What is a WHO guideline?
And what is the Guideline Review Committee, anyway?

Dr Suzanne Hill, HTP/PSM/PAR
Problem:

- WHO guidelines are insufficiently transparent and not evidence based
  - Lack of use of systematic reviews
  - Lack of transparency about judgements
  - Too much dependence on expert opinion
  - Lack of emphasis on adapting global guidelines to end users' needs
  - Tension between time taken and when advice needed
  - Lack of resources
Process for developing recommendations

- More about expert consensus
- Less about evidence and specifically less about systematic reviews
- Relatively little use of central support function and guidelines for guidelines to ensure methodological rigour
- Relatively little formalization of the recommendation development process
- Constrained often by a lack of resources and the urgency of the need
More about expert consensus

- Expert committees – heavy reliance on ‘world experts’
- Sometimes little insight into shortcomings, variable preparation
- Consensus process can be dominated by strong person at the table
- Variation in consultation processes – who is consulted, why
- Cultural change from GOBSAT needs support
Is this unfair? Or wrong?
'Critical appraisal of the JNC VI, WHO/ISH and BHS guidelines for essential hypertension.'

'These differing recommendations between JNC VI and BHS, and WHO/ISH cannot be reconciled and they are of such magnitude as to carry serious implications for clinical practice, not least among which is that acceptance of the WHO/ISH levels of 'normality' for blood pressure would result in some 45% of the population of all ages and nearly 60% of elderly people being classified as 'hypertensive'.

O'Brien & Staessen, 2000
## Internal Review of WHO primary drug treatment guidelines, N=71

<table>
<thead>
<tr>
<th>Feature</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Printed in 1995 or later</td>
<td>52/71</td>
<td>(73%)</td>
</tr>
<tr>
<td>Linked references</td>
<td>39/71</td>
<td>(55%)</td>
</tr>
<tr>
<td>Target audience clearly identified</td>
<td>36/71</td>
<td>(51%)</td>
</tr>
<tr>
<td>Affiliations of contributors described</td>
<td>31/71</td>
<td>(44%)</td>
</tr>
<tr>
<td>Development process described</td>
<td>27/71</td>
<td>(38%)</td>
</tr>
<tr>
<td>Index present</td>
<td>12/71</td>
<td>(17%)</td>
</tr>
<tr>
<td>Evidence based according to current standards</td>
<td>1/71</td>
<td></td>
</tr>
</tbody>
</table>
Critics slam draft WHO report on homoeopathy

A WHO group that caused controversy with a 2003 report on acupuncture has now turned its attention to homoeopathy. But if the allegations of bias levelled at a draft version of the report are anything to go by, the group has once again put itself in the firing line. Michael McCarthy reports.
And 2005…

- Internal review, about 60 'guideline type' documents
- <20% had a reference….
Why?

- Time
- Money
- Skills
- Quantity
- Responding to needs
Solution 2 (2007)

- A review committee
- Minimum standards for:
  - Reporting
  - Processes
  - Use of evidence
- Revised WHO guidelines for guidelines
- Different types of documents to fit different purposes:
  - Emergency
  - Focussed
  - Full
  - 'Books'
  - Joint guidelines?
What is a WHO Guideline?

- Systematic statements/recommendations to aid decision making about health interventions
- Applies to clinical and public health interventions
- Does not apply to standards (e.g., pharmacopoeial, food), standard operating procedures, evidence synthesis without recommendations
- Grey area – compilations of clinical information without clear recommendations

- The NAME is irrelevant
Guideline Review Committee – terms of reference

1. Defining appropriate and standardized processes related to guideline development, including developing standard formats and templates for different levels of recommendations (e.g. rapid policy advice or full guidelines) and different stages of preparation of such recommendations.

2. Ensuring that all guidelines prepared by WHO comply with the previously prepared Guidelines for Guidelines (GFG) by providing advice, guidance, and initial and final approval of WHO guideline documents.

3. Developing and implementing a plan to ensure that WHO guidelines committee members have appropriate knowledge of the approved methods for guideline development and to identify opportunities in collaboration with the Global Learning Committee to build capacity of WHO staff in guidelines development.

4. Develop collaboration and cooperation with other organizations and international networks that have methodological expertise and skills in relation to guidelines development, adaptation and implementation (e.g. National Institute for Clinical Excellence, United Kingdom, the Guidelines International Network, the Cochrane Collaboration, etc.).
1. Scoping the document: reasons for choosing the topic, problems with existing guidelines, variations and gaps,

2. Group composition (or consultations)

3. Conflict of interest

4. Formulations of the questions and choice of the relevant outcomes

5. Evidence retrieval, evaluation and synthesis (balance sheet, evidence table)

6. Benefit/risk profile: integrating evidence with values and preferences, equity and costs

7. Formulation of the recommendations

8. Implementation and evaluation of impact

9. Research needs or areas of further research

10. Peer-review process and updating

Title, responsible person, WHO Department
- responsible of the clearance process, WHO Departments involved, CC involved,

Reporting standard and process

Standards for evidence: GRADE system

Reporting standard and process
Minimum standards for *reporting* in WHO guidelines

- Who was involved and their declaration of interests
- How the guideline was developed, including
  - how the evidence was identified
  - how the recommendations were made
- Use by date (review by date)
Guideline types

- **Emergency advice**
  - Response to acute need, evidence informed, limited consultation, short use-by date

- **Standard /focused advice**
  - Limited topic area, 10-20 'questions', evidence-based, 1 guideline group meeting

- **Comprehensive advice**
  - Disease/policy area, evidence-based, 3-4 meetings

- **Textbooks**

- **Joint guidelines**
Guideline 'products'

- Protocols, algorithms, wall charts, booklets.....
- Aids to implementation
- Should be linked to a 'guideline'
And what about evidence?
Problem

- Too many systems
- Concentrating only on study design
- Not including other factors that influence judgements and recommendations
Some myths about evidence

- Evidence has to be an RCT
- A systematic review of RCTs is all you need
- If you don't have an RCT you can't do evidence based guidelines
- Public health interventions are different
- Policy interventions are different
- The expert know the evidence anyway… so we will just call it expert opinion….
Why bother about grading?

- People draw conclusions about the
  - quality of evidence
  - strength of recommendations

- Systematic and explicit approaches can help
  - protect against errors
  - resolve disagreements
  - facilitate critical appraisal
  - communicate information
Definitions

Quality of evidence

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

- Although the degree of confidence is a continuum, four categories suggested:
  - High
  - Moderate
  - Low
  - Very low
# WHO guidelines

## Quality of evidence

<table>
<thead>
<tr>
<th>Quality of evidence</th>
<th>Study design</th>
<th>Lower if *</th>
<th>Higher if *</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Randomized trial</td>
<td>Study quality:</td>
<td>Strong association:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- 1 Serious limitations</td>
<td>+1 Strong, no plausible confounders, consistent and direct evidence**</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- 2 Very serious limitations</td>
<td>+2 Very strong, no major threats to validity and direct evidence***</td>
</tr>
<tr>
<td>Moderate</td>
<td></td>
<td>- 1 Important inconsistency</td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>Observational study</td>
<td>Directness:</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- 1 Some uncertainty</td>
<td>+1 Evidence of a Dose response gradient</td>
</tr>
<tr>
<td>Very low</td>
<td>Any other evidence</td>
<td>- 2 Major uncertainty</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- 1 Sparse data</td>
<td>+1 All plausible confounders would have reduced the effect **</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- 1 High probability of Reporting bias</td>
<td></td>
</tr>
</tbody>
</table>

---

* 1 = move up or down one grade (for example from high to intermediate)

2 = move up or down two grades (for example from high to low)

** A statistically significant relative risk of $>2$ ($<0.5$), based on consistent evidence from two or more observational studies, with no plausible confounders.

*** A statistically significant relative risk of $>5$ ($<0.2$) based on direct evidence with no major threats to validity.
Practicalities

- Synthesis of ALL available evidence
- Evidence summaries for group meetings using standard template
QUESTION: Should active management of the third stage of labour be used by skilled providers for all women to prevent postpartum hemorrhage (PPH)?

<table>
<thead>
<tr>
<th>Quality assessment</th>
<th>Summary of findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>No of studies (Ref)</td>
<td>No of patients</td>
</tr>
<tr>
<td>Design</td>
<td>Limitations</td>
</tr>
</tbody>
</table>

**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 limitation2,3,17 |-1 | no important inconsistency | some uncertainty about directness8,5 |-1 | 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) | low quality ++oo | 6.3 |

**Blood loss ≥ 1000 ml**

| 4 PW 00:1 Ad 97 Br 88 Du 90 Hi 98 RCT | serious limitation2,3,17 |-1 | no important inconsistency | some uncertainty about directness8,5 |-1 | 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) | low quality ++oo | 7.7 |
QUESTION: Should active management of the third stage of labour be used by skilled providers for all women to prevent postpartum hemorrhage (PPH)?

<table>
<thead>
<tr>
<th>Benefits:</th>
<th>Maternal deaths</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>No of studies (Ref)</td>
<td>Design</td>
<td>Limitations</td>
</tr>
<tr>
<td>0</td>
<td>RCT</td>
<td>serious limitation</td>
</tr>
</tbody>
</table>

| Admission to intensive care unit | 0 | - | - | - | - |

| Blood loss ≥ 500 ml | 4 \(\text{PW 00}^1\) \(\text{Ad 97}\) \(\text{Br 88}\) \(\text{Du 90}\) \(\text{Hi 98}\) | RCT | serious limitation | no importance | some uncertainty about directness | none |

| Blood loss ≥ 1000 ml | 4 \(\text{PW 00}^1\) \(\text{Ad 97}\) \(\text{Br 88}\) \(\text{Du 90}\) \(\text{Hi 98}\) | RCT | serious limitation | no importance | some uncertainty about directness | none |
**QUESTION:** Should active management of the third stage of labour be used by skilled providers for all women to prevent postpartum hemorrhage (PPH)?

<table>
<thead>
<tr>
<th>Benefits:</th>
<th>Maternal deaths</th>
<th>Admission to intensive care unit</th>
<th>Blood loss ≥ 500 ml</th>
<th>Blood loss ≥ 1000 ml</th>
</tr>
</thead>
<tbody>
<tr>
<td>No of studies (Ref)</td>
<td>4 PW 00¹ Ad 97 Br 88 Du 90 Hi 98</td>
<td>4 PW 00¹ Ad 97 Br 88 Du 90 Hi 98</td>
<td>4 PW 00¹ Ad 97 Br 88 Du 90 Hi 98</td>
<td>4 PW 00¹ Ad 97 Br 88 Du 90 Hi 98</td>
</tr>
<tr>
<td>Design</td>
<td>RCT</td>
<td>RCT</td>
<td>RCT</td>
<td>RCT</td>
</tr>
<tr>
<td>Limitations</td>
<td>serious limitation²,³,¹⁷ -1</td>
<td>serious limitation²,³,¹⁷ -1</td>
<td>serious limitation²,³,¹⁷ -1</td>
<td>serious limitation²,³,¹⁷ -1</td>
</tr>
<tr>
<td>Consistency</td>
<td>no important inconsistency</td>
<td>no important inconsistency</td>
<td>no important inconsistency</td>
<td>no important inconsistency</td>
</tr>
<tr>
<td>Directness</td>
<td>some uncertainty about directness⁴,⁵ -1</td>
<td>some uncertainty about directness⁴,⁵ -1</td>
<td>some uncertainty about directness⁴,⁵ -1</td>
<td>some uncertainty about directness⁴,⁵ -1</td>
</tr>
<tr>
<td>Other considerations</td>
<td>none</td>
<td>none</td>
<td>none</td>
<td>none</td>
</tr>
<tr>
<td>No of patients</td>
<td>3126</td>
<td>3158</td>
<td>3126</td>
<td>3158</td>
</tr>
<tr>
<td>Effect</td>
<td>Baseline Risk (95%CI)</td>
<td>Relative risk (95%CI)</td>
<td>NNT (95%CI)</td>
<td>Quality</td>
</tr>
<tr>
<td>Summary of findings</td>
<td>Active management</td>
<td>Standard procedures</td>
<td>Blood loss ≥ 500 ml</td>
<td>Blood loss ≥ 1000 ml</td>
</tr>
<tr>
<td>Maternal deaths</td>
<td>0 - - - -</td>
<td>0 - - - -</td>
<td>0 - - - -</td>
<td>0 - - - -</td>
</tr>
<tr>
<td>Admission to intensive care unit</td>
<td>0 - - - -</td>
<td>0 - - - -</td>
<td>0 - - - -</td>
<td>0 - - - -</td>
</tr>
<tr>
<td>Blood loss ≥ 500 ml</td>
<td>0 - - - -</td>
<td>0 - - - -</td>
<td>0 - - - -</td>
<td>0 - - - -</td>
</tr>
<tr>
<td>Blood loss ≥ 1000 ml</td>
<td>0 - - - -</td>
<td>0 - - - -</td>
<td>0 - - - -</td>
<td>0 - - - -</td>
</tr>
</tbody>
</table>

---

**Quality assessment**

- **Importance**
- **Quality**
- **Effect**
- **No of patients**
- **Baseline Risk (95%CI)**
- **Relative risk (95%CI)**
- **NNT (95%CI)**

---

**Summary of findings**

- **Baseline Risk (95%CI)**: 0.38 (0.32, 0.46)
- **Relative risk (95%CI)**: 8.3% (6.3, 10.3)
- **NNT (95%CI)**: 4,5,17

---

**Blood loss ≥ 500 ml**

- **Quality**: low quality
- **Importance**: ++oo
- **NNT**: 6.3

---

**Blood loss ≥ 1000 ml**

- **Quality**: low quality
- **Importance**: ++oo
- **NNT**: 7.7

---

**Admission to intensive care unit**

- **Quality**: low quality
- **Importance**: ++oo
- **NNT**: 8.5

---

**Maternal deaths**

- **Quality**: low quality
- **Importance**: ++oo
- **NNT**: 6.4
### Table 3: Deciding on strength of a recommendation

<table>
<thead>
<tr>
<th>Issue</th>
<th>Recommended process</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Quality of evidence</strong></td>
<td>Strong recommendations usually require higher quality evidence for all the critical outcomes. The lower the quality of evidence the less likely is a strong recommendation.</td>
</tr>
<tr>
<td><strong>Balance of benefits and downsides</strong></td>
<td></td>
</tr>
<tr>
<td>1. Quality of evidence</td>
<td>Seek evidence about the relative values that patients place on outcomes and the actual value they place on them (critical; important but not critical; not important). Seek evidence about variability in preferences and values in patients and other stakeholders. It should be upfront that the relative importance of the outcomes should be included in the considerations before you make recommendations. If values and preferences vary widely, a strong recommendation becomes less likely.</td>
</tr>
<tr>
<td>2. Relative importance of the outcomes</td>
<td>Consider the baseline risk of an outcome. Is the baseline risk going to make a difference? If yes, then consider making separate recommendations for different populations. The higher the baseline risk, the higher the magnitude of benefit and the more likely the recommendation is strong.</td>
</tr>
<tr>
<td>a. benefits of therapy</td>
<td></td>
</tr>
<tr>
<td>b. harm of treatment</td>
<td></td>
</tr>
<tr>
<td>c. burdens of therapy</td>
<td></td>
</tr>
<tr>
<td>3. Baseline risks of outcomes</td>
<td>Consider the relative magnitude of the net effect. Large relative effects will lead to a higher likelihood of a strong recommendation if the balance of benefit, harms, and burdens go in the same direction. If they go in opposite directions and the relative magnitude of effects is large (large benefits coming with large risk of adverse effects), the recommendation is more likely to be weak.</td>
</tr>
<tr>
<td>a. benefits of therapy</td>
<td></td>
</tr>
<tr>
<td>b. harm of treatments</td>
<td></td>
</tr>
<tr>
<td>c. burdens of therapy</td>
<td></td>
</tr>
<tr>
<td>4. Magnitude of relative risk</td>
<td></td>
</tr>
<tr>
<td>a. benefits (reduction in RR)</td>
<td>Large absolute effects are more likely to lead to strong recommendation.</td>
</tr>
<tr>
<td>b. harms (increase in RR)</td>
<td></td>
</tr>
<tr>
<td>c. burden</td>
<td></td>
</tr>
<tr>
<td>5. Absolute magnitude of the effect</td>
<td>The greater the precision the more likely the recommendation is strong.</td>
</tr>
<tr>
<td>a. benefits</td>
<td></td>
</tr>
<tr>
<td>b. harms</td>
<td></td>
</tr>
<tr>
<td>c. burden</td>
<td></td>
</tr>
<tr>
<td>6. Precision of the estimates of the effects</td>
<td>The more similar the setting and patients for which one is making a recommendation to the setting and patients generating the evidence, the more likely the recommendation is strong.</td>
</tr>
<tr>
<td>a. benefits of therapy</td>
<td></td>
</tr>
<tr>
<td>b. harms of treatments</td>
<td></td>
</tr>
<tr>
<td>c. burdens of therapy</td>
<td></td>
</tr>
<tr>
<td>7. Factors that modify effects in specific settings/Local factors that may affect translating of the evidence into practice</td>
<td>Consider that important benefits should come at a reasonable cost. The higher the incremental cost, all else being equal, the less likely that the recommendation in favour of an intervention is strong.</td>
</tr>
</tbody>
</table>
Although the degree of confidence is a continuum, two categories are used: strong and weak.

A **strong recommendation** is one for which the panel is confident that the desirable effects of adherence to a recommendation outweigh the undesirable effects.

A **weak recommendation** is one for which the panel concludes that the desirable effects of adherence to a recommendation probably outweigh the undesirable effects, but the panel is not confident about these trade-offs. Reasons for not being confident can include:

- absence of high quality evidence;
- presence of imprecise estimates of benefits or harms;
- uncertainty or variation in how different individuals value the outcomes;
- small benefits;
- the benefits may not be worth the costs (including the costs of implementing the recommendation).
guidelines for WHO guidelines
Strength of a recommendation
Strong vs weak?

Examples of implications of a strong recommendation are:

- **For patients**: Most people in your situation would want the recommended course of action and only a small proportion would not.
- **For clinicians**: Most patients should receive the recommended course of action. Adherence to this recommendation is a reasonable measure of good quality care.
- **For policy-makers**: The recommendation can be adapted as a policy in most situations. Quality initiatives could use this recommendation to measure variations in quality.

Examples of implications of a weak recommendation are:

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

- Synthesis of ALL available evidence
- Evidence summaries for group meetings using standard template
- Explicitly linked to recommendations, explaining reasons for judgements
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