Report

Medical tests (assessment of established medical science and medical practice)

Approved by CVZ on 20th January 2011 and subsequently presented to the Minister of VWS

Publication number 293

Publication College voor zorgverzekeringen
P. O. Box 320
1110 AH Diemen
Fax +31 (0) 20 797 85 00
E-mail info@cvz.nl
Internet www.cvz.nl

Case number 2010078493
Department Care Advice
Authors J.T.M. Derksen, gynaecologist, P.C. Staal and Dr. G. Ligtenberg
Direct line Tel. +31 (0) 20 797 85 55

Contents:

Realisation of this report:

The report was drawn up by:
J.T.M. Derksen, gynaecologist
P.C. Staal
Dr. G. Ligtenberg
J. Heymans, M.D., MPH
Dr. I.M. Verstijnen

A draft version of the report was presented during a work conference to a number of external experts, to whom we extend our warmest appreciation for their critical comments. These were the following experts (listed alphabetically):
Prof. Dr. W.J.J. Assendelft
Prof. Dr. P.M.M. Bossuyt
Dr. A. van den Bruel
Prof. Dr. Y. van der Graaf
Prof. Dr. K.G.M. Moons
Dr. A.J. Rijnsburger
Prof. Dr. R.J.P.M. Scholten
Prof. Dr. E.W. Steyerberg

The draft report was also discussed within CVZ's *Adviescommissie Pakket* [ACP, Insured Package Advisory Committee] and *Duidingscommissie Pakket* [DCP, Insured Package Clarification Committee].

Summary

Coverage

A key aspect of the Dutch Health Insurance Act (*Zorgverzekeringswet*) is that compliance with 'established medical science and medical practice' co-determines whether care is covered. General basic health insurance only includes care that is considered effective.

Assessment

In 2007, the Dutch Health Care Insurance Board (*college voor zorgverzekeringen,* CVZ) described their method for determining whether care complies with this criterion in a report entitled "Established medical science and medical practice". In principle, the points of departure described in that report also apply to medical tests. In this report, however, CVZ provides further details of the assessment framework for medical tests. This is necessary, partly due to rapid developments in the methodology for evaluating tests scientifically and CVZ's desire to keep abreast of these. Furthermore, a second aim of this report is to inform other parties about the principles and methods CVZ uses when assessing medical tests.

Medical tests

The term medical tests covers all interventions used for the diagnosis, prognosis, prediction or course of a patient's disease.

Principles of the framework

When assessing medical tests, CVZ is interested in the effects tests have on the health status of the people who undergo them. CVZ feels that a test should be assessed not only for its ability to provide a pretty picture or a correct diagnosis. Obviously, it is extremely important that a test is reliable. However, a positive recommendation from CVZ will be based in particular on proven positive consequences for health-related outcomes on people who undergo the test. In general, the results of a test will determine the subsequent course of action to be followed. When evaluating a test, CVZ assesses the effectiveness (also referred to as the clinical utility) of this entire path, i.e., the combined test-plus-treatment strategy. Clinical utility can be determined based on a comparative study of the usual ('old') and the proposed ('new') test-plus-

treatment strategy. Such direct evidence is, however, often lacking. In that case, CVZ constructs a comparative analytic framework for examining whether indirect evidence can be useful in providing a reply to the question of clinical utility.

Further development

CVZ uses the working method described in this report. This method may eventually need to be adjusted or refined on the basis of experience or scientific developments.

1. Introduction

Key aspect of the Zvw

A key aspect of the Health Care Insurance Act (Zvw) is that 'established medical science and medical practice' contributes to determining what is covered by the basic health insurance. What this amounts to is that care will only be included in the basic insured package if it is considered effective.

Assessment framework

CVZ has developed an assessment framework for determining whether care fulfils the criterion 'established medical science and medical practice'. This framework appeared in the report sent to the Minister of VWS in 2007: "Assessment of established medical science and medical practice".[1] The method described in that report, which is based on the principles of evidence-based medicine (EBM), applies to all forms of care. In principle, CVZ also bases its assessment of medical tests on the points of departure described in that report.

Definition

Medical tests refers to all interventions used for the diagnosis, prognosis, prediction or follow-up/course of a disease in a patient. The nature of such tests can obviously vary; ranging from questionnaires or surgical exploration to advanced diagnostics.

Limited experience

To date, CVZ has had limited experience in assessing medical tests. In the past few years emphasis was in particular on the assessment of therapeutic interventions and medicines. The steadily increasing number of (expensive) medical tests and the question this raises as to whether they belong in the insured package formed a reason for CVZ to explicitly elaborate upon the examination of new tests, and possibly existing tests as well, in relation to the criterion 'established medical science and medical practice'. This detailed elaboration is particularly necessary because of rapid developments in the methodology for scientifically assessing

[1] CVZ. Assessment of established medical science and medical practice. Diemen, 2007. Report no. 254. Available via www.cvz.nl.

tests and CVZ wants to keep abreast of these developments.

External expertise

To this end, at the request of CVZ, Prof. Dr. P.M.M. Bossuyt carried out an investigation of the available scientific literature and of the way in which other organisations develop recommendations for tests within the framework of EBM. The results of his investigation are presented in the report 'Evidence-Based medical testing'.[2]

Aim of the current report

The current report contains an elaboration of the method for assessing medical tests on the basis of EBM. How does CVZ plan to select the available scientific evidence for tests and weigh them in a structured manner? The method presented in this report is in keeping with the findings from the report on 'Evidence-Based medical testing'.[2]

Realisation

A draft version of the report was discussed with a number of external experts. A summary of that discussion is included as appendix. The draft report was also discussed within CVZ's *Adviescommissie Pakket* [ACP, Insured Package Advisory Committee] and *Duidingscommissie Pakket* [DCP, Insured Package Clarification Committee]. The basic assumptions of the report are endorsed by the external experts as well as the two above-mentioned committees.

Obtaining practical experience

CVZ is using the approach to medical tests described in this report in practice. The method may eventually require (some) refinements. CVZ will examine whether this is necessary at the appropriate time, once the necessary experience has been obtained using this working method.

Scientific questions over methodology

It is important to mention that the method for evaluating test strategies is still undergoing rapid developments.[2,3] It is also the subject of international research. CVZ is aware that scientific insights in this field are becoming increasingly

[2] Bossuyt PMM. Evidence-Based medical testing. Amsterdam, 2010. Available via www.cvz.nl.

[3] Ludwig Bolzmann Institut 2010. http://eprints.hta.lbg.ac.at/898/

comprehensive and may alter as a result. In certain cases, CVZ will put their evaluations of tests before external methodology experts.
This is in addition to the usual consultation of relevant professionals regarding content.

Report structure

The structure of the report is as follows. Section 2 contains a short description of the relevant statutory framework.
Sections 3 and 4 examine this statutory framework in practice.
Section 5 describes the method for assessing medical tests.
Section 6 completes the report with a brief discussion of the importance of transparency and support.

2. Statutory framework

2.a. Provisions covered

Summary and elaboration

Summary in Zvw

Article 10 of the Zvw contains a summary of the risks insured. This is a global description of provisions, the rights to which are covered by health insurance.[4]

The following insured risks are involved, i.e., the need of:

a. medical care;
b. dental care;
c. pharmaceutical care;
d. care in the form of medical devices;
e. nursing;
f. caring;
g. residence;
h. transport.

Elaboration in Bzv and Rzv

The Health Insurance Decision (*Besluit zorgverzekering,* Bzv) and the Health Insurance Regulation (*Regeling zorgverzekering,* Rzv) provide an elaboration of the contents and quantities of the forms of care listed in article 10 of the Zvw.

The elaboration of forms of care varies per item. Some forms of care are described in general terms (generic). This applies for example to medical care. When describing this form of care, the law makes use of the phrase 'normally provided'. For example, it stipulates that medical care includes care that is normally provided by G.P.s, medical specialists, clinical psychologists and obstetricians (article 2.4, para. 1, of the Bzv). Other forms of care are described in more detail (specific) and a limiting factor sometimes applies. This is the case, for example, for extramural pharmaceutical care.

[4]On the grounds of the Zvw ,Health insurers are obliged (see article 11) to include and convert the health insurance products they offer into provisions insured under the health insurance acts (insurance coverage provided by the insurance agreement).

Established medical science and medical practice

The content and amount of all forms of care is determined in part by 'established medical science and medical practice'. This is regulated in article 2.1, second para., of the Bzv.[5] This norm is discussed in more detail below.

2.b. Open and closed system

Open system

A generic description of insured provisions usually lead to an open system for the insured provisions. Basically it means the fairly automatic inclusion and removal of provisions. For example, a generic description exists for care provided by medical specialists. The basic insurance covers care that is normally provided by medical specialists and which complies with established medical science and medical practice. Innovative care that (eventually) starts to fulfil these conditions (i.e., falls under this generic heading), tends to automatically get included among the insured provisions. No prior examination is required nor adjustments in the regulations. Care that eventually comes to be regarded as obsolete, and is no longer used by medical specialists in practice, disappears from the insured package. In fact, the chosen statutory text ensures that the insured package is always current and in accordance with the most recent developments.

Closed system

Specific (detailed) sweeping descriptions (such as positive lists) form a closed system of insured provisions. This does not lead to automatic inclusions and removals. An alteration in the insured package can only be achieved by altering the legislation. As a result, with a closed system the insured package will not always be up to date.

[5] The norm 'established medical science and medical practice' does not apply to sedentary medical transport. See section 3 of the above-mentioned report for further clarification: CVZ. Assessment of established medical science and medical practice. Diemen, 2007. Report no. 254. Available via www.cvz.nl.

3. Examining compliance with the statutory framework in practice

3.a. Method for assessing established medical science and medical practice

Norm for all forms of care

The content and amounts of all forms of care are determined in part by 'established medical science and medical practice' (see article 2.1, second para. of the Bzv).[5] In other words: only care that complies with 'established medical science and medical practice' – which can be regarded as effective – is covered by basic insurance.

General method

CVZ described its methods for determining what can be regarded as 'established medical science and medical practice' in the report "Assessing medical science and medical practice".[1] CVZ's method is based on the principles of evidence-based medicine (EBM). The EBM method focuses on "the meticulous, explicit and judicious use of the current best evidence". Furthermore, CVZ's general assumption is that a positive decision on the 'established medical science and medical practice' criterion will require medical-scientific data with the highest possible level of evidence. CVZ can depart from this requirement if there are grounds.

3.b. Method for assessing 'normally provided'

'Normally provided'

In paragraph 2.b we commented on the fact that the phrase 'normally provided' is used for forms of care that are described in generic terms. For example, medical care is said to cover care that is normally provided by G.P.s, medical specialists, clinical psychologists and obstetricians (article 2.4, para. 1 of the Bzv).

Concept definition

In its report "The meaning of the 'normal provision' criterion and its assessment"[6], CVZ explained how to determine whether this criterion has been fulfilled. In brief, 'normally provided' care is care that the professional group of the care-provider named in the legislation regards as part of the accepted arsenal of care and it is provided in a way that the professional group concerned deems to be professionally correct. As a rule, guidelines and standards of the professional group help to determine whether the care involved is 'normally provided' by the professional group. The same documents can also serve to determine whether/when care is being provided in a way that is 'professionally correct'.

In order to be accepted into the basic insurance package, care that falls under the criterion 'normally provided' must also (among other things) fulfil the criterion 'established medical science and medical practice'.

3.c. CVZ's assessment activities

Prior assessment not always necessary

The assessment of care – can the care be regarded as an insured provision? – is not always carried out prior to its introduction into daily practice. As indicated above, those forms of care that are described in fairly general, generic terms (with an open system) are automatically included in the insured package as long as the description is fulfilled. In general, this is either tacitly assumed, or it is given no attention whatsoever. Patients are provided with the care and, if there is a cost item (tariff) that can be used, the costs of the care are charged to and paid by the health insurers at the expense of the basic insurance. Obviously this does not lead to problems as long as the care does fulfil the statutory criteria and should actually be regarded as an insured provision. The point of departure with an open system is the confidence that professionals will only want to provide care

[6] CVZ. Meaning and assessment of the 'normal provision' criterion. Diemen, 2008. Report no. 268. Available via www.cvz.nl.

that is effective, appropriate and safe.

Assessment by CVZ

Nevertheless, CVZ does regularly assess – upon request or at CVZ's own initiative – whether (innovative) care should (actually) be included in the basic package. This could be for a variety of reasons. New (expensive) interventions that require the establishment of a new tariff and which (will) consume a large proportion of the total health care costs, do not generally slip into the insured package unnoticed. Another question that may arise with respect to care provided for many (groups of) patients (large volumes) and care that may be deemed as unsafe, is whether this really is insured care and whether health insurers are justified in funding the care at the expense of the basic insurance. Scientific publications could also lead to CVZ assessing a given form of care in more detail.

CVZ policy

CVZ's policy focuses on, among other things:

- recognising, as far as possible, undesired inclusions of care in the insured package (because the generic descriptions of insured provisions have been fulfilled), so that CVZ can advise the Minister of VWS to prevent the inclusion of a new form of care in the insured package by means of legislation. Undesired inclusion in the package is inclusion that is contrary to CVZ's package principles. For example, an unfavourable cost-effectiveness ratio could be a reason for CVZ to advise the Minister of VWS to explicitly exclude a newly included form of care from the insured package;
- preventing, as far as possible, the reimbursement of care at the expense of the health insurance, in the event there is doubt regarding whether it is care that should be insured, by providing clarity on the question as to whether it is – in view of the statutory conditions – insured care or not.

Risk-oriented package management

In order to be able to implement this policy adequately, CVZ has to actively follow developments in medical practice, and in particular focus on fields that inherently involve the danger of undesired inclusion and unjustified reimbursement. CVZ is

currently elaborating upon this policy, which is referred to as 'risk-oriented package management'. For a more detailed explanation of 'risk-oriented package management', see Package Agenda 2011-2012 and the reports 'Package management in practice, parts 1 and 2.[7]

[7] CVZ. Package agenda 2011-2012. Available via www.cvz.nl. CVZ. Package management in practice. Diemen, 2006. Report no. 245 and Package management in practice 2. Diemen, 2009. Report no. 277. Available via www.cvz.nl.

4. Medical tests

4.a. Relationship to the statutory framework

Medical tests and Zvw

In medical practice, research is carried out using medical tests on patients in order to determine a diagnosis, prognosis, prediction or the course of a disease. These tests, which can vary from questionnaires to advanced conceptualisation or combinations thereof, are an integral aspect of the care provided for patients. This means that tests are covered by the Zvw basic insurance, at least if they can be included under an item described in the basic package. In this respect, many medical tests will go by the name of 'medical care' or 'medical aid care' (see paragraphs 2.a. and 3.b.). If the tests also comply with the ‘established medical science and medical practice’ requirement, then they can be regarded as provisions insured under the Zvw.

AWBZ

The framework described in the previous paragraphs relates to the Zvw. In the Bzv based on that law, ‘established medical science and medical practice’ is explicitly included as a requirement. Although this is not the case for insurance regulated in the AWBZ, nevertheless, for this law also, the assumption is that the right to care only exists if it is effective. Therefore, in principle, the following assessment framework also applies to assessments within the framework of the AWBZ.

4.b. Reason for developing a method for assessing medical tests according to established medical science and medical practice

Necessity of developing a method

As stated in the introduction, CVZ still has limited experience in assessing medical tests according to the criterion ‘established medical science and medical practice’. The reality is, however, that new technological developments continually, and increasingly lead to new tests being used when treating patients. However, tests are not automatically worthwhile or safe, which means they do not always belong in the basic

insurance. For this reason – from the point of view of risk-oriented package management – it is important that CVZ pays attention to the assessment of new and existing medical tests. The development of a working method geared to the assessment of medical tests would contribute to this.

The next section discusses the working method for assessing medical tests.

5. Assessment of medical tests according to established medical science and medical practice

5.a. EBM and medical tests

Emphasis on accuracy in the past

For some time the emphasis within the concept of EBM – with respect to determining the value of medical tests – has been on determining the accuracy of a test.
In brief, this is the question of whether a test actually measures that which it is supposed to measure. A distinction can be drawn here between analytical accuracy and diagnostic accuracy.[8] In medical scientific literature, the term accuracy usually refers to diagnostic accuracy; the extent to which the test is capable of demonstrating or precluding a disorder. In this report the term accuracy is used to mean diagnostic accuracy.

Discussion over focussing on accuracy

Over the course of time, the focus on the diagnostic accuracy of tests has been increasingly open to discussion.
It is becoming increasingly clear that it is not just about accuracy, but also in particular about the effects of using tests on the health of patients and on the means available within (health) care. Acceptable diagnostic accuracy is usually not sufficient to demonstrate the utility of using a test.

Fineberg

As early as 1978 Fineberg wrote the following: "Diagnosis is not an end in itself. (…) In general, medicine is directed toward the goal of improved health outcome.(...) The ultimate value of the diagnostic test is that difference in health outcome resulting from the test: in what ways, to what extent, with what frequency, in which patients is health outcome improved because of this test?" At the time he was already drawing attention to the criticism of the diagnostic accuracy paradigm. He was interested in the question of whether we assess tests

[8] The **analytical accuracy,** or reproducibility. Does repeating the test in a controlled test environment lead to the same outcome? The **diagnostic accuracy**. Does the test measure what it is supposed to measure? How often is the test positive (discrepant) on people who have (or will develop) the disorder concerned and how often is it negative (not discrepant) on people who do not have the disorder?

for what they do (do they provide representative portrayals, do they lead to test results that are congruent with reality), or for their value in improving health outcomes.

Effect on health

No-one can deny the importance of a test's reliability. Currently, however, the prevailing outlook in the literature is that the judgement on a test should be based on an evaluation of the results of using the test on the health of the people who are tested. This applies in particular when decisions are involved on test recommendations for use in medical guidelines or within the insured package. In such cases a positive test recommendation cannot be made based solely on good accuracy. The point of departure is that tests will not be used if they have no positive effect on health or if they do more harm than good in comparison with the alternative of not testing or using a different test. Irrespective of whether the test results themselves are valid.

Example of a test with a high validity but no clinical utility

CA 125 is a tumour marker that can be measured in the blood of patients with a given form of ovarian cancer. An increased level of CA 125 in controls after completing a successful initial treatment (usually) indicates a return of the ovarian cancer. This increase in the CA 125 in the blood usually occurs several months earlier than the clinical signs or symptoms. A recent RCT shows that early treatment of ovarian cancer based on an increase in the level of CA 125 is not associated with any benefit to health in comparison with later treatment based on clinical examination or symptoms. This study demonstrated the inefficacy of routinely determining CA 125 in control women who had undergone successful initial treatment for ovarian cancer, with the aim of possible treatment before the occurrence of (clinical) symptoms of the disease. This is in spite of the fact that the CA 125 is an extremely valid test for the early diagnosis of relapsed ovarian cancer.[9] Furthermore, the scores of the women in the group treated based on CA 125 showed an earlier deterioration in quality of life.

[9] Rustin GJS, van der Burg MEL, Griffin CL, et al. Early versus delayed treatment of relapsed ovarian cancer: a randomised trial. Lancet 2010;376:1155-1163.

Screening

This debate on screening within the framework of the Population Screening Act was settled decades ago and the assessment of any possible new screening tests is based purely on outcomes relating to health. Screening means systematically testing individuals who have no symptoms of the disease that is being researched.
Screening is used to prevent, cure or delay the disease concerned. An implicit requirement is that treatment is linked to screening. The Wilson and Jungner criteria which are used for screening[10] show that it is not just about having a reliable and acceptable test, but also about the availability of treatment that is more effective in the early (pre-symptomatic) stage of the disease that one is trying to detect.

International

Our inventory for this report shows that many international organisations in the field of insured care and guideline development assume health-related outcomes of tests when assessing tests.

5.b. Clinical utility is point of departure

Clinical utility

As package manager, CVZ assesses interventions on the basis of the health outcomes for patients. This applies not only to therapeutic interventions but also to tests.

We assume that not only the intrinsic value of medical tests must be assessed, but also, in particular, their consequences for the health of patients. As package manager, CVZ feels that funding via the basic insurance (collective means), which demands the solidarity of all insured persons, is only justified if the intervention (in this case, the medical test) really is useful to the health of those who undergo the intervention. This means that, as CVZ sees it, a medical test can only be regarded as complying with 'established medical science and medical practice' if there is evidence, or it has been made plausible, that using the test leads to health benefits for

[10] Wilson JMG, Jungner G. Principles and practice of screening for disease. Geneva: WHO; 1968.

patients. The medical test must – in short – have clinical utility.

Test-plus-treatment strategy

Clinical utility refers to an improvement in the health of patients who undergo the test. Whether clinical utility exists will depend in part on the treatment – in the broadest sense of the word – that the patient receives after having undergone the test. This means that an assessment of clinical utility will also involve the treatment – in the broadest sense of the word – that follows the test. In fact, the subject to be assessed by CVZ is: the test-plus-treatment-strategy. The term 'treatment in the broadest sense of the word' includes all interventions that will be used based on the test and that will have an effect on the final outcome for the patient. These interventions can be extremely diverse and could include, for example, therapeutic surgery, medication, additional tests or periods of waiting.

5.c. Phased steps in assessing medical tests

Accuracy and clinical utility

The above means that the clinical utility of a test-plus-treatment-strategy is what eventually determines the value of the test. One of the important factors here is the accuracy of the test. CVZ assumes that an assessment of the accuracy will not always precede an assessment of the clinical utility. The following is an explanation of why we feel this way.

Fixed assessment sequence?

Is there a fixed phased/hierarchic sequence for evaluating medical tests or is it even considered necessary? An article by Lijmer et al.[11] presents a systematic search for articles about schemes for the phased evaluation of tests. The authors found 19 different models for the phased assessment of tests and concluded that no international standard exists for the assessment of tests.
One advantage of a phased assessment of tests could be that more expensive tests would only be done if there were sufficient evidence for the previous steps. However, the authors made a few critical comments about an obligatory

[11] Lijmer JG, Leeflang M, Bossuyt PMM. Proposals for a phased evaluation of medical tests. Med Decis Making 2009;29(5):E13-21.

	phased assessment of tests. For example, If the focus is on diagnostic accuracy, problems may arise in comparing with the existing test if there is no golden standard, or the new test is expected to be better than the present reference test. Furthermore, it may lead to problems with tests that are not for diagnostics, but which are used for such matters as determining the prognosis, predicting the response to treatment, making a choice for a certain treatment or tracking the disease or its treatment. In these situations a reference test is not always available, nor is it clear how to define the desired reference test. Based on this, the authors concluded that the assessment and development of tests is actually a
Cyclic process	cyclic process rather than a series of phases. Therefore, studies that focus on demonstrating improved outcomes for patients (clinical utility) do not necessarily have to precede studies that provide information on the accuracy of a test.

Example of an RCT into clinical utility without prior accuracy research

Jochen Cals carried out research among G.P.s into the effect of (among other things) a rapid test for determining CRP[12] on the prescription of antibiotics to patients with an infection of the lower airways. Antibiotics are often prescribed to these patients, even though research shows that this is of little or no value to this group of patients. However, it is difficult to determine, on the basis of an anamnesis and clinical research, whether a patient has pneumonia (for which antibiotics are effective) or acute bronchitis. In such cases, due to uncertainty about the diagnosis, antibiotics are often prescribed 'for safety's sake'. CRP is known to be a good marker for an infection. However, there are no scientific data on the accuracy of the CRP test when used in primary care with a view to distinguishing between infections of the lower airways that do or do not require antibiotics. Cals showed that adding CRP determination, in comparison with anamnesis and clinical research alone, leads to a significant reduction in the prescription of antibiotics for infections of the lower airways without this having any disadvantages for health outcomes. He demonstrated

[12] C Reactive Protein is an acute-phase-protein in the blood, related to inflammation.

this clinical value without determining in advance what accuracy the CRP test needs in order to demonstrate an infection of the airways that needs to be treated with antibiotics.[13]

Possible phases in researching tests

Models for the phased assessment of tests generally include (among others) the following phases[11]:

1. evaluation of the analytical accuracy;
2. evaluation of the diagnostic accuracy;
3. evaluation of the clinical efficacy (determining the clinical value) of the test-plus-treatment-strategy;
4. evaluation of cost-effectiveness and other additional intended and unintended effects.

CVZ is assuming that in many cases studies into the clinical value of a medical test will only be done after clarity has been obtained about its accuracy (the first 2 phases). This is possible in particular when there is a clear standard reference test. As described above, however, research into the clinical value of tests does not always have to be preceded by research into their accuracy.

Professionals or manufacturers

When requested to assess a test (plus-treatment-strategy), CVZ expects applicants to supply data (except in cases where there are grounds for exclusion) that will form a basis on which CVZ can form an opinion. In addition to a description of the patients on whom the test is used, we also expect a description of the claimed test-plus-treatment-strategy (expressed in health outcomes) and data on the analytical and diagnostic accuracy of the test.

The following is a discussion of clinical utility and our method for determining it.

[13] Cals JWL, Butler CC, Hopstaken RM. Effect of point of care testing for C reactive protein and training in communication skills on antibiotic use in Lower respiratory tract infections: cluster randomised trial. BMJ 2009;338:b1374.

5.d. Clinical utility of medical tests

Benefits to health

The clinical utility of a test-plus-treatment-strategy refers to an improvement in health-related outcomes for patients who have undergone the test. Clinical utility can also become apparent in effects that may be important to those undergoing the test, e.g., a positive effect of the test on ease of use for the patient. A point worth mentioning is that in many cases the health outcomes relating to patients are not the only effects of tests. The deployment of means (technique, personnel, money) or other persons than those who being tested (e.g., with tests for infectious diseases) may also be affected by the test-plus-treatment-strategy.
The occurrence of additional (undesired) effects may only become apparent after using a given test-plus-treatment-strategy in daily practice.
This could form a reason for reassessment by CVZ.

Generally, the effect a test has on health-related outcomes occurs as a result of the treatment (in the broadest sense of the word[14]) that is given after the test results have been made known.

Test's own Influence on health

However, sometimes the test itself can directly affect the health of those being tested. This is sometimes a positive effect: it seems that women with fertility problems who undergo a photo of the fallopian tubes with oil-based contrast subsequently become pregnant more frequently than women who received a water-soluble contrast. Obviously, however, there is more chance that the direct effect of a test will be negative, for example, due to medical complications of the test itself, such as perforated intestines during a coloscopy (camera examination of the large intestines).
These effects of the tests themselves must be taken into account when assessing a test-plus-treatment-strategy.

Patients' influence on health outcomes

Furthermore, the outcome of a given test-plus-treatment-strategy can also be influenced by potential changes in

[14] See paragraph 5b for a description of the concept 'broadly-based treatment'.

patients. These changes may become apparent on an emotional, social, cognitive or behavioural level.

This is illustrated by Fig. 6.2[2].

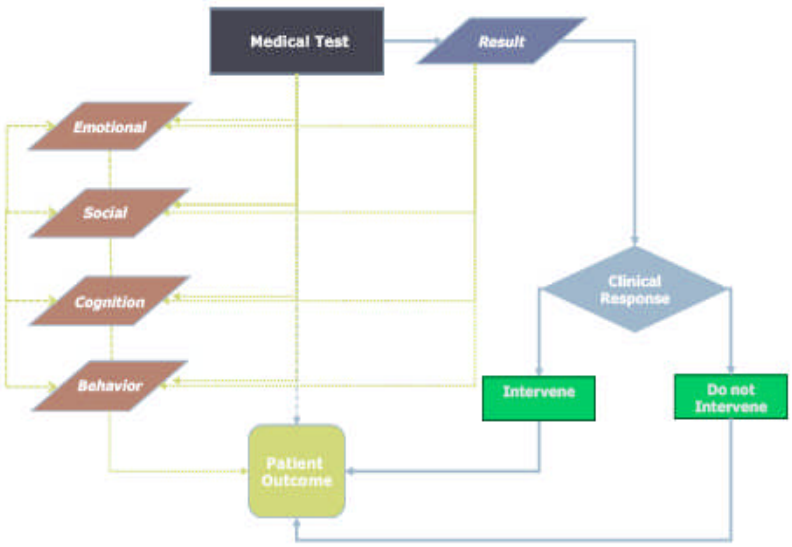


Figure 6.2 - How tests affect patient outcome: alternative pathways and additional outcomes

Outcome instruments

In each case it will be necessary to determine, per test-plus-treatment-strategy, the outcome indicators according to which it can be assessed. It is extremely important that claims are described using health-related outcome indicators.
Intermediate outcome indicators, such as, for example, the therapy choice of doctors or the number of days admitted to hospital, can lead to erroneous conclusions on the utility of a test-plus-treatment-strategy. After all, when the number of days of admission is reduced, there is no certainty about the inference that this is an actual health outcome.
The relevant outcome indicators will have to be weighed against one another. For example, a sizeable clinical health effect of a test-plus-treatment-strategy could form a reason for accepting an additional (unintended) negative effect. In the long run, it is all about finding a balance between desired and undesired health-related outcomes of the test-plus-treatment-strategy. Weighing these up is essentially the same as what CVZ does when assessing therapeutic interventions.

EBRO classification

When assessing the scientific literature obtained from a systematic search strategy and when formulating its conclusions, CVZ makes use of the well-known EBRO classification[15], which focuses on levels of evidence. This is described in the report "Established medical science and medical practice".[1]

Modified Quadas

Based on the experience that, when assessing the diagnostic accuracy of tests, the EBRO classification does not provide a good solution for questions regarding the quality and applicability of a test, CVZ has attempted to find other approaches that may prove more suitable.
It became clear that there had been international initiatives to reach consensus on the approach to assessing the quality of accuracy studies. An important milestone was reached with the

[15] Relates to classification developed within the platform for Evidence-based Guideline Development, the so-called EBRO-platform.

development of the QUADAS instrument in 2003.[16] The Cochrane Collaboration included a section on assessing methodological quality in its manual for systematic reviews of the accuracy of diagnostic tests.[17] This made use of a modified QUADAS[18] instrument.
CVZ wants to use this instrument when assessing the accuracy of tests. It is better suited to the questions we encounter as package manager when assessing tests. What is the risk of a distortion in the results (quality) and does this study provide a reply to the question at hand (applicability)?

5.e. Determining clinical utility

Direct evidence

RCTs for test-plus-treatment-strategies

Randomised controlled trials (RCTs) of test-plus-treatment-strategies, if they are high-quality and of a sufficiently long duration, supply direct – and potentially the best – evidence of the clinical utility of tests. Furthermore, they are capable of indicating not only the intended effects on patients, but also the unintended effects.

RCTs not present

However, there are often no RCTs of test-plus-treatment-strategies and even if they are available, they do not always provide an answer to the question posed. This is partly because it can be more difficult to set up such RCTs for tests than for therapeutic interventions. For example, it may be necessary to include large numbers of patients in the study, because the advantages of the test only apply to a small portion of the group of patients studied. After all, it may be the case that only a small proportion of the studied population shows a 'positive' test result and goes on to be treated according to the test-plus-treatment-strategy that is being

[16] Whiting P, Rutjes AWS, Reitsma JB, et al. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Medical Research Methodology. 2003.
[17] Reitsma JB, Rutjes AWS, Whiting P, Vlassov VV, Leeflang MMG, Deeks JJ,. Chapter 9: Assessing methodological quality. In: Deeks JJ, Bossuyt PM, Gatsonis C (editors), *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* Version 1.0.0. The Cochrane Collaboration, 2009. Available from: http://srdta.cochrane.org/.
[18] QUADAS is currently being revised and amended. Version 2.0 will be ready in a few months.

studied. Another reason for more problems when setting up RCTs for test-plus-treatment-strategies is, for example, the complexity of the research protocol. For example, it may be necessary to provide a detailed description of many steps that follow use of the test. Adhering to all points of such a study protocol can be problematic. Furthermore, it can compromise the external validity of an RCT for a test.

Approach in the absence of RCTs

In fact, RCTs of test-plus-treatment-strategies with a long-term follow-up are not always necessary. Different evidence may be sufficient for determining or demonstrating the plausibility of the clinical utility of the proposed test-plus-treatment-strategy. This applies, for example, in the situation in which a new test has the same accuracy as the old test, but is easier to use, apply or interpret. In this case, it is unnecessary to carry out an RCT in which the two entire test-plus-treatment-strategies are directly compared with one another in order to demonstrate that the new strategy is just as effective as the old one. In this case, research into the test's accuracy is sufficient.
After all, the subsequent clinical strategy remains unaltered and its efficacy has already been demonstrated earlier. Direct evidence can also be obtained from a non-randomised comparative study, a cohort study or similar.
These are studies with a lower level of evidence than an RCT. In some cases, these may be sufficient for assessing the clinical utility of a given test-plus-treatment-strategy. In that case, CVZ will substantiate its reasons for accepting a lower level of evidence.[1]

Indirect evidence

Comparative analysis framework for obtaining indirect evidence

However, matters are not usually this simple and when direct evidence is not available, it is necessary to analyse the indirect evidence that is available.
The literature mentions various approaches for doing this. CVZ chooses to apply a comparative analysis framework, within which the current test-plus-treatment-strategy is compared with the new one. Constructing a comparative analysis

framework can help to reply to the question of whether the clinical utility of a test-plus-treatment-strategy has been demonstrated or made plausible. The usual ('old') test-plus-treatment-strategy is compared with the proposed ('new') one in this comparative analysis framework.
The comparative analysis framework provides points of departure for collecting the available indirect evidence, thereby revealing whether crucial data are missing. In such cases, a comparative analysis framework will help to find sufficient evidence for taking a position on clinical utility, or – as the case may be – a (positive or negative) position on 'established medical science and medical practice', or clarity will be provided over which essential data are lacking. The missing data can form the basis for additional scientific research.

5.f. Working method for assessing clinical utility

Step-by-step assessment

Because of what was discussed in the previous paragraphs, in principle, CVZ adopts a number of assessment steps when examining the test-plus-treatment-strategy with respect to the 'established medical science and medical practice' criterion. First of all, a research question is formulated, based on the PICO diagram. After this, we start looking for direct evidence. Where there is no evidence or it is insufficient, we search for indirect evidence on the basis of a comparative analysis framework.

PICO

1. Formulating a PICO question. The P stands for the patient and the setting in which he/she is being tested; the I is for the test-plus-treatment-strategy being studied; the C is for the comparative test-plus-treatment-strategy (the current best/usual strategy); and the O is for the relevant outcome indicators relating to the patient's health. This PICO is formulated on the basis of the claim made for the test. It is important to formulate precisely for which patients the test will be used and within which setting.

Direct evidence

2. Determining whether direct evidence is available. Preferably in the form of RCTs within which the proposed

claim is studied as test-plus-treatment-strategy in comparison with the usual strategy.

No direct evidence

3. If no direct evidence is available (or it is insufficient) for the claim(s), the next step is to address the question of whether indirect evidence can sufficiently demonstrate the clinical utility of the proposed test-plus-treatment-strategy. This involves making an analysis framework based on a comparison of the usual test-plus-treatment-strategy and the proposed strategy.

Direct evidence

As with other forms of EBM, the strongest form of direct evidence for a given test-plus-treatment-strategy is formed by a set of two or more RCTs that provide consistent results; the assumption being that these RCTs provide the exact same outcome indicators as those mentioned in the claim.
Direct evidence can also be obtained from studies with a lower level of evidence than an RCT. In some cases this can be sufficient for assessing clinical utility. CVZ will have to explain why it is willing to accept a lower level of evidence.[1]
All studies that provide intermediate outcomes must examine the degree to which these are connected with the final health-related outcome indicators.

No direct evidence

The search for indirect evidence takes place according to the analysis framework within which the test-plus-treatment-strategies are compared with one another.
Based on this comparative analysis framework, the critical differences between the new and the usual test-plus-treatment-strategy are identified and the PICO questions are formulated for them. In order to answer these questions, a systematic literature search is carried out, in accordance with the principles of EBM, for each individual question.

Formulating conclusions

Obviously, the formulation of conclusions on the basis of which we, as package manager, can make a decision about whether or not to include a given test (-plus-treatment-strategy) in the package is least complicated when direct evidence is available. A good basis is provided by the levels of

evidence classification, supplemented with the findings of any relevant accuracy data. Assessing indirect evidence means providing separate answers to questions about the crucial differences between the test-plus-treatment-strategies based on the comparative analysis framework. In general, these answers cannot be weighed up solely on the basis of the levels of evidence classification. In order to be able to draw a conclusion, it is often necessary to estimate the strength, the amount and the uncertainty of the evidence obtained. There is no simple recipe for this.

5.g. Details of constructing a comparative analysis framework

Examples of analysis frameworks

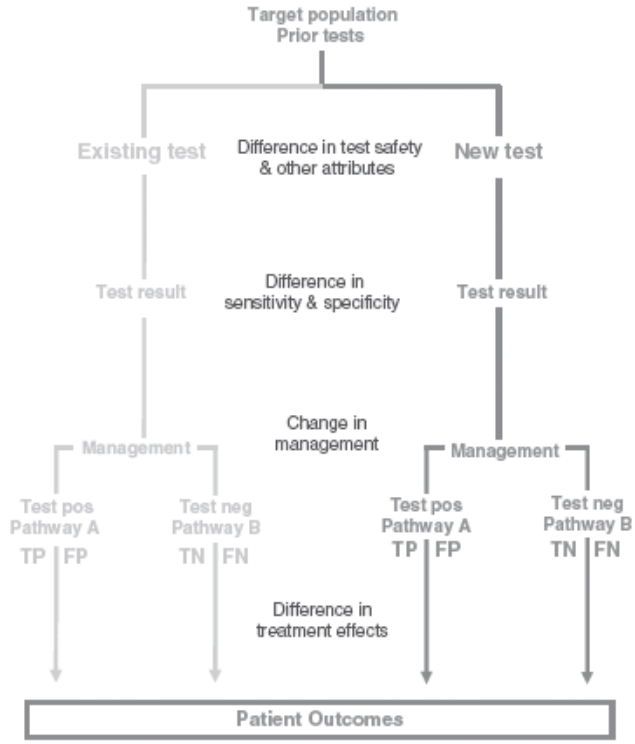
The literature provides examples of comparative analysis frameworks for the various situations that can be distinguished, which are as follows. The new test acts or will act as:

a. a replacement of the usual 'old' test (replacement test);
b. an addition to another test currently in use (add-on test);
c. a triage for another test currently used (triage test).

Examples of analysis frameworks: [19]

[19] Lord SJ, Irwig L, Bossuyt PMM. Using the principles of randomised controlled trial design to guide test evaluation. Med Decis Making 2009; 29; E1.

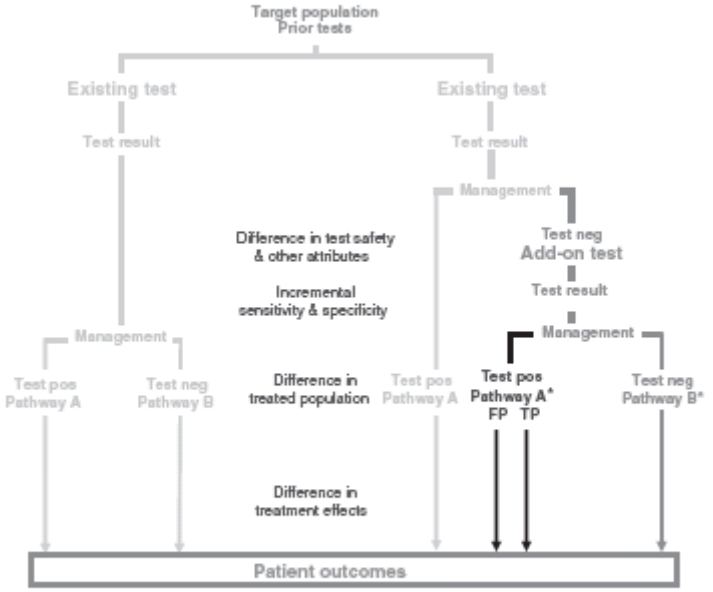
a. The replacement test



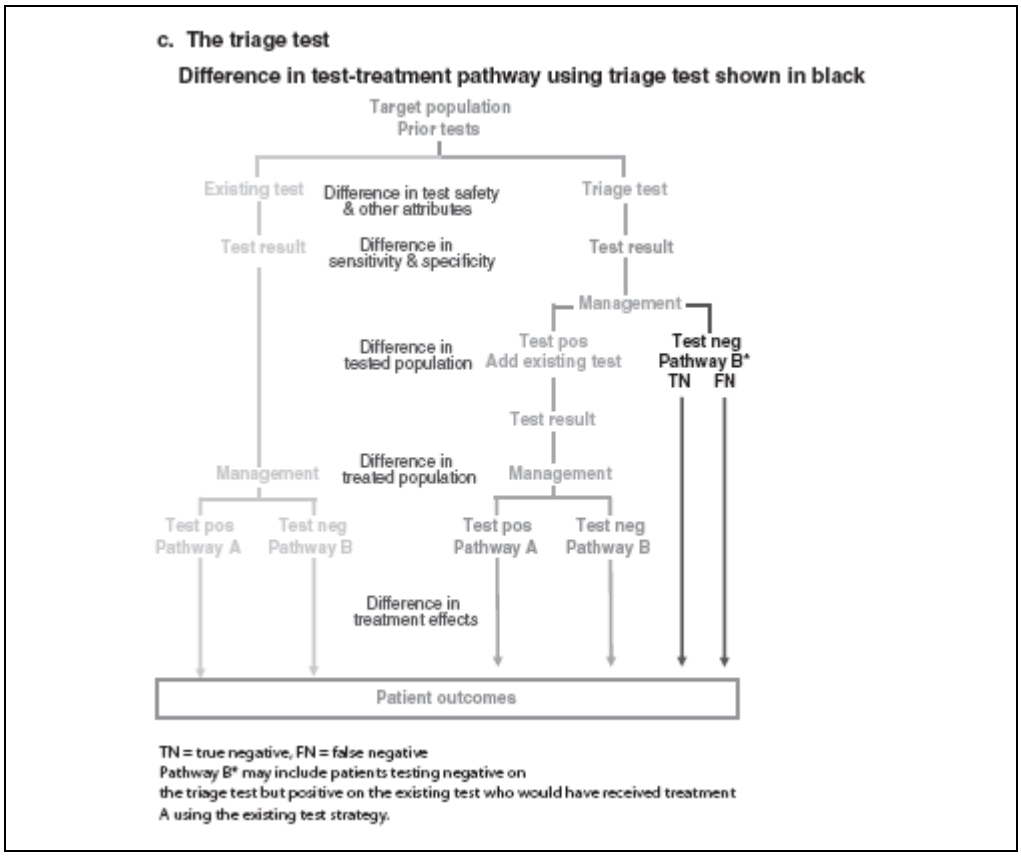
TP = true positive, FP = false positive, TN = true negative, FN = false negative

b. The add-on test

Difference in test-treatment pathway using add-on test shown in black



TP = true positive, FP = false positive
Pathway A* includes patients testing positive on the add-on test but negative on the existing test who would not have been assigned to treatment A using the existing test strategy.



These examples can form the point of departure for CVZ when setting up an analysis framework for assessing a test-plus-treatment-strategy.

Setting up an analysis framework step-by-step

The various steps for setting up an analysis framework on the basis of a comparison between strategies are as follows:

1. determine how the test is to be used, as a replacement test, an add-on test or as a triage test;
2. determine for which patients the test will be used, how and at which stage in the care process (intended use);
3. establish the claim in terms of health outcomes;
4. set out the current (best) test-plus-treatment- strategy;

5. set out the new test-plus-treatment-strategy;
6. Name - as critical comparisons - all (advantageous and disadvantageous) differences between both strategies (these determine the efficacy of the new test and, where necessary, the questions for further research). The following are important in all cases:
   - differences in accuracy and safety of the test;
   - other consequences of the test, such as improved access for patients, increasing the test's prognostic value, improved patient compliance during treatment or prevention;
7. identify and prioritise all differences between the tests in crucial aspects, in order to be clear about the questions posed;
8. on the basis of that prioritisation, carry out literature research to collect replies to the questions mentioned in point 7;
9. determine whether the questions have been answered sufficiently and – where applicable – determine which crucial data are lacking;
10. formulate a conclusion.

Prioritising crucial questions

It will not always be necessary to reply to all questions on the basis of differences in the comparison analysis framework. What is extremely important, as mentioned above, is not only to identify the differences, but also to classify them according to importance. If it proves to be impossible to answer a certain prioritised crucial question in the framework, that may be the moment at which it becomes clear that the test-plus-treatment-strategy does not comply with 'established medical science and medical practice' and there is no need to elaborate any further on the other questions.

5.h. Substantiating a request

Professionals or manufacturers

Part of the working method proposed by CVZ is that whoever asks for CVZ's opinion over a test(-plus-treatment-strategy)

must substantiate their request. After all, when asked to assess a test(-plus-treatment-strategy), CVZ will expect – except in cases of substantiated exclusions – applicants to provide data on the basis of which CVZ can form an opinion. Apart from a description of the patients on whom the test will be used and the setting in which the test will be used, we will also expect a description of the claim of the test-plus-treatment-strategy (defined in terms of health outcomes) and data on the test's analytical and diagnostic accuracy.

5.i. Consulting external experts

Consultation

When determining 'established medical science and medical practice', CVZ always consults relevant experts, depending on the subject. This will also be the case when assessing test-plus-treatment-strategies.
When making their assessment, CVZ will also consult – where necessary – methodological experts, in addition to the experts on the subject matter.

Timing of consultation

Up till now the consultation of experts on the subject matter has generally taken place at the moment that CVZ has already completed a draft report on the 'established medical science and medical practice' of a given intervention. However, in view of the potentially complex nature of the required comparative analysis framework, in some cases it may be necessary to consult experts on the subject matter and methodological experts during the design phase in order to obtain an (initial) critical assessment.

The assessment of test-plus-treatment-strategies on the basis of the available scientific literature will inevitably involve a degree of estimation and assessment of these data, in particular where indirect evidence is involved. It may therefore be necessary to consult the said external experts in this evaluation phase as well.

6. Transparency and support

Transparency and support

CVZ feels that, in order to promote the quality of assessments, it is particularly important to ensure that assessments take place transparently and, where necessary, to seek the allegiance of external experts on the subject matter and methodology. The latter is of particular importance, in view of the fact that the methodology is still under development. Furthermore, obtaining the help of professionals may increase their support for the points of departure of the assessment method and the resulting conclusions.

CVZ is already using the working method described in this report. The working method described (or parts of it) may eventually require adjustments/refinements on the basis of experience or scientific developments. In developing the working method further, CVZ will observe transparency and, where necessary, obtain external expertise.

College voor zorgverzekeringen

Dep. Chairman of the Executive Board

Ms. H.B.M. Grobbink CCMM

2010134727

APPENDIX

Report of the work conference on the working method for assessing medical tests held on 9th November 2010 at the offices of CVZ

The subject of discussion – under the chairmanship of Dr. A. Boer (member of CVZ's Executive Board) – is the Draft Report on Medical Tests (assessment of established medical science and medical practice). This draft report contains an elaboration of the method CVZ will use to examine whether medical tests comply with the established medical science and medical practice criterion. Before finally approving the report, CVZ wants to hold a content-related discussion of the proposed working method with a number of external experts. This is the reason for the work conference. The following experts attended the work conference (in alphabetical order):

- Prof. Dr. W.J.J. Assendelft
- Prof. dr. P.M.M. Bossuyt
- Dr. A. van den Bruel
- Prof. Dr. Y. van der Graaf
- Prof. Dr. K.G.M. Moons
- Dr. A.J. Rijnsburger
- Prof. Dr. R.J.P.M. Scholten
- Prof. Dr. E.W. Steyerberg

The outcomes of the discussion can be summarised as follows.

Consensus over the points of departure in the draft report

The participants in the work conference subscribe unanimously to the choice in the draft report to base assessments of the effectiveness of tests on their clinical utility. They also unanimously confirmed the importance of comparing a new test-plus-treatment strategy with the best current strategy. The participants also all agree with using the method that was developed by the Cochrane Collaboration for assessing accuracy.

Comments on the draft report itself

Carrying out a hypothetical RCT can be an enormous and exceedingly time-consuming task. In some cases it is important to allow room to be able to carry out a more pragmatic assessment of tests that concentrates on the case in hand.

Essentialism versus consequentialism. This discussion has for the main part already been settled: i.e., when the guidelines were being developed. As a result this discussion required less attention in this report.

When assessing tests it is important to bear in mind the level at which a test is being used (1st, 2nd, or 3rd line) and for which indication.

Expertise relating to the subject and methodological expertise is indispensible when setting up a comparison between the current and the new test-plus-treatment strategy, the hypothetical RCT. When interpreting the evidence obtained, it is essential to estimate the risk of distortion, the weight of the evidence, the degree of uncertainty and relevance to the question. Here also, it would be wise to obtain subject-related and methodological expertise.

Although GRADE has started to develop a method for assessing the clinical utility of tests, they are by no means finished. A plan for accuracy does already exist. For the moment, when interpreting indirect evidence it is important to be meticulous in correlating the data obtained with one another and in estimating the bias, cogency, accuracy and magnitude of the evidence.
In practice this means that CVZ must safeguard against encyclopaedic assessments that take years to complete. The main thing is to remain as close as possible to the context in which a test will be used and to weigh the data obtained in dialogue with the professionals.

The term 'hypothetical RCT' can be confusing. What is important is the principle of comparing the existing strategy with the new test-plus-treatment strategy.

Comments on research in the field of tests
Experience has taught that the assessment of new tests sometimes goes no further than an accuracy study, e.g., in cases involving a known test-plus-treatment strategy and a new test with comparable accuracy. In such a case the new test will replace the existing test, e.g., because the latter is more invasive or more expensive. An example is the replacement of invasive venography with non-invasive echography for diagnosing deep venous thrombosis, based only on an accuracy study.

There is often a lack of research into the clinical utility of tests. Research reports involving tests often focus on sensitivity and specificity. This must be taken into account during a systematic search, which will require a broad definition of the terms.

Limited research has been carried out in the field of tests. For example, there is a lack of clarity about the consequences for determining indications for new tests that are much less invasive than the 'old' ones. There is the possibility that the indication for a test will be ‘stretched’, resulting in entirely different effects than those originally expected. This makes matters extremely complex.

Tests are often developed by small businesses without any in-house clinical expertise. The requirements for bringing tests onto the market are extremely limited. The CE hallmark is basically about safety and the ‘plausibility’ of a test’s effects. These (small) businesses do not carry out clinical research (or very little) prior to market introduction; unlike the pharmaceutical industry, for example, which is regulated by many rules. The lack of an equivalent to trial registers means that extra vigilance is required to safeguard against publication bias, with particular attention to sponsoring and confusion of interests.

Follow-up appointments
CVZ will provide a summarised report of the contents of the discussions during the work conference. The report will be sent to the external experts who will be asked if they subscribe to its contents. The report will be enclosed as an appendix to the definitive report.

CVZ will pay due consideration to any comments and observations made during the work conference and – where necessary by way of clarification/refinement – incorporate them in the proposed report.