Holger Schünemann, MD, MSc, PhD
Professor and Chair, Dept. of Clinical Epidemiology & Biostatistics
Professor of Medicine
Michael Gent Chair in Healthcare Research
McMaster University, Hamilton, Canada

IQWiG Herbsttagung, 25. Nov. 2011

DIAGNOSTIK UND LINKED EVIDENCE – WIE ROBUST MUSS DIE KETTE SEIN?
Disclosure

• Co-chair GRADE Working Group
• Leitlinienprojekte - GRADE
  – American College of Physicians (ACP) Clinical Practice Guidelines Committee
  – American College of Chest Physicians (ACCP)
• Weltgesundheitsorganisation (WHO): Advisory Committee for Health Research, Leitlinien, Drittmittel für systematische Übersichtsarbeiten
• Keine direktes Einkommen von profitorientierten Unternehmen/Organisationen
• Dank an Kollegen (Drs. Jan Brozek & Reem Mustafa)
 Übersicht

Diagnostische Fragestellungen

### Pulmonary rehabilitation compared to usual community care for COPD with recent exacerbation

#### Bibliography:

#### Outcomes

<table>
<thead>
<tr>
<th>Outcome</th>
<th>No of Participants</th>
<th>Follow up</th>
<th>Quality of the evidence (GRADE)</th>
<th>Relative effect (95% CI)</th>
<th>Anticipated absolute effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospital admission</td>
<td>250 (6 studies)</td>
<td>3-18 months</td>
<td>⊕⊕⊕⊕ HIGH</td>
<td>OR 0.22 (0.08 to 0.58)</td>
<td>405 per 1000, 275 fewer per 1000 (from 122 fewer to 353 fewer)</td>
</tr>
<tr>
<td>Mortality</td>
<td>110 (3 studies)</td>
<td>3-48 months</td>
<td>⊕⊕⊕⊝ MODERATE</td>
<td>2 due to imprecision</td>
<td>OR 0.28 (0.1 to 0.84)</td>
</tr>
<tr>
<td>Quality of life (CRQ) dyspnea</td>
<td>258 (5 studies)</td>
<td>12 and 76 weeks</td>
<td>⊕⊕⊕⊝ MODERATE</td>
<td>4 due to imprecision</td>
<td>The mean quality of life (crq) dyspnea in the intervention groups was 0.97 higher (0.35 to 1.58 higher)</td>
</tr>
<tr>
<td>Quality of life (SGRQ) total</td>
<td>127 (3 studies)</td>
<td>12 and 26 weeks</td>
<td>⊕⊕⊕⊝ MODERATE</td>
<td>4 due to imprecision</td>
<td>The mean quality of life (sgrq) total in the intervention groups was 9.88 lower (5.37 to 14.4 lower)</td>
</tr>
<tr>
<td>Ambulation (as measured by 6 min walking distance)</td>
<td>299 (6 studies)</td>
<td>1-208 weeks</td>
<td>⊕⊕⊕⊝ MODERATE</td>
<td>4,8 due to imprecision</td>
<td>The mean ambulation (as measured by 6 min walking distance) in the intervention groups was 77.7 higher (12.21 to 143.2 higher)</td>
</tr>
</tbody>
</table>

#### Resource use - not reported

See footnote

---

Einführung

Evidenz & Beurteilungen

Empfehlungen & Implementierung
Übersicht

Diagnostische Fragestellungen

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>No of Participants</th>
<th>Follow up</th>
<th>Quality of the evidence</th>
<th>Relative effect (95% CI)</th>
<th>Anticipated absolute effects</th>
<th>Risk with Usual community care</th>
<th>Risk difference with</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospital admission</td>
<td>250</td>
<td>3-18 months</td>
<td>⊕⊕⊕⊕ HIGH</td>
<td>OR 0.22 (0.08 to 0.58)</td>
<td>405 per 1000</td>
<td>275 fewer per 1000</td>
<td>(from 122 fewer to 353 fewer)</td>
</tr>
<tr>
<td>Mortality</td>
<td>110</td>
<td>3-48 months</td>
<td>⊕⊕⊕⊝ MODERATE</td>
<td>OR 0.28 (0.1 to 0.84)</td>
<td>Low 100 per 1000</td>
<td>70 fewer per 1000</td>
<td>(from 15 fewer to 89 fewer)</td>
</tr>
<tr>
<td>High</td>
<td>1</td>
<td>500 per 1000</td>
<td>281 fewer per 1000</td>
<td>(from 43 fewer to 409 fewer)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality of life (CRQ)</td>
<td></td>
<td></td>
<td>⊕⊕⊕⊝ MODERATE</td>
<td>4 due to imprecision</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dyspnea</td>
<td>258</td>
<td>12 and 76 weeks</td>
<td>3.1</td>
<td>The mean quality of life (crq) dyspnea in the intervention groups was 0.97 higher (0.35 to 1.58 higher)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality of life (SGRQ)</td>
<td></td>
<td></td>
<td>50</td>
<td>The mean quality of life (sgrq) total in the intervention groups was 9.88 lower (5.37 to 14.4 lower)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>127</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ambulation</td>
<td>299</td>
<td>1-208 weeks</td>
<td>⊕⊕⊕⊝ MODERATE</td>
<td>77.7 higher (12.21 to 143.2 higher)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Einführung

Evidenz & Beurteilungen

Empfehlungen & Implementierung
Anwendung eines Tests

• Kliniker benutzen eine Reihe von Tests (oder Teststrategien), die “diagnostisch” genannt werden:
  – Symptome und Zeichen, bildgebende Verfahren, Laborparameter, pathologische und psychologische Befunde

• Wenige Tests sind wirklich diagnostisch (positiv oder negativ - Schwangerschaftstest)
  – Im allgemeinen verbunden mit Wahrscheinlichkeiten

• Für diesen Vortrag: vereinfachtes Modell (positiv und negativ)
HUNTINGTONs
CHOREA
Morbus Huntington
18 years

33 years
Sensitivität = 98.8%
Spezifität = 100%
“pre-test” Wahrscheinlichkeit in Kindern = 50%

<table>
<thead>
<tr>
<th>Erkrankung</th>
<th>vorliegend</th>
<th>nicht vorliegend</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA Test +</td>
<td>494</td>
<td>0</td>
</tr>
<tr>
<td>DNA Test -</td>
<td>6</td>
<td>500</td>
</tr>
</tbody>
</table>

500 500
Würden Sie diesen genetischen Test für Kinder von betroffenen Patienten empfehlen?
Keine Prävention
Keine effektive Behandlung
Würden Sie diesen genetischen Test für Kinder von betroffenen Patienten empfehlen?
Sensitivität
Spezifität
Lebensverlängerung
Weniger Symptome
Komplikationen
Lebensqualität

Sensitivität
Spezifität
Test accuracy ist ein Surrogatparameter für patientenrelevante Endpunkte

- Kliniker konzentrieren sich typischerweise auf ‘test accuracy’/Testgüte

- Annahme: Diagnose führt zu besserer Behandlung oder endpunktübergreifendem Zusatznutzen
Patientennutzen sollte vorliegen

• Die Annahme auf der Basis von ‘accuracy’ Daten, dass ein Test patientenrelevante Endpunkte verbessert, erfordert das Vorhandensein von effektiven Behandlungsstrategien = linked evidence

• Inklusive:
  – Verringerung von testgebundenen Nebenwirkungen
  – Ausschluss von Erkrankungen oder Verminderung von Angst
  – Bestätigung einer Diagnose verbessert Lebensqualität durch die prognostische Information, die vermittelt wird
Studiendesigns in der Diagnoseerstellung

• Wenn ein Test patientenrelevante Endpunkte nicht verbessert, gibt es keinen Grund für seine Anwendung (unabhängig von seiner ‘accuracy’)

• Vernünftigste Verfahren, um ein Testverfahren zu evaluieren: randomisierte, kontrollierte Studien die Tests (mit Behandlung) gegeneinander vergleichen
Studiendesign I

Endpunkte:
Mortalität
Morbidität
Nebenwirkungen
QoL
GRADE für Interventionen und Behandlungen: Qualitätsbeurteilung Nutzen/Schaden/Werte/Ressourcen
Empfehlung

Example
Randomised control trials (RCTs) explored a diagnostic strategy guided by the use of B type natriuretic peptide (BNP)—designed to aid diagnosis of heart failure—compared with no use of BNP in patients presenting to the emergency department with acute dyspnoea. As it turned out, the group randomised to receive BNP spent a shorter time in the hospital at lower cost, with no increased mortality or morbidity.
Studiendesign II

Accuracy study

Target population

New test or strategy:
Triage
Replacement
Add-on

Reference test

New test positive
True and false positives

New test negative
True and false negatives

Judgments about outcomes with new test

Judgments about outcomes with reference test

Two-step inference
Figure 2-2. Example of an analytical framework within an overarching conceptual framework in the evaluation of breast biopsy techniques*

Women referred for biopsy after detection of a breast abnormality

Development and testing of biopsy method feasibility
- Open: Surgical procedure
- Marking techniques
- CNB:
  - Needles
  - Automation
  - Imaging quality
  - Vacuum
  - Specimen collection

Core-needle or open biopsy?

Diagnosis classification of breast abnormality

Change in clinical decisions

Results of additional testing
- Clear surgical margins
- Response to treatment
- Cosmetic results

Survival
- Recurrence
- Quality of life
- Total number of surgical procedures required

Cost per QALY from societal perspective

The numbers in the figure depict where the three key questions are located within the flow of the analytical framework.
Sensitivität & Spezifität

Patientenrelevante Konsequenzen

TP (behandelt…)
TN (vergewissert…)
FP (unnötigerweise behandelt…)
FN (nicht behandelt…)

Unklare Resultate
Komplikationen durch Test
Ressourcenverbrauch
Lebensverlängerung
Weniger Symptome
Komplikationen?
Lebensqualität?

Sensitivität
Spezifität
Lebensverlängerung
 Weniger Symptome
 Komplikationen?
 Lebensqualität?

Sensitivität
Spezifität

surrogat
Übersicht

Diagnostische Fragestellungen

Bibliography:

Outcomes

<table>
<thead>
<tr>
<th>No of Participants</th>
<th>Follow up</th>
<th>Quality of the evidence (GRADE)</th>
<th>Relative effect (95% CI)</th>
<th>Anticipated absolute effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospital admission</td>
<td>6 studies</td>
<td>3-18 months</td>
<td>⊕⊕⊕⊕ (HIGH)</td>
<td>OR 0.22 (0.08 to 0.58)</td>
</tr>
<tr>
<td>Mortality</td>
<td>3 studies</td>
<td>3-48 months</td>
<td>⊕⊕⊕⊝ (MODERATE)</td>
<td></td>
</tr>
<tr>
<td>Quality of life (CRQ) dyspnea</td>
<td>5 studies</td>
<td>12 and 76 weeks</td>
<td>⊕⊕⊕⊝ (MODERATE)</td>
<td></td>
</tr>
<tr>
<td>Ambulation (as measured by 6 min walking distance)</td>
<td>6 studies</td>
<td>1-208 weeks</td>
<td>⊕⊕⊕⊝ (MODERATE) 4,8 due to imprecision</td>
<td></td>
</tr>
</tbody>
</table>

Einführung

Evidenz & Beurteilungen

Empfehlungen & Implementierung
‘Linked’ Evidenz

Systematische Übersichtsarbeiten
GRADE für ‘diagnostic accuracy’:
8 Qualitätsdomänen
Vertrauen in die Effektschätzer

Vertrauen in die Konsequenzen
‘Linked’ Evidenz

Hohe/gute Qualität

Directness: Surrogat – patientenrelevante Endpunkte?
Herunterstufen der Qualität?
Systematische Übersichtsarbeiten:
Therapie, ‘natural history’
## Domains, sub-domains and items in a conceptual framework for decision modeling in diagnostic test studies

<table>
<thead>
<tr>
<th>Domain</th>
<th>Sub-domain</th>
<th>Items *</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purpose</td>
<td>Triage</td>
<td>Screening</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Diagnosis</td>
</tr>
<tr>
<td></td>
<td>Replacement</td>
<td>Staging of disease</td>
</tr>
<tr>
<td></td>
<td>Add-on</td>
<td>Monitoring of treatment</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Monitoring of disease</td>
</tr>
<tr>
<td>Population</td>
<td></td>
<td>Pretest probability of a condition</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Any subgroups with different baseline risk or</td>
</tr>
<tr>
<td></td>
<td></td>
<td>prevalence (comorbidities, patients’</td>
</tr>
<tr>
<td></td>
<td></td>
<td>characteristics .. etc)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Stage of the disease</td>
</tr>
<tr>
<td>Intervention (test of interest,</td>
<td></td>
<td>Test’s accuracy characteristics</td>
</tr>
<tr>
<td>aka index test)</td>
<td></td>
<td>Test’s side effects</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Test benefits</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cut-off points</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Resources required</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inconclusive results</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Values and preferences</td>
</tr>
<tr>
<td>Comparison (reference test or</td>
<td></td>
<td>Test’s accuracy characteristics</td>
</tr>
<tr>
<td>alternative test)</td>
<td></td>
<td>Test’s side effects</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Test benefits</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cut-off points</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Resources required</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inconclusive results</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Values and preferences</td>
</tr>
<tr>
<td>Diagnostic test accuracy</td>
<td>Test +ve, sensitivity</td>
<td>TP &amp; FP</td>
</tr>
<tr>
<td>outcomes</td>
<td>Test –ve, specificity</td>
<td>TN &amp; FN</td>
</tr>
<tr>
<td>Patient outcomes</td>
<td>Treatment 1</td>
<td>Efficacy of available treatment</td>
</tr>
<tr>
<td></td>
<td>Treatment 2</td>
<td>Rate of side effects of available treatment</td>
</tr>
<tr>
<td></td>
<td>Treatment 3</td>
<td>Resource use with available treatment</td>
</tr>
<tr>
<td></td>
<td>No treatment</td>
<td>Values and preferences</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Prognosis/natural course of condition</td>
</tr>
<tr>
<td>Quality of evidence</td>
<td>Criteria for downgrading</td>
<td>Risk of bias</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inconsistency</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Imprecision</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Indirectness</td>
</tr>
</tbody>
</table>

*Items listed are not exhaustive and can be expanded based on specific study needs.*
Decision modelling

Figure 2-3. Replacement test example: full-field digital mammography versus screen-film mammography

* Figure taken from Blue Cross and Blue Shield Association Technology Evaluation Center, 2002.
World Allergy Organization

COWS MILK ALLERGY GUIDELINES

Workshop summary

Diagnosis and Rationale for Action against Cow's Milk Allergy (DRACMA): A summary report

Alessandro Fiocchi, MD, Holger J. Schünemann, MD, PhD, Jan Brozek, MD, Patrizia Restani, PhD, Kirsten Beyer, MD, Riccardo Troncone, MD, Alberto Martelli, MD, Luigi Terracciano, MD, Sami L. Bahna, MD, Fabienne Rancé, MD, Motohiro Ebisawa, MD, Ralf G. Heine, MD, FRACP, Amal Assa’ad, MD, Hugh Sampson, MD, Elvira Verduri, MD, G. R. Bouygue, MSc, Carlos Baena-Cagnani, MD, Walter Canonica, MD, and Richard F. Loekey, MD

Milan, Naples, and Genoa, Italy, Hamilton, Ontario, Canada, Berlin, Germany, Strasbourg, La, Toulouse, France, Kanazawa, Japan, Melbourne, Australia, Cincinnati, Ohio, New York, NY, Cordoba, Argentina, and Tampa, Fla

1120 FIOCCHI ET AL

http://www.implementationscience.com/content/6/1/62

Application of GRADE: Making evidence-based recommendations about diagnostic tests in clinical practice guidelines

Jonathan Hsu, Jan L Brozek, Luigi Terracciano, Julia Kreis, Enrico Compalati, Airtorn Tetelborn Stein, Alessandro Fiocchi and Holger J. Schünemann
Systematic review

Guideline development

Grade recommendations
- For or against (direction) ↓↑
- Strong or conditional/weak (strength)

By considering balance of:
- Quality of evidence
- Balance benefits/harms
- Values and preferences

Revise if necessary by considering:
- Resource use (cost)

Formulate Recommendations (↓↑ | ⊕...)
- “We recommend using...” | “Clinicians should...”
- “We suggest using...” | “Clinicians might...”
- “We suggest not using...” | “Clinicians ... not...”
- “We recommend not using...” | “Clinicians should not...”

Grade overall quality of evidence across outcomes based on lowest quality of critical outcomes
Darf sie Milch trinken?
Food challenge test

Skin prick test
Food challenge test

Skin prick test
Sollten ‘skin prick tests’ zur Diagnose von Kuhmilchallergien (KMA) angewandt werden?
Population (Wer?)
Intervention (Welcher test)
Comparison (Anstatt?)
Outcomes (Wofür?)
niedrig  mittel  hoch

0%  Initiale Wahrscheinlichkeit KMA  100%
current test
triage

current test

current test

P. Bossuyt et al. BMJ 2006
triage  current test  replacement
Sollten ‘skin prick tests (3 mm Reizreaktion)’ als triage test bei Patienten mit Verdacht auf KMA zur Diagnose von Kuhmilchallergien (KMA) benutzt werden?
Population (Wer?)

Intervention (Welcher test)

Comparison (Anstatt?)

Outcomes (Wofür?)
  – Anaphylaxis
  – Umstände
  – Benutzung von Kuhmilchersatz
  – Korrekte Diagnose verzögert
  – Ressourcen
GENERATE CLINICAL QUESTIONS

IDENTIFY OUTCOMES CRITICAL TO THE RECOMMENDATION

SYSTEMATICALLY GATHER CURRENT EVIDENCE ADDRESSING EACH OF THE QUESTIONS

ESTIMATE PRETEST PROBABILITIES (BASED ON LITERATURE REVIEW) AS WELL AS TEST AND TREATMENT THRESHOLDS

PREPARE SUMMARIES OF EVIDENCE INFORMING GUIDELINE PANEL’S DECISIONS ABOUT EACH QUESTION ASKED

• IDENTIFY CLINICAL PROBLEMS REQUIRING GUIDANCE
• GENERATE FOCUSED QUESTIONS (PICO)
• REACH CONSSENSUS AMONG PANEL MEMBERS ON THE FINAL QUESTIONS (REFINE THEM IF NECESSARY)

• IDENTIFY ALL PATIENT IMPORTANT OUTCOMES
• DEFINE THE CONSEQUENCES OF BEING CLASSIFIED IN EACH OF THE CATEGORIES (TP FP FN TN)
• EXPLICITLY RATE IMPORTANCE OF OUTCOMES

• PERFORM A SYSTEMATIC REVIEW
• USE EXISTING HIGH QUALITY UP-TO-DATE SYSTEMATIC REVIEW
• PERFORM AS SYSTEMATIC A SEARCH AS POSSIBLE AND TRANSPARENTLY SUMMARISE IDENTIFIED EVIDENCE

IF JUSTIFIED AND NECESSARY DEFINE DISTINCT SUB-POPULATIONS WITH A DIFFERENT BASELINE RISK OF THE DISEASE (PRE-TEST PROBABILITY)

FOR EACH CRITICAL OUTCOME:
• ASSESS THE QUALITY OF THE SUPPORTING EVIDENCE
• SUMMARISE THE EXPECTED EFFECTS
**TP:** the child will undergo oral food challenge which will turn out positive with risk of anaphylaxis, albeit in controlled environment; burden on time and anxiety for family; exclusion of milk and use of special formulae. Some children with high pre-test probability of disease and/or at high risk of anaphylactic shock during the challenge will not undergo challenge test and be treated with the same consequences of treatment as those who underwent food challenge.

**TN:** the child will receive cow’s milk at home with no reaction, no exclusion of milk, no burden on family time and decreased use of resources (no challenge test, no formulae); anxiety in the child and family may depend on the family; looking for other explanation of the symptoms.

**FP:** the patient will undergo an oral food challenge which will be negative; unnecessary burden on time and anxiety in a family; unnecessary time and resources spent on oral challenge. Some children with high pre-test probability of CMA would not undergo challenge test and would be unnecessarily treated with elimination diet and formula that may lead to nutritional deficits (e.g. failure to thrive, rickets, vit D or calcium deficiency); also stress for the family and unnecessary carrying epinephrine self injector which may be costly as well as delayed diagnosis of the real cause of symptoms.

**FN:** the child will be allowed home and will have an allergic reaction (possibly anaphylactic) to cow’s milk at home; high parental anxiety and reluctance to introduce future foods; may lead to multiple exclusion diet. The real cause of symptoms (i.e. CMA) will be missed leading to unnecessary investigations & treatments.

**Inconclusive results:** (either negative positive control or positive negative control): the child would repeat SPT which may be distressing for the child and parent; time spent by a nurse and a repeat clinic appointment would have resource implications; alternatively child would have sIgE measured or undergo food challenge

**Complications of a test:** SPT can cause discomfort or exacerbation of eczema which can cause distress and parental anxiety; food challenge may cause anaphylaxis and exacerbation of other symptoms.

**Resource utilization (cost):** SPT adds extra time to clinic appointment however; oral food challenge has much greater resource implications.
Konsequenzen der Fehldiagnose: KMA nicht diagnostiziert
(False Negative SPT result)

• Allergische (anaphylaktische) Reaktion auf Kuhmilch
• Angst der Eltern
• Verminderte Einführung von anderen Nahrungsstoffen
• Unnötige andere Untersuchungen und Behandlungen
Sensitivität  
0,81 (95% CI: 0,77 to 0,85)  

Specifizität  
0,72 (95% CI: 0,68 to 0,76)  

SPT 3 mm  
Initiale Wahrscheinlichkeit ~10%
SPT 3 mm
Initiale Wahrscheinlichkeit ~10%
Based on combined sensitivity of 81% (95% CI: 77 to 85) and specificity of 72% (95% CI: 68 to 76)

1, 2 Most studies enrolled highly selected patients with atopic eczema or gastrointestinal symptoms, no study reported if an index test or a reference standard were interpreted without knowledge of the results of the other test, but it is very likely that those interpreting results of one test knew the results of the other; all except for one study that reported withdrawals did not explain why patients were withdrawn.

3 Estimates of sensitivity ranged from 10% to 100%, and specificity from 14% to 100%; we could not explain it by quality of the studies, tests used or included population

4 There is uncertainty about the consequences for these patients; in some a diagnosis of other potentially serious condition may be delayed

One study in a different population (children younger than 12 months) reported 8% inconclusive challenge tests but did not report number of inconclusive skin prick tests.
<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Illustrative Risks (95% CI)</th>
<th>Number of participants (studies)</th>
<th>Quality of the Evidence</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>True positives (Patients correctly classified as having coronary artery disease)</td>
<td>192 per 1000</td>
<td>1570 (21)</td>
<td>Moderate</td>
<td>Benefit from treatment and fewer complications.* Some patients will have to undergo angiography.</td>
</tr>
<tr>
<td>True negatives (Patients correctly classified as not having coronary artery disease)</td>
<td>592 per 1000</td>
<td>1570 (21)</td>
<td>Moderate</td>
<td>Benefit from reassurance and fewer complications</td>
</tr>
<tr>
<td>False positives (Patients incorrectly classified as having coronary artery disease)</td>
<td>208 per 1000</td>
<td>1570 (21)</td>
<td>Moderate</td>
<td>Harm from unnecessary treatment</td>
</tr>
<tr>
<td>False negatives (Patients incorrectly classified as not having coronary artery disease)</td>
<td>8 per 1000</td>
<td>1570 (21)</td>
<td>Low</td>
<td>Detriment from delayed diagnosis or myocardial insult</td>
</tr>
<tr>
<td>Complications (MI, allergic reactions, renal failure)</td>
<td>99 per 1000</td>
<td>1570 (21)</td>
<td>Low</td>
<td>There is a higher rate of rare complications (infarction and death) and higher cost with angiography - a full profile would be required.</td>
</tr>
</tbody>
</table>

| Resource use* | See comment | See comment | See comment | Cost are higher for angiography, |

---

1- Quality rated from 1 (very low quality) to 4 (high quality), 2- Cross sectional studies. Indirectness of outcomes in a wide spectrum of patients and indirect comparison of tests, 3- there is greater uncertainty whether these patients will have negative outcomes.

*Assumed efficacy of: 1) aspirin daily = 20% RRR; 2) beta-blockage = 18% RRR.
<table>
<thead>
<tr>
<th>Guideline Panel</th>
<th>Problem</th>
<th>Evidence</th>
<th>Solution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guideline Panel</td>
<td>Question (PIC)</td>
<td>Systematic Review</td>
<td>Values and Preferences</td>
</tr>
<tr>
<td>Outcomes (Os)</td>
<td>Estimates of Effects</td>
<td>Quality of Evidence</td>
<td>Balance Benefits &amp; Harms</td>
</tr>
<tr>
<td></td>
<td>Evidence Table</td>
<td></td>
<td>Recommendation and Its Strength</td>
</tr>
</tbody>
</table>
Pulmonary rehabilitation compared to usual community care for COPD with recent exacerbation

**Outcomes**

<table>
<thead>
<tr>
<th>No of Participants</th>
<th>Follow up</th>
<th>Quality of the evidence</th>
<th>Relative effect (95% CI)</th>
<th>Anticipated absolute effects</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hospital admission</strong></td>
<td>250 (6 studies)</td>
<td>3 - 18 months</td>
<td>HIGH</td>
<td>OR 0.22 (0.08 to 0.58)</td>
</tr>
<tr>
<td><strong>Mortality</strong></td>
<td>110 (3 studies)</td>
<td>3 - 48 months</td>
<td>MODERATE</td>
<td>OR 0.28 (0.1 to 0.84)</td>
</tr>
<tr>
<td><strong>Quality of life (CRQ) dyspnea</strong></td>
<td>258 (5 studies)</td>
<td>12 and 76 weeks</td>
<td>MODERATE</td>
<td>The mean quality of life (crq) dyspnea in the intervention groups was 0.97 higher (0.35 to 1.58 higher)</td>
</tr>
<tr>
<td><strong>Quality of life (SGRQ) total</strong></td>
<td>127 (3 studies)</td>
<td>12 and 26 weeks</td>
<td>MODERATE</td>
<td>The mean quality of life (sgrq) total in the intervention groups was 9.88 lower (5.37 to 14.4 lower)</td>
</tr>
<tr>
<td><strong>Ambulation (as measured by 6 min walking distance)</strong></td>
<td>299 (6 studies)</td>
<td>1 - 208 weeks</td>
<td>MODERATE</td>
<td>The mean ambulation (as measured by 6 min walking distance) in the intervention groups was 77.7 higher (12.21 to 143.2 higher)</td>
</tr>
</tbody>
</table>


---

**Übersicht**

**Diagnostische Fragestellungen**

**Einführung**

**Evidenz & Beurteilungen**

**Empfehlungen & Implementierung**
Recommendation 1.4.

In patients with low pre-test probability of CMA we suggest using a skin prick test with a cut-off value of ≥3 mm as a triage test to avoid oral food challenge in those in whom the result of a skin prick test turns out negative. (weak recommendation | low quality evidence)

Underlying values and preferences

This recommendation places a relatively high value on avoiding risk of anaplylaxis, burden and resource use with an OFC test (~67% challenges avoided). It places a lower value on avoiding an allergic reaction in around 1 in 25–50 patients misclassified as not having CMA while they would actually be allergic to cow’s milk (2–4% false negative results).
Other examples of GRADE in diagnostic reviews and guidelines
Annahmen und Beurteilungen

<table>
<thead>
<tr>
<th>Example of new test and reference test or strategy</th>
<th>Putative benefit of new test</th>
<th>Diagnostic accuracy</th>
<th>Patient Outcomes and expected impact on management for the following test outcomes</th>
<th>Balance between presumed patient outcomes, test complications and cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>A shorter version of a dementia test compared with the original Mini Mental State Exam for diagnosis of dementia</td>
<td>Simpler test, less time</td>
<td>equal</td>
<td>Presumed influence on patient important outcomes</td>
<td>Evidence of shorter time and similar test accuracy (and thus patient outcomes) would generally support the new test’s usefulness</td>
</tr>
<tr>
<td></td>
<td>Equal</td>
<td>True positives</td>
<td>True negatives</td>
<td>False positives</td>
</tr>
<tr>
<td></td>
<td>Sensitivity</td>
<td>Specificity</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Uncertain benefit from earlier diagnosis and treatment</td>
<td>Almost certain benefit from reassurance</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Directness of the evidence (test results) for patient important outcomes</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Evidence uncertainty</td>
<td>No uncertainty</td>
</tr>
<tr>
<td>Example of new test and reference test or strategy</td>
<td>Putative benefit of new test</td>
<td>Diagnostic accuracy</td>
<td>Patient Outcomes and expected impact on management for the following test outcomes</td>
<td>Balance between presumed patient outcomes, test complications and cost</td>
</tr>
<tr>
<td>--------------------------------------------------</td>
<td>-----------------------------</td>
<td>---------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------</td>
</tr>
<tr>
<td>Helical CT for renal calculus compared with intravenous pyelogram</td>
<td>Detection of more (but smaller) calculi</td>
<td>greater</td>
<td>equal</td>
<td>Presumed influence on patient important outcomes</td>
</tr>
</tbody>
</table>

**Directness of the evidence (test results) for patient-important outcomes**

- Some
- No
- No
- Major
Zusammenfassung

• ‘Diagnostic accuracy’/Testgüte bedarf Evaluierung im Zusammenhang mit Konsequenzen
  – TP, FP, TN, FN, Ressourcen, Testnebenwirkungen
• Qualitätsbeurteilung muss sich auf alle Glieder in der Kette beziehen
  – Explizite Bewertung der Evidenz – Konsequenzen
    • Systematische Übersichtsarbeiten – Transparenz in den Annahmen
• Ansätze vorhanden, Pilotprojekte
PELIGRO FUERTE OLEAJE
DANGER STRONG WAVES
GEFAHR HEAVY WELLENGANG