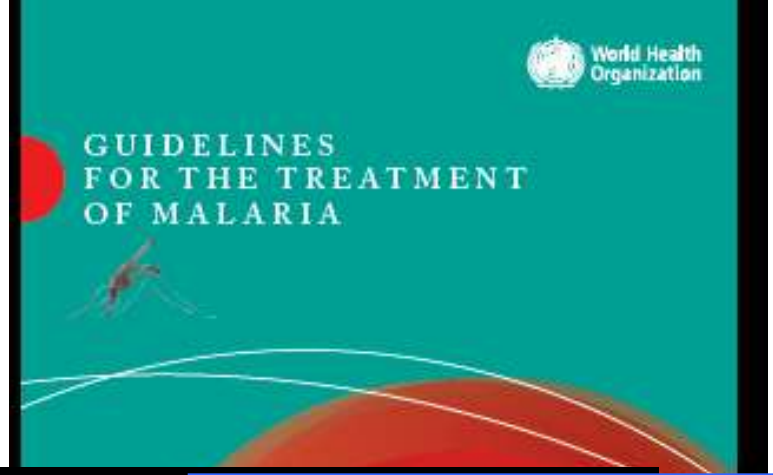
WHO Guidelines: Grading evidence and developing evidence-based guidelines

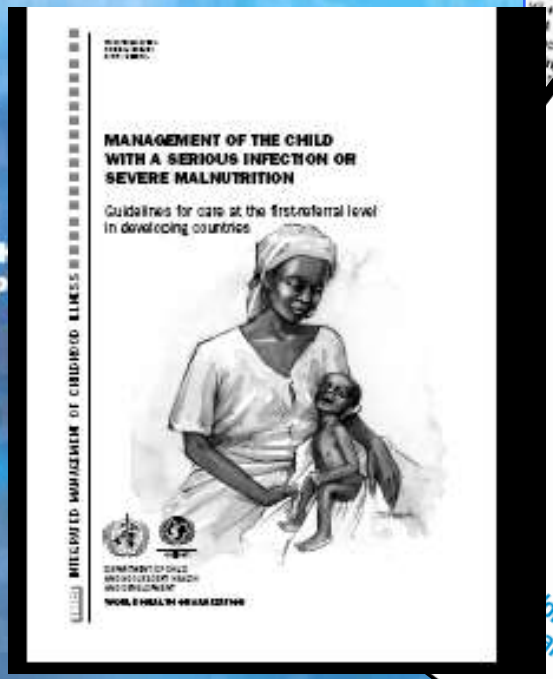
Suzanne Hill

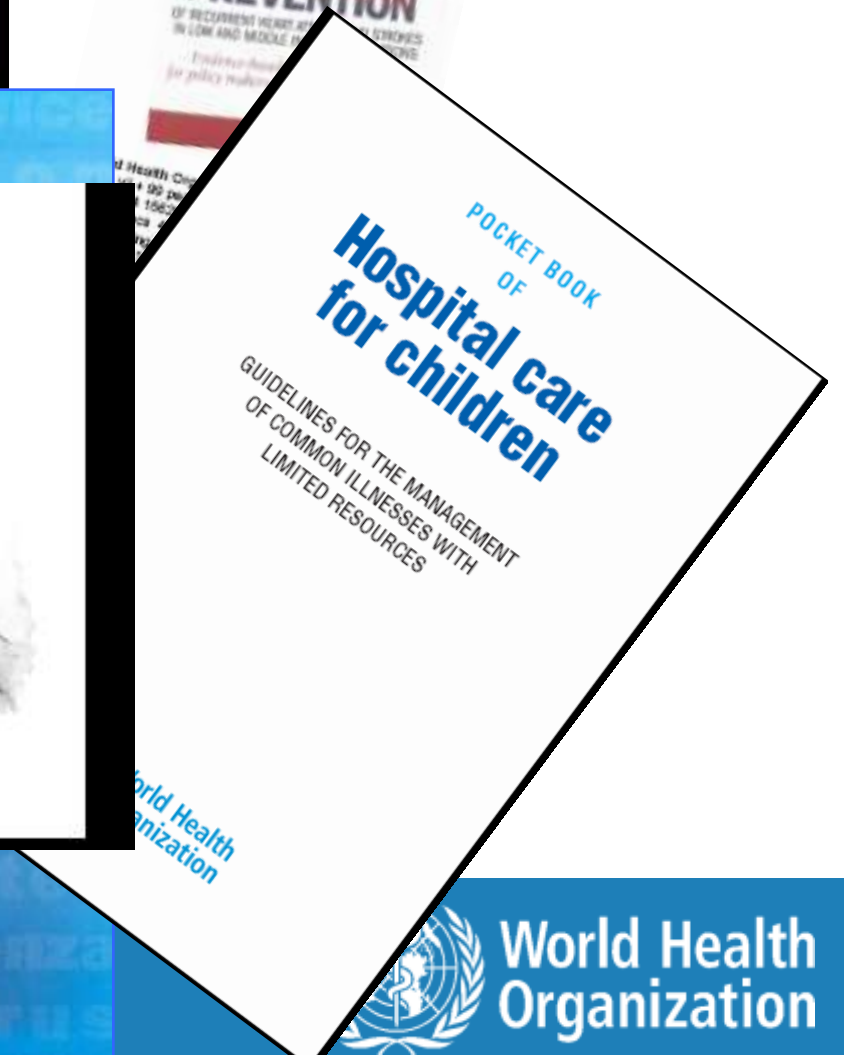
Training Course in Sexual and Reproductive Health Research

Geneva, February 2009

World Health Organization







World Health Organization

What is a guideline?

- **Medical guideline/clinical guideline, clinical practice guideline**

  "Systematically developed statements to assist practitioner and patient decisions about appropriate healthcare for specific clinical circumstances''

Field & Lohr, 1992

Other terms

- **Clinical protocol/critical pathway/integrated care pathway**

"Management recommendation based on a programmed description of the policy, containing well-defined choices regarding the policy to be followed, based on agreements between the disciplines involved."

Altena et al 1994

Why are they important?

We know what to do…

- **Prescribe beta-blockers for patients after myocardial infarction**
- **Wash hands between patients**
- **Not prescribe antibiotics for viral upper respiratory tract infections**
- **Stop smoking….**


GUIDELINES
HONEYSETT

So to improve clinical decision making and quality of care…

- **Clinical guidelines with :**
  - **Concrete aims and objective**
  - **Sufficient evidence to support most of the recommendations**
  - **Clear structure and attractive layout**
  - **Clear and specific recommendations**
  - **Taking account of norms and values**
  - **Applicable in different settings**

So what is evidence?

- *Stanford Encyclopedia of Philosophy*
- *The concept of evidence is central to both epistemology and the philosophy of science. Of course, 'evidence' is hardly a philosopher's term of art: it is not only, or even primarily, philosophers who routinely speak of evidence, but also lawyers and judges, historians and scientists, investigative journalists and reporters, as well as the members of numerous other professions and ordinary folk in the course of everyday life.*

- *And when we try to define 'evidence' … we find it very difficult.*
  —R.G. Collingwood, The Idea of History

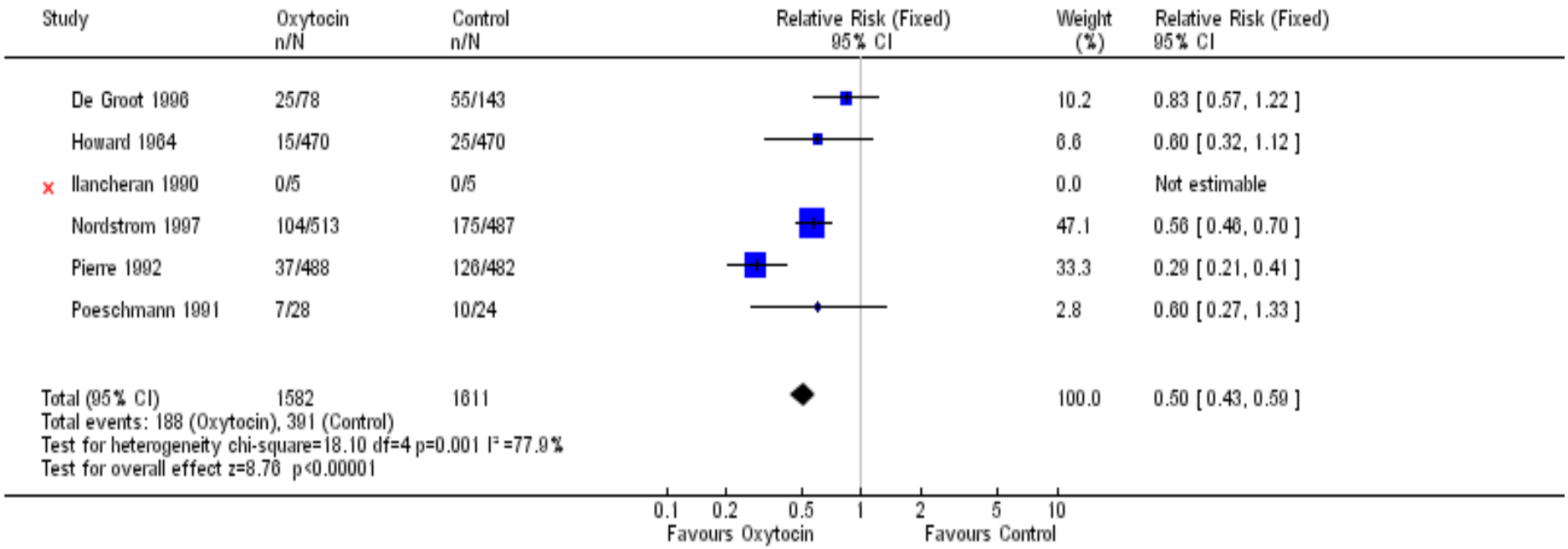
- Professional good intentions and plausible theories are insufficient for selecting policies and practices for protecting, promoting and restoring health.
- We will serve the public more responsibly and ethically when research designed to reduce the likelihood that we will be misled by bias and the play of chance has become an expected element of professional and policy making practice, not an optional add-on.

Iain Chalmers

How do we judge that we are sure that adherence to a recommendation will do more good than harm?

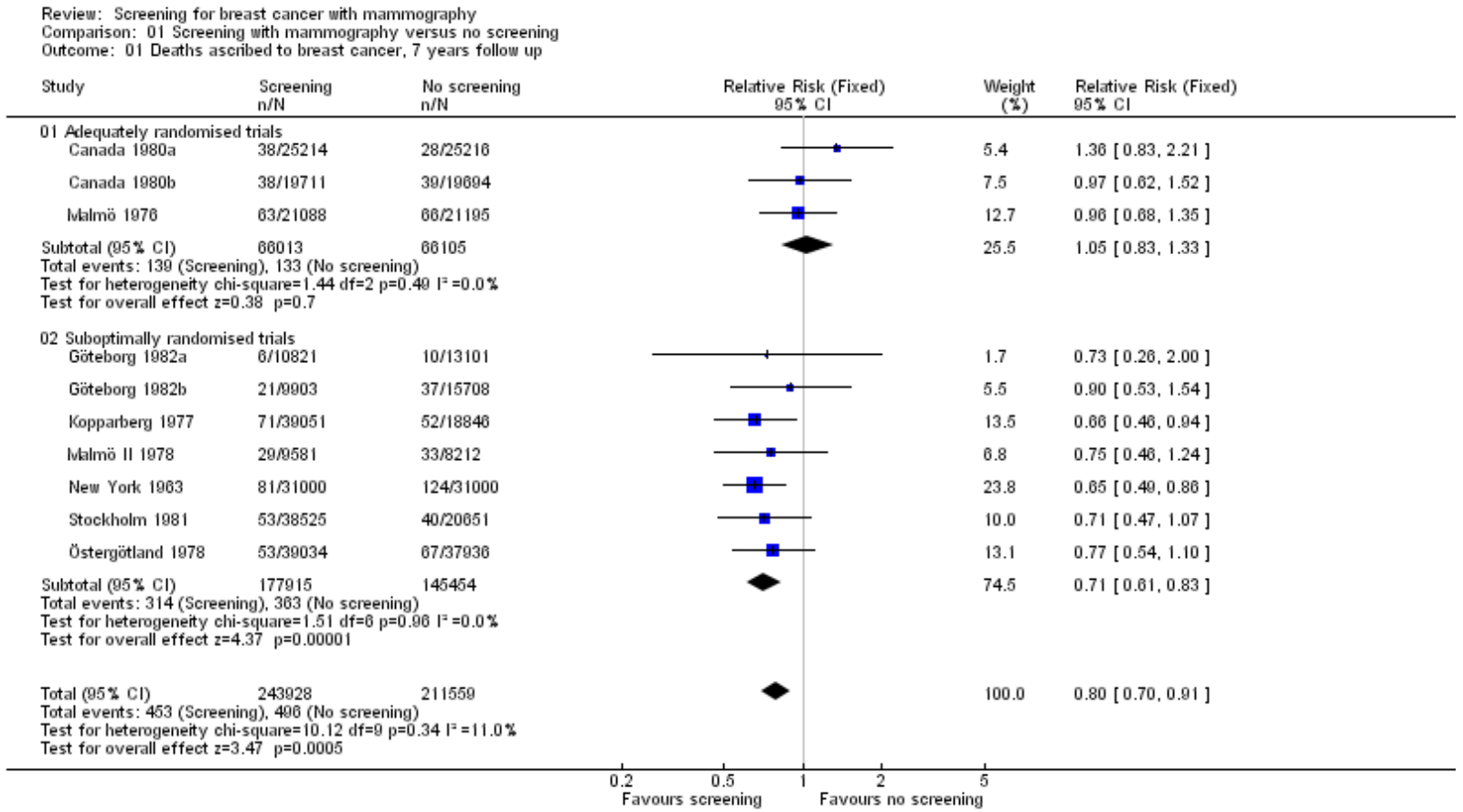
Does an intervention work?

Review: Prophylactic oxytocin for the third stage of labour
Comparison: 01 Oxytocin versus no uterotonics (all trials)
Outcome: 01 PPH (clinically estimated blood loss > or = 500 ml)



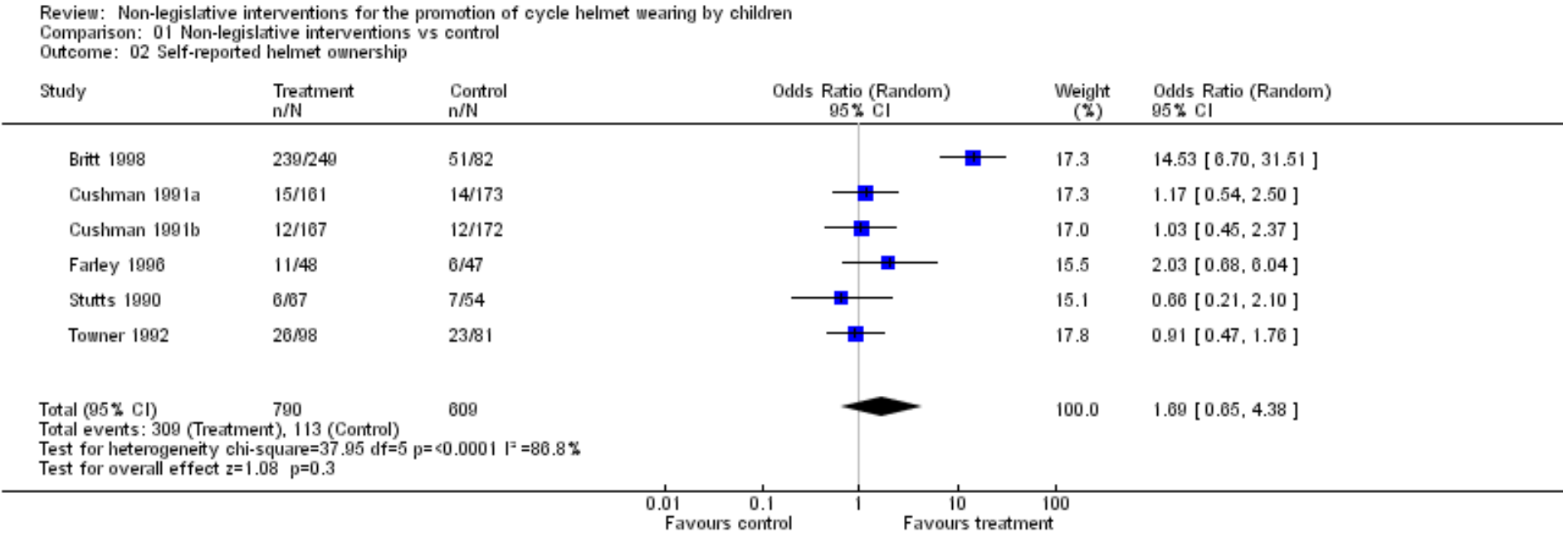
World Health Organization

Does a screening test save lives?



World Health Organization

What makes children wear a bicycle helmet?



World Health Organization

How would you grade that evidence?

Evidence	Recommendation	Organization
II-2	B	USPSTF
C+	1	ACCP
Strong	Strongly recommended	SIGN

Recommendation for use of oral anticoagulation in patients with atrial fibrillation and mitral valve disease

Evidence	Recommendation	Organization
II-2	B	USPTSTF
C+	1	ACCP
Strong	Strongly recommended	SIGN

Problem

- **Too many systems**
- **They only evaluate design**
- **No consideration of other important factors that influence judgements and recommendations**

Why bother about grading?

- **People draw conclusions about**
  - **Quality of evidence**
  - **Strength of recommendations**
- **Systematic explicit approaches help**
  - **Protect against errors**
  - **Resolve disagreements**
  - **Facilitate critical appraisal**
  - **Communicate information**

What about WHO guidelines?

GRADE

Grades of Recommendation Assessment Development and Evaluation

Definitions

Quality of evidence

the extent to which one can be confident that an estimate of effect or association is correct

Four categories

- **High** ++++

  Further research is very unlikely to change our confidence in the estimate of effect

- **Moderate** +++

  Further research is likely to have an important impact…

- **Low** ++

  Further research is very likely to have an important impact…..

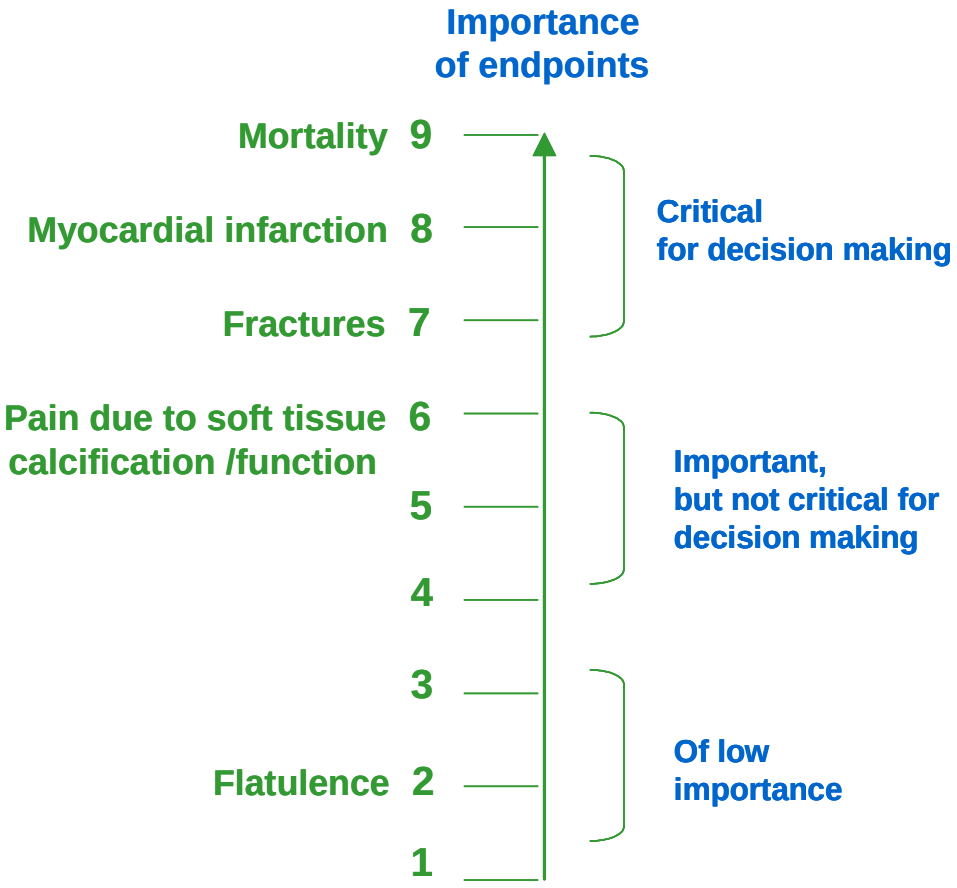
- **Very low** +

  Any estimate of effect is very uncertain

The quality of evidence needs to be considered for each important outcome

- **A review or guideline needs a clearly formulated question**
  - **Patients or population**
  - **Intervention**
  - **Comparison**
  - **Outcomes**
- **The quality of evidence may be different for different outcomes**
- **Decision makers (and review authors) need to consider the relative importance of outcomes**

Hierarchy of outcomes according to their importance to assess the effect of phosphate lowering drugs in patients with renal failure and hyperphosphatemia



Study design is important

- **Early systems of grading the quality of evidence focused almost exclusively on study design**
- **Randomised trials provide, in general, far stronger evidence than observational studies.**
  - **Randomised trials start out at High**
  - **Observational studies start out at Low**
- **However, other factors may decrease or increase the quality of evidence**

Factors that may decrease the quality of evidence

- Study limitations
- Inconsistency of results
- Indirectness of evidence
- Imprecise results
- Reporting bias

Factors that may increase the quality of evidence

- Large magnitude of effect
- A dose response relationship

Quality assessment criteria

Quality of evidence	Study design	Lower if	Higher if
High	Randomised trial	**Study quality**: Serious limitations Very serious limitations Important **inconsistency** **Directness**: Some uncertainty Major uncertainty **Sparse or imprecise data** High probability of **reporting bias**	**Strong association**: Strong, no plausible confounders Very strong, no major threats to validity Evidence of a **Dose response gradient** All plausible **confounders** would have reduced the effect
Moderate			
Low	Observational study		
Very low			

Strength of recommendation

The degree of confidence that the desirable effects of adherence to a recommendation outweigh the undesirable effects.

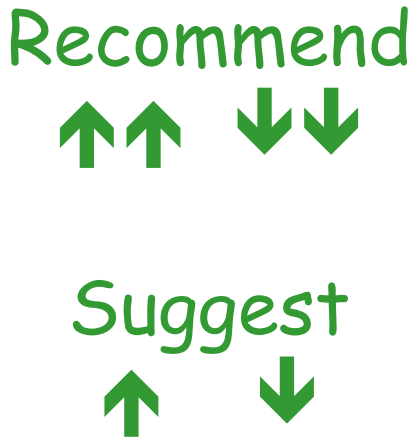
Desirable effects
- health benefits
- less burden
- savings

Undesirable effects
- harms
- more burden
- costs

Categories of recommendations

Although the degree of confidence is a continuum, we suggest using two categories: strong and weak.

- **Strong recommendation:** the panel is confident that the desirable effects of adherence to a recommendation outweigh the undesirable effects.
- **Weak recommendation:** the panel concludes that the desirable effects of adherence to a recommendation probably outweigh the undesirable effects, but is not confident.



Determinants of strength of recommendation

Factors	Impact on the strength of a recommendation
Balance between desirable and undesirable effects	**Larger the difference between the desirable and undesirable effects, more likely a strong recommendation warranted. Narrower the gradient, more likely weak recommendation warranted.**
Quality of the evidence	**Higher the quality of evidence, more likely a strong recommendation warranted.**
Values and preferences	**More variability in values and preferences, or more uncertainty in values and preferences, more likely weak recommendation warranted.**
Costs (resource use)	**Higher the costs of an intervention – that is, the more resources consumed – less likely a strong recommendation warranted.**

Judgements about the strength of a recommendation

- No precise threshold for going from a strong to a weak recommendation
- The presence of important concerns about one or more of these factors make a weak recommendation more likely.
- Panels should consider all of these factors and make the reasons for their judgements explicit.
- Recommendations should specify the perspective that is taken (e.g. individual patient, health system) and which outcomes were considered (including which, if any costs).

Implications of a strong recommendation

- **Patients:** **Most people in your situation would want the recommended course of action and only a small proportion would not**
- **Clinicians:** **Most patients should receive the recommended course of action**
- **Policy makers:** **The recommendation can be adapted as a policy in most situations**

Implications of a weak recommendation

- **Patients:** The majority of people in your situation would want the recommended course of action, but many would not
- **Clinicians:** Be prepared to help patients to make a decision that is consistent with their own values
- **Policy makers:** There is a need for substantial debate and involvement of stakeholders

Example

- **Post partum haemorrhage is the major cause of maternal mortality**
- **Effective interventions are available – active management**
- **Which ones?**
- **Is one better than the other?**
- **Who should use them?**

Should active management of the third stage of labour be used by skilled providers for all women to prevent post-partum haemorrhage?

Quality assessment						Summary of findings						
						No of patients		Effect			Quality	Importance
No of studies (Ref)	Design	Limitations	Consistency	Directness	Other considerations	Active management	Standard procedures	Baseline Risk (95%CI)	Relative risk (95%CI)	NNT (95%CI)		
Benefits:												
Maternal deaths												
0	-	-	-	-	-	-	-	-	-	-	-	8.5
Admission to intensive care unit												
0	-	-	-	-	-	-	-	-	-	-	-	6.4
Blood loss ≥ 500 ml												
4 PW 00[1] Ad 97 Br 88 Du 90 Hi 98	RCT	serious limitation[2,3,17]	no important inconsistency	some uncertainty about directness[4,5]	none	3126	3158	min 8.3% (6.3, 10.3) max 17.9% (15.3, 20.5)	0.38 (0.32, 0.46)	min 8 (6.7, 11.2) max 16 (11.7, 24.7)		6.3
Blood loss ≥ 1000 ml												
4 PW 00[1] Ad 97 Br 88 Du 90 Hi 98	RCT	serious limitation[2,3,17]	no important inconsistency	some uncertainty about directness[4,5]	none	3126	3158	min 1.5% (0.6-2.4) max 3.2% (2.0-4.4)	0.33 (0.21, 0.51)	min 41 (26.5, 90.1) max 73 (43.3, 225.5)		7.7
Need for blood transfusion												
5 PW 00[1] Ad 97 Br 93 Br 88 Du 90 Hi 98	RCT	minor limitation[3,8]	no important inconsistency	some uncertainty about directness[7]	none	3229	3248	5.7% (4.1-7.2)[16]	0.34 (0.22, 0.53)	28 (18.7, 59,1)[16]		7.8

Quality assessment					
No of studies (Ref)	**Design**	**Limitations**	**Consistency**	**Directness**	**Other considerations**
Benefits:					
Maternal deaths					
0	-	-	-	-	-
Admission to intensive care unit					
0	-	-	-	-	-
Blood loss ≥ 500 ml					
4 PW 00[1] Ad 97 Br 88 Du 90 Hi 98	RCT	serious limitation[2,3,17]	no important inconsistency	some uncertainty about directness[4,5]	none
Blood loss ≥ 1000 ml					
4 PW 00[1] Ad 97 Br 88 Du 90 Hi 98	RCT	serious limitation[2,3,17]	no important inconsistency	some uncertainty about directness[4,5]	none
Need for blood transfusion					
5 PW 00[1] Ad 97 Br 93 Br 88 Du 90 Hi 98	RCT	minor limitation[3,8]	no important inconsistency	some uncertainty about directness[7]	none

Summary of findings						
No of patients		Effect			Quality	Importance
Active management	Standard procedures	Baseline Risk (95%CI)	Relative risk (95%CI)	NNT (95%CI)		
-	-	-	-	-	-	8.5
-	-	-	-	-	-	6.4
3126	3158	min 8.3% (6.3, 10.3) max 17.9% (15.3, 20.5)	0.38 (0.32, 0.46)	min 8 (6.7, 11.2) max 16 (11.7, 24.7)		6.3
3126	3158	min 1.5% (0.6-2.4) max 3.2% (2.0-4.4)	0.33 (0.21, 0.51)	min 41 (26.5, 90.1) max 73 (43.3, 225.5)		7.7
3229	3248	5.7% (4.1-7.2)[16]	0.34 (0.22, 0.53)	28 (18.7, 59,1)[16]		7.8

Quality assessment						Summary of findings						
						No of patients		Effect			Quality	Importance
No of studies (Ref)	Design	Limitations	Consistency	Directness	Other considerations	Active management	Standard procedures	Baseline Risk (95%CI)	Relative risk (95%CI)	NNT (95%CI)		
Benefits:												
Maternal deaths												
0	-	-	-	-	-	-	-	-	-	-	-	8.5
Admission to intensive care unit												
0	-	-	-	-	-	-	-	-	-	-	-	6.4
Blood loss ≥ 500 ml												
4 PW 00[1] Ad 97 Br 88 Du 90 Hi 98	RCT	serious limitation[2,3,17]	no important inconsistency	some uncertainty about directness[4,5]	none	3126	3158	min 8.3% (6.3, 10.3) max 17.9% (15.3, 20.5)	0.38 (0.32, 0.46)	min 8 (6.7, 11.2) max 16 (11.7, 24.7)		6.3
Blood loss ≥ 1000 ml												
4 PW 00[1] Ad 97 Br 88 Du 90 Hi 98	RCT	serious limitation[2,3,17]	no important inconsistency	some uncertainty about directness[4,5]	none	3126	3158	min 1.5% (0.6-2.4) max 3.2% (2.0-4.4)	0.33 (0.21, 0.51)	min 41 (26.5, 90.1) max 73 (43.3, 225.5)		7.7
Need for blood transfusion												
5 PW 00[1] Ad 97 Br 93 Br 88 Du 90 Hi 98	RCT	minor limitation[3,8]	no important inconsistency	some uncertainty about directness[7]	none	3229	3248	5.7% (4.1-7.2)[16]	0.34 (0.22, 0.53)	28 (18.7, 59,1)[16]		7.8

What would you recommend?

- **Rate the importance of outcomes**
- **Check the quality of evidence**
- **Decide on your recommendation**

What would you recommend?

- **Active management of third stage of labour should be offered by skilled attendants to all women. (Strong recommendation, moderate quality evidence)**

Expertise needed for guideline development

- Literature search and analysis
- Epidemiology and biostatistics
- Healthcare research
- Clinical expert knowledge
- Social group processes
- Writing and editing of texts
- Production of guideline products

Benefits and limitations of clinical guidelines

- Improving quality of care
- Improving information about optimal care
- Summary of research findings
- External accountability
- Basis for teaching and education
- Basis for interdisciplinary cooperation
- Contributing to efficient care
- Setting health care priorities

- Cookbook medicine
- Unrealistic expectations
- Loss of clinical autonomy
- Professional resistance and concern for legal consequences
- Misuse by governmental authorities
- Uncertainty about cost-effectiveness
- Hidden political motives

Grol et al 2005

Implementation and behaviour change

Usually effective	Sometimes effective and sometimes not	Of little or no effect	Effectiveness unknown
Outreach visits	Audit and feedback	Educational materials	Financial stimuli
Decision support, reminders	Efforts of opinion leaders	Courses, conferences	Administrative or organisational interventions
Interactive education	Local consensus meetings		
Multifaceted interventions	Patient oriented interventions		
Mass media interventions			

Bero et al, 1998

Summary

- **Evidence is a tough taskmaster**
- **Systematic reviews and critical appraisal essential**
- **Content experts alone insufficient**
- **Transparent system required**
- **Judgements should be explicit**
- **To make it worth while, implementation and evaluation have to be integral to process**