Antivirals for Influenza:
Review of the Evidence from Observational Studies

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

Seoul, August 30, 2011
10:30 am – 12:00 pm
Collaborators


Disclosure

• Co-chair GRADE Working Group
• Work with various guideline groups using GRADE
• American College of Physicians (ACP) Clinical Practice Guidelines Committee
• WHO: Expert Advisory Panel on Clinical Practice Guidelines and Clinical Research Methods and Ethics & chair of various guideline panels
• WHO funding
• No direct/personal for profit payments
Outline

• Observational studies and systematic reviews
• Quality of evidence and GRADE in observational studies
• Feasibility of synthesizing information from observational studies
• Reconciling inconsistent evidence from randomized trials and observational studies

Background

• Little is known about the proper synthesis of observational study evidence in systematic reviews
• Special challenges:
  – Assessing quality of a body of evidence from observational studies
  – Developing recommendations based on observational research
• But necessary:
  – Safety, public health, surgical specialties, policy making
Determinants of quality

- RCTs ⊕⊕⊕⊕
- observational studies ⊕⊕○○
- 5 factors that can lower quality
  1. limitations in detailed design and execution (risk of bias criteria)
  2. Inconsistency (or heterogeneity)
  3. Indirectness (PICO and applicability)
  4. Imprecision (number of events and confidence intervals)
  5. Publication bias
- 3 factors can increase quality
  1. large magnitude of effect
  2. plausible residual bias or confounding
  3. dose-response gradient

  • Valid – yes!

GRADE and observational studies

- Users of GRADE have expressed concern that GRADE places greater confidence on the results of randomized studies (RCTs)
  - population or public health interventions and environmental health, health policy making and often surgery, where conducting RCTs is either challenging or unethical
  - Consequently, the best quality of evidence for these questions will come from observational studies
- Some argue that it would be unreasonable to grade such “best quality” evidence, typical of most public health questions, as low
GRADE and observational studies

- Argument is not valid for several reasons:
  - inability to obtain RCT data does not eliminate or minimize the bias associated with observational data
  - quality of evidence from observational studies can lead to moderate and even high quality evidence within the GRADE framework – why is this forgotten?
  - need to be able to compare confidence in estimates of effect across healthcare questions

GRADE and WHO guidelines

- GRADE is approach for WHO
- New guideline on pharmacological management of influenza
  - Previously few randomized trials
    - Criticized
    - Low quality evidence for many outcomes (imprecision)
    - Industry sponsored – publication bias
    - Not all outcomes
- Review of observational studies
  - To inform guidelines
Methods

• Standard systematic review
  – MEDLINE, EMBASE, CENTRAL, CINAHL, SIGLE, the Chinese Biomedical Literature Database, Panteleimon and LILACS for relevant studies up to November 2010
  – contacted pharmaceutical companies and international agencies
  – RevMan 5.1

• 10 PICO → recommendations approach
  – Outcomes determined through Delphi process previously

• QoE according to GRADE approach
  – GRADEpro (www.gradeworkinggroup.org)
  – Risk of bias using Ottawa Newcastle scale
Methods

Types of participants
- We included studies in all populations with influenza or influenza like-illness.

Types of intervention
- Oseltamivir, zanamivir, amantadine or rimantadine in any dose or by any route.

Type of outcome measures
- We determined a priori to report on the following outcomes because they were judged to be important or critical for decision making:
  - Mortality, Hospitalisation, ICU Admission, mechanical ventilation and respiratory failure, Duration of hospitalization, Time to alleviation of symptoms, Time to return to normal activity, Complications
  - Critical adverse events (e.g. major psychotic disorders, encephalitis, stroke and seizure),
  - Important adverse events (e.g. pain in extremities, clonic twitching, body weakness, dermatological changes such as urticaria and rash)
  - Viral shedding and Resistance

Results - PRISMA

Records identified through database searching (all study designs)
EMBASE = 9873
CINAHL = 1062
LILACS = 19
COCHRANE = 301
Chinese Biomedical Literature Database = 914
Pantanal = 12
(Total n = 12178)

Records after duplicates removed
(n = 7463)

Records screened
(n = 7483)

Records excluded
(n = 6563)

Full-text articles assessed for eligibility
(n = 920)

Full-text articles excluded
(n = 825)

Records included
(n = 89)

Studies awaiting assessment
(n = 6)
- Studies awaiting translation (1)
- Papers could not obtain in full (5)

Studies included
N = 89
- 51 + 5 studies
- 7 studies
- 6 studies
- 0 studies
- 8 studies
- 5 studies
- 2 studies
- 1 study
- 0 studies
- This study may be relevant to multiple questions
### Results

#### Should oseltamivir versus no treatment be used to treat influenza?

**Mortality (adjusted)**

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Events</th>
<th>Total</th>
<th>Weight</th>
<th>Odds Ratio (95% CI)</th>
<th>Test for overall effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Favours oseltamivir</td>
<td>0.1 0.2 0.5 1 2 5 10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favours no treatment</td>
<td>0.001 0.1 1 10 1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Mortality (unadjusted)**

<table>
<thead>
<tr>
<th>Study or Subgroup</th>
<th>Events</th>
<th>Total</th>
<th>Weight</th>
<th>Odds Ratio (95% CI)</th>
<th>Test for overall effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Favours oseltamivir</td>
<td>0.1 0.2 0.5 1 2 5 10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Favours no treatment</td>
<td>0.001 0.1 1 10 1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Question:** Should oseltamivir versus no antiviral treatment be used for influenza (follow-up: 30 days)?

<table>
<thead>
<tr>
<th>Quality assessment</th>
<th>Summary of findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall mortality</td>
<td>Study event rates (%)</td>
</tr>
<tr>
<td>No antiviral treatment</td>
<td>With no antiviral treatment</td>
</tr>
</tbody>
</table>

### Results

#### Question: Should oseltamivir versus no antiviral treatment be used for influenza (follow-up: 30 days)?

<table>
<thead>
<tr>
<th>Quality assessment</th>
<th>Summary of findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall mortality</td>
<td>Study event rates (%)</td>
</tr>
<tr>
<td>No antiviral treatment</td>
<td>With no antiviral treatment</td>
</tr>
</tbody>
</table>

**Study event rates (%)**

- No antiviral treatment: 1074/10000(10.74%)
- Oseltamivir: 992/10000 (9.92%)

**Risk ratio (95% CI)**

- No antiviral treatment: 1.14 (95% CI: 1.03 to 1.26)
- Oseltamivir: 0.83 (95% CI: 0.72 to 0.96)

**Anticipated absolute effects**

- No antiviral treatment: 7.47 (95% CI: 6.34 to 8.61)
- Oseltamivir: 5.50 (95% CI: 4.38 to 6.83)

**Absolute effect with Oseltamivir**

- 0.64 (95% CI: 0.53 to 0.76)

**Relative effect**

- 0.23 (95% CI: 0.13 to 0.43)

**Mortality**

- 172 fewer deaths per 1000 (from 220 to 210 fewer)
- 101 fewer deaths per 25 more
- 3 fewer hospitalisations per 1000 (from 1 to 4 fewer)
- 3 fewer hospitalisations per 1000 (from 2 to 4 fewer)

**Hospitalisations**

- 127 hospitalisations per 1000 (from 1 to 4 fewer)
- 127 hospitalisations per 1000 (from 2 to 4 fewer)

**ICU admissions in mechanical ventilation/respiratory failure**

- 3 fewer hospitalisations per 1000 (from 1 to 4 fewer)
- 3 fewer hospitalisations per 1000 (from 2 to 4 fewer)

**Anticipated absolute effects**

- 0.51 (95% CI: 0.23 to 1.14)
- 0.03 (95% CI: 0.00 to 0.69)

**Relative effect**

- 0.34 (95% CI: 0.13 to 0.88)
- 0.08 (95% CI: 0.03 to 0.23)

**Absolute effect with Oseltamivir**

- 0.39 (95% CI: 0.16 to 0.97)

**Relative effect**

- 0.13 (95% CI: 0.04 to 0.41)
- 0.03 (95% CI: 0.01 to 0.11)

**Anticipated absolute effects**

- 2.14 (95% CI: 0.60 to 7.64)
- 0.08 (95% CI: 0.02 to 0.39)

**Relative effect**

- 0.49 (95% CI: 0.19 to 1.25)
- 0.13 (95% CI: 0.04 to 0.43)

**Absolute effect with Oseltamivir**

- 0.34 (95% CI: 0.13 to 0.88)
- 0.08 (95% CI: 0.03 to 0.23)

**Relative effect**

- 0.13 (95% CI: 0.04 to 0.41)
- 0.03 (95% CI: 0.01 to 0.11)
Results

• We successfully used the GRADE approach to assess the quality of evidence for observational studies
• GRADE evidence profiles for most PICO questions to inform the WHO essential medicine list and the WHO committee that prepares guidelines for the pharmacological management of influenza
• WHO guideline panel to use the evidence profiles for decision making
• Very low to low quality evidence for four major pharmacological interventions
• However, this evidence of equal or higher quality compared to that of RCTs for some of the interventions and outcomes
• Evidence for harms!

Results/Discussion

• Large Team – work completed in 6 months
  – 10 PICOs
  – Complete evidence profiles
• Challenges with using existing risk of bias tool
• Indirect comparison and internally controlled studies
  – Should we use these studies?
• Publication bias?
• Upgrade mortality: adj OR 0.23 (0.13 to 0.43)
  – 3 studies, fairly narrow CI
• Easy to use information for guideline panels
Discussion

• The quality of evidence from observational studies may be equivalent to that of RCTs
• Do we need rating of quality within categories
  – Within low quality
  – Probably under certain circumstances
• What if body of evidence inconsistent from RCTs and observational studies but of perceived similar quality
  – Further downgrade for inconsistency? E.g. adverse events
    • RCT Odds Ratio: odds ratio 1.79 (1.10 to 2.93) – nausea
    • Obs. Adj Rate ratio: 0.76 (0.7 to 0.81) – low quality
  – Further upgrade for consistency? E.g. Rx complications:
    • RCT Odds Ratio 0.55 (0.22 – 1.35)
    • Obs adj OR 0.58 (0.31 to 1.1); OR 0.45 (0.25 to 0.81)
Agenda

• Introduction to GRADE
  – Presentation (20 min)
• Small group work
  – Evidence profile (5 min)
  – PICO (5 min)
  – Complete domain and subdomain for each question (20 min)
  – Formulate recommendation (5 min)
  – Feedback (10 min)

GRADE Working Group meeting

Date & Time: August 30, 2011 at 13:00h (1:00 pm) to 17:00 (5:00 pm)
Location: Room 6

www.gradeworkinggroup.org
Content

GRADE – 20 minute overview
• Quality of evidence
• Going from evidence to recommendations

GRADE Uptake
- World Health Organization
- Allergic Rhinitis in Asthma Guidelines (ARIA)
- American Thoracic Society
- American College of Physicians (ACP)
- Canadian Task Force for the Preventive Services
- European Respiratory Society
- European Society of Thoracic Surgeons
- British Medical Journal
- Infectious Disease Society of America
- UpToDate®
- National Institutes of Health and Clinical Excellence (NICE)
- Scottish Intercollegiate Guideline Network (SIGN)
- Cochrane Collaboration
- Clinical Evidence
- Agency for Health Care Research and Quality (AHRQ)
- Partner of GIN
- Over 60 major organizations (over 250 members)
Guideline development Process (for WHO)

Key issues
1. Guidelines for guidelines
2. Priority setting
3. Group composition and consultation process
4. Managing conflicts of interest
5. Group processes
6. Determining which outcomes are important
7. Deciding which evidence to consider
8. Synthesis and interpretation of evidence
9. Grading evidence and recommendations
10. Integrating values and consumer involvement
11. Incorporating considerations of cost-effectiveness, affordability, and acceptability
12. Adaptation, applicability and transferability
13. Reporting guidelines
14. Disseminating and implementing guidelines
15. Evaluation

Healthcare problem

“Healthy people”
“Herd immunity”
“Long term perspective”
“Few RCTs”
“Lots of other things”

recommendation
Hierarchy of evidence based on quality

STUDY DESIGN
- Randomized Controlled Trials
- Cohort Studies and Case Control Studies
- Case Reports and Case Series, Non-systematic observations
- Expert Opinion

BIAS

Relative risk reduction:
...> 99.9 % (1/100,000)
U.S. Parachute Association reported 821 injuries and 18 deaths out of 2.2 million jumps in 2007
Simple hierarchies are (too) simplistic

STUDY DESIGN

- Randomized Controlled Trials
- Cohort Studies and Case Control Studies
- Case Reports and Case Series, Non-systematic observations
  Expert Opinion

BIAS

Expert Opinion

Schünemann & Bone, 2003

GRADE: recommendations & quality of (a body of) evidence

Clear separation, but judgments required:

1) Recommendation: 2 grades – conditional (aka weak) or strong (for or against an action)?
   - Balance of benefits and downsides, values and preferences, resource use and quality of evidence

2) 4 categories of quality of evidence:
   - @@@@ (High), @@@@@(Moderate), @@@@ (Low), @@@@@@ (Very low)?
   - methodological quality of evidence
   - likelihood of bias related to recommendation
   - by outcome and across outcomes

*www.GradeWorking-Group.org
Meta-analyses of several critical and important outcomes (one PICO)

**Mortality** (critical)  
High (⊕⊕⊕⊕)

**Myo. Infarct.** (critical)  
Moderate (⊕⊕⊕⊙)  
Due to imprecision

**Nausea** (important)  
Low (⊕⊕⊙⊙)  
Due to imprecision and risk of bias

**SAE** (critical)  
High (⊕⊕⊕⊕)

Overall Quality of Evidence:  
Moderate (⊕⊕⊕⊙)  
Based on critical outcomes

---

Meta-analyses of several critical outcomes (one PICO)

**Mortality**  
High (⊕⊕⊕⊕)

**Dis. Specific QoL**  
Moderate (⊕⊕⊕⊙)  
Due to imprecision

**Stroke**  
High (⊕⊕⊕⊕)

**SAE**  
High (⊕⊕⊕⊕)

Overall Quality of Evidence:  
High (⊕⊕⊕⊕)

Threshold of acceptable harm for strong recommendation based on sure benefit in mortality and stroke
Meta-analyses of several critical outcomes (one PICO)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Relative Risk</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mortality</td>
<td>High</td>
<td>⊗⊗⊗⊗</td>
</tr>
<tr>
<td>Dis. Specific QoL</td>
<td>High</td>
<td>⊗⊗⊗⊗</td>
</tr>
<tr>
<td>Stroke</td>
<td>Moderate</td>
<td>⊗⊗⊗ ⊗</td>
</tr>
<tr>
<td>SAE</td>
<td>High</td>
<td>⊗⊗⊗⊗</td>
</tr>
</tbody>
</table>

Overall Quality of Evidence: Moderate ⊗⊗⊗⊗

GRADE evidence profile
Content

• Quality of evidence
• Going from evidence to recommendations

Strength of recommendation

“The strength of a recommendation reflects the extent to which we can, across the range of patients for whom the recommendations are intended, be confident that desirable effects of a management strategy outweigh undesirable effects.”

• Strong or conditional/weak
## Determinants of the strength of recommendation

### Factors that can strengthen a recommendation

<table>
<thead>
<tr>
<th>Factors that can strengthen a recommendation</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality of the evidence</td>
<td>The higher the quality of evidence, the more likely is a strong recommendation.</td>
</tr>
<tr>
<td>Balance between desirable and undesirable effects</td>
<td>The larger the difference between the desirable and undesirable consequences, the more likely a strong recommendation warranted. The smaller the net benefit and the lower certainty for that benefit, the more likely weak recommendation warranted.</td>
</tr>
<tr>
<td>Values and preferences</td>
<td>The greater the variability in values and preferences, or uncertainty in values and preferences, the more likely weak recommendation warranted.</td>
</tr>
<tr>
<td>Costs (resource allocation)</td>
<td>The higher the costs of an intervention – that is, the more resources consumed – the less likely is a strong recommendation warranted</td>
</tr>
</tbody>
</table>

### Factors that can weaken the strength of a recommendation

<table>
<thead>
<tr>
<th>Factors that can weaken the strength of a recommendation.</th>
<th>Decision</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower quality evidence</td>
<td>□ Yes</td>
<td>□ No</td>
</tr>
<tr>
<td>Uncertainty about the balance of benefits versus harms and burdens</td>
<td>□ Yes</td>
<td>□ No</td>
</tr>
<tr>
<td>Uncertainty or differences in values</td>
<td>□ Yes</td>
<td>□ No</td>
</tr>
<tr>
<td>Uncertainty about whether the net benefits are worth the costs</td>
<td>□ Yes</td>
<td>□ No</td>
</tr>
</tbody>
</table>

**Table. Decisions about the strength of a recommendation**

Frequent “yes” answers will increase the likelihood of a weak recommendation.
**Recommendation:** In patients with HIV and drug resistant TB requiring second line drugs, the expert panel recommends/suggests to (not) administer ART (? recommendation, ? quality evidence).

**Population:** HIV positive individuals with drug resistant TB requiring second line drugs

**Intervention:** ART use during TB treatment vs ART non-use

<table>
<thead>
<tr>
<th>Factor</th>
<th>Decision</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>High or moderate quality evidence (is there high quality evidence?)</td>
<td>☐ Yes ☐ No</td>
<td>There is limited evidence from published studies to evaluate ART use in HIV-TB coinfected patients receiving second line drugs for XDR-TB and MDR-TB. However, using IPD from longitudinal cohort studies, we found moderate quality evidence from observational studies that there</td>
</tr>
<tr>
<td>Certainty about the balance of benefits versus harms and burdens (is there certainty?)</td>
<td>☐ Yes ☐ No</td>
<td>Although there is some uncertainty about cure, there is a significant decrease in hazards ratio for death even after controlling for initial CD4 count</td>
</tr>
<tr>
<td>Certainty or similarity in values (is there certainty?)</td>
<td>☐ Yes ☐ No</td>
<td>Little uncertainty regarding the outcomes of cure and survival. Significant uncertainty regarding effects of ART on other outcomes, including adverse events, default, time to smear and culture conversion and timing of ART initiation.</td>
</tr>
<tr>
<td>Resource implications (are the resources consumed worth the expected benefit)</td>
<td>☐ Yes ☐ No</td>
<td>More resources required for concomitant ART use</td>
</tr>
</tbody>
</table>

| Overall strength of recommendation | Strong or conditional |

---

**Balancing desirable and undesirable consequences**

- **Conditional**
  - ↑ herd immunity
  - ↓ Morbidity
  - ↑ QoL
  - ↓ Death
- **Strong For**
- **Against**
  - ↑ Resources
  - ↑ Allergic reactions
  - ↑ Local skin reactions
  - ↓ Resources
  - ↓ Allergic reactions
  - ↓ Local skin reactions
Balancing desirable and undesirable consequences

- For:
  - Conditional
  - Strong
  - ↓ herd immunity
  - ↓ Morbidity
  - ↓ QoL
  - ↓ Death

- Against:
  - Conditional
  - Strong
  - ↑ Resources
  - ↑ Nausea
  - ↑ Allergic reactions
  - ↑ Local skin reactions

05.10.2011
Balancing desirable and undesirable consequences

Conditional

Strong

For

Against

↑ herd immunity
↑ Resources
↑ Allergic reactions
↑ Local skin reactions

↑ QoL
↑ Morbidity
↑ Nausea
↑ Local skin reactions

↓ Death
Systematic review

Guideline development

Formulate recommendations:
- For or against (direction)
- Strong or weak/conditional (strength)

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

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

GRADE Grid
Implications of a strong recommendation

- Patients: Most people in this 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/conditional recommendation

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

- Practice guidelines should be based on the **best available** evidence to be evidence based
- GRADE combines what is known in health research methodology and provides a structured approach to improve communication
- Criteria for evidence assessment across questions and outcomes
- Criteria for moving from evidence to recommendations
- **Systematic**
  - four categories of quality of evidence
  - two grades for strength of recommendations
- Transparency in decision making and judgments is key