Children’s competence to consent to medical treatment or research
Hein, Irma

Citation for published version (APA):
Hein, I. M. (2015). Children’s competence to consent to medical treatment or research Amsterdam: Amsterdam University Press

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
3 Accuracy of assessment instruments for patients’ competence to consent to medical treatment or research.


Abstract

Background

The informed consent model assigns patients autonomy over medical interventions and research participation. Where vulnerable patient groups including mentally comprised adults, elderly patients and children and adolescents are concerned, their degree of competence in exercising such autonomy is questionable. Therefore, the determination of patients’ competence is critical in striking a proper balance between respecting the autonomy of people who are capable of making informed decisions, and protecting those whose abilities are impaired. However, clinicians may not know which standard to apply, and probably use many factors that are not formally recognized in law when assessing their patients’ ability to consent to treatment or clinical research. It is not known which instruments are best qualified to assess patients’ competence to consent to clinical care or research.

Objectives

To assess the reliability and validity of assessment instruments for patients’ competence to consent to medical treatment or clinical research, versus a reference standard, in patients of any age with and without developmental or mental deficits, or both.
Search methods

We searched MEDLINE, EMBASE, and PsycINFO from inception to 9 July 2012. No language or publication restrictions were applied.

Selection criteria

We included all prospective diagnostic accuracy studies that assessed instruments covering the four criteria relevant for competence in most jurisdictions: to communicate a choice, to understand the relevant information, to appreciate the medical consequences of the situation, and to reason about treatment choices. Clinicians who are aware of these relevant criteria should be able to assess a patient’s competence, and are used as the generally accepted reference standard. We excluded studies with a time period between index test and reference test of more than 2 weeks.

Data collection and analysis

We screened all titles generated by electronic database searches and abstracts of all potentially relevant studies were reviewed. Full papers were assessed for eligibility and data extracted by two independent assessors. Quality assessment (risk of bias and applicability) was determined using the QUADAS-2 tool.

Main results

The results of the review indicate that with current knowledge, the MacCAT-CR is the most supported instrument for assessing competence to consent to clinical research, with proved reliability and validity in populations of mentally compromised adults and older persons. For a treatment context there is no competence assessment instrument supported by reliability and validity affirmation. The MacCAT-T showed good interobserver agreement, the ACCT showed preliminary indications of good interobserver agreement and good internal consistency, but only moderate validity. Research on the remaining instruments lacked underpinning of reliability and validity.

Authors’ conclusions

Our findings suggest that the MacCAT-CR has good accuracy and could be used for assessment of (in)competence in populations of mentally compromised
adults or older people in a clinical research context. In a treatment context, accuracy of competence assessment tools have not been confirmed yet, however the MacCAT-T is promising and the ACCT shows moderate results. The validity and generalizability of our findings are limited because they are based on few, heterogeneous studies with methodological flaws, with important concerns on the reference standards used. Hence, this review highlights the need to conduct more accuracy studies on competence assessment tools. Especially regarding specific populations, including children, and regarding a treatment setting, more research on assessment tools is needed.

Background

The informed consent model assigns patients autonomy over medical interventions and research participation. Where vulnerable patient groups including children and adolescents are concerned, their degree of competence in exercising such autonomy is questionable. Partly because of concerns about the adequacy of consent procedures in vulnerable populations, interest in research into decisional capacity has grown in recent years.

Consent is required for all aspects of medical care, for preventive, diagnostic or therapeutic interventions and research participation. Competence to consent, as we use it in this review, is the clinical concept of the ability of a person to consent to medical interventions or clinical research. The clinical concept of competence may be distinct from the legal one. By law clinicians are required to determine whether patients are competent to give their consent. Strictly speaking, incompetence denotes a legal status that in principle should be determined by a court. Resorting to judicial review in every case of suspected incompetence, however, would very heavily burden both the medical and legal systems; there is therefore good reason to continue the traditional practice of having clinicians determine patients’ competence.(38)

Within the context of daily medical practice, competence is usually assessed implicitly. However, in some clinical settings competence regularly becomes problematic, especially when concerns arise about a person’s capacity to make well-considered medical decisions. For example, if a 14-year old boy with acute lymphatic leukemia is eligible for drug trial participation, how should the researcher examine the competence of his decision to participate in the trial? If a 15-year old boy with germ cell
cancer and anemia expresses his wish not to receive blood transfusion during planned surgery, for religious reasons, is this decision a competent one? A 63-year old man with type 2 diabetes mellitus and schizophrenia is recommended a below-the-knee amputation for peripheral vascular disease but declines; how can his physician assess whether his choice is based on competent decision-making? A woman of 72, living with dementia and anemia, is recommended to undergo an investigation to trace a location of blood loss, but refuses – is she competent to give informed refusal?

The most extensive and influential research on patients’ competence to consent was conducted by the MacArthur Research Network on Mental Health and the Law,(39) examining competence standards identified by the legal system as relevant to decision-making competence in the USA, UK and many other nations. The four legally-relevant abilities that were addressed were: the ability to state a choice; to understand relevant information; to appreciate the nature of one’s own situation; and to reason with information. These four abilities have been generally accepted as the standard for patients’ competence to consent in clinical treatment and research practice. Several interview procedures to operationalize these standards and to measure abilities have been developed in recent decades.

The determination of patients’ competence is critical in striking a proper balance between respecting the autonomy of people who are capable of making informed decisions, and protecting those whose abilities are impaired. There is an undeniable need for standardized and accurate competence assessment methods.

Target condition being diagnosed

The target condition we will focus on in this review is patients’ competence to consent to medical treatment or clinical research. We will cover groups of patients with and without developmental or mental deficits, or both.

Index test(s)

A variety of methods for competence assessment exist, mostly consisting of a structured or semi-structured interview format. Some instruments examine a real-life medical decision, while others are based on clinical vignettes. The instruments include the MacArthur Competence
Assessment Tool for Clinical Research (MacCAT-CR), the MacArthur Competence Assessment Tool for Treatment (MacCAT-T), and the Hopkins Competency Assessment Test (HCAT). Content, rating and cut-offs vary between index tests and there is no systematic study on their reliability and validity.

Clinical pathway

The process of obtaining informed consent starts with providing appropriate information about the proposed medical intervention and alternative options or the scientific research project. An essential element of consent is that informed choices are made voluntarily without coercion or force, and that the person is competent to make such a decision. Starting-points for competence are task and context specificity. This means that competence should not be conceived as an all-or-nothing judgment implying that the patient is generally competent or generally incompetent. Instead, assessment of competence should be regarded as a specific judgment at a specific moment of whether the patient is able to complete the concrete task that they are facing. The law imposes a dichotomy (competent versus incompetent) on what, from a clinical perspective, is a spectrum of capacities.

In law, competence has traditionally been regarded as a function of age. The statutory age of majority is generally set at 18 years, although exceptions exist in states or countries worldwide. Competence is presumed present in all adults and is rarely examined as long as the outcomes of decisions concur with the physician’s recommendations.

Competence to consent may be reduced by several influences such as cognitive impairment, developmental immaturity, certain psychiatric symptoms, and situational factors such as the complexity of the information. Children are deemed competent if they appear to understand information designed for their level of comprehension to an extent appropriate to the nature and scope of the decision. Internationally the statutory age limits differ for clinical research: the lower age limit varies from 7 to 15 years, while the upper age limit is set at 17 or 18 years. Also, for treatment decisions various age limits exist: in some countries autonomous decision-making is lawful only from 18 years onwards, but in other countries minors are allowed to take healthcare decisions from a fixed age below legal majority starting from 12 years. Parents decide for children who are younger than the lower age limit, as these children are considered by definition incompetent to act for themselves. For these children, no actual assessment of competence is necessary. For children between the two age limits,
informed consent is required both from children and parents, provided the child is judged competent to decide. Above the designated upper age limit, children are deemed adult in medical decision-making.

In case of incompetence in adults, usually family is allowed to make a proxy decision for general treatment decisions.\(^{(43)}\) In the context of research, laws are less clear, but generally some kind of proxy decision-making is allowed.\(^{(43)}\) In emergency contexts, a person's stress, pain or diminished consciousness may impair their competency. For treatment decisions the most likely result of the informed consent process under these circumstances is 'uninformed trust'.\(^{(44)}\) Decisions on research participation in the emergency context are a matter of debate; however, formally some kind of proxy decision-making by a legally authorized person is generally allowed.\(^{(43;44)}\)

Possibilities to advance patients' competence rely greatly on improved information provision. Decision aids are developed to provide adult patients and their families with the relevant information about the available options and possible outcomes, to support them in making a decision that is aligned with their preferences. In children decision-making can be facilitated by breaking the process down into smaller but linked choices. Communication difficulties can be overcome by innovative and age-appropriate techniques to convey information.\(^{(4)}\) In conversation, children need clearly-worded information tailored to their comprehension level. It has been found in current practice that communication with parents and children is often flawed, and even that children are not fully informed.\(^{(4)}\)

At present, clinicians tend to make intuitive assessments of children's and adolescents' competence, because no standardized method is available to test it objectively. Currently clinicians base their competence judgments on information such as age or school type. It is recognized that age is, at best, a proxy for developmental capacity, and that experience, maturity and psychological state are key determining factors.\(^{(4)}\) In adults, clinicians may not know which standard to apply, and probably use many factors that are not formally recognized in law when assessing their patients' ability to consent to treatment or research.\(^{(45)}\) In older adults and psychiatric patients, clinicians might find it difficult to distinguish between mental status examinations and competence assessment.\(^{(46)}\) Several methods of structured competence assessment exist for adults, varying widely in procedure, reliability and validity. However, in the research context few investigators assess understanding of the research protocol and competence prior to accepting consent, and the use of standardized tools is the exception rather than the rule.\(^{(47)}\) Data suggest that the performance of
competence assessments is often sub-optimal and hence the reliability of unstructured judgments has been poor. Providing clinicians with the generally accepted legal standards for competence improves their judgments and increases significantly the inter-rater agreement. These legal standards embody the four capacities: to communicate a choice, to understand the relevant information, to appreciate the medical consequences of the situation, and to reason about treatment choices. Clinicians who are aware of these relevant criteria should be able to assess a patient’s competence.

Alternative test(s)

An assessment measure for competence to consent should have close conceptual relationships with the relevant standards of competence. This implies that more general measures of cognitive abilities like reading ability or Mini Mental State Examination (MMSE) would not be appropriate for a valid test of the specific context-dependent competence to consent to the research or treatment on offer.

Rationale

This review addresses the following question: which instruments are best qualified to assess patients’ competence to consent to clinical care or research? We will analyze the validity and reliability of the assessment instruments.

Objectives

To assess the reliability and validity of the index tests for competence assessment versus the reference standard in people of any age. Although a gold standard for competence does not exist and the reference standard may be imperfect, we will examine whether using a structured assessment instrument instead of an expert judgment would be possible without compromising accuracy.

Secondary objectives

To examine the accuracy of standardized competence assessment instruments in the subpopulation of patients under 18 years of age, with deficits in competence due to developmental stage.
Methods

Criteria for considering studies for this review

Types of studies

We will include prospective observational studies on test accuracy. We will exclude comparative test accuracy studies, in view of the expected challenges related to the imperfect reference standard. We will also exclude diagnostic case studies.

Participants

This review deals only with assessment of competence regarding consent to treatment and consent to scientific research programs. Most assessment instruments can be applied to heterogeneous patient populations. We will include children, adolescents, adults and elderly populations with different conditions, including medical conditions, cognitive impairment, psychiatric disorders and co-morbidities. The target condition, competence to consent, can vary over time and be influenced by many factors, so the time period between index test and reference test must be short enough to be reasonably sure that the target condition did not change between the two tests.

Index tests

In an effort to standardize and hence increase the reliability and validity of competence evaluations, several formal assessment instruments have been developed. Some instruments offer a reported cut-off point. Other instruments do not provide a cut-off point, stating that a serious deficit in any of the tested domains may translate to a clinical opinion of incompetence. We anticipate variation in threshold and lack of a consensus about thresholds as challenges for this review. Data on validity and reliability of the index tests are not available in standardized summaries.

We will exclude instruments that do not cover all four relevant criteria (T for treatment context and CR for clinical research context): Aid to Capacity Evaluation (T), Brief Informed Consent Test (CR), California Scale of Appreciation (CR), Competency Assessment Interview (T & CR), Deaconess Informed Consent Comprehension Test (CR), Direct Assessment of Decision-Making Capacity (T), Evaluation to Sign Consent
Hopemont Capacity Assessment Interview (T),
Hopkins Competency Assessment Test (T),
Informed Consent Survey (CR),
Ontario Competency Questionnaire (T),
Quality of Informed Consent questionnaire (CR),
Two-Part Consent Form (T & CR),
University of California San Diego Brief Assessment of Capacity to Consent (T).

Most vignette methods will not be included, as we judged assessment of appreciation to be insufficient.

We will include the following index tests, comprising all four relevant criteria:

- Assessment of Capacity to Consent to Treatment (T), a structured interview and three vignettes, studied in adults with dementia, schizophrenia and controls.
- Assessment of Consent Capacity for Treatment (T), three vignettes taking 45 minutes to administer and used in adults with and without mild and moderate retardation.
- Competency Interview Schedule (T), a structured interview, applied in people with major depression.
- Competency to Consent to Treatment Instrument (T), hypothetical vignettes and a structured interview, taking 20 to 25 minutes administration time. It is used in people with Alzheimer’s disease, dementia, Parkinson’s disease and controls.
- Competency Questionnaire (CQ), a 15-item questionnaire covering all four capacities developed by Appelbaum in 1979. Each question is rated a 0 or 1, and added up for an overall score. Several modified versions were developed:
  - CQ – Child Psychiatric (T), a 17-item questionnaire to test competence in children to consent to psychiatric hospital care and treatment.
  - CQ-Peds (T), a 19-item questionnaire for use in pediatrics, used for inpatients and outpatients between 5 and 18 years of age, and CQ-Med (T), for assessing competence of general medical patients to consent to hospitalization.
- MacArthur Competence Assessment Tool for Clinical Research (MacCAT-CR) (CR), a semi-structured interview format developed by Appelbaum and Grisso in 2001, that guides clinicians and patients through the process of information disclosure required for informed consent, combined with an assessment of the patient’s capacities, in approximately 15 to 20 minutes. The instrument provides scores for each sub scale: 0 – 26 for understanding, 0 – 6 for appreciation, 0 – 8 for reasoning, and 0 – 2 for expressing a choice. MacCAT-CR is based on the structure of the MacArthur Competence Assessment Tool for Treatment (MacCAT-T).
MacArthur Competence Assessment Tool for Treatment (MacCAT-T) (T) was preceded by the original MacArthur instruments (Understanding of Treatment Disclosures, Perception of Disorder, Thinking Rationally About Treatment, Expressing a Choice). MacCAT-T was developed by Grisso and Appelbaum in 1998 and is a semi-structured interview format taking approximately 15 to 20 minutes to administer.(79) The instrument provides scores for each sub scale: 0 – 6 for understanding, 0 – 4 for appreciation, 0 – 6 for reasoning and 0 – 2 for expressing a choice. The MacCAT scales do not offer a total score or a cut-off for competence, but the scores on the sub scales need to be weighed by the interviewer. The MacCAT scales receive the most empirical support out of the variety of assessment instruments. The MacCAT scales have been tested in particular in samples of people with dementia, mental disabilities, schizophrenia and other psychiatric disorders.(80-83)

Older Adults’ Capacity to Consent to Research (CR), a brief tool for assessing older adults’ capacity to consent, consisting of four items each testing one of the four capacities, and used in nursing home residents and community-dwelling older adults.(84)

Structured Interview for Competency and Incompetency Assessment Testing and Ranking Inventory (T), a 20-minute structured interview examining all four aspects of competence, and used in psychiatric and medical patients.(85)

Target conditions

The target condition is the patient’s competence at that moment to consent to treatment or research participation. The outcome is binary: yes or no.

Reference standards

Agreement is poor between unstructured clinical competence judgments by physicians, and no better than chance. Providing clinicians with information regarding the legal standards improves their judgments and significantly increases the inter-rater agreement to moderate (kappa 0.46). (2;86) These legal standards embody the four previously-mentioned capacities: to communicate a choice, to understand the relevant information, to appreciate the medical consequences of the situation, and to reason about treatment choices. Clinicians who are aware of these relevant criteria should be able to assess a patient’s competence, and they are generally considered to establish the reference standard.(2;33;59;87) However, limitations of this
approach include the frequent discordance of expert competence judgment, which would lead to inconsistencies in the reference standard. Where the experts give their clinical judgment based on their knowledge of the four relevant criteria, the index tests consist of an operationalization of these criteria into interview questions and do not necessarily have to be administered by an expert. When agreement between reference standard and index test is high, the index test performs as well as the reference standard. Poor performance of the index test could be interpreted in different ways: it could result from imperfections in the reference standard, or the index test may not offer an accurate assessment of competence. Slightly different reference standards in each study would preclude a quantitative meta-analysis, and it would not be possible to demonstrate superiority of any of the index tests.

Search methods for identification of studies

We will use a single search strategy for this review.

Electronic searches

To identify all relevant studies, we will search the following databases: MEDLINE, EMBASE, PsycINFO. We will use the search terms and strategy summarized in Appendix 1. We will restrict the searches to human studies. We will not limit the search by language or publication status. We will perform cited reference checking of the initially included articles in Web of Science.

Data Collection And Analysis

Selection of studies

Two review authors (IH and MM) will initially screen all the titles and abstracts identified by the search strategies for eligibility. Two review authors (IH, MM and/or LG) will retrieve potentially relevant papers in full and assess them independently using a screening form developed for this review. We will list the excluded studies in the Characteristics of Excluded Studies Table. We will resolve any discrepancies by discussion. Where two review authors cannot reach agreement, we will consult the third review author. We will include study reports in English, Dutch and German; if
we find suitable studies in other languages we will attempt to get them translated. We will identify studies by the surname of the first author and the year of publication.

Data extraction and management

Two review authors (IH and LG) will independently extract a standard set of data from each study using a tailored data extraction form (see Appendix 2), resolving any discrepancies by discussion. In cases where only a subgroup of participants meet the review inclusion criteria, we will extract and present data for that particular subgroup only.

We will extract information about the capacity domains assessed.

Results of the reference test for competence to consent are dichotomous; positive is defined as ‘competent’ and negative as ‘not competent’. Results of the index tests may be mixed: some may offer a reported cut-off point and provide dichotomous results, while others may not provide a cut-off point but report that a serious deficit in any of the tested domains translates to a clinical opinion of incompetence. In that case we will compare the various given cut-offs with the reference standard. For each comparison of index test with reference test, we will extract data on the number of true positives, true negatives, false positives and false negatives in the form of a two-by-two table.

Assessment of methodological quality

Two review authors (IH and LG) will independently assess the quality of each individual study using the checklist adapted from the QUADAS-2 tool; the criteria are summarized in Appendix 3. We will answer each question on the checklist with a yes/no response, or note it as unclear if insufficient information was reported to allow us to make a judgment; we will document the reasons for this judgment. We will assess whether the study design and patient selection was appropriate, and whether the reference test and index test were rated blind to the results of each another. We will also assess if cut-off values were prespecified. Given that most instruments will not provide this information, we will not exclude studies that do not offer a prespecified cut-off but will report it as a possible risk of bias. Furthermore, we will assess whether the reference standard consisted of judgment by an expert aware of the relevant criteria, which must be accounted for by a description of the training the expert received. We will assess whether all patients received both index
test and reference test, and will exclude patient groups who did not. If an atypical reference test was used, for example competence judgment by a panel of clinicians who are not aware of the relevant criteria, we will also exclude these patient groups. We will assess the time interval between reference standard and index test; we consider two weeks as justifiable. Although in cases of acute illness the level of competency could change markedly over two weeks, we do not intend to exclude studies designed for a population of patients with chronic diseases where researchers need some time to enroll them. We will summarize the methodological quality judgments graphically.

**Statistical analysis and data synthesis**

We will examine the validity and reliability of the competence assessment instruments. Content validity (the degree to which the instrument’s content reflects the universe of content relevant to the constructs being measured) is usually determined on the basis of expert consensus. In this review we will examine whether each instrument’s construct is consistent with the widely-accepted four-capacities model.

We will assess criterion validity, the degree to which scores on a scale are associated with the standard, in terms of inter-correlations. Other useful values are the scale’s sensitivity (valid positive, in this case competent) and specificity (valid negative, incompetent) rate. The accepted standard against which criterion validity is evaluated may be an established measure. In the absence of a gold standard for measuring competence, we will use the expert judgment by clinicians aware of the four relevant criteria as a reference standard.

If we have sufficient data, we will summarize results for the patient population under 18 years of age with deficits due to developmental stage only.

Marked inconsistency in the reference standards in the included studies will be particularly difficult to deal with quantitatively. In this systematic review we will demonstrate the descriptive elements of the index tests without performing a meta-analysis, which may be equally important in this first attempt at reviewing evidence on this topic.

We anticipate that within some index tests various cut-off values will have been applied to individuals. In that case we will perform the above-mentioned analyses for subgroups of studies that report similar cut-offs for that index test.
Investigations of heterogeneity

Not applicable.

Sensitivity analyses

Not applicable.

Assessment of reporting bias

We judge it acceptable not to assess reporting bias, given that very little is known about publication bias in test accuracy studies and that extrapolating from publication bias in effectiveness research may not be appropriate.

Results

Results of the search

We identified 3304 references through electronic searches of MEDLINE (Ovid SP) (N = 829), EMBASE (Ovid SP) (N = 600), and PsycINFO (Ovid SP) (N = 1875). After the exclusion of 753 duplicates, 2551 references remained; we found 2523 to be irrelevant references. Twenty-eight references on studies seemed to fulfill the inclusion criteria. We excluded seventeen studies after reading the full text. Finally, we included 11 studies and considered them for data analyses.

Included studies

Eleven studies are included, of which four studied participants’ competence to consent to clinical research(33;34;84;89) and 7 studied participants’ competence to consent to treatment.(69;71;78;85;90;91) Six studies reported on MacCAT scales,(33;34;81;89-91) other assessment instruments were reported on in one study each: Competency Interview Schedule (CIS),(71) Competency Questionnaire-Med (CQ-Med),(78) Older Adults’ Capacity to Consent to Research (OACCR),(84) Assessment of Capacity to Consent to Treatment (ACCT),(69) and Structured Interview for Competency and Incompetency Assessment Testing and Ranking Inventory (SICIATRI). (85) Five studies reported on competence to consent in a geriatric population,(33;78;84;90;91) four studies in a population of patients with psychiatric
disorders, and two studies in a mixed population (geriatric/psychiatric, psychiatric/medical conditions). There were no studies in a population of minors.

**Excluded studies**

Seventeen studies are excluded from the review. The reasons for exclusion were: the index test did not cover all four relevant criteria for competence in ten studies (Hopkins Competency Assessment Test, Competency Questionnaire, University of California San Diego Brief Assessment of Capacity to Consent, Hopemont Capacity Assessment Interview, and vignette methods). Four studies did not use a reference standard. Two studies used an invalid reference standard (a Community Treatment Order, and competence judgments of untrained clinical teams).

**Methodological quality of included studies**

Overall, concerns were high on risk of bias and applicability of the reference tests used in the included studies (Figure 1). Only two studies were considered to be of good quality, and one study to be of moderate quality. The other included studies raised high or unclear concerns on at least two out of three quality dimensions concerning patient selection, reference standard and index test (Figure 2).
Figure 2. Risk of bias and applicability concerns summary: review authors’ judgments about each domain for each included study

<table>
<thead>
<tr>
<th></th>
<th>Risk of Bias</th>
<th>Applicability Concerns</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Patient Selection</td>
<td>Index Test</td>
</tr>
</tbody>
</table>

Findings

Accuracy of assessment tools
All disagreements between evaluators were resolved through discussion. The data extraction was difficult mainly because of the above mentioned heterogeneity in the definition of the reference standard. Very few studies used the preset reference standard being competence judgment by clinicians who are aware of and trained in assessing the four domains of competence (to communicate a choice, to understand the relevant information, to appreciate the medical consequences of the situation, and to reason about
treatment choices). Data of only two studies of good quality on MacCAT-CR could be combined, and outcomes of one study on CQ-Med can be presented as reliable. Eight studies show shortcomings concerning suitability for establishing accuracy, and will be presented and discussed separately. The results are presented according to the index test used.

**Index test 1: MacCAT-CR**

There were two studies of good quality using MacCAT-CR. In the first study, participants were 37 patients with Alzheimer’s disease and 15 elderly controls. The gold standard was based on expert judgments on competence to consent of the patients with Alzheimer’s disease. Results show a high interrater reliability of the MacCAT-CR, ICC=0.94. Optimal cutoffs for the sub scale scores of the MacCAT-CR were determined based on the reference standard: understanding 18 (sensitivity=90%, specificity=88%); appreciation 5 (sensitivity=80%, specificity=100%); and reasoning 6 (sensitivity=100%, specificity=75%). Using this cutoff, 31 of the 37 patients with Alzheimer’s disease were found incompetent and all 15 comparison subjects found competent. These results suggest accurate cutoff scores on sub scales of the MacCAT-CR.

The second study, performed by the same first author, examined competence to consent in 91 people with severe mental illness (55 participants in a schizophrenia study, and 36 people recruited at in- and outpatient clinics serving people with severe and persistent mental illness) and 40 people in the community comparison group. The reference standard was established by three psychiatrists with experience in assessing decisional capacity. These experts gave their categorical judgment (competent/incompetent) based on video-taped MacCAT-CR interviews with the participants. A majority (2 or 3 out of 3) agreement among the three experts was used to determine the final competence status. The reference standard was established for 90 people with severe mental illness and for 11 controls. A disadvantage of this approach is that the expert judgment was influenced by the information from the MacCAT-CR interview. Results demonstrate that 25 of the 101 people judged, were deemed incompetent. Pairwise kappa coefficients among the experts establishing the reference standard ranged from 0.56 to 0.90, the group kappa coefficient was 0.69. Intra-class correlations for total scores of MacCAT-CR subscales were 0.93 for understanding, 0.89 for appreciation, 0.84 for reasoning. MacCAT-CR sub-scale scores were analyzed for determining the optimal cutoff with the highest sensitivity and specificity. The area under the ROC curve was higher for understanding, 0.94,
than for appreciation, 0.85, and reasoning, 0.80. For understanding, the optimal cutoff corresponded with a score of 18. This is the second study indicating that an accurate cutoff on a subscale of the MacCAT-CR, in this case understanding, can be determined.

One study on MacCAT-CR did not meet the quality criteria for the index test and reference test,(89) taking into account that the reference standard consisted of the clinical judgment of competence by the Principal Investigator, and the index test MacCAT-CR was interpreted with knowledge of the reference standard. Results are hard to interpret because of the deficient quality of the measures. The study population consisted of 106 forensic psychiatric patients, and results showed that psychiatric symptoms were modestly related to decision-making competence.

**Index test 2: CQ-Med**
The study on the CQ-Med is considered to be of sufficient quality.(78) The study population consisted of 29 older patients admitted to a geriatric medicine unit, who were assessed on their competence to consent to hospitalization and treatment. The CQ was developed by Appelbaum in 1981 and was further modified into the CQ-Med by the research group of Billick and colleagues, for assessing the competence of general medical patients to consent to hospitalization. The reference standard consisted of a competence assessment using a clinical exam by a geriatric psychiatry fellow. Results show that the reference standard correlated with the CQ-Med, r=0.58. Validation of the CQ-Med to the reference standard showed a sensitivity of 0.80 (8/10) and a specificity of 0.60 (3/5). Reported limitations of the study include the small sample size and the lack of comparison of the CQ-Med with other competence assessment instruments, such as the MacCAT-T developed by Appelbaum in 1997. Concerns rise on covering all four capacities by the CQ-Med, as the study does not describe exactly how the CQ-Med is composed. The original CQ was left out of analysis because we judged that the emphasis was on understanding, CQ-Med may struggle with the same issue.

**Index test 3: MacCAT-T**
One study on MacCAT-T was a cross-sectional survey conducted in a sample of patients with mild cognitive impairment and mild dementia. (90) One-hundred patients were included, their relatives (99) and their referring physicians. Decision-making capacity concerning a hypothetical medical decision was examined, making use of the MacCAT-T as index test and referring physician’s judgment on patient’s medical decision-making.
competence as a reference standard. Some sections of the MacCAT-T were not included because patients had different diagnoses and the treatment was hypothetical. Results demonstrated that physician’s confidence in patient’s decision-making correlated significantly with the MacCAT-T sub scale understanding, correlation coefficient, 0.24, and appreciation of treatment, 0.34. Physician’s confidence in patient’s decision-making did not correlate with the sub scales reasoning, 0.14, and expressing a choice, 0.10. Both the reference standard as the index test, however, do not fulfill all the quality requirements.

Lui(91) studied competence to make treatment decisions in 66 community dwelling Chinese older people, of whom 33 with very mild dementia and mild Alzheimer disease, using MacCAT-T as an index test. The reference standard used was a competence judgment by three psychiatrists based on four audio taped questions. Results show that 17 of the 33 patients were judged incompetent on the reference standard, and 15 on MacCAT-T. Participants who were incompetent on the reference standard, scored significantly lower on the MacCAT-T sub scales. The authors describe that the MacCAT-T scores correlate significantly with the competence judgment of a clinician based on that same MacCAT-T interview, which however obviously must be influenced by it. Interrater agreement of MacCAT-T among three raters was good: 0.83 for understanding, 0.78 for appreciation, 0.77 for reasoning, and 0.64 for expressing a choice.

The MacCAT-T was studied in a large sample of 350 psychiatric patients referred for admission and treatment.(81) There was not a clear reference standard, as the study aimed at estimating the prevalence of mental competence to make treatment decisions in people from different diagnostic and legal groups admitted to psychiatric hospital. The study compared the opinion of the psychiatric trainee with the outcome of the MacCAT-T, however, the study population was partly assessed by trainees (n=138) and partly assessed by MacCAT-T (n=200). The prevalence of incompetence was estimated to be 60% (95% CI 55-65).

Index test 4: ACCT
Moye and colleagues(69) studied the ACCT in 59 adult participants: 20 people with dementia, 20 with schizophrenia, and 19 controls from a primary care clinic. The ACCT is developed by Moye and starts with a value interview, followed by 3 hypothetical vignettes. The reference standard consisted in 27 cases of the rating based on the clinical opinion of the primary care provider, and in 12 cases of the retrospective rating by three experienced clinicians not trained in the four relevant criteria. Internal consistency
reliability was alpha = 0.96 based on all items in 56 participants, and for sub scale understanding a = 0.91, appreciation, a = 0.88, reasoning, a = 0.82, and communicating a choice, a = 0.66. For 10 cases, interobserver agreement was analyzed for the total score, r = 0.90, and for sub scale understanding, r = 0.90, appreciation, r = 0.89, reasoning, r = 0.68, and communicating a choice, r = 0.98. Validity of the ACCT based on the reference standard of primary care providers’ judgment, showed kappa = 0.44. Validity of the ACCT based on the reference standard of experienced clinicians’ judgment, showed kappa = 0.50.

Index test 5: OACCR

OACCR is a four-item instrument for assessing competence to consent to clinical research developed and studied by Lee and colleagues in 203 nursing home residents and 201 community dwelling older adults in South Korea. (84) The reference standard consisted of a competence judgment by the capacity to consent screen, a 10-item instrument covering all four relevant criteria. Internal reliability of OACCR was demonstrated with high item-total correlations, ranging from 0.59 to 0.76 and an overall reliability coefficient of 0.85. Interobserver agreement is not reported. To establish validity, the relationship between the OACCR and the capacity to consent screen was analyzed using Pearson correlation coefficient, r = 0.94. In 401 participants cross-tab analysis showed sensitivity of 99% and specificity of 73%.

Index test 6: CIS

The CIS consists of 15 items: 4 on evidencing a choice, 3 on understanding information, 2 on rational reasoning and 5 on appreciation. (71) Responses are rated on a 7 point Likert scale: 1-3 adequate, 4 marginal, 5-7 inadequate. The outcomes of competence on the CIS are not reported. The reference standard was based on judgment by a third person who reviewed medical records to determine whether the patient was found competent or incompetent by the attending physician. Participants were 96 admitted patients with severe psychiatric disorders referred for electroconvulsive therapy. Results describe that 75 patients were found competent on the reference standard. In 13 patients, test-retest reliability was analyzed for the average correlations based upon pooled items: intra class, r = 0.79. In 11 patients, interrater reliability was examined, intra class correlation coefficient for the average correlations based upon pooled items was r = 0.95. Internal consistency of the instrument was examined using inter-item correlation coefficients, which ranged from 0.39 to 0.85. Validity
was tested by calculating correlations of individual item scores with the reference standard, which ranged from 0.35 to 0.73, showing a significant association. Further analysis of sensitivity, specificity or predictive value was not performed.

**Index test 7: SICIATRI**

Competence to consent to treatment was tested using SICIATRI in 48 inpatients, of which 25 had a psychiatric disorder and 23 a medical condition. The SICIATRI consists of 12 items, each rated on a 3-point scale, after administration the patient was classified into 5 categories ranging from completely incompetent to completely competent. The reference standard consisted of a competence rating by the attending physician using the Disclosure Consent Checklist, which focuses on understanding the nature and purpose of hospitalization and treatment. Interrater reliability of the SICIATRI is expressed by kappa’s ranging form 0.14 to 0.82, with 6 out of 12 items showing a kappa of 0.60 or higher. On the reference standard, 13% of the participants were incompetent. Thirty-five (73%) patients were competent on both reference standard and SICIATRI, in cross-tab analysis two cells have numbers under 5. A total agreement of competence classification between SICIATRI and reference standard was 81.3%. Sensitivity of the SICIATRI was 0.83 and specificity 0.67.

**Discussion**

Our review demonstrated that there are few reported studies on competence assessment tools that fulfill the criteria of solid diagnostic test accuracy studies. Considering that competency issues are relatively frequent in clinical practice with mentally comprised patients, it is surprising that we found only three studies with analyzable data of sufficient quality.

**Feasibility**

Feasibility was not a problem for most instruments under study. Feasibility of MacCAT-CR, MacCAT-T, ACCT, OACCR, and CIS is established. Doubts rise on feasibility of CQ-Med because of concerns regarding the coverage of all four domains relevant for competent decision-making, therefore CQ-Med will not be considered for recommendation until these doubts can be dissolved.
Reliability

Overall, the included studies lacked in reporting on internal consistency of the instruments, and only 3 studies showed that the instrument met the standard. Good interobserver agreement was demonstrated for MacCAT-CR and MacCAT-T only.

Internal consistency

The studies on MacCAT-T, CQ-Med, and SICIATRI did not report on internal consistency of the scale. MacCAT-CR,(34) OACCR,(84) and ACCT(69) showed good internal consistency. For CIS, inter-item correlation coefficients ranged from low to good.(71)

Interobserver agreement

The studies on MacCAT-CR by Kim and colleagues reported good interobserver agreement in a sample of sufficient participants.(33;34) Interobserver agreement of MacCAT-T was described by Lui and colleagues to be good.(91) For the CIS, good interobserver agreement was reported for a small sample, consisting of 11 participants, which can be considered a preliminary indication.(71) SICIATRI showed a wide range in interobserver reliability between the items, which renders the interobserver reliability for the total scale with some weaknesses.(85) For ACCT good interobserver agreement was reported in a small sample of 10 patients,(69) also forming a preliminary indication of interobserver reliability. No reports on interobserver agreement were found for CQ-Med(78) and OACCR.(84)

Validity

The only two studies that met the quality criteria for feasibility, index test and reference standard, study the MacCAT-CR. The studies by Kim validated the MacCAT-CR scale against an expert judgment standard.(33;34) Depending on the prevalence of incompetence in the studied population, these studies allow for estimation of optimal cutoffs of the subscale scores. Both studies report on an optimal cutoff for subscale understanding of 18 points. The studies explain how positive predictive value (PPV, the probability that a person performing below the cutoff score is indeed incompetent) and negative predictive value (NPV, the probability that a person performing above the cutoff is in fact competent) depend on the prevalence of (in) competence in the given population (e.g. use of a high cutoff score when prevalence of incompetence is low, would lead to excluding many persons...
unjustly as incompetent). Therefore the authors suggest that for establishing optimal cutoff scores, knowledge of prevalence of (in)competence must be incorporated. The validity of a cutoff on a MacCAT-CR subscale is demonstrated.

The other studies do not comply with the requirements. Studies on MacCAT-T do not offer data for estimating validity: either data are lacking,(81;90) or the reference standard is insufficient for analysis.(91) The ACCT demonstrated moderate correlation with the reference standard, although the reference standard was of insufficient quality.(69) The OACCR showed good construct validity, however the reference standard demonstrated similarities with the index test and therefore the high correlation coefficient may present an overestimation. For the CIS only associations between reference standard and index test were demonstrated, but no further relevant measures of validity, which does not allow for establishing validity.(71) The study on SICIATRI reported high sensitivity and adequate specificity, however cross-tab analysis shows insufficient distribution amongst the cells.(85)

Summary of Main Results

In this review, 11 studies on instruments for assessing competence to consent to treatment or clinical research were included, all concerning the adult or older population, together reporting on 7 different instruments. Overall, quality of the studies under review raised many concerns, mostly regarding insufficient quality of the reference standard. Initially, three studies were considered of sufficient quality. Closer review of the studies revealed that CQ-Med produced doubts on feasibility, leaving two studies of sufficient quality, both on MacCAT-CR.

The results of the review indicate that with current knowledge, the MacCAT-CR is the most supported instrument for assessing competence to consent to clinical research, with proved reliability and validity in populations of mentally comprised adults and older persons. The OACCR, also applicable in a research setting, lacks this underpinning. For a treatment context there is no competence assessment instrument supported by reliability and validity affirmation. The MacCAT-T showed good interobserver agreement, the ACCT showed preliminary indications of good interobserver agreement and good internal consistency, but only moderate validity. Results of research on the CIS and the SICIATRI renders little support for these instruments until now. Taking into account the similarities between the MacCAT-T and the MacCAT-CR, use of the MacCAT-T may be supported by the positive accuracy outcomes of the MacCAT-CR.
Strengths and weaknesses of the review

Despite an extensive and thorough search, we retrieved only 11 studies with varying sample sizes, of which 7 were assessed with considerable risk of bias due to sub-optimal study design. In the end only 2 studies assessed a competence assessment tool on reliability and validity. No study assessed competence to consent in children.

Applicability of findings to the review question

The accuracy of instruments for assessing patients’ competence to consent could be addressed regarding a clinical research context, in populations of mentally compromised adults and older people. For a treatment context, results are less outspoken, however a recommendation for an instrument that will frequently be applicable is possible. In settings with lower prevalence of the target condition, validity of cutoffs must be estimated beforehand. The applicability to other specific participant groups, e.g. children, is even more uncertain.

Authors’ Conclusions

Implications for practice

The accuracy of competence assessment instruments partly depends on the prevalence of (in)competence in the study population. Our findings suggest that the MacCAT-CR has good diagnostic accuracy and could be used for assessment of (in)competence in populations of mentally compromised adults or older people in a clinical research context. In a treatment context, accuracy of competence assessment tools have not been confirmed yet, however the MacCAT-T is most promising and ACCT shows preliminary indications for reliability but moderate validity. The validity and generalizability of our findings are limited because they are based on few, heterogeneous studies with methodological flaws, with important concerns on the reference standards used.

Implications for research

Accuracy of assessment studies on patients’ competence to consent may be difficult, because a systematic study design requires establishing a firm
reference standard. Hence, this review highlights the need to conduct more accuracy studies on competence assessment tools. Especially regarding specific populations, including children, and regarding a treatment setting, more research on assessment tools is needed.

Appendix 1   Search Strategy

Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE(R) 1946 to Present
Search Date: 9 July 2012:

1. informed consent.ab,sh,ti.
2. (decision* adj3 capacit*).ab,ti.
3. (capacit* adj3 consent).ab,ti.
4. disclosed information.ab,ti.
5. (information adj3 reasoning).ab,ti.
6. (express* adj3 choice).ab,ti.
7. consent form?.ab,ti.
8. mental capacity.ab,sh,ti.
9. mental competen*.ab,ti.
10. mental competency/
11. mental incompeten*.ab,ti.
12. patient participation.ab,sh,ti.
13. or/1-12 [competence to consent]
14. decision making/ or decision making.ab,ti.
15. consensus/ or consensus.ab,ti.
16. or/13-15 [competence to consent sensitive]
17. (macarthur adj3 tool).ab,ti.
18. mccat.ab,ti.
19. consent questionnaire.ab,ti.
20. deaconess informed.ab,ti.
21. two part consent.ab,ti.
22. (california adj3 appreciation).ab,ti.
23. vignette method?.ab,ti.
24. informed consent survey.ab,ti.
25. competency interview schedule.ab,ti.
26. assessment of consent capacity for treatment.ab,ti.
27. Hopemont capacity assessment interview.ab,ti.
28. aid to capacity evaluation.ab,ti.
29. direct assessment of decision making capacity.ab,ti.
30. cq peds.ab,ti.
31. competency questionnaire.ab,ti.
32. SICIATRI.ab,ti.
33. structured interview for competenc*.ab,ti.
34. (hopkins adj2 assessment).ab,ti.
35. brief informed consent.ab,ti.
36. or/17-35 [all relevant psychological tests]
37. (psychiatric adj3 scale?).ab,ti.
38. (psychiatric adj3 test?).ab,ti.
39. (psychologic* adj3 scale?).ab,ti.
40. (psychologic* adj3 test?).ab,ti.
41. (neuropsycholog* adj3 test?).ab,ti.
42. (neuropsycholog* adj3 scale?).ab,ti.
43. (neuropsychiatric adj3 test?).ab,ti.
44. (neuropsychiatric adj3 scale?).ab,ti.
45. psychological tests/ or exp aptitude tests/ or language tests/ or exp
   neuropsychological tests/
   or exp personality tests/
46. exp Psychiatric Status Rating Scales/
47. or/37-46 [psychological tests general]
48. 36 or 47 [psychological tests sensitive]
49. 16 and 48

OVIDSP PsycINFO, 1806 to Present
Search Date: 11 July 2012

1. (macarthur adj3 tool).ab,id,ti,tm.
2. maccat.ab,ida,ti,tm.
3. consent questionnaire.ab,ida,ti,tm.
4. deaconess informed.ab,ida,ti,tm.
5. two part consent.ab,ida,ti,tm.
6. (california adj3 appreciation).ab,ida,ti,tm.
7. vignette method?.ab,ida,ti,tm.
8. informed consent survey.ab,ida,ti,tm.
9. competency interview schedule.ab,ida,ti,tm.
10. assessment of consent capacity for treatment.ab,ida,ti,tm.
11. Hopemont capacity assessment interview.ab,ida,ti,tm.
12. aid to capacity evaluation.ab,ida,ti,tm.
13. cq peds.ab,ida,ti,tm.
14. competency questionnaire.ab,id,ti,tm.
15. SICIATRI.ab,id,ti,tm.
16. structured interview for competenc*.ab,id,ti,tm.
17. (hopkins adj2 assessment).ab,id,ti,tm.
18. brief informed consent.ab,id,ti,tm.
19. (competency adj3 interview).ab,id,ti,tm.
20. (Hopemont adj4 interview).ab,di,ti,tm.
21. or/1-20 [specific psych. tests]
22. ("2220" or "2222" or "2223" or "2224" or "2225" or "2226").cc.
23. exp testing/
24. 22 or 23 [psych. tests / testing]
25. decision making.ab,idi,sh,ti.
26. informed consent.ab,idi,sh,ti.
27. voluntary consent.ab,idi,ti.
28. (decision* adj1 capacit*).ab,idi,ti.
29. or/25-28 [informed consent specific]
30. 24 and 29
31. ("3400" or "3410" or "3430" or "3450" or "3470").cc.
32. 30 not 31
33. 21 or 32
34. limit 33 to ("0100 journal" or "0110 peer-reviewed journal" or "0400 dissertation abstract")

OVIDSP Embase, 1947 to Present
Search Date: 9 July 2012

1. informed consent.ab,sh,ti.
2. (decision* adj3 capacit*).ab,ti.
3. (capacit* adj3 consent).ab,ti.
4. disclosed information.ab,ti.
5. (information adj3 reasoning).ab,ti.
6. (express* adj3 choice).ab,ti.
7. consent form?.ab,ti.
8. mental capacity.ab,sh,ti.
9. mental competen*.ab,ti.
10. mental incompeten*.ab,ti.
11. patient participation.ab,sh,ti.
12. or/1-11 [competence to consent]
13. decision making/ or decision making.ab,ti.
14. consensus/ or consensus.ab,ti.
15. or/12-14 [competence to consent sensitive]
16. (macarthur adj3 tool).ab,ti.
17. maccat.ab,ti.
18. consent questionnaire.ab,ti.
19. deaconess informed.ab,ti.
20. two part consent.ab,ti.
22. vignette method?.ab,ti.
23. informed consent survey.ab,ti.
24. competency interview schedule.ab,ti.
25. assessment of consent capacity for treatment.ab,ti.
27. aid to capacity evaluation.ab,ti.
28. direct assessment of decision making capacity.ab,ti.
29. cq peds.ab,ti.
30. competency questionnaire.ab,ti.
31. SICIATRI.ab,ti.
32. structured interview for competenc*.ab,ti.
33. (hopkins adj2 assessment).ab,ti.
34. or/16-33 [all relevant psychological tests]
35. psychometric?.ab,ti.
36. psychometry.ab,sh,ti.
37. (psychiatric adj3 scale?).ab,ti.
38. (psychiatric adj3 test?).ab,ti.
39. (psychologic* adj3 scale?).ab,ti.
40. (psychologic* adj3 test?).ab,ti.
41. (neuropsycholog* adj3 test?).ab,ti.
42. (neuropsycholog* adj3 scale?).ab,ti.
43. (neuropsychiatric adj3 test?).ab,ti.
44. (neuropsychiatric adj3 scale?).ab,ti.
45. neuropsychological test/
46. psychological rating scale/
47. or/35-46 [psych. tests sensitive]
48. 15 and 47
49. or/37-46
50. 15 and 49
51. 50 not 34
52. 34 or 50
Appendix 2  Data Extraction Form

Study ID  First author, year of publication
Clinical features and settings  Presenting conditions, clinical setting
Participants  Sample size, age, sex, ethnicity, country.
Study design  Were patients enrolled retrospectively or prospectively?
             Was the sampling method consecutive or random?
             Duration between reference test and index test.
Target condition  Competence to consent to treatment or research participation.
Reference standard  The reference standard test(s) used
Index tests  The index test used. Details of the test content, operators, including any special training provided. Cut-off point used.
Results  Number of true positives, true negatives, false positives, false negatives
Notes  Source of funding.

Appendix 3  Quality Assessment Form

1.  Patient selection
   A.  Risk of bias
   Was a consecutive or random sample of patients enrolled?  Yes/No/Unclear
   Was a case control-design avoided?  Yes/No/Unclear
   Did the study avoid inappropriate exclusions?  Yes/No/Unclear
   Could the selection of patients have introduced bias?
   Risk: high/low/unclear
   B.  Concerns regarding applicability
   Is there a concern that the included patients (prior testing, presentation, intended use of index test and setting) do not match the review question?
   Concern: high/low/unclear

2.  Index test(s)
   A.  Risk of bias
   Were the index test results interpreted without knowledge of the results of the reference standard?  Yes/No/Unclear
   If a threshold was used, was it pre-specified?  Yes/No/Unclear
Could the conduct or interpretation of the index test have introduced bias?  
Risk: high/low/unclear

B. Concerns regarding applicability
Is there concern that the index test, its conduct, or interpretation differ from the review question?  
Concern: high/low/unclear

3. Reference standard
A. Risk of bias
Is the reference standard likely to correctly classify the target condition?  
Yes/No/Unclear
Were the reference standard results interpreted without knowledge of the results of the index test?  
Yes/No/Unclear
Could the reference standard, its conduct, or its interpretation have introduced bias?  
Risk: high/low/unclear

B. Concerns regarding applicability
Is there concern that the target condition as defined by the reference standard does not match the review question?  
Concern: high/low/unclear

4. Flow and timing
A. Risk of bias
Was there an appropriate interval between index test(s) and reference standard?  
Yes/No/Unclear
Did all patients receive a reference standard?  
Yes/No/Unclear
Did patients receive the same reference standard?  
Yes/No/Unclear
Were all patients included in the analysis?  
Yes/No/Unclear
Could the patient flow have introduced bias?  
Risk: high/low/unclear