Aims of course/workshop

Despite the growing evidence in the field of lower urinary tract symptoms, the development and interpretation of guidelines remains difficult. This workshop aims to provide ICS members (both guideline-developers and users) with important background knowledge to enhance the quality of future guidelines.

Within the allotted time, we will focus on the following aspects:
- GRADE methodology and systematic reviews & meta-analyses
- What is a risk factor?
- Interpretation of odds ratios for common conditions.
- Statistical significance vs. clinical relevance for treatment outcomes?
- The impact of the setting from which evidence arises

Learning Objectives

After this workshop participants should be able to:
1. To know how to interpret odds ratios for common conditions.
2. To know the difference between statistical significance and clinically relevant outcomes.
3. To know about the background of the GRADE methodology and how this is applied to modern guidelines.

Learning Outcomes

After the course, the student will be able to:
- Know the difference between associated factors and true risk factors;
- Interpret odds ratios for common conditions;
- Compare odds ratios to relative risks (or rate ratios);
- Make the difference between statistical significance and clinical relevance of outcomes;
- Estimate the absolute risk difference based on relative risk reductions and prevalence rates;
- Interpret findings that result from the GRADE methodology.

Target Audience

All delegates

Advanced/Basic

Basic

Conditions for learning

This will be an interactive workshop in which participants are encouraged to have an active role. Speakers will invite participants to ask questions and respond to the presentations.

Suggested Learning before workshop attendance

http://www.gradeworkinggroup.org/#pub

Website with synopsis for:
- Explanation about The GRADE working group;
- Why rate the certainty in the evidence and strength of recommendations;
- Criteria for applying or using GRADE
**Marco Blanker**

Will discuss the qualifications of risk factors. Many patient characteristics are mentioned as risk factors, even from studies in which no causal associations can be distinguished. What are the requisites for a characteristic to become a “true” risk factor? The association between lower urinary tract symptoms and cardiovascular disease will illustrate this topic, by means of discussion of the (in)ability to define risk factors based on cross sectional studies.

**Take home message:** A risk factor is any attribute, characteristic or exposure of an individual that increases the likelihood of developing a disease or injury (WHO definition). Therefore, longitudinal data are required to find risk factors for diseases; from cross sectional studies, at most characteristics can be defined as ‘associated to’ some disease.

---

**Kari Tikkinen & Rufus Cartwright**

Will compare statistical considerations and patient-importance. What do p-values tell us about the clinical relevance of a described risk difference, or risk reduction? Relative risk reductions can result in large differences in absolute risk reductions, depending on the baseline risk of patients. Ultimately, patients are interested in absolute risk (reductions), and physicians should also be. The topic is illustrated with clinical scenarios, including examples from cancer screening and pharmacological prophylaxis. Epidemiological aspects covered in this part include the interpretation of a p-value, relative risk reduction, absolute risk reduction, risk difference, number needed to treat (NNT).

**Take home message:** When considering treatment, patients are interested in their absolute risk reduction, which depend on their baseline risk; for a proper estimation of an absolute risk reduction, both baseline risk and relative risk reduction are needed.

---

**Ilse Hofmeester**

Will elaborate on the interpretation of odds ratios for common conditions. Often, results from epidemiological studies present large odds ratios (ORs), or at least large ORs get much attention. Many physicians regard such high ORs as relevant for their patients. As a consequence, advises may enter guidelines, but is that always relevant? From what kind of study were the ORs derived? How should ORs be interpreted for different conditions with different prevalence? Ilse Hofmeester will take the association between nocturia and nocturnal polyuria as an example.

**Take home message:** for the sound interpretation of odds ratios, information about the prevalence of the disease/outcome is needed; only for conditions with low prevalence, odds ratios may be interpreted as relative risks.

---

**Rufus Cartwright & Kari Tikkinen**

Many systematic reviews fail to adequately assess the quality of the evidence they synthesize, and many clinical guidelines lack transparency about their methods for deriving recommendations from that evidence. This talk will apply basic principles of clinical epidemiology to assessment of the quality of evidence, and explain the main tenets of the GRADE methodology, as the cornerstone of modern guideline development.

**Take home message:** GRADE provides a systematic way to assess both the quality of evidence (that is, certainty in estimates), and interpret the size of a pooled effect based on that evidence. The GRADE approach separately considers the impact of bias from design factors, inconsistency in results, indirectness, imprecision, and publication bias. GRADE allows guideline authors to reach “strong” or “weak” recommendations, reflecting the extent to which we can be confident that desirable effects of an intervention outweigh the undesirable effects, and the extent to which that balance will apply for most patients, or vary with patients’ own values and preferences.
Marco H. Blanker

Affiliations to disclose†:
University of Groningen, University Medical Center Groningen, Department of General practice, Groningen, The Netherlands

Funding for speaker to attend:
- Self-funded
- Institution (non-industry) funded
- Sponsored by:

How to build an evidence-based guideline
important epidemiological principles

ICS Annual meeting 2016
Kari Tikkinen
Ilse Hofmeester
Rufus Cartwright
Marco Blanker

Schedule

08:30-08:40 Introduction
- Marco Blanker
- Ilse Hofmeester
- Rufus Cartwright
- Kari Tikkinen
- Marco Blanker

General introduction

Most physicians have difficulties in interpreting effect sizes 1

This may hamper
- sound interpretation of literature
- sound interpretation of guidelines
- sound development of guidelines

1. Johnston et al. CMAJ 2015

Guidelines intended for patients with symptom / disease, e.g. incontinence

Guideline developers AND users need to be aware of pitfalls when interpreting guidelines

We will address some (certainly not all) pitfalls

At 10:00 you will be able to:

Interpret and distinguish different outcome measures for associations, especially Odds Ratios for common conditions

Discuss the differences between statistical significance and clinical relevance of treatment outcomes

Discuss different aspects of risk factors

Tell others about the GRADE methodology
Faculty

Kari Tikkinen, MD PhD, adjunct professor of clinical epidemiology & urology resident
Ilse Hofmeester, MD, epidemiologist & urology resident
Rufus Cartwright, MD PhD, urogynaecologist
Marco Blanker, MD PhD, general practitioner & epidemiologist

Who are you?

Personal introduction impossible, but please rise if you are a:

- nurse
- urologist
- researcher
- GP
- (pelvic) physiotherapist
- (uro)gynaecologist
- other:...

Who are you?

How do you rate your epidemiological knowledge/skills?

(please provide honest answer....)

Less than average

Average

Better than average

(What’s average?)

Who are you?

Your input is more than welcome in this workshop

The interpretation of odds ratios for common conditions

Ilse Hofmeester

ICS Annual meeting 2016 – workshop

How to build an evidence-based guideline

Important epidemiological principles
What’s a risk factor?

Marco H. Blanker
ICS Annual meeting 2016 – workshop
How to build an evidence-based guideline
Important epidemiological principles

What’s a risk factor?

True or false?
Smoking is a risk factor for lung cancer
Vaginal delivery is a risk factor for Pelvic Organ Prolapse
Smoking is a risk factor for bladder cancer
Smoking is a risk factor for cardiovascular disease (CVD)

Lower urinary tract symptoms are a risk factor for CVD

What’s a risk factor?

Lower urinary tract symptoms (LUTS) – an independent risk factor for cardiovascular disease (CVD)
G. Jackson, M.G. Kirby, R. Rosen, BJU Int 2015

Editorial comment on

Increase of Framingham cardiovascular disease risk score is associated with severity of lower urinary tract symptoms
Giuseppe G. Russo, Tommaso Costi, Salvatore Pizzocara, Eugenio Fragola, Vincenzo Favilla, Giorgio G. Ralli, Dieter Lillic, Sandra La Vignera, Rostislav A. Andoroli, Aldo E. Cologno, Sebastiano Cimino and Giuseppe Morgo
Department of Urology, and *Department of Medical and Molecular Sciences, Section of Endocrinology, Andrology and Intermural Medicine, University of Catania, Catania, Italy

What’s a risk factor?

What’s a risk factor?

Lower urinary tract symptoms (LUTS) – an independent risk factor for cardiovascular disease (CVD)
G. Jackson, M.G. Kirby, R. Rosen, BJU Int 2015

Developing disease (in the future)
Causal association between risk factor & disease
True association (not explained by other variables)

Ask yourself “why would LUTS cause CVD?”

World Health Organization:
A risk factor is any attribute, characteristic or exposure of an individual that increases the likelihood of developing a disease or injury.
What’s a risk factor?

Increase of Framingham cardiovascular disease risk score is associated with severity of lower urinary tract symptoms

Crossectional study
336 Consecutive patients with BPH-related LUTS
Assessment of Framingham Heart Risk score
(based on age, HDL, total cholesterol level, systolic blood pressure, anti-hypertensive medication use, diabetes and current smoking status)

Risk of having moderate/severe LUTS for high CVD-risk group: OR 5.9 (age-adjusted)

Comments:
Crossectional study
No CVD but ‘risk-for CVD score’

No firm conclusion can be drawn

What’s a risk factor?

Rosso-study no evidence of LUTS as risk factor for CVD

More information is needed

Objective: To evaluate whether LUTS severity can be considered as a significant risk factor of major adverse cardiac events (MACE) in the male population.

Authors included all cross-sectional & longitudinal trials enrolling men, comparing prevalence/incidence of MACE in men with moderate to severe LUTS and those without LUTS or with mild LUTS.

Objective: To evaluate whether LUTS severity can be considered as a significant risk factor of major adverse cardiac events (MACE) in the male population.

Authors included all cross-sectional & longitudinal trials enrolling men, comparing prevalence/incidence of MACE in men with moderate to severe LUTS and those without LUTS or with mild LUTS.
What’s a risk factor?

Male Lower Urinary Tract Symptoms and Cardiovascular Events: A Systematic Review and Meta-analysis

5 studies with 25,494 patients and 2,291 MACE.

Authors included all cross-sectional and longitudinal trials enrolling men, comparing prevalence/incidence of MACE in men with moderate to severe LUTS and those without LUTS or with mild LUTS.

Presence of moderate to severe LUTS associated with increased incidence of MACE compared with the rest of the sample (OR: 1.68; 1.13–2.50)

BUT:

No adjustment for confounders
No exclusion of patients with MACE/CVD at baseline

What’s a risk factor?

Do lower urinary tract symptoms predict cardiovascular diseases in older men? A systematic review and meta-analysis

Iris J. Boomsma1 · Maarten J. H. Nolkaamp3 · Bondeson J. · Kolonen1 · Rijn J. M. Nijman1 · Wouter K. van der Heide2 · Marco H. Blokker3

5 studies with 6,027 (LUTS) & 18,993 (no LUTS) men
All without CVD at baseline
Follow-up period 5 - 17 years
2,780 CVD events

No clear association between CVD and LUTS [pooled effect size: hazard ratio 1.09 (95 % CI 0.90–1.31)].

What’s a risk factor?

Term might lead to confusion, as definitions differ
Most often used in epidemiology:
- particular outcome will occur after particular exposure
- an exposure that is statistically related to an outcome

Risk factors may be immutable or modifiable

Uncertainty about what strength of association is needed

Related terms:
Risk marker: attribute/exposure associated with increased probability of outcome, but not necessarily a causal factor
Determinant: attribute/exposure that increases probability of outcome
Modifiable risk factor: a determinant that can be modified by intervention, thereby reducing the probability of disease
What’s a risk factor?

In case of LUTS & CVD
- In those with CVD: LUTS seems to be associated
  BUT: CVD history itself is major predictor of new CVD
- In those without CVD: no association

Most probably: LUTS and CVD share common risk factors
If so, LUTS might be a risk marker

Thank you for your attention
ICS Annual meeting 2016 – workshop

How to grade the quality of evidence?

Kari Tikkinen & Rufus Cartwright
ICS Annual meeting 2016 – workshop
How to build an evidence-based guideline
important epidemiological principles
The interpretation of odds ratios for common conditions

Ilse Hofmeester
Urologist in training – epidemiologist
The Netherlands

Which risk estimates do you know?

Absolute risk
Relative risk
Odds ratio
Risk ratio
Hazard ratio

Example if (mis)use

72% of alpha-blocker users experience improvement of symptoms
61% of placebo users experience improvement of symptoms

Use of 5-alpha reductase inhibitors reduces the risk of acute urinary retention (AUR) by 50%
Absolute risk reduction of AUR after 5 years: 2.5%
Nocturia & nocturnal polyuria

In-depth example of interpretation of OR

Risk of having nocturnal polyuria based on nocturia status
Results of meta-analyses (Hofmeester, J Urol 2014)

What’s your interpretation of this OR?

People with nocturia have nocturnal polyuria 5 times more often than those without nocturia

Don’t know

Interpretation of odds ratios

Back to basics!

Relative risk estimates are based on absolute risk estimates in 2 or more groups

Absolute risk estimates important for interpretation

Interpretation of odds ratios

Prevalence of disease – influence on OR

<table>
<thead>
<tr>
<th>Prevalence NP 5%</th>
<th>NP+</th>
<th>NP-</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nocturia +</td>
<td>20</td>
<td>230</td>
<td>250</td>
</tr>
<tr>
<td>Nocturia -</td>
<td>30</td>
<td>720</td>
<td>750</td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
<td>750</td>
<td>1000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Prevalence NP 25%</th>
<th>NP+</th>
<th>NP-</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nocturia +</td>
<td>100</td>
<td>150</td>
<td>250</td>
</tr>
<tr>
<td>Nocturia -</td>
<td>150</td>
<td>600</td>
<td>750</td>
</tr>
<tr>
<td>Total</td>
<td>250</td>
<td>750</td>
<td>1000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Prevalence NP 60%</th>
<th>NP+</th>
<th>NP-</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nocturia +</td>
<td>160</td>
<td>210</td>
<td>370</td>
</tr>
<tr>
<td>Nocturia -</td>
<td>140</td>
<td>590</td>
<td>730</td>
</tr>
<tr>
<td>Total</td>
<td>300</td>
<td>800</td>
<td>1100</td>
</tr>
</tbody>
</table>

Prevalence

<table>
<thead>
<tr>
<th>Prevalence</th>
<th>5%</th>
<th>25%</th>
<th>60%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Odds ratio</td>
<td>2.09</td>
<td>2.67</td>
<td>26.00</td>
</tr>
<tr>
<td>Relative risk</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
</tr>
</tbody>
</table>
Interpretation of odds ratio’s

Association between OR and RR depends on prevalence of condition.

Odds ratio’s look like relative risks, but only if prevalence of condition is small.

ORs may be interpreted as RR under the Rare disease assumption.

Nocturia & nocturnal polyuria

What’s your interpretation of this OR?

People with nocturia have nocturnal polyuria 5 times more often than those without nocturia.

Don’t know.

Prevalence of nocturnal polyuria 63.9% (well above 10%)

Relative risk: 1.41
In summary

Relative risk estimates most often used

Absolute risk estimates are important for interpretation

For proper interpretation of odds ratio's, information on prevalence of condition is vital
How to grade quality of evidence

Rufus Cartwright (@roofus)
Department of Epidemiology and Biostatistics
Imperial College London, UK

Kari Tikkinen (@KariTikkinen)
Departments of Urology and Public Health,
Helsinki University Hospital, Academy of Finland and University of Helsinki, Finland

W16: How to Build an Evidence-Based Guideline – Important Epidemiological Principles
46th Annual Scientific Meeting of the International Continence Society
September 14, 2016 – Tokyo, Japan
Guidelines and clinicians

• increasingly, clinicians rely on formal guidelines

• strong recommendations
  – strong methods
  – large precise effect
  – few down sides of therapy

• weak recommendations
  – weak methods
  – imprecise estimate
  – small effect
  – substantial down sides
Proliferation of systems 😞

Common international grading 😊

- **GRADE** (*Grades of recommendation, assessment, development and evaluation*)

- International group
  - Australian NMRC, SIGN, USPSTF, WHO, NICE, Oxford CEBM, CDC, CC

- ~35 meetings over last 14 years
  - (~10 – 80 attendants – now 300 contributors)
80+ Organizations
What are we grading?

two components

strength of recommendation: strong and weak
Grading system – for what?

• interventions
  – management strategy 1 versus 2

• what grade is **not** about
  – individual studies (body of evidence)
What GRADE is not primarily about

• diagnostic accuracy questions
  – in patients with a sore leg, what is the accuracy of a blood test (D-Dimer) in sorting out whether a deep venous thrombosis is the cause of the pain

• prognosis

• what it is about: diagnostic impact
  – are patients better off (improved outcomes) when doctors use the d-dimer test
Determinants of quality

- RCTs start high
- Observational studies start low
- What can lower confidence?
What can lower confidence?

• clue 1
  – lack of blinding in an RCT

• clue 2
  – RCT loses $\frac{1}{2}$ patients to follow-up

• high risk of bias in RCTs lowers confidence
Clue: Have a look at the forest plot below – Infections with short and long term antibiotics after open fractures

<table>
<thead>
<tr>
<th>Study</th>
<th>3-5 days</th>
<th>1 day</th>
<th>Relative Risk (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dellinger risk 1988</td>
<td>25/172</td>
<td>17/91</td>
<td>0.78 (0.44, 1.36)</td>
</tr>
<tr>
<td>Dellinger duration 1988</td>
<td>21/169</td>
<td>10/79</td>
<td>0.98 (0.49, 1.98)</td>
</tr>
<tr>
<td>Carsenti-Etesse 1999</td>
<td>24/300</td>
<td>21/316</td>
<td>1.20 (0.68, 2.12)</td>
</tr>
</tbody>
</table>

Random Effects Estimate (p=0.86), I²=0%

0.97 (0.69, 1.37)

Any concerns?
Another reason for rating down: imprecision
Clue: Have a look at the forest plot below

Aspirin in primary prophylaxis

<table>
<thead>
<tr>
<th>1.2.2 Myocardial infarction</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>BDT</td>
<td>169 3429</td>
</tr>
<tr>
<td>PHS</td>
<td>139 11037</td>
</tr>
<tr>
<td>HOT</td>
<td>82 9399</td>
</tr>
<tr>
<td>TPT</td>
<td>69 1268</td>
</tr>
<tr>
<td>PPP</td>
<td>19 2226</td>
</tr>
<tr>
<td>WHS</td>
<td>198 19934</td>
</tr>
<tr>
<td>JPAD</td>
<td>12 1262</td>
</tr>
<tr>
<td>POPADAD</td>
<td>76 638</td>
</tr>
<tr>
<td>AAA</td>
<td>90 1675</td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>50868</td>
</tr>
<tr>
<td>Total events</td>
<td>854</td>
</tr>
</tbody>
</table>

Any concerns?

Another reason for rating down: inconsistency
1.2.2 Myocardial infarction

<table>
<thead>
<tr>
<th>Study</th>
<th>Events</th>
<th>Events Exact</th>
<th>Events Total</th>
<th>Odds Ratio</th>
<th>95% CI</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDT</td>
<td>169</td>
<td>3429</td>
<td>1710</td>
<td>0.96</td>
<td>[0.75, 1.23]</td>
<td>1988</td>
</tr>
<tr>
<td>PHS</td>
<td>139</td>
<td>11037</td>
<td>11034</td>
<td>0.58</td>
<td>[0.47, 0.72]</td>
<td>1989</td>
</tr>
<tr>
<td>HOT</td>
<td>82</td>
<td>9399</td>
<td>9391</td>
<td>0.65</td>
<td>[0.49, 0.85]</td>
<td>1998</td>
</tr>
<tr>
<td>TPT</td>
<td>69</td>
<td>1268</td>
<td>1272</td>
<td>0.71</td>
<td>[0.52, 0.95]</td>
<td>1998</td>
</tr>
<tr>
<td>PPP</td>
<td>19</td>
<td>2226</td>
<td>2269</td>
<td>0.69</td>
<td>[0.39, 1.23]</td>
<td>2001</td>
</tr>
<tr>
<td>WHS</td>
<td>198</td>
<td>19934</td>
<td>19942</td>
<td>1.03</td>
<td>[0.84, 1.25]</td>
<td>2005</td>
</tr>
<tr>
<td>JPAD</td>
<td>12</td>
<td>1262</td>
<td>1277</td>
<td>0.87</td>
<td>[0.40, 1.87]</td>
<td>2008</td>
</tr>
<tr>
<td>POPADAD</td>
<td>76</td>
<td>638</td>
<td>638</td>
<td>1.10</td>
<td>[0.81, 1.50]</td>
<td>2008</td>
</tr>
<tr>
<td>AAA</td>
<td>90</td>
<td>1675</td>
<td>1675</td>
<td>1.05</td>
<td>[0.78, 1.40]</td>
<td>2010</td>
</tr>
<tr>
<td>Subtotal (95% CI)</td>
<td>50868</td>
<td>49208</td>
<td>854</td>
<td>0.83</td>
<td>[0.69, 1.00]</td>
<td></td>
</tr>
<tr>
<td>Total events</td>
<td>854</td>
<td>942</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Heterogeneity: \( \tau^2 = 0.05; \ Chi^2 = 27.51, \ df = 8 \ (P = 0.0006); \ I^2 = 71\% \\
Test for overall effect: \( Z = 1.99 \ (P = 0.05) \)
More reasons to lose confidence

- RCTs show less UI after new intervention
  - patients in RCTs 40 to 70
  - your patient 90

- are you confident?

- indirectness of population
  - older, sicker or more co-morbidity
More reasons to lose confidence

- operation for lap mesh prolapse repair
- technically challenging
  - frequent complications
- RCTs: lap surgery decreases recurrence
  - only top surgeons participate in the RCTs
- are you confident?
- indirectness of intervention
Directness

interested in A versus B
available data A vs C, B vs C
Another reason to lose confidence

• some trials never get published

• “negative” studies more likely

• biased sample of studies
  – overestimates of treatment effect
Positive results more likely to get published

<table>
<thead>
<tr>
<th>Ethics committee</th>
<th>No of research proposals</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>285</td>
</tr>
<tr>
<td>B</td>
<td>342</td>
</tr>
<tr>
<td>C</td>
<td>172</td>
</tr>
<tr>
<td>D</td>
<td>198</td>
</tr>
<tr>
<td>E</td>
<td>218</td>
</tr>
</tbody>
</table>

**Fig 1** Meta-analysis of five studies examining association of significant results and publication among research proposals submitted to ethics committees. The unadjusted odds ratios were combined by using a fixed effects model.
How to demonstrate?

Funnel plot

Precision of estimate of treatment effect

Magnitude of the effect size
How to demonstrate?

Publication bias

![Graph showing publication bias with favor intervention and favor control on the x-axis, and precision on the y-axis. The graph illustrates the distribution of data points favoring one condition over the other.](image-url)
Determinants of quality

- RCTs start high
- Observational studies start low
- What can lower quality?
- Risk of bias
- Inconsistency
- Indirectness
- Imprecision
- Publication bias
What can raise quality?

- Large magnitude can rate up one level
  - Very large two levels

- Common criteria
  - Everyone used to do badly
  - Almost everyone does well
  - Quick action

- Hip replacement for hip osteoarthritis

- Mechanical ventilation in respiratory failure
Confidence assessment criteria

<table>
<thead>
<tr>
<th>Study Design</th>
<th>Confidence in estimates</th>
<th>Lower if</th>
<th>Higher if</th>
</tr>
</thead>
<tbody>
<tr>
<td>Randomised trial</td>
<td>High</td>
<td>Risk of bias</td>
<td>Large effect</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-1 Serious</td>
<td>+1 Large</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-2 Very serious</td>
<td>+2 Very large</td>
</tr>
<tr>
<td></td>
<td>Moderate</td>
<td>Inconsistency</td>
<td>Dose response</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-1 Serious</td>
<td>+1 Evidence of a gradient</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-2 Very serious</td>
<td></td>
</tr>
<tr>
<td>Observational study</td>
<td>Low</td>
<td>Indirectness</td>
<td>All plausible confounding</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-1 Serious</td>
<td>+1 Would reduce a demonstrated effect</td>
</tr>
<tr>
<td></td>
<td>Very low</td>
<td>-2 Very serious</td>
<td>or</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Imprecision</td>
<td>+1 Would suggest a spurious effect when</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-1 Serious</td>
<td>results show no effect</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-2 Very serious</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Publication bias</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-1 Likely</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>-2 Very likely</td>
<td></td>
</tr>
</tbody>
</table>
Strength of Recommendation

• **strong recommendation**
  – benefits clearly outweigh risks/hassle/cost
  – risk/hassle/cost clearly outweighs benefit

• what can downgrade strength?

• low confidence in estimates

• close balance between up and downsides
Risk/Benefit tradeoff

- **aspirin after myocardial infarction**
  - 25% reduction in relative risk
  - side effects minimal, cost minimal
  - benefit obviously much greater than risk/cost

- **warfarin in low risk atrial fibrillation**
  - warfarin reduces stroke vs ASA by 50%
  - but if risk only 1% per year, ARR 0.5%
  - increased bleeds by 1% per year
Conclusion

• clinicians, policy makers need summaries
  – quality of evidence
  – strength of recommendations

• explicit rules
  – transparent, informative

• GRADE
  – simple, transparent, systematic
  – increasing wide adoption
  – great opportunity for teaching EBHC
Statistical considerations versus patient-importance

Rufus Cartwright (@roofus)
Department of Epidemiology and Biostatistics
Imperial College London, UK

Kari Tikkinen (@KariTikkinen)
Departments of Urology and Public Health,
Helsinki University Hospital, Academy of Finland and University of Helsinki, Finland

W16: How to Build an Evidence-Based Guideline – Important Epidemiological Principles
46th Annual Scientific Meeting of the International Continence Society
September 14, 2016 – Tokyo, Japan
Calibrating Your Enthusiasm

Your flight is cancelled due to bad weather

Your flight will arrive earlier than scheduled due to very good weather and nice tailwind
Interpreting the Evidence

Willingness to fund mammography screening

- *program A* reduces the rate of dying from breast cancer by 33% ($p=0.001$)

- *program B* increases the rate of patients *not* dying from breast cancer from 99.82% to 99.88% ($p=0.001$)

- *program C* means that 1,667 women needed to be screened yearly for 7 years to prevent one death from breast cancer ($p=0.001$)
Breast Cancer Screening

Breast cancer death rates (p=0.001)

- unscreened: 0.18% (18 out of 10,000)
- screened: 0.12% (12 out of 10,000)

Relative risk reduction: \( \frac{0.18\% - 0.12\%}{0.18\%} = 33\% \)

Breast cancer death rates:

- unscreened: 0.18% means 99.82% don’t die
- screened: 0.12% means 99.88% don’t die

Absolute risk reduction: 0.18% - 0.12% = 0.06%

Number needed to screen: \( \frac{100}{0.06} = 1,667 \)
Example: VA hypertension study

Mortality after 5 years of treatment

<table>
<thead>
<tr>
<th></th>
<th>Controls</th>
<th>Treated</th>
<th>RRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>DBP (90 – 104)</td>
<td>0.074</td>
<td>0.059</td>
<td>0.074 - 0.059</td>
</tr>
</tbody>
</table>

DBP, diastolic blood pressure

RRR: Relative Risk Reduction
### Relative risk reduction (RRR)

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>Treatment</th>
<th>RRR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TOD+</strong></td>
<td>0.20</td>
<td>0.16</td>
<td>20%</td>
</tr>
<tr>
<td><strong>TOD-</strong></td>
<td>0.057</td>
<td>0.045</td>
<td>21%</td>
</tr>
</tbody>
</table>

TOD, target organ damage
## Absolute risk reduction (ARR)

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>Treatment</th>
<th>RRR</th>
<th>ARR</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOD+</td>
<td>0.20</td>
<td>0.16</td>
<td>20%</td>
<td>4%</td>
</tr>
<tr>
<td>TOD-</td>
<td>0.057</td>
<td>0.045</td>
<td>21%</td>
<td>1.2%</td>
</tr>
</tbody>
</table>

TOD, target organ damage
<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>Treatment</th>
<th>RRR</th>
<th>ARR</th>
<th>NNT</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOD+</td>
<td>0.20</td>
<td>0.16</td>
<td>20%</td>
<td>4%</td>
<td>25</td>
</tr>
<tr>
<td>TOD-</td>
<td>0.057</td>
<td>0.045</td>
<td>21%</td>
<td>1.2%</td>
<td>83</td>
</tr>
</tbody>
</table>

TOD, target organ damage
Patient with DVT

Completes 6 months prophylaxis

Question: continue or not?

Doctor: continuing reduces risk of recurrence by 33%

- chance unlikely to explain the difference (p=0.001)

What does patient understand?

Is there something missing?
Patient with DVT

Constant Relative Risk With Varying Risk Differences

- Population 1: RR 0.67, RD 10%
- Population 2: RR 0.67, RD 3.3%
- Population 3: RR 0.67, RD 1%
**Patients with atrial fibrillation**

CHADS$_2$: congestive heart failure; hypertension; age >75; diabetes; prior stroke

**Risk of stroke varies**

<table>
<thead>
<tr>
<th>CHADS$_2$</th>
<th>Risk per 1,000 per year</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>1</td>
<td>22</td>
</tr>
<tr>
<td>2</td>
<td>45</td>
</tr>
<tr>
<td>3</td>
<td>96</td>
</tr>
</tbody>
</table>

**Warfarin constant 2/3 relative risk reduction**

<table>
<thead>
<tr>
<th>CHADS$_2$</th>
<th>Risk per 1,000 per year</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>2</td>
<td>40</td>
</tr>
<tr>
<td>3</td>
<td>64</td>
</tr>
</tbody>
</table>
Measures of Relative Effect

- Relative risk
- Relative risk reduction
- Odds ratio
- Relative odds reduction
- Hazard ratio
Small, medium or large?

VTE prophylaxis in 65 year old man, COPD exacerbation, anticipated walking in hall day 3, hospitalization

RRR 50%

Baseline risk 4/1,000

Risk difference 2/1,000 so, NNT 500

Balance in favour of treatment?

VTE, venous thromboembolism
VTE prophylaxis in 65 year old man, disseminated cancer, severe pneumonia, likely bed-bound for at least 3 days

RRR 50%
Baseline risk 100/1,000
Risk difference 50/1,000 so, NNT 20
Balance in favour of treatment?
Summary

Relative estimates: RR, OR, HR

Absolute estimates: RD (ARR), NNT

Ultimately patients interested in absolute risk (reductions)

Patients not interested in p-values or relative estimates

Relative risk reductions constant across patients, absolute risk reductions not

So, to get absolute risk reductions, need baseline risk and relative risk reductions