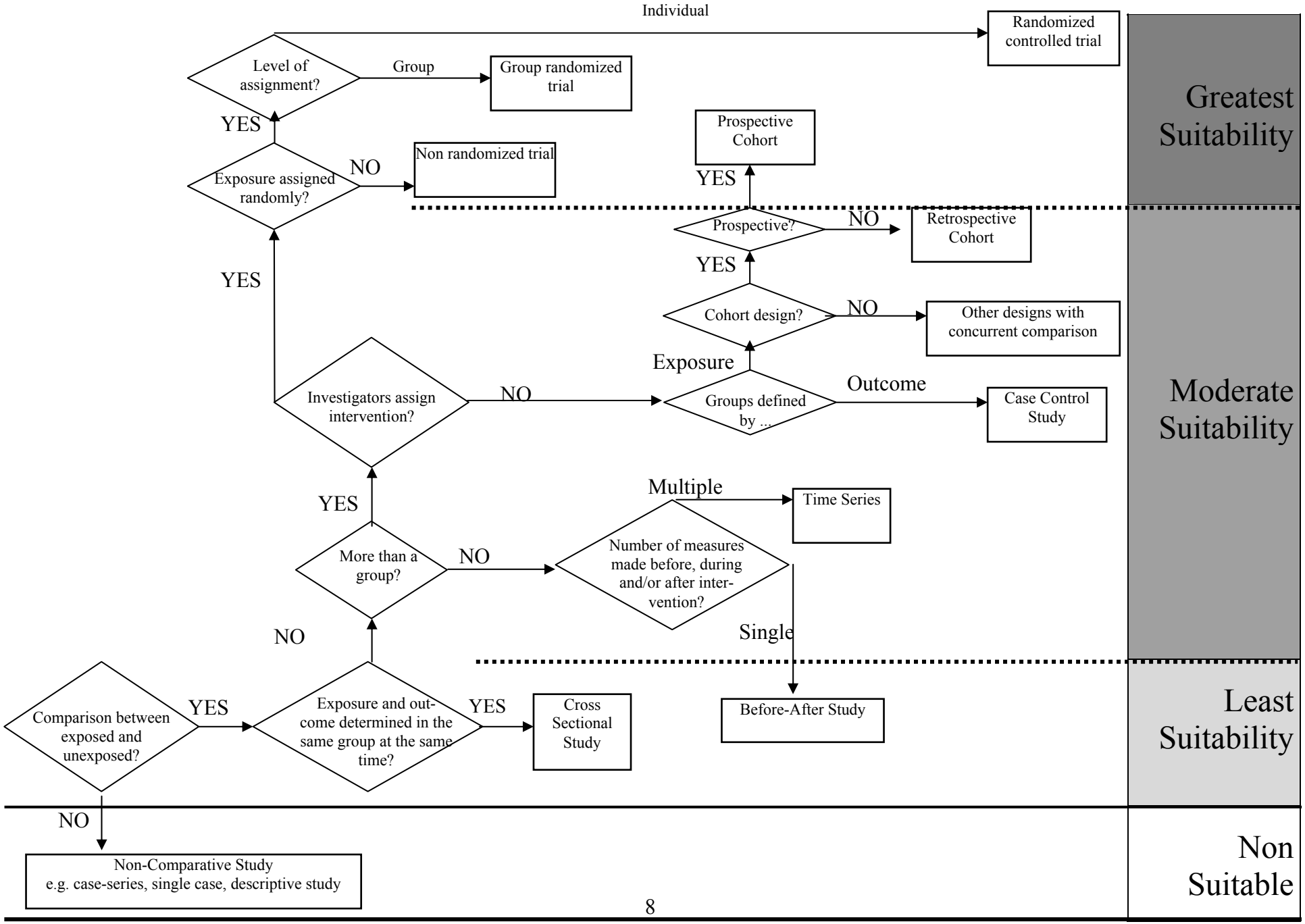
In summary, pieces of evidence are considered better or worse according to their internal validity and to their relevance to the questions to be answered.

The approaches shown above allow for a gross classification of the available evidence into different levels of *goodness* (i.e. validity + relevance). This gross approach can be used to limit the types of studies which are taken into account when evaluating the effectiveness of an intervention. So, with the help of the hierarchy of research designs one can set a threshold concerning the types of research to be considered in the evaluation. For example, in the systematic reviews of the Cochrane Collaboration the threshold is usually set on the RCT design. Generally, these kinds of review do not account for evidence from other study designs, as they consider that these have too high risk of bias, so producing misleading results. In contrast, the TFCPS sets the threshold based on the existence of a comparison group. As a result of this, studies including any kind of explicit comparison group are considered to be suitable (and thus are taken into account), whereas studies which do not include any kind of comparison (e.g. case description) are consequently excluded from the review (see Figure 2).

[3] Animal or test-tube research has usually high internal validity and undoubtedly it contributes to the understanding of physiology and pathophysiology, and serves in the development of diagnostic and therapeutic interventions. However, findings from this kind of research can not be extrapolated to individuals, thus, it is not relevant when assessing the benefit from an intervention as measured for example by reductions in mortality.

[4] RCTs are expensive. Further, RCTs follow very strict protocols for the selection of participants, and for the delivery of the interventions, in order to warrant the high level of internal validity stated above. This, however, makes the results from RCTs to be difficult to generalize to the situation in every-day-care.

Figure 2. Algorithm of Study Design and Levels of Suitability (adapted from Briss et al. 2000)



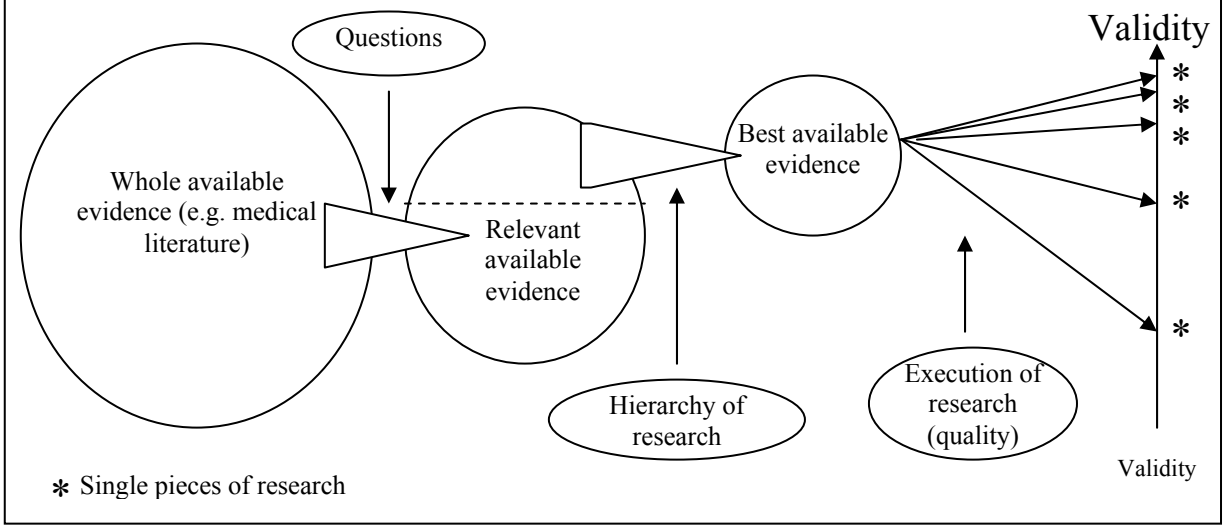
Thus, the way such thresholds are applied, determines the interpretation of statements like "*no evidence was found*", "*there is no conclusive evidence*" etc., which can be often found as the result from systematic evaluations of interventions.

However, within the same level (e.g. "moderate suitability" or "RCT"), there are also differences on the validity, due to differences in the *execution* of the study. For example, a RCT which has a big number of patients who left prematurely the study (so called *lost to follow up*), or in which no measures were taken to avoid different kinds of bias, may have a very limited validity. The limitations to validity can be so relevant, that even though the design put the study on the highest level of the hierarchy, the results have to be considered as lower level evidence due to the actual conduction of the study.

Similarly, very well conducted cohort studies, in which measures were taken to avoid the influence of bias or confounding, may present a very high validity, which in some cases may make them comparable to RCT. Several tools have been developed to assess and score the **quality of execution/conduction** of single studies, following the same rationale of the gross hierarchy (i.e. *higher quality = higher internal validity = higher level of evidence*). A recent review identified 67 systems to assess and/or score the quality of different study designs, most of them have been developed for assessment of the quality of RCT (West et al. 2002). Using this systems, one can order a group of studies with the same design according to the quality of their conduction, obtaining again a hierarchy. This approach allows to organise the available evidence before conclusions and recommendations can be drawn.

In a very simplified way, Figure 3 shows the process of selection and organisation of the evidence, as it takes place in HTA (or in the conduction of systematic reviews).

Figure 3. Simplified process of selection and organisation of the evidence.



The **body of evidence** (that is the group of studies selected as the best available) will be then characterised using a combination of both factors discussed above, i.e. the hierarchy of research design and the quality of execution. In addition, other factors such as the number of studies, the size of the effect and the homogeneity/consistency of results across the group of studies are also relevant when judging the **strength of the evidence**. The matter here is to judge the evidence (saying for example "there is strong evidence that…") in order to give answers to the research questions and, in the end, to the policy questions (saying for example

"thus it is strongly recommended to…"). Several approaches have been developed to standardise the way researchers make their judgements about the strength of the evidence which will underlie their recommendations. A recent review has identified 40 different systems to rate the strength of the evidence, which differ in the combination of factors stated above used to rate the strength of evidence (West et al. 2002).

Besides rating the strength of evidence these systems also establish a link between the strength of the evidence and the **grade of recommendation**, say the **strength of the recommendation**. Strong evidence allows for strong recommendations, and the other way round, based on weak evidence only weak recommendations can be made. The use of an standardised system for grading recommendations allows the readers to quickly identify the underlying strength of evidence, given that the reader knows how the system works. We will now exemplary present three of these systems (for a detailed description of most of the systems published in the English-speaking literature, confer to West et al. 2002).

(1) The first approach to describe the strength of evidence in a standard way was the one of the former Canadian Task Force on the Periodical Health Examination[5] in 1979 (Table 3). The system is mainly used to decide whether to include or not to include interventions (e.g. screening) in the periodical health examination (PHE). This system has served as the base for the development of many of the subsequent ones. The strength of the evidence is judged taking into account the hierarchy of research designs and the number of studies. The strength of evidence is classified into five levels.

Table 3. Strength of evidence of the Canadian Task Force on Preventive Health Care (1979)

Strength of Evidence	*Description*
I	Evidence from at least one 1 properly randomized controlled trial
II-1	Evidence from well-designed controlled trials without randomization
II-2	Evidence from well-designed cohort or case-control analytic studies, preferably from than 1 center or research group
II-3	Evidence from comparisons between times or places with or without the intervention. Dramatic results in uncontrolled experiments could also be included here.
III	Opinions of respected authorities, based on clinical experiences, descriptive studies or reports of expert committees.

Based on the strength of evidence, five different grades of recommendation can be achieved (Table 4), although no explicit link between both tables has been stated. The system leaves a category (C) for inconclusive evidence, and two grades of strength of the evidence for each recommending for (A=stronger, B=weaker) and against (D=weaker, E=stronger).

Table 4. Grades of Recommendation of the Canadian Task Force on Preventive Health Care (1979)

Grade of recommendation	*Description*
A	Good evidence to support the recommendation that the condition be specifically considered in a Periodical Health Examination.
B	Fair evidence to support the recommendation that the condition be specifically considered in a Periodical Health Examination.
C	Poor evidence regarding inclusion or exclusion of a condition in a Periodical Health Examination, but recommendations may be made on other grounds.
D	Fair evidence to support the recommendation that the condition be specifically excluded from consideration in a Periodical Health Examination.
E	Good evidence to support the recommendation that the condition be specifically excluded from consideration in a Periodical Health Examination.

[5] Today called Canadian Task Force on Preventive Health Care (CTFPHC).

(2) A very popular system is the one developed by researchers of the Oxford Centre for Evidence Based Medicine (Table 5). The strength of the evidence is rated taking into account the hierarchy of research design, the quality of execution, the number of studies available and the consistency of results across studies. Thus systematic reviews with homogeneity of results are considered the highest level of evidence. Systematic reviews per se represent a summary of the evidence on one topic, and if they found consistent results from different studies, with a sufficient amount of studies included and a big effect size one can conclude that the evidence they provide is strong.

This system has an explicit link to the grade of recommendations, which is directly derived from the strength of evidence. The system has several types of hierarchies depending on the type of question (therapy, properties of diagnostic tests, prognostic factors, etc.). The one presented in Table 5 refers to interventions (preventive or therapeutic) and is intended to be useful for the evaluation of the benefit or harm from the intervention. The grading of recommendations is purely hierarchical (A=strongest category, E= weakest category).

Table 5. Levels of Evidence and recommendations from the Oxford Centre for Evidence Based Medicine (www.eboncall.co.uk/content/levels.html, accessed May 2003)

Grade of Recommendation	*Strength of Evidence*	*Description*
A	1a	Systematic Review of Randomised Controlled Trials (with homogeneity)
	1b	Individual Randomised Controlled Trial (with narrow confidence interval)
	1c	All or none (i.e. spectacles to correct myopia)
B	2a	Systematic Review of cohort studies
	2b	Individual cohort study / low quality randomised controlled trial
	2c	"Outcomes" research/Ecological studies
	3a	Systematic review of case-control studies
	3b	Individual case-control study
C	4	Case-series, poor quality cohort and case-control studies
D	5	Expert opinion without explicit critical appraisal, or results from animal experiments, etc.

Using these two examples, it is easy to show, that in the language of recommendations the letters do not always mean the same. If we have strong evidence that an intervention is harmful and want to strongly recommend to exclude it from our health care package, and we are using the system of Oxford we will grade our recommendation as "A" – in contrast, if we are following the system of the CTFPHC we will grade it as "E". It is important to note that in the CTFPHC system "A" indicates a strong recommendation in favour and in the Oxford system "E" indicates very weak evidence!

(3) The Task Force on Community Preventive Services from the USA (TFCPS) has recently also developed a system to rate the strength of evidence (Table 6). The rating of the strength of the evidence on effectiveness is based on the hierarchy of research designs (see Figure 2), the quality of execution, the amount of evidence, the consistency and the size of observed effects. The strength of recommendations is directly derived from the strength of evidence.

Table 6. Strength of evidence and Strength of Recommendations (Briss et al. 2000)

Strength of Recommendation	*Strength of evidence*	*Quality of execution*	*Design suitability*	*Number of studies*	*Con-sistency*	*Effect Size*	*Expert Opinion*
Strongly Recommended (Discouraged, if harmful)	Strong	Good	Greatest	At least 2	Yes	Sufficient	Not used
		Good	Greatest or Moderate	At least 5	Yes	Sufficient	
		Good or Fair	Greatest	At least 5	Yes	Sufficient	
		Evidence meets criteria for "Sufficient" but effect is …				Large	
Recommended	Sufficient	Good	Greatest	1	NA	Sufficient	Not used
		Good or Fair	Greatest or Moderate	At least 3	Yes	Sufficient	
		Good or Fair	Greatest, Moderate or Least	At least 5	Yes	Sufficient	
Recommended based on expert opinion	Insufficient empirical informa-tion, supported by expert opinion	varies	varies	varies	varies	varies	Supports a recom-menda-tion
Available studies do not provide sufficient evidence to assess	Insufficient	A: Insufficient designs or execution		B: too few studies	C: incon-sistent	D: Small	Not used

We have briefly presented three approaches to judge the strength of the evidence in a standardised manner. However, all of them leave some room for subjectivity. For example, the systems of the CTFPHC and Oxford do not define in detail what is meant by "poor" or "good" quality of execution. The TFCPS does not explicitly define what is to be understood under "large" or "sufficient" effect size. Similarly, the Oxford system does not clearly state how wide a confidence interval may be to still being considered to be "narrow".

The examples also show that the "strength" of "strong evidence", and thus of "strong recommendations", varies according to the system used is not always the same. Therefore it has been recommended that, whenever such a system is used in technology assessment reports, the authors should provide enough information for the reader to interpret the grading of recommendations used (Busse et al. 2002). The following Table 7 summarises the factors taken into account to rate the strength of the evidence, in the systems presented above.

Table 7. Factors used to judge the strength of evidence in three different systems

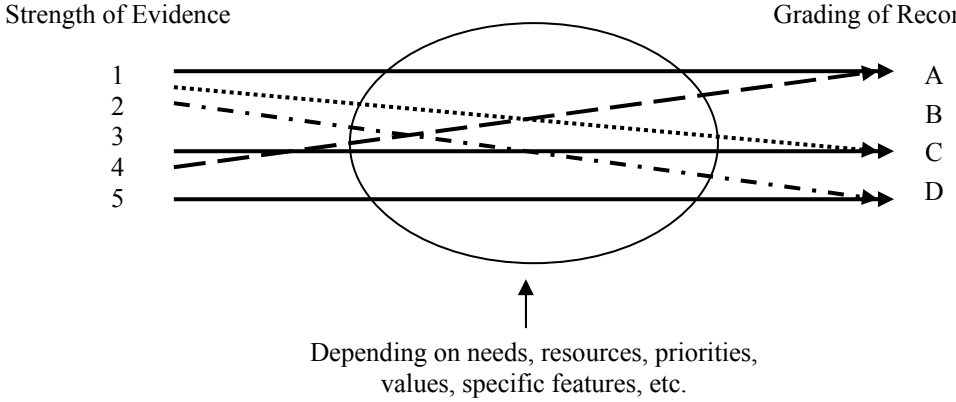
System	*Hierarchy of research design*	*Quality of execution*	*Number of studies*	*Consistency*	*Effect size*
CTFPHC	+	+/-	+	-	-
Oxford	+	+/-	+/-	+	-
TFCPS	+	+	+	+	+/-

The systems presented above are thought to base recommendations on the evidence about the benefit (or harm) derived from an intervention. However, when decisions are made in the health care system, especially those concerning allocation of resources, other important factors should be taken into account.

So, even with strong evidence of benefit from an intervention, the recommendations for policy may vary as the burden of disease, needs and priorities, cost and cost effectiveness

issues, special features of the system, barriers to implementation etc. need to be considered (Gray et al. 1997, Council of Europe 2001). These factors may modify the translation from strong evidence to recommendation grades. This issue has been stressed and presented graphically in the report on guidelines development compiled under the auspices of the Council of Europe (Figure 4). Given any system to rate the strength of evidence (left column) the recommendations made based on that evidence may vary in their strength (right column), depending in other factors such as resources available, priorities, need for decision or issues of transferability. Research evidence below a certain threshold, however, should never be used for strong(er) recommendations (in the figure, this is symbolised by the arrow leading from 5 to D only).

Figure 4. Relation between Strength of Evidence and Grade of Recommendations (Council of Europe 2001).



Health technology assessment explicitly takes such factors into account when evaluating an intervention. Other approaches, like the systematic reviews (e.g. Cochrane Collaboration), limit themselves usually to issues of benefit and harm. Health technology assessment can use findings from systematic reviews to assess some aspects. However, to provide useful support for decision-making the factors stated above also need to be considered.

Summary of Sections 1+2

- Health Technology Assessment (HTA) is an approach with the explicit aim of supporting the process of decision making at the policy level of health care with best evidence.
- Ideally, HTA is conducted commissioned by policy-/ decision-makers to support a particular decision process (policy question).
- Researchers doing HTA should translate the policy question in close co-operation with commissioners into research questions answerable with scientific evidence.
- Evidence is understood in this context as the product of systematic observation or experiment and it is inseparable from the notion of data.
- The concept of best evidence implies a hierarchy of research designs.
- Pieces of evidence are considered better or worse according to their internal validity, other methodological factors and to their relevance to the questions to be answered.
- The strength of the available evidence needs to be judged before making conclusions or giving any recommendations.
- The factors that determine the strength of the evidence are:
  - Validity of research

    - hierarchy of research designs
    - quality of execution
  - Number of studies
  - Size of the effect
  - Consistency of effect across studies
  - Applicability
- Different systems have been developed to assess the strength of evidence in a standardised way.
- The strength of evidence *plus* other factors like
  - Burden of disease
  - Trade-offs harm-benefit (safety-effectiveness)
  - Resources (issues of cost and cost-effectiveness)
  - Priorities
  - Values
  - System-specific features determine the grading of recommendations.
- Due to the diversity of systems used to judge the strength of the evidence and the grades of recommendations, any report using such an approach should clearly state the way the grading system is designed – otherwise, there is a risk of misinterpretations.

Section 3. Evidence basis for two selected interventions

In this section, we will present two scenarios: Prostate Cancer Screening and Nicotine Replacement Therapy. For each we will state a possible policy question and translate it into research questions, as one would do prior to the conduction of an assessment. Then, we will present the conclusions and recommendations from existing assessments and/ or systematic reviews. We will discuss the available evidence and the way it was handled to give recommendations concerning these topics. The exercise does not intend to be exhaustive. We focused our search on HTA reports from INAHTA-members[6] and systematic reviews from the Cochrane Collaboration[7]. We also searched in the recommendations of the "Community Guide"[8]. We will refer only to the evidence collected and synthesised in the reports we analysed. We did not collect any further evidence (e.g. from primary research). The aim is to show what decision makers can expect from such reports.

[6] INAHTA is the International Network of Agencies for Health Technology Assessment. Currently INAHTA has 40 member agencies from 20 countries. The internet site of INAHTA (www.inahta.org) has links to many of them and provides also access to the HTA-Database, a searchable collection of HTA reports, and similar documents.

[7] The Cochrane Collaboration is an international, not-for-profit organisation that aims to help people make well-informed decisions about health care by preparing, maintaining and promoting the accessibility of systematic reviews of the effects of healthcare interventions. The reviews of the Cochrane Collaboration are mostly based on evidence from RCTs.

[8] The "Community Guide" is a collection of evidence based recommendations on public health interventions, which has been launched by the Task Force on Community Preventive Services (USA).

Scenario 1. Prostate Cancer Screening

- *Policy Question*: You are planning to establish a "health check-up" programme for the population of your community. The main aim of the check-up is to reduce mortality from avoidable causes, and reduce the burden of disease (suffering, economical) by systematically identifying people at risk, or by early diagnosis of potentially treatable conditions (like cancer). You are shaping now the contents of the health check-up and would like to know whether to include prostate cancer screening with the help of a simple and cheap test for prostate specific antigen (PSA).
    - *Should the health check-up in my community include PSA-testing for screening of all men for prostate cancer?*

- *Research Questions*: You contact the researchers of your HTA unit and ask them for advice. In co-operation with you they formulate the following questions:
    - *Is PSA-testing an effective intervention?*
        - Does PSA-testing contribute to a reduction of mortality from prostate cancer?
        - Does PSA-testing contribute to a reduction of overall mortality?
        - Does PSA-testing contribute to improving the quality of life of men suffering from prostate cancer?
    - *Besides effectiveness, which other issues should be taken into account to consider when deciding upon the inclusion of PSA-testing?*

- *The assessment process*: The researchers of your HTA unit search systematically for evidence to answer the research questions. They appraise the validity of the evidence, and judge the strength of the evidence. They write a report in which they (attempt to) give you an answer to your question.

- *Your decision* after reading the report: _____

We will present the ten most recent reports on prostate cancer screening that we found by systematically searching in the sources stated above. Seven full reports were available and three were executive summaries. The three most recent documents were from Canada (two) and the USA. Two documents were from Spain (one of them representing a joint project with Sweden), two documents were from UK, and one each from France, Germany and Norway respectively. The recommendations given in the reports and the underlying evidence basis is presented in the following tables. The full references can be found at the end of the document.

The reports range from 1997 to 2002. Only the recommendations from the USTFPS used an explicit, standardised system to rate the strength of evidence and grade its recommendations. Although the other did not follow such an standardised approach, the underlying notion of evidence and the judgements about the strength of overall evidence followed implicitly the principles explained in Section 2. In Table 8, we have rated the evidence using the system of Oxford to allow comparisons between the reports. The evidence basis underlying the conclusion and/or recommendations was similar in all of the reports. Some reports summarised the conclusions from previous assessments.

All reports agree on the fact that there is **no conclusive evidence** regarding the effectiveness of prostate cancer to reduce mortality. None of the reports claimed to have found evidence

from the highest level showing that PSA-screening contributes to reduction of mortality from prostate cancer. However, there was some evidence of less strength, indicating the possibility of some benefit from screening. Two ongoing RCTs are expected to provide evidence of the highest level around 2008.

Most of the reports recommended **against** introducing **mass screening** for prostate cancer. Some limited themselves to acknowledge the lack of evidence without given any clear recommendations. The recommendation was mainly founded on the **lack of evidence** that the intervention could contribute to reduce mortality from prostate cancer. Here is to be noted that *lack of evidence of benefit does not mean evidence of no benefit*. The manual for reviewers of the Cochrane Collaboration warns against what they considered a common mistake, "to confuse 'no evidence of effect' with 'evidence of no effect' (Clarke et al. 2003). With the available evidence it is not possible to say whether the intervention is effective or ineffective.

Thus any recommendations given have to be based on factors other than benefit (effectiveness). This was the case here. One of these factors was the fact that two ongoing RCTs are expected to shed some light on the topic. The recommendation to await for such results implies judgements about the urgency to take action, which in this case seems to be determined by, for example, the burden of disease or the existence of an effective therapy for prostate cancer. In the reports reviewed here, the burden of disease has been considered not to justify the intervention. Incidence and prevalence of prostate cancer seem to indicate an important public health problem, however due to the natural history of the condition, most men will not die of it. The lack of an effective therapy, and especially the harms associated with the existing therapeutic alternatives, seems to have been given also an important weight in the reports. Some of the reports explicitly checked the suitability of screening for prostate cancer against the criteria of Wilson and Jungner (1968), i.e.:

(1) The condition being screened for should be an important health problem.
(2) The natural history of the condition should be well understood.
(3) There should be a detectable early stage.
(4) Treatment at an early stage should be of more benefit than at a later stage.
(5) The risks, both physical and psychological, should be less than the benefits.
(6) The costs should be balanced against the benefits.

They acknowledged that, independently of the lack of evidence concerning the effectiveness of screening, most of the criteria were not fulfilled. The fulfilment of these criteria was assessed with evidence from research, however it was not always clear how these evidence was collected (i.e. systematic search) or appraised. Some authors recognised that evidence, for example, on harm from the screening intervention was indirect. It can be understood, that even if screening for prostate cancer could be shown to be effective, it could be recommended against it, due to other factors (refer also to Section 2).

Some of the reports recognised however, the difficulty to control the diffusion of the intervention and pleaded for assuring that men undergoing the test are fully informed about the potential consequences of the test.

Prostate Cancer Screening

- There is no evidence supporting effectiveness (a reduction in mortality from prostate cancer) of PSA-screening.
- There is also no evidence supporting ineffectiveness of PSA-screening.
- Most HTA agencies have recommended against PSA-test mass screening.

- These recommendations are based on the one side in the lack of evidence of effectiveness and on the other side on other factors which indicate that results from ongoing trials should be awaited.

Table 8. Prostate Cancer Screening				
Source	**Year**	**Target Group**	**Recommendations***	**Comments**
ICES (1) (full report)	2002	Asymptomatic men	"The potential for over-treatment of some prostate cancers not destined to cause future mortality, the uncertainty about the benefits of aggressive treatment of screen-detected cancers, the lack of evidence from randomized trials of the effectiveness of PSA screening and the relative high costs of prostate cancer screening programs combine to suggest that *a program of PSA screening of asymptomatic men should not be introduced at this time*." Two policy options were given: o "Continue the status quo: Health care resources are limited, and many believe that resources should be preferentially directed to tests and therapies that have been shown to be effective. Continuing the policy of not covering PSA screening for asymptomatic men is consistent with this evidence-based approach, and with other MOHLTC policies such as only paying for drugs that have demonstrated to be cost-effective." o "Provide PSA testing on request, with informed consent: The lack of evidence about the effectiveness of PSA screening is not the same as knowing that PSA screening is ineffective. The biological rationale for PSA screening is reasonably strong, and the decrease in prostate cancer mortality short after PSA screening was introduced is intriguing. Many tests and therapies are paid for by the MOHLTC without definitive proof of benefit, and it could be argued, that a PSA screening test should be paid for if men are fully informed about its potential benefits and risks. Men should be given a decision aid describing the options and their consequences, and should indicate that they have fully understood the information provided."	The authors did not find results from evidence of the highest level. However, two ongoing studies of this level were identified which will provide results in 2008. Based on this lack of highest level evidence and considering the potential harmful consequences of screening, the authors concluded that PSA screening should not be introduced. However, they could not agree on the recommendations for policy, and two options were given. The disagreement seems to be due to the different weight given to the results from the "best available evidence" which was not classified according to any hierarchy of evidence.
	Country	**Intervention**		**Evidence Body**
	Canada (Ontario)	PSA Screening		o Lack of conclusive evidence from RCTs (two ongoing RCTs identified) o Two 2b studies (a prospective cohort and a low quality RCT) o 15 studies ranging 2c to 4 o Recommendations from 18 professional societies, similar bodies and HTA agencies, 13 of them recommending against screening of asymptomatic men.
		Main Outcome		
		Mortality (overall and prostate cancer specific)		

<table>
<tr><th>Source</th><th>Year</th><th>Target Group</th><th>Recommendations*</th><th>Comments</th></tr>
<tr><td rowspan="4">USPSTF (2)

(based on a full report from AHRQ)</td><td>2002</td><td>Asymptomatic men</td><td rowspan="4">“The U.S. Preventive Services Task Force (USPSTF) concludes that evidence is insufficient to recommend for or against routine screening for prostate cancer using prostate specific antigen (PSA) testing or digital rectal examination (DRE).

I[nsufficient] recommendation.

The USPSTF found good evidence that PSA screening can detect early-stage prostate cancer but mixed and inconclusive evidence that early detection improves health outcomes. Screening is associated with important harms, including frequent false positive results and unnecessary anxiety, biopsies, and potential complications of treatment of some cancers that may never have affected a patient’s health. The USPSTF concludes that evidence is insufficient to determine whether the benefits outweigh the harms for a screened population.”</td><td rowspan="2">The authors did not find any evidence from RCT, the evidence from other study designs was inconclusive concerning mortality benefits with prostate screening. They stated that it was not possible to judge whether the benefits outweigh the harms.
The authors did not give any recommendations (neither for nor against) because evidence was insufficient (“I”).
The strength of evidence was classified as good, fair or poor according to the own system of the USTFPS. The system used was explained in the paper.</td></tr>
<tr><td>Country</td><td>Interventions</td><td>Evidence Body</td></tr>
<tr><td rowspan="2">USA</td><td>PSA Screening, DRE Screening</td><td rowspan="2">o Lack of conclusive evidence from RCTs (two ongoing RCTs)
o One 2b study (a poor quality RCT)
o Several studies ranging 2c to 3b
o Recommendations from 7 professional societies considered, none of them gave any recommendations regarding prostate cancer mass screening. The focus was on individual screening.</td></tr>
<tr><td>Main Outcome</td></tr>
<tr><td></td><td>Prostate cancer mortality</td><td></td><td></td></tr>
</table>

<table>
<tr><th>Source</th><th>Year</th><th>Target Group</th><th>Recommendations*</th><th>Comments</th></tr>
<tr><td rowspan="4">HSURC (3)

(full report)</td><td>2000</td><td>Asymptomatic men</td><td rowspan="4">“Key questions remain to be answered for prostate cancer screening and treatment. Generally, this literature review did not uncover any research advancements that were inconsistent with the current HSRUC guideline.”

”Key questions remain for prostate cancer and treatment:
1. Mortality reductions must be shown in randomized trials.
2. Treatment still has considerable effects.
3. Quality of life should be evaluated for prostate cancer screening. since years of life saved are at older ages.
4. Cost effectiveness of prostate cancer screening needs to be established.”

“Prostate cancer screening using the PSA test still lacks evidence of mortality reduction from prostate cancer […]”</td><td>This report did not contain a section “Conclusions” or “Recommendations”. Such statements were spread out through the report text, especially in the “Summary” and “Discussion” sections.
The report was an update of a previous one, which recommended against mass screening in asymptomatic men.</td></tr>
<tr><td>Country</td><td>Intervention</td><td>Evidence Body</td></tr>
<tr><td rowspan="2">Canada (Saskat-chewan)</td><td>PSA Screening</td><td rowspan="2">o Lack of evidence on effectiveness from RCTs (two ongoing studies identified)
o One RCT of poor quality (2b)
o Recommendations and guidelines from 11 professional and learned societies were compared (10 of them against mass screening).
o Results from 8 systematic reviews and HTA reports (all of them were against mass screening)</td></tr>
<tr><td>Main Outcome</td></tr>
<tr><td></td><td></td><td>Prostate cancer mortality</td><td></td><td></td></tr>
</table>

<table>
<tr><th>Source</th><th>Year</th><th>Target Group</th><th>Recommendations*</th><th>Comments</th></tr>
<tr><td rowspan="4">ANAES (4)

(executive summary)</td><td>1999</td><td>Asymptomatic men</td><td rowspan="4">“So, in 1998, the benefit of mass screening for prostatic cancer has not been established.”

“In conclusion, current knowledge does not support any recommendation for mass screening for prostatic cancer.”</td><td>The authors did not find any evidence from RCT, and thus concluded that there was no demonstrated benefit from prostate cancer screening. Moreover, the authors checked the justification of screening against a set of well accepted, theory-driven criteria from WHO (Wilson & Jungner 1968). These criteria were not fulfilled.</td></tr>
<tr><td>Country</td><td>Intervention</td><td>Evidence Body</td></tr>
<tr><td rowspan="3">France</td><td>PSA Screening</td><td rowspan="3">o Lack of evidence on effectiveness from RCTs (three ongoing studies identified)
o Reports and recommendations from 15 HTA organisations or similar were accounted for.</td></tr>
<tr><td>Main Outcome</td></tr>
<tr><td>Prostate cancer mortality</td></tr>
</table>

Source	Year	Target Group	Recommendations*	Comments
DIMDI (5) (full report)	1999	Asymptomatic men	"[…] yet there is no evidence supporting routine PSA-screening for prostatic cancer. Generalised access to the PSA-test should not be warranted, because of the considerable problems linked to the interpretation of the PSA-test and because of the generally accepted need for comprehensive information of the patient previous to the test. However, it could be considered to link the use of the PSA-test to quality assurance […]" "Given the international consensus that both screening and therapy trials should have a follow-up of at least 10 years, it does not seem appropriate to start other evaluation studies till the results of ongoing trials are available." [Translated from German by MVG]	The author analysed the results of previous HTA reports. The recommendation not to provide mass screening seems to be in disagreement with the recommendation to link PSA-screening to quality assurance provisions. With this recommendation, the author seems to face the fact that PSA-screening is being offered on an individual basis in Germany (occasional screening) without any control. His recommendation is intended to, at least, assure that patients are well informed about the consequences, and to reduce potential harm. This recommendation is compatible with the statements from guidelines, which focused on individual screening.
	Country	**Intervention**		**Evidence Body**
	Germany	PSA Screening		o Recommendations from 8 HTA reports were analysed (all of them recommending against mass screening). o Guidelines from two professional societies were also considered (both did focus on individual screening).
		Main Outcome		
		Prostate cancer mortality		

<table>
<tr><th>Source</th><th>Year</th><th>Target Group</th><th>Recommendations*</th><th>Comments</th></tr>
<tr><td rowspan="4">INAHTA (6)

(OSTEBA SBU joint project, full report)</td><td>1999</td><td>Asymptomatic men</td><td rowspan="4">“Mass screening for prostate cancer is not recommended because of lack of evidence regarding the benefits and the considerable risks of adverse effects of subsequent treatment.”</td><td>The authors summarized the findings from other HTA reports and updated the literature search. They did not identify any studies which could change the recommendations of previous reports.
They also checked for the suitability of any kind of screening for prostate cancer using the criteria of Wilson & Jungner (1968).</td></tr>
<tr><td>Countries</td><td>Interventions</td><td>Evidence Body</td></tr>
<tr><td rowspan="3">Spain (Basque Country) & Sweden</td><td>PSA Screening, DRE Screening, TRUS Screening</td><td rowspan="3">o Lack of Evidence on effectiveness from RCTs (two ongoing studies identified)
o Reports from 9 HTA organisations (all of them recommending against mass screening) were summarized and updated.</td></tr>
<tr><td>Main Outcome</td></tr>
<tr><td>Prostate cancer mortality</td></tr>
</table>

Source	Year	Target Group	Recommendations*	Comments
SMM (7) (executive summary)	1999	Asymptomatic men	"This group of experts agrees with the main conclusion from INAHTA-report; routine screening of asymptomatic males for early diagnosis of prostate cancer is not recommended. Any clinical benefit (such as reduced mortality) must be considered in the light of unnecessary medical examinations and treatment as well as the considerable risk of adverse effects following medical interventions (surgical removal of the prostate gland) which in many patients reduce the quality of life." "Widespread screening without clinical indication as alluded to above, is not effective and does not contribute to improved health of the population." "SMM would like to emphasize that, due to the insufficiencies in the current level of knowledge on prostate cancer, we lack the fundamental scientific knowledge to on which to base decisions regarding those medical interventions best suited to prevent, diagnose and treat this disease."	Given the existing amount of systematic works, the authors did not conduct further research. A group of experts analysed the evidence from the other reports and based on them its recommendations.
	Country	**Intervention**		**Evidence Body**
	Norway	PSA Screening		○ Reports from 9 HTA organisations were summarised.
		Main Outcome		
		Prostate cancer mortality		

<table>
<tr><th>Source</th><th>Year</th><th>Target Group</th><th>Recommendations*</th><th>Comments</th></tr>
<tr><td rowspan="5">AETSA (8)

(full report)</td><td>1998</td><td>Asymptomatic men</td><td rowspan="5">“[…] the reports considered consistently do not recommend the introduction of this type of programs, because of the high level of uncertainty about the problem, and because, in the best case, a very low level of effectiveness in reducing incidence and disease specific mortality can be achieved, at the expense of a considerable rate of adverse effects, and high costs.”

“According to the facts stated above, considering the effectiveness of the available test, and as long as results from controlled trials are not available, it is not recommended to implement a program for early detection of prostate cancer. However it is necessary to develop clinical criteria for the indication of the test and for information of the patient in the form of clinical practice guidelines.”

[Translated from Spanish by MVG]</td><td>The authors summarised the findings from other HTA reports and discussed it in the light of the situation in Andalusia.
The authors recommended against mass screening. The authors seem to have also recognised the problem of uncontrolled occasional screening and advocate for arrangements to control this.</td></tr>
<tr><td>Country</td><td>Interventions</td><td>Evidence Body</td></tr>
<tr><td rowspan="3">Spain (Andalusia)</td><td>PSA Screening, DRE Screening, TRUS Screening</td><td rowspan="3">o Reports from 9 HTA organisations were summarised.
o Recommendations from two clinical guidelines were also analysed.</td></tr>
<tr><td>Main Outcome</td></tr>
<tr><td>Prostate cancer mortality</td></tr>
</table>

Source	Year	Target Group	Recommendations*	Comments
NCCHTA (9) (full report)	1997	Asymptomatic men	"There is no justification for the routine use of PSA testing in primary care. GPs should be actively discouraged from using PSA tests for the purposes of early detection." "Younger men with a strong family history form a distinct group and such men may warrant a selective approach to PSA and DRE, although such men should be fully counselled as to the uncertainties of treatment effectiveness for localised disease" "Major questions remain concerning the efficacy and effectiveness of treatments and until these are resolved there is no justification for the introduction of a screening programme."	This is a comprehensive report summarising the evidence on screening, diagnosis and treatment of prostate cancer. We focused only on the screening part. The authors recommended against mass screening, but identified a group which can be considered of high risk in which it could be indicated to screen. The justification of the condition for screening was checked against a set of well accepted criteria (Wilson & Jungner 1968). These criteria were not fulfilled, and this fact was also considered to support the recommendation against screening.
	Country	**Interventions**		**Evidence Body**
	United Kingdom	PSA Screening, DRE Screening		o Lack of evidence from RCTs (two ongoing studies were identified). o 28 observational studies (ranging form 2c to 4).
		Main Outcome		
		Prostate cancer mortality		

<table>
<tr><th>Source</th><th>Year</th><th>Target Group</th><th>Recommendations*</th><th>Comments</th></tr>
<tr><td rowspan="5">CRD (10)

(executive summary)</td><td>1997</td><td>Asymptomatic men</td><td rowspan="5">”Routine testing of men to detect prostate cancer should be discouraged, irrespective of family history.”

”Purchasers should not fund screening services for prostate cancer.”

”Evidence from randomised controlled trials of prostate cancer screening using PSA (or similar tests) and treatment are needed before consideration is given to funding prostate screening.”

”Patients enquiring about prostate screening should be clearly informed about the current state of evidence on the benefits and harms of screening and treatment.”</td><td>The recommendations given in this report are clearly against mass screening. The authors were specially concerned with the lack of evidence of benefit, and on the concerns about potential harm, together with the lack of an effective treatment for the condition.</td></tr>
<tr><td>Country</td><td>Intervention</td><td>Evidence Body</td></tr>
<tr><td rowspan="3">United Kingdom</td><td>PSA Screening</td><td rowspan="3">○ Based on the results of 12 previous HTA reports and systematic reviews.</td></tr>
<tr><td>Main Outcome</td></tr>
<tr><td>Prostate cancer mortality</td></tr>
</table>

*Italics or bold characters as used by the authors of the original reports

Scenario 2. Nicotine Replacement Therapy

- *Policy Question*: In order to achieve the health target of reducing smoking in the population of your country you are considering different strategies to raise the rate of smoking cessation among smokers. In your country, different nicotine replacement products are available at the market, however they are not included in the positive list.
    - *Should we include nicotine replacement therapy in the positive list?*

- *Research Questions*: You contact the researchers of your HTA unit and ask them for advice. In co-operation with you they formulate following questions:
    - *Is nicotine replacement therapy effective for increasing the rate of smoking cessation?*
    - *Is it a safe intervention?*

- *The assessment process*: The researchers of your HTA unit search systematically for evidence to answer the research questions. They appraise the validity of the evidence, and judge the strength of the evidence. They write a report in which they (attempt to) give you an answer to your question.

- *Your decision* after reading the report: _____

We present six reports on the effectiveness of nicotine replacement therapy (NRT). Four of them were from the UK, one from the USA and one was a systematic review of the Cochrane Collaboration. The conclusions and/or recommendations given in the reports are summarised in Table 9. Full references are given at the end of this paper.

The reports ranged from 1998 to 2002. A systematic review of the Cochrane Collaboration (first conducted in 1997) and updated in 2002 gives **strong evidence** (level 1a) that any kind of preparation of NRT is effective in facilitating smoking cessation. The work of the Cochrane Collaboration is the base of all of the other documents, except the one from the TFCPS. Two of the documents are recommendations for clinical practice. One of the reports was a cost-effectiveness analyses, which showed NRT to be cost-effective. Overall it was recommended to prescribe NRT for motivated smokers, as the intervention helps achieve success in smoking cessation. None of these documents or reports rated the strength of evidence or the grade of recommendation in a standardised way. However all based the results on the Cochrane Review which itself represents evidence of the highest level. The target group of those documents was mainly clinical practitioners, thus no clear recommendations are given to policy-makers.

The TCPFS has assessed the value of facilitating access to the NRT (i.e. by reducing co-payments), since the intervention has been proven to be effective. Following the system described in Section 2, the authors found that there is enough ("sufficient") evidence to recommend the reduction of out-of-pocket expenditure for NRT, as such a reduction increases the number of people using an effective therapy and the number of smokers who quit.

None of the reports mentioned mayor safety threats from the use of NRT.

The evidence on effectiveness is strong enough to derive recommendations for policy, even if the reviews analysed targeted providers.

Nicotine Replacement Therapy

- There is strongest evidence that NRT is effective.
- Only one of the documents reviewed gives clear recommendations for policy-makers.
- The others give recommendations more oriented to providers.

<table>
<tr><th colspan="5">Table 9. Nicotine Replacement Therapy</th></tr>
<tr><th>Source</th><th>Year</th><th>Target Group</th><th>Recommendations*</th><th>Comments</th></tr>
<tr><td rowspan="5">Cochrane Collaboration (1)

(full report)</td><td>2002</td><td>Smokers with different levels of dependence and motivation to quit, of any age or gender</td><td rowspan="5">“All of the commercially available forms of NRT (nicotine gum, transdermal patch, nicotine nasal spray, nicotine inhaler and nicotine sublingual tablet) are effective as part of a strategy to promote smoking cessation. They increase long term quit rates approximately 1.5 to 2 fold regardless of setting. Use of NRT should be preferentially directed to smokers who are motivated to quit (as demonstrated by their initiative to request assistance) and have high levels of nicotine dependency. There is little evidence about the role of NRT for individuals smoking less than 10-15 cigarettes/day.”

“The effectiveness of NRT appears to be largely independent of the intensity of additional support provided to the smoker. Provision of more intense levels of support, although beneficial in facilitating the likelihood of quitting, is not essential to the success of NRT.”</td><td>Strictly speaking, this work did not give any recommendations, but presented results which can be translated into recommendations for policy.
The evidence produced by this systematic review can itself be considered of highest level (level 1a), as it relies on a big number of valid studies, with consistent results.</td></tr>
<tr><td>Country</td><td>Intervention</td><td>Evidence Body</td></tr>
<tr><td>Internati-onal collabora-tion</td><td>NRT (gum, transdermal patch, intranasal spray, inhalers and oral preparations)</td><td>• Evidence from 110 RCTs: 37 were classified as having Quality A, and the rest as having Quality B, following a simple classification of quality proposed in the Cochrane Reviewers Manual. “A” means low risk of bias, “B” means moderate risk of bias, none of the studies were classified as “C” high risk of bias.)</td></tr>
<tr><td></td><td>Main Outcome</td></tr>
<tr><td></td><td>Long-term cessation rate (>6 month)</td></tr>
</table>

<table>
<tr><th>Source</th><th>Year</th><th>Target Group</th><th>Recommendations*</th><th>Comments</th></tr>
<tr><td rowspan="5">NICE (2)

(full report)</td><td>2002</td><td>Smokers of any age or gender</td><td rowspan="5">“Nicotine replacement therapy (NRT) and bupropion are recommended for smokers who have expressed a desire to quit smoking.”

“NRT or bupropion should normally only be prescribed as part of an abstinent-contingent treatment (ACT), in which the smoker makes a commitment to stop smoking on or before a particular date (target stop date). Smokers should be offered advice and encouragement to aid their attempt to quit.”

“NHS Trusts, primary care teams, local health groups, community pharmacists, hospital-based clinical services and health authorities should review policies and practices regarding smoking cessation to take account of the guidance set out in Section 1 [above].”</td><td>This document based its recommendations mainly on the findings from the Cochrane Review and on a rapid assessment form CRD. The results of the systematic review were appraised by a committee and translated into a recommendation for the NHS, after assessing economic and service organisation consequences. The prescription recommendation (that is coverage by the NHS) is limited only to the group of smokers motivated to quit. The guidance given is directed mainly to clinical practice, however, the implementation of policies to take account of it is strongly encouraged.</td></tr>
<tr><td>Country</td><td>Interventions</td><td>Evidence Body</td></tr>
<tr><td>United Kingdom</td><td>NRT (gum, transdermal patch, intranasal spray, inhalers and oral preparations), Bupropion (antidepressant drug)</td><td>• Systematic Review (highest level of evidence)
• Submissions from stakeholders
• Experts</td></tr>
<tr><td></td><td>Main Outcome</td><td></td></tr>
<tr><td></td><td>Long-term cessation rate (>6 months)</td><td></td></tr>
</table>

<table>
<tr><th>Source</th><th>Year</th><th>Target Group</th><th>Recommendations*</th><th>Comments</th></tr>
<tr><td rowspan="5">HTBS (3)

(full report)</td><td>2002</td><td>Smokers with different levels of dependence and motivation to quit</td><td rowspan="5">”HTBS supports the NICE Technology Appraisal Guidance which states in summary that:
NRT and bupropion (Zyban) should be available on prescription for smokers who have said they want to quit smoking.
NRT or bupropion (Zyban) should normally only be prescribed when a smoker has made a commitment to stop smoking on or before a certain date (which is called your ”target stop date”). Healthcare professionals should offer you advice and encouragement to help you quit.”</td><td>This document is an explanation of the guidance from NICE. No new evidence is appraised. HTBS limits itself to check the transferability of the NICE guidance to the Scottish NHS, and to make it more understandable for patients.</td></tr>
<tr><td>Country</td><td>Intervention</td><td>Evidence Body</td></tr>
<tr><td>United Kingdom</td><td>NRT (gum, transdermal patch, intranasal spray, inhalers and oral preparations)</td><td rowspan="3">Guidance based on a Systematic Review (highest level of evidence)</td></tr>
<tr><td></td><td>Main Outcome</td></tr>
<tr><td></td><td>Long-term cessation rate (>6 months)</td></tr>
</table>

Source	Year	Target Group	Recommendations*	Comments
NCCHTA (4) (full report)	2002	Smokers of any age or gender	”Both NRT and bupropion are effective interventions to assist smoking cessation.” ”Information on how to maximize effectiveness in practice is still lacking, but motivational support is probably involved.” ”Irrespective of the methods used or the assumptions involved, the results of economic evaluations and the model developed in this review consistently suggest that smoking-cessation interventions, including the use of NRT and/or bupropion, are relatively cost-effective in terms of the cost per life-year saved. The worst-case scenarios still provide estimates of cost-effectiveness better than many other medical interventions.”	The authors updated a previous Cochrane Review, accounting also for unpublished results. Strictly, the authors did not gave any recommendations, but stated conclusions. Overall the evidence provided by this work can be considered of the highest level (level 1a), and can be translated into recommendations
	Country	**Interventions**		**Evidence Body**
	United Kingdom	NRT (gum, transdermal patch, intra-nasal spray, inhalers and oral preparations), Bupropion (antidepressant drug)		• A high quality Systematic Review (level 1a) • Five RCTs of level 1b-2b • Two unpublished RCT.
		Main Outcome		
		Long-term cessation rate (6, 12 or more months after starting therapy)		

<table>
<tr><th>Source</th><th>Year</th><th>Target Group</th><th>Recommendations*</th><th>Comments</th></tr>
<tr><td rowspan="5">TFCPS (5)

(full report)</td><td>2001</td><td>Smokers of any age or gender</td><td rowspan="5">“Reducing patient out-of-pocket costs for effective cessation therapies: recommended. This intervention includes efforts to reduce the financial barriers to patient use of cessation therapies that have previously demonstrated evidence of effectiveness. Techniques include providing the services within the health care system, or providing coverage to or reimbursement of patients for expenditures on cessation groups or on nicotine replacement or other pharmacologic therapies. Reducing patient out-of-pockets costs for effective cessation therapies is recommended by the TFCPS on the basis of sufficient scientific evidence of effectiveness in (1) increasing use of the effective therapy, and (2) increasing the total number of tobacco-using patients who quit.”</td><td>This is a comprehensive collection of recommendations regarding public health interventions targeting tobacco use and control. Concerning NRT - given its effectiveness -, the Task Force evaluated the evidence for system interventions (reducing out-of-pocket- costs of NRT) to facilitate the use of effective therapy. According to the strength of evidence, the intervention was ”recommended” following the rating system of the TFCPS.</td></tr>
<tr><td>Country</td><td>Intervention</td><td>Evidence Body</td></tr>
<tr><td rowspan="3">USA</td><td>Efforts to reduce financial barriers to patient use of effective cessation therapies (NRT)</td><td rowspan="3">• Five studies of “fair quality of execution” and “greatest” or ”moderate” suitability
• The strength of the evidence body was qualified as “sufficient”</td></tr>
<tr><td>Main Outcomes</td></tr>
<tr><td>Cessation rate, rate of patients using the effective therapy</td></tr>
</table>

<table>
<tr><th>Source</th><th>Year</th><th>Target Group</th><th>Recommendations*</th><th>Comments</th></tr>
<tr><td rowspan="5">CRD (6)

(full report)</td><td>1998</td><td>Smokers of any age or gender</td><td rowspan="5">”Smoking cessation interventions are highly cost-effective, therefore:
• Health professionals should encourage the use of NRT in those smokers who are motivated to quit.”
• […]
• Health authorities and primary care groups should work together to ensure the implementation of these effective interventions
• Health authorities and other health service commissioners should develop co-ordinated smoking cessation strategies and fund their implementation.”</td><td>This bulletin summarizes existing evidence on interventions to reduce smoking. The recommendations are not rated but the evidence basis is strong.</td></tr>
<tr><td>Country</td><td>Intervention</td><td>Evidence Body</td></tr>
<tr><td rowspan="3">United Kingdom</td><td>NRT and advice from health professional</td><td rowspan="3">• A Cochrane Systematic Review (1998 Version), which can be considered high level evidence (1a)
• A systematically developed Guideline from AHRQ, based on a comprehensive and systematic literature review.</td></tr>
<tr><td>Main Outcome</td></tr>
<tr><td>Cessation rate</td></tr>
</table>

*Italics or bold characters as used by the authors of the original reports

References

General References

Briss PA, Zaza S, Pappaioanou M, et al. Developing an evidence-based Guide to Community Preventive Services –Methods. *Am J Prev Med* 18(1S):35-43 (2000).

Busse R, Orvain M, Velasco M, et al. Best practice in undertaking and reporting health technology assessments. *Int J Technol Assess Health Care* 18:361-422 (2002).

Clarke M, Oxman AD (eds). *Cochrane Reviewers' Handbook 4.1.6*. In: The Cochrane Library, Issue 1, 2003. Oxford: Update Software

Council of Europe. *Developing a methodology for drawing up guidelines on best medical practices* (Recommendation (2001)13 and explanatory memorandum). Paris: Council of Europe, 2001.

Hennekens CH, Buring JE. Epidemiology in Medicine. Boston – Toronto: Little, Brown and Company, 1987

Innvaer S, Vist G, Trommald M, Oxman A. Health policy-makers' perceptions of their use of evidence: a systematic review. *J Health Serv Res Policy* 7:239-244 (2002).

Lohr KN, Carey TS. Assessing "best evidence": issues in grading the quality of studies for systematic reviews. *Jt Comm J Qual Improv* 25:470-479 (1999).

McQueen D, Anderson LM. What counts as evidence: issues and debates. In: Rootman I, Goodstadt M, Hyndman B, McQueen D, Potvin L, Springett J, Ziglio E (eds.). *Evaluation in health promotion: Principles and perspectives*. Copenhagen: World Health Organization, 2001, pp. 63-82.

Gray JAM, Haynes RB, Sackett DL et al. Transferring evidence from research into practice: 3. Developing evidence-based clinical policy. *ACP J Club* 126:A-14 (1997).

Perleth M, Jakubowski E, Busse R. What is 'best practice' in health care? State of the art and perspectives in improving the effectiveness and efficiency of the European health care systems. *Health Policy* 56:235-250 (2001).

Sackett DL, Rosenberg WMC; Gray JAM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *Br Med J* 312:71-72 (1996).

West S, King V, Carey TS, et al. *Systems to rate the strength of scientific evidence*. Evidence Report / Technology Assessment No.47. Rockville: Agency for Healthcare Research and Quality, 2002.

Wilson JMG, Jungner G. *Principles and practices of screening for disease*. Geneva: World Health Organisation, 1968 (Public Health Papers, No. 34).

Prostate Cancer Screening Reports (as numbered in Table 8).

1. Slaughter P M, Pinfold S P, Laupacis A. Prostate-specific antigen (PSA) screening in asymptomatic men. Toronto: Institute for Clinical Evaluative Sciences (ICES), 2002.
2. U.S. Preventive Service Task Force (USPSTF). Screening for prostate cancer. Recommendations and Rationale. Rockville: Agency for Healthcare Research and Quality (AHRQ) Pub. No. 03-515A, 2002.
3. Gander L. The PSA Test in early detection of prostate cancer. Guidelines Review. Saskaton: Health Services Utilization and Research Commision (HSURC), 2000.
4. Agence Nationale d'Accréditation et d'Évaluation en Santé (ANAES). Appropriateness of systematic screening for prostatic cancer by prostate specific antigen (PSA). Paris: ANAES, 1999.
5. Pientka L. PSA-Screening beim Prostatakarzinom. Baden-Baden: Nomos Verlagsgesellschaft, 1999 (Schriftenreihe Health Technology Assessment 11).
6. Schersten T, Baile M A, Asua J, Jonsson E. Prostate cancer screening: evidence synthesis and update (INAHTA Joint Project). Vitoria-Gasteiz, Spain: Osteba (Basque Office for Health Technology Assessment, Health Department of the Basque Government), 1999.
7. The Norwegian Centre for Health Technology Assessment (SMM). Screening for prostate cancer, Norwegian review of international studies.. Oslo: The Norwegian Centre for Health Technology Assessment (SMM), 1999 (SMM-Report 3/1999).

8. Briones E. Effectiveness of prostate cancer screening and its application in Andalucia -systematic review. Sevilla: Agencia e Evaluacion de Tecnologias Sanitarias de Andalucia (AETSA), 1998.
9. Selley S, Donovan J, Faulkner A, Coast J, Gillat D. Diagnosis, management and screening of early localised prostate cancer. *Health Technology Assessment* Vol. 1: No.2 (1997).
10. NHS Centre for Reviews and Dissemination. Screening for prostate cancer (includes patient leaflet 'Information for men considering or asking for PSA tests'). York: Centre for Reviews and Dissemination, 1997 (Effectiveness Matters 2(2)).

Nicotine Replacement Therapy Reports (as numbered in Table 9).

1. Silagy C, Lancaster T, Stead L, Mant D, Fowler G. Nicotine replacement therapy for smoking cessation (Cochrane Review). In: The Cochrane Library, Issue 2, 2003. Oxford: Update Software.
2. Health Technology Board of Scotland (HTBS). Nicotine replacement therapy (NRT) and bupropion (Zyban) to help quit smoking. Understanding HTBS advice. Glasgow: Health Technology Board of Scotland, 2002.
3. National Institute for Clinical Excellence (NICE). Guidance on the use of nicotine replacement therapy (NRT) and bupropion for smoking cessation. Technology Appraisal Guidance No. 39. London: National Institute for Clinical Excellence, 2002.
4. Woolacott NF, Jones L, Forbes CA, Mather LC, Sowden AJ, et al. The clinical effectiveness and cost-effectiveness of bupropion and nicotine replacement therapy for smoking cessation: a systematic review and economic evaluation. *Health Technology Assessment* Vol. 6, No. 16 (2002).
5. Task Force on Community Preventive Services (TFCPS). Recommendations regarding interventions to reduce tobacco use and exposure to environmental tobacco smoke. *Am J Prev Med* 20(2S):10-15 (2001).
6. Centre for Reviews and Dissemination. Smoking Cessation: What the health service can do. York: Centre for Reviews and Dissemination, 1998 (Effectiveness Matters 3(1)).

Abbreviations

AETSA	Andalusian Agency of Health Technology Assessment
AHRQ	Agency for Healthcare Research and Quality
ANAES	Agence National d'Accréditation et d'Évaluation en Santé
CRD	Centre for Reviews and Dissemination
DIMDI	German Institute for Medical Documentation and Information
DRE	Digital Rectal Examination
HSURC	Health Services Utilization and Research Commission
HTA	Health Technology Assessment
HTBS	Health Technology Board of Scotland
ICES	Institute for Clinical Evaluative Sciences
INAHTA	International Network of Agencies for Health Technology Assessment
NCCHTA	National Coordinating Centre for Health Technology Assessment
NHS	National Health Service
NICE	National Institute for Clinical Excellence
NRT	Nicotine Replacement Therapy
OSTEBA	Basque Office for Health Technology Assessment
PSA	Prostate Specific Antigen
RCT	Randomised Controlled Trial
SBU	Swedish Council on Health Technology Assessment in Health Care
SMM	Norwegian Centre for Health Technology Assessment
TFCPS	Task Force on Community Preventive Services
TRUS	Transrectal ultrasound imaging
USPSTF	United States Preventive Services Task Force