Appendix 3. Shared methodologies

**Important notice:** This appendix represents auxiliary content of the HTA Core Model. It is under construction and not fully updated for the HTA Core Model version 2.0. It is not yet a comprehensive presentation of useful methodologies for core HTA information producers. It will be updated and amended during Joint Action 2 by September 2015.

### Diffusion and translation models

The relation between technology and organisation can be tackled in different ways. At least two different and incompatible views on causality and transferability can be differentiated with respect to the organisational issues: the diffusion model and the translation model (Kristensen 2001, Latour 1987). Parallel viewpoint is seen in the social domain.

**Diffusion model**
- bases on a linear, unidirectional conception of causality
- considers technology as an exogenous and independent entity
- seen as a given object which stands outside or above the society, its organisations and actors
- supposes that technology stays constant
- sees technology be diffused and transferred from the innovator to different users (Leavitt 1965)

**Translation model (Leavitt 1965):**
- sees technology as endogenous, as a part of the organisational and use process
- technology can't be separated from the organisation and its users
- technology does not stay constant during the implementation process
- human activity is a part of the technology in question
- asks "how many and what kind of resources (material entities, time, money, people, etc.) must be mobilised and organized in order to produce satisfactory results from a health technology."
- technology does not causally affect the organisation and change its social structures
- organisation and its work processes and social structures have to be organized so that good results can be produced from the technology.
(Leavitt 1965)

### References:


General guidance to critical appraisal of published studies and other information

Critical appraisal of HTAs
[to be added]

Critical appraisal of systematic reviews
AMSTAR

Critical assessment of indirect comparisons
AMSTAR

Critical appraisal of guidelines
- AGREE is an international collaboration improving the quality of clinical practice guidelines by establishing a shared framework for development, reporting and assessment
  http://www.agreecollaboration.org

Critical appraisal of trials
[to be added]

Critical appraisal of observational studies
There are several checklists or scales on quality available but no consensus about using those. The most appropriate are:
- Newcastle Ottawa Scale http://www.cochrane.org/training/cochrane-handbook
- Checklist of items that should be included in reports of observational studies (actually not meant for assessing quality): STROBE http://www.strobe-statement.org

Critical appraisal of diagnostic accuracy studies
QUADAS-2

Critical appraisal of modelling studies
International Society for Pharmacoeconomics and Outcomes Research (ISPOR) has published a useful article describing the basic guidelines for conducting and reporting modelling studies (Weinstein 2003). It can be used also as guidance for using and critically appraising modelling studies. Furthermore, ISPOR is developing more specific guidelines on different modelling methods.
Critical appraisal of economic evaluation

There are several methodological characteristics to consider, when assessing the quality of an economic evaluation. Several checklists have been published for reporting an economic evaluation, but also to help in identifying the strengths and weaknesses of different studies (e.g. Drummond 1996, Drummond 2005). An example of a checklist (by Drummond 2005) is:

1. Was a well-defined question posed in answerable form?
2. Was a comprehensive description of the competing alternatives given?
3. Was the effectiveness of the programmes or services established?
4. Were all the important and relevant costs and consequences for each alternative identified?
5. Were costs and consequences measured accurately in appropriate physical units?
6. Were costs and consequences valued credibly?
7. Were costs and consequences adjusted for differential timing?
8. Was an incremental analysis of costs and consequences of alternatives performed?
9. Was allowance made for uncertainty in the estimates of costs and consequences?
10. Did the presentation and discussion of study results include all issues of concern to users?

References:


Critical appraisal of qualitative studies

Examples of quality assessment instruments:

- Critical Appraisal Skills Programme – CASP
- EPPI-review by the EPPI Centre. http://eppi.ioe.ac.uk/eppireviewer/login.aspx
- Quality Framework UK Cabinet Office
- Checklist of items that should be included in reports of qualitative studies (not checklist for assessing quality) COREQ http://www.aaz.hr/dokumenti/odjel-raz-ist-i-zdra-teh/edukativni-materijali/smjernice/7.%20Guidelines%20for%20qualitative%20research.pdf
- Popay et al (1998)
- The Mays & Pope criteria (2000)
Quality assessment of routine collected statistics and administrative data

Routine collected administrative data (e.g. DRG, discharge databases, reimbursement claims databases) can be useful too, when available. For example sickness funds collect great amounts of information which could be used to analyse utilisation of technology etc. However, analysis of this kind of data might be very time consuming, since data need to be “prepared” before analysis. By definition, these data has been collected for other purposes than research and they cannot be used to answer scientific questions without previous processing. This might not be feasible in the context of an HTA project, due to resource constraints.

The use of routine collected statistics has several limitations. The reliability of the diagnosis varies and usually it is not possible to differentiate between different stages of the disease. Even the validity of the coding of causes of death may be variable, and in some countries it is known to be very limited.

Own analysis of administrative data often requires authorization from the data owner, which in some countries might be difficult to obtain due to issues of privacy protection and confidentiality.

Critical appraisal of register data

ISPOR is developing guidelines for patient registry data:
http://www.ispor.org/sigs/PR_analysis_data_mgt.asp

General guidance to conducting own research

Guidance for modelling

International Society for Pharmacoeconomics and Outcomes Research (ISPOR) has published an article describing the basic guidelines for conducting and reporting modelling studies. ISPOR is also developing more specific guidelines on different modelling methods.

References

http://www.ispor.org/workpaper/healthscience/TFModeling.asp

Guidance for conducting a register study

[to be added]

Guidance for conducting survey (questionnaire, interview)

[to be added]

General guidance for synthesis

Meta analyses of accuracy studies

No heterogeneity

A forest plot of sensitivity versus specificity with 95 % confidence intervals can be used whenever the results from two or more comparable studies are included in the review. Forest plot illustrates
the range of results, enables the reader to assess heterogeneity, and possible trade-off between sensitivity and specificity, and may show the summary estimate where pooling is appropriate.

Another option is to plot pairs of sensitivity and 1 - specificity from original studies on a ROC plane. If sensitivity or specificity is constant or if there is linear relationship between them, simple summary measures for sensitivity, specificity, or likelihood are adequate.

When pooling pairs of sensitivity and specificity, the statistical model used depends on the studies selected. Fixed effect model assumes the studies to represent a random sample of one large common study. The differences between study outcomes are considered to be the result of random error. The model weights individual studies based on the inverse variance of the accuracy or the number of participants. Random effects model assumes the differences between studies to be due to real differences between the study populations and procedures. A more complex mathematical model is used to weight studies. Separate estimates of mean sensitivity and specificity underestimate test accuracy.

**Heterogeneity present**

When forest plot and heterogeneity testing shows that there is significant heterogeneity in sensitivities and specificities across studies, it is not appropriate to report the pooled values of sensitivity and specificity as a summary estimate. Instead, further analysis of the heterogeneity detected is needed, and it starts with examining of threshold effect. Threshold effect can be seen in forest plot if there is an inverse relationship between sensitivity and specificity. If this is not apparent the results should be plotted to a ROC plane to examine the threshold effect further.

Paired estimates of sensitivity and 1 - specificity in original studies are plotted in a ROC plane. Regression model is used to fit the SROC curve (Moses 1993). If the SROC curve is symmetrical around the line where sensitivity equals specificity, the studies share one common DOR, and any variability is due to differences in the test threshold. In statistical terms, if in the model the slope b (estimated regression coefficient) is not statistically significant and approaches zero, The SROC will be symmetrical.

Spearman’s test for a nonparametric distribution has also been used to test for a threshold effect. Using this method, the correlation between sensitivity and 1-specificity for each study is measured and a Spearman rank correlation coefficient > 0.6 is used to confirm variation across studies due to a threshold effect (Moses 1993). If the correlation is poor (Spearman rank correlation coefficient < 0.6) the variation between studies is attributed to other differences. This is a crude measure and is not generally recommended.

**Threshold effect only**

If there is symmetry in the SROC curve, DOR is constant regardless of the diagnostic threshold, and any variability in the paired sensitivity and specificity between different studies is due to differences in the test threshold. In this case, SROC curve represents the most informative synthesis of evidence about test accuracy and the pooled DOR is a useful single summary measure.

SROC curve does not provide one summary estimate of sensitivity and specificity but it allows assessment of their interdependence. Summary DOR (SDOR) of the test and a comparator test can be presented with 95% CI:s to compare differences in diagnostic performance. The area under SROC curve and its 95% confidence interval provides a global summary of overall test accuracy. The point on the curve where sensitivity equals specificity, the Q* statistics, can also be used as a summary measure of the accuracy of the test. These summary measures can also be used to
compare the accuracy of two test strategies. Software for diagnostic meta-analysis include MetaTest, Meta-Disc, Stata and SAS.

**Heterogeneity that is more than just threshold effect**

If the slope $b$ in the SROC model is statistically significant, the SROC will be asymmetrical and the DOR changes along the threshold. In such cases advanced methods for fitting the SROC is used. Advanced methods to pool are indicated if heterogeneity in the results can be attributed to known sources of variation (see above Chapter Assessing heterogeneity). Otherwise the interpretation of the summary estimate is not possible (Lijmer 2002).

Possible sources of variation include

1. Chance
2. Different threshold
3. Different study designs, methods, biases: different reference standard, different versions of the technology
4. Variation by clinical subgroups in terms of age, severity or stage of disease, prevalence of the target condition, differential diagnoses, and setting
5. Unexplained heterogeneity

If differences in the results can not be attributed to these known sources of heterogeneity, then pooling of the results to one summary estimate should not be attempted, because its interpretation will be impossible (Lijmer 2002).

Methods to test for heterogeneity (Medical Services Advisory Committee 2005):

1. Plot the sensitivity and specificity from each study with their 96% confidence interval in a table and/or forest plot to illustrate the range of estimates and identify outliers.
2. If sufficient data are available, plot the paired sensitivity and 1-specificity results for each study on the ROC plane to detect heterogeneity and identify outliers. A small number of studies will limit the power of regression to detect heterogeneity.
3. Use a chi-square test for heterogeneity (Cochran's Q test) or Fischer's exact test for small studies to test the hypothesis that there is no statistically significant difference in the sensitivity and specificity reported.

Advanced models enable incorporation of covariates, e.g. population subtype in the meta-regression analysis. Poor reporting of primary studies may though lead to biased estimates. The two main advanced models are hierarchical SROC and bivariate meta-regression, and they are mathematically identical (Harbord 2007). Syntax to run these models in SAS, STATA, WINBUGS, S-PLUS and R are or will be available. Hierarchical SROC (HSROC) produces informative summary measures with confidence ellipses (Reitsma 2005). Model is infrequently used, probably due to complex fitting.

References:

General guidance for interpretation

Guidance for assessing applicability
Atkins et al. (2011):

- Step 1. Determine the most important factors that may affect applicability
- Step 2. Systematically abstract and report key characteristics that may affect applicability in evidence tables (highlight studies with a pragmatic approach and data on effect size of effect modification).
- Step 3. Make and report judgements about major limitations to applicability of individual studies.
- Step 4. Consider and summarize the applicability of a body of evidence