asim. Academy of Swiss Insurance Medicine

MINI SYMPOSIUM - EUMASS - UEMASS
European Union of Medicine in Assurance and Social Security

Quality of Evidence and Grades of Recommendations in guidelines

A role for insurance medicine?

Prof. Regina Kunz

asim, Academy of Swiss Insurance Medicine
Basel

asim.

Outline

Definitions and clarifications

Deficiencies in current evidence assessment

The GRADE approach

What determines "quality of evidence"

How to move from the evidence to recommendations

Definitions and clarifications (I)

Insurance medicine

All kind of activities in social and private insurance where medicine links to insurance issues

➔ disability, accident, health care, liability, life insurance

Interfaces of Insurance Medicine

- Clinical medicine
- Rehabilitation medicine
- Occupational medicine
- Social-/ preventive medicine
- Forensic medicine

Insurances
- Accident
- Disability
- Health care
- Medical liability
- Life insurance

- Clinical epidemiology
- Population epidemiology
- Psychology / soziology
- Ethics

Social law
Liability law
Health economics
Insurance mathematics

Definitions and clarifications (II)

Guidelines vs. pathways vs. standards vs. protocols

Clinical guidelines are **systematically** developed statements **to assist** practitioner and patient **decisions** about **appropriate health care** for specific clinical circumstances.

Institute of Medicine 1990

MJ Field / KN Lohr

AWMF online

AWMF Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften

**Leitlinien der Deutschen Gesellschaft für Neurologie (DGN),
der Deutschen Gesellschaft für Orthopädie und Orthopädische Chirurgie (DGOOC),
der Deutschen Gesellschaft für Psychosomatische Medizin und Psychotherapie (DGPM),
des Deutschen Kollegium für Psychosomatische Medizin (DKPM),
der Deutschen Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde (DGPPN) und
der Deutschen Gesellschaft zum Studium des Schmerzes (DGSS)**

„Interdisciplinary Guideline
for performing a medical expertise / IME on patients with chronic pain“

Medical societies of neurology, of orthopaedics, of psychosomatic medicine and psychotherapy, of psychiatry, for the study of pain

der Deutschen Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde (DGPPN) und

GOAL of the GUIDELINE

To standardize the process and the content of doing an independent medical evaluation (IME) in pts with chronic pain

To take into account the complexity of pain …

To improve the quality of IMEs

To enable more uniform assessments

To improve the understanding of the situation between physicians and lawyers

NOT: to make decisions about appropriate management

asim.

Inhalt der Leitlinie

Grundlage der Leitlinie sind einerseits das Grundlagenwissen um Schmerzentstehung, Schmerzverarbeitung und Schmerzchronifizierung sowie die fachgebietsspezifischen Einschätzungen zu schmerzkranken Probanden. Andererseits werden Kenntnisse der Begutachtungsgrundlagen für verschiedene Rechtsbereiche zugrunde gelegt. Wesentlich war die Zusammenführung fachgebietsspezifischer Erkenntnisse zu einer interdisziplinären Leitlinie. Bestehende Publikationen und Leitlinien zum Thema wurden berücksichtigt [1,3,4,16,19,22].

Content of the guideline:

- **To compile basic knowledge** about the etiology of pain, about coping with pain …
- **To compile the basics** about an IME in various legal contexts
- **To integrate the judgements** from the various disciplines on patients with pain
- **To compile and harmonize the views** from the various disciplines

My definition for guidelines in the remaining presentation

Medical guidelines **provide advice** on alternative management strategies

➔ **Recommendations**

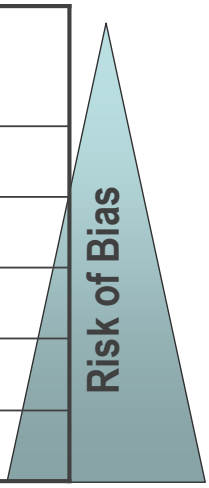
Medical guidelines are based on a systematic assessment of the literature

➔ **quality of the evidence**

Quality of the evidence

Levels of Evidence (Sackett, Chest 1986)

Level of evidence	Study Design
I	Large RCTs with unambiguous results
II	Small RCTs with uncertain results
III	Non-RCTs with concurrent controls
IV	Non-RCTs with historical controls
V	Case series without controls



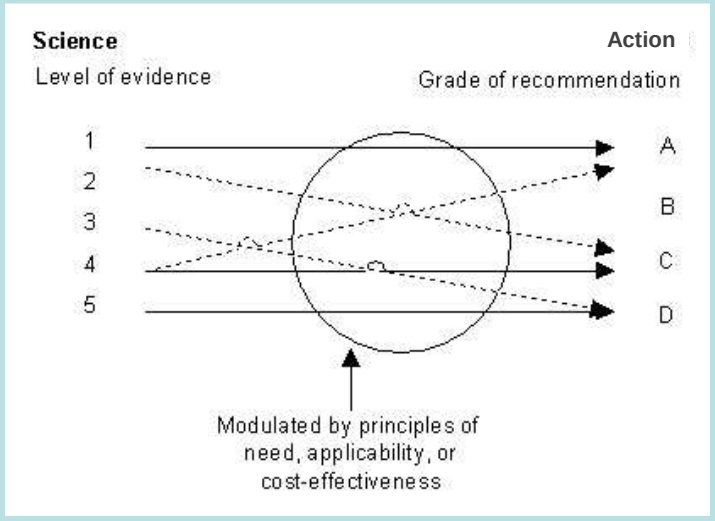
Why bother about rating the quality of evidence?

People draw conclusions about the quality of evidence

Systematic and explicit approaches can help

- protect against errors
- resolve disagreements
- facilitate critical appraisal
- communicate information

European Council on guidelines (2001)



However, there is wide variation in currently used approaches

Tower of Babel

Comparison of the categorisation
of evidence and recommendations
of 3 guidelines on fibromyalgia

Häuser et al.
Guidelines on the management of fibromyalgia syndrome.
A systematic review.
European Journal of Pain 2010

Guidelines on the Management of Fibromyalgia Syndrome

Comparison of the level of evidence

Evidence	American Pain Society	European League against Rheumatism	AWMF / Germany *(„Oxford scheme")*
Level I	Meta-analysis of multiple well-designed controlled studies	double-blind RCTs	Ia – SR of RCTs Ib – single RCT Ic – All or none
Level II	Well-designed experimental studies	blinded crossover RCT	IIa – SR of cohorts IIb –single cohort IIc – "Outcomes" Research; Ecological studies
Level III	Quasi-experimental studies *(non-RCTs, single-group pre-post, cohort, time series)*	single blind RCT	IIIa – SR of case-control studies IIIb – single case-control Study
Level IV	Non-experimental studies *(comparative / correlational / descriptive / case studies)*	open RCT; single blind non-RCT	Case-series
Level V	Case reports/ clinical examples	open non-RCT	Expert opinion, based on physiology / lab research

Comparison of the strength of recommendations

Recommen-dation	American Pain Society	European League against Rheumatism	AWMF *Germany*
Strength A	**Level I** or **consistent findings** from multiple studies of **level II – IV**	Directly: **level I**	**Level I**
Strength B	**Level II - IV** with **generally consistent** findings	Directly: **level II** or Extrapolated: **level I**	**Level II**
Strength C	**Level II - IV** with **inconsistent** findings	Directly: **level III** or extrapolated: **levels I - III**	**Levels III - V**
Strength D	**Level V** or **little/no evidence**	Directly: **level IV** or Extrapolated: **level I - III**	

The recommendations

	American Pain Society		European league		AWMF / Germany	
	LoE	Rec	LoE	Rec	LoE	Rec
Cogn. behavioral therapy	**I**	**A**	**IV**	**D**	**Ia**	**A**
Tramadol	**II**	**B**	**Ib**	**A**	**IIb**	**C**
Biofeedback	**II**	**B**	**⊗**	**⊗**	**IIb**	**Not / B**
Anticonvulsants	**II**	**B**	**Ib**	**A**	**IIb**	**B**
Acupuncture	**II**	**C**	**⊗**	**⊗**	**Ia**	**Not / A**

Who is the GRADE working group?

International collaboration of methodologists, guideline developers, policy makers with an interest in making guidelines more transparent and explicit

Leading figures: Gordon Guyatt, Andy Oxman, Holger Schünemann

Exists since 2000

Grading of Recommendations Assessment, Development and Evaluation

What is different in the GRADE approach?

Distinction between *Quality of evidence* **and** *Strength of recommendations*

Quality of Evidence

4 categories:

– High

– Moderate

– Low

– Very low

Strength of Recommendation

2 categories:

– Strong recommendations

and

– Weak recommendations

… *in favour* or *against* an intervention

Quality of the „*body* ***of evidence***“

GRADE Perspective: **„Confidence in the evidence“**

High quality: We are **very confident** that the true effect lies close to that of the estimate of the effect

Moderate quality: We are **moderately confident** in the effect estimate:

The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.

Low quality Our **confidence** in the effect estimate is **limited:**

The true effect may be substantially different from the estimate of the effect.

Very low quality We have **very little confidence** in the effect estimate**:**

The true effect is likely to be substantially different from the estimate of effect.

Quality of the „*body* ***of evidence***“

Alternative definition

High quality: Further research is **very unlikely to change** our confidence in the estimate of effect.

Moderate quality: Further research is **likely to have an important impact** on our confidence in the estimate of effect and may **change the estimate.**

Low quality: Further research is **very likely to have an important impact** on our confidence in the estimate of effect and **is likely to change the estimate.**

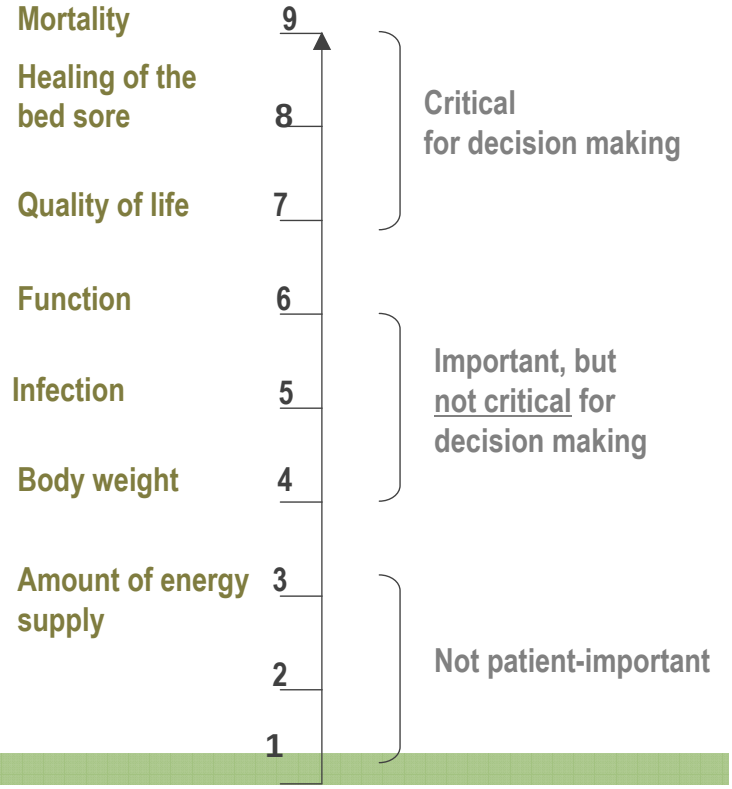
Very low quality: **Any estimate** of effect is **very uncertain**.

Hierarchy of outcomes
according to relevance for patients

Example:

Effect of enteral supplement nutrition

to improve healing of bed sores in geriatric patients



The GRADE approach to overall quality assessment is more comprehensive

Quality of evidence	Study design	Lower if …	Higher if …
High	**Randomised trial**	**Risk of bias**:	**Strong association**
Moderate		**Inconsistency**	
Low	**Observational study**		**Evidence of a dose response gradient**
Very low	Any other evidence	**(In-)Directness**:	
		Imprecise data	
		Publication bias	

Inconsistency of results (*„heterogeneity“*)

Look for explanations

- patients, intervention(s), outcome, methods

Judgment of consistency

- variation in size of effect
- overlap in confidence intervals
- statistical significance of heterogeneity I^2

(In-) Directness of Evidence

Differences in

- patients
- interventions
- comparators

Differences in outcomes

- surrogates

Indirectness

2 situations

- A) Difference: **Our own question** *vs.* **available evidence**

  own question: **„long-term effect of antidepressants in fibromyalgia“**

  available information: **studies with short-term follow-up** (6 weeks)

- B) **Differences in the patients**:

  Patients with whiplash who filed a claim *(e.g. for liability)* **vs. patients with whiplash who *did not* file a claim**

Differences in endpoints: Surrogat vs. patients- /insurance relevant endpoints

„radiologic fracture healing“ ***versus*** „shortening of time span to full weight bearing or full function“

What can raise quality?

particularly relevant for observational studies

large magnitude of the effect

common criteria

- **every pt used to do badly**
- **almost all pts. do well**
- **(quick action)**

Examples:

mechanical ventilation in respiratory failure

dialysis vs no dialysis for prolonging life

ABO incompatible renal transplantation

Summary Part I

Distinction **between** Rating Quality of Evidence vs. Grading Strength of Recommendation

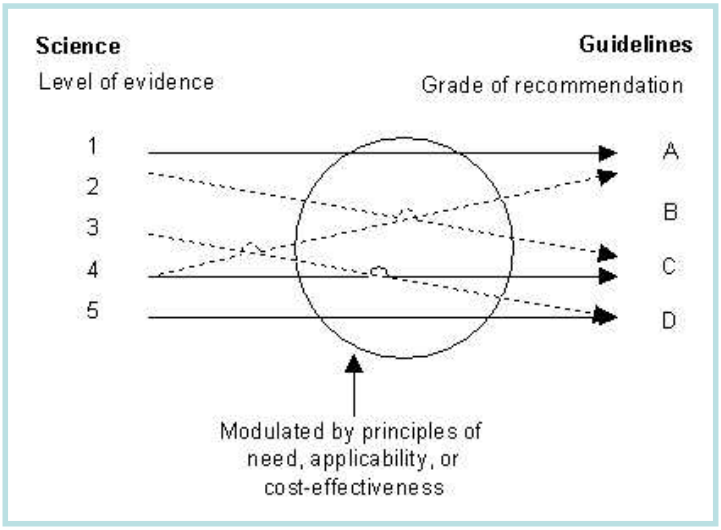
Hierarchy of **patient**-relevant / **clinically** relevant / **insurance relevant** outcomes

Explicit criteria for rating quality ↓ and ↑

Judgement **in each step** ➔ **transparency and explicitness**

Useful for systematic reviews, technology assessment, guidelines

Part II: Moving to recommendations



European Council 2001

What the GRADE system does

A guideline system should

have simple and clear messages
be transparent
include a comprehensive assessment

Only 2 grades of recommendations

and clarity what they imply

Considerations in making the judgment

Key message of a recommendation

- Do the desirable effects of an intervention clearly outweigh the undesirable effects?

OR

- Is there a close or uncertain balance?

What GRADE does ….

- **Two strengths of recommendations**
  "Strong" and **"Weak"**

What do we mean by saying „strong" or „weak" ?

Definition:

" … reflects the extent to which we can be confident that desirable effects **of an intervention** outweigh undesirable effects**"**

Recommendations in GRADE

Strong recommendations:

The GL panel **is confident**

that **desirable effects** of an intervention **clearly outweigh undesirable effects"**

Weak **recommendations:**

The GL panel is **less confident**

that **desirable effects** of an intervention **outweigh undesirable effects"**

Strong and weak recommendation Implications

	Strong	Weak
Patient	Most **patients** would want **the recommended course of action; only a small proportion would not**	Most patients would want the recommended course of action, but **many would not**
Medical Decision maker	Most patients should receive **the recommended course of action – extended discourse not necessary,** JUST DO IT **!**	Different choices **are** appropriate **for different patients.**
Policy Maker	**Recommendation** can be **adopted as a policy** in most situations.	Policy making requires substantial **debate & involvement of stake-holders**

What factors should determine the strength of a recommendation?

Factor	Comment
1) Balance between **desirable and undesirable effects**	Large difference **between desirable and undesirable effects** ➔ **high likelihood of a** strong **recommendation** Narrow gradient ➔ **high likelihood for** weak **recommendation**
2. Quality of evidence	**The** higher the quality**, the higher the likelihood of a** strong **recommendation**
3. Values & preferences	Large variation **in values & preferences, or great uncertainty in values & preferences** ➔ **the** higher **the likelihood of a** weak **recommendation**
4. Costs (resource allocation)	**The** higher the costs **of an intervention, the** lower **the likelihood of a** strong **recommendation**

1. Balance: desirable and undesirable effects

(examples)

In pts. with acute low back pain, advice to stay active

BUT: short-term discomfort

Middlekoop, Eur Spine J 2010

➔ **Strong recommendation in favour**

In pts. with acute low back pain, NSAID and opioids have a small effect compared to placebo,

BUT: they show more adverse effects

Kuijpers, Eur Spine J 2010

➔ **Weak recommendation in favour**

2. Quality of evidence

(example)

Compression stockings **in people making long plane journeys**

Benefit: **Reduction in DVT: RR = 0.10**
(95%CI: 0.04 - 0.25)

Harm: **„Inconvenience“, but stockings well tolerated**

Assessment of evidence:

RCTs with serious flaws

(Lack of concealment, of reproducibility in measuring DVT, of blinding; asymptomatic DVTs)

➔ **More challenging to make a STRONG recommendation**

(but not impossible !)

3. Values & preferences matter

Water fluoridation to prevent tooth decay in the general population

Benefit: strong effect; low quality evidence

Undesirable effects: harmless discoloration of teeth, very low quality evidence

Values & Preferences:

Enforced fluoridation of an essential element versus self-determination of the people

Take Home Messages
What's new with the GRADE-System?

Distinction **between** Rating Quality of Evidence vs. Grading Strength of Recommendation

Hierarchy of patient-relevant / clinically relevant / insurance relevant outcomes

Quality assessment according to outcome

Explicit criteria for rating quality ↓ and ↑

Judgement **in each step → transparency and explicitness**

- **Explicitness and Transparency: how to move from evidence to recommendations for the practice**
- **Integration of Values & Preferences**
- **Useful for technology assessment and guidelines**

Role of GRADE for insurance medicine

- Helps to determine knowledge gaps
- Makes judgement more transparent
- Helps to distinguish between informative and less informative diagnostic tests and effective treatments
- Helps to communicate

- Transparency increases credibility

BMJ SERIES
2008; 336: 924-926 // 995-998 // 1049-1051

RATING QUALITY OF EVIDENCE AND STRENGTH OF RECOMMENDATIONS

GRADE: an emerging consensus on rating quality of evidence and strength of recommendations

Guidelines are inconsistent in how they rate the quality of evidence and the strength of recommendations. This article explores the advantages of the GRADE system, which is increasingly being adopted by organisations worldwide

Guideline developers around the world are inconsistent in how they rate quality of evidence and grade strength of recommendations. As a result, guideline users face challenges in understanding the messages that grading systems try to communicate. Since 2006 the *BMJ* has requested in its "Instructions to Authors" on bmj.com that authors should preferably use the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system for grading evidence when submitting a clinical guidelines article. What was behind this decision?

In this first in a series of five articles we will explain why many organisations use formal systems to grade

Gordon H Guyatt professor, Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, ON, Canada L8N 3Z5
Andrew D Oxman researcher, Norwegian Knowledge Centre for the Health Services, PO Box 7004, St Olavs Plass, 0130 Oslo, Norway
Gunn E Vist researcher, Norwegian Knowledge Centre for the Health Services, PO Box 7004, St Olavs Plass, 0130 Oslo, Norway
Regina Kunz associate professor, Basel Institute of Clinical

advantages and disadvantages but also by their confidence in these estimates. The cartoon depicting the weather forecaster's uncertainty captures the difference between an assessment of the likelihood of an outcome and the confidence in that assessment (figure). The usefulness of an estimate of the magnitude of intervention effects depends on our confidence in that estimate.

Expert clinicians and organisations offering recommendations to the clinical community have often erred as a result of not taking sufficient account of the quality of evidence.[2] For a decade, organisations recommended that clinicians encourage postmenopausal women to use hormone replacement therapy.[3] Many primary care phy-

www.gradeworkinggroup.org

Organizations

www.gradeworkinggroup.org