Eastern Association for the Surgery of Trauma

So You Want to Write a Practice Management Guideline?

January 14, 2014
Waldorf Astoria Naples
Naples, Florida

Accreditation Statement
This activity has been planned and implemented in accordance with the Essential Areas and Policies of the Accreditation Council for Continuing Medical Education through the joint sponsorship of the American College of Surgeons and the Eastern Association for the Surgery of Trauma (EAST). The American College of Surgeons is accredited by the ACCME to provide continuing medical education for physicians.

AMA PRA Category 1 Credits™
The American College of Surgeons designates this live activity for a maximum of 4.25 *AMA PRA Category 1 Credits™**.* Physicians should claim only the credit commensurate with the extent of their participation in the activity.

**This workshop qualifies for Self-Assessment Credit.

.

American College of Surgeons
Division of Education

GRADE: FROM EVIDENCE TO RECOMMENDATIONS

east || 27th Annual Scientific Assembly

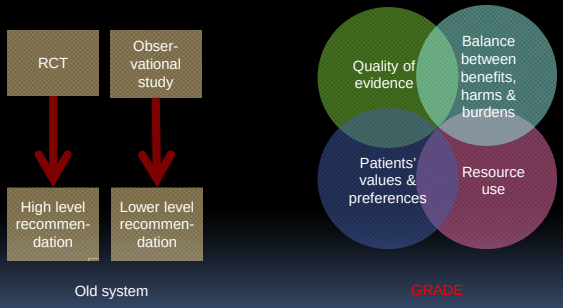
So you want to write a practice management guideline?

Naples, FL, February 8, 2014

Yngve Falck-Ytter, MD, AGAF for the GRADE team
Associate Professor, Case Western Reserve University, Case & VA Medical Center
Chief, Gastroenterology & Hepatology, VA Medical Center, Cleveland

1

From evidence to recommendations



2

Desirable and undesirable effects

- Desirable effects
  - Mortality reduction
  - Improvement in quality of life, fewer hospitalizations/infections
  - Reduction in the burden of treatment
  - Reduced resource expenditure
- Undesirable effects
  - Deleterious impact on morbidity, mortality or quality of life, increased resource expenditure

3

Values and preferences

- Implicit value judgments in recommendations
- Trade-offs: example prevention of VTE in surgery
  - Thrombotic events
    - Deep vein thrombosis, pulmonary embolism
  - Bleeding events
    - Gastrointestinal bleeds, operative site bleeds
  - Inconvenience of injections
- Variability in values and preferences

Case

- 77 y/o patient with atrial fibrillation, mild CHF, HTN, DM and history of stroke (fully recovered)
- Meds: warfarin, antihypertensives, statin, glyburide
- Admitted with nausea/vomiting, then hematemesis; INR 2.5; 1 U blood transfused; EGD: no active bleed, possible Mallory Weiss
- This is his second major bleed since he started warfarin one year ago

CHADS2 score

CHADS2 item	Points
Congestive heart failure	1
Hypertension (systolic >160 mmHg)	1
Age greater than 75 years	1
Diabetes	1
Prior cerebral ischaemia	2

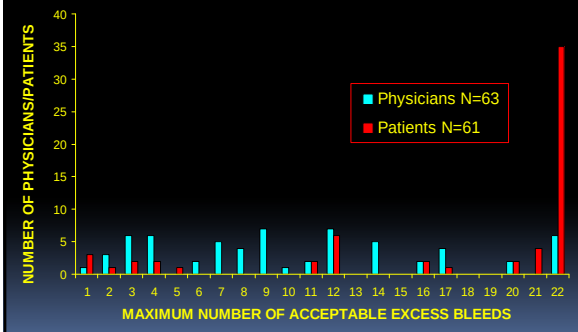
	Number of:		
CHADS2 score	Patients	Strokes	Adjusted annual stroke rate (95%CI)
0	120	2	1.9 (1.2 to 3.0)
1	463	17	2.8 (2.0 to 3.8)
2	523	23	4.0 (3.1 to 5.1)
3	337	25	5.9 (4.6 to 7.3)
4	220	19	8.5 (6.3 to 11.1)
5	65	6	12.5 (8.2 to 17.5)
6	5	2	18.2 (10.5 to 27.4)

Acceptable additional bleeds?

- Study: Patients at high risk for atrial fibrillation and high risk of stroke (h/o CHF/MI); internists and cardiologists
- Warfarin decreases risk at cost of increased GI bleeds
- Without treatment 100 patients will suffer:
  - 12 strokes (six major, six minor), 3 serious GI bleeds in 2 years
- Warfarin would decrease strokes in 100 patients to 4 per 2 years (8 fewer strokes, 4 major, minor)
- How many additional bleeds would you accept in 100 patients over a year, and still be willing to administer/take warfarin?

Slide courtesy of: G. Guyatt, Study: Devereaux et. al., 2001

PHYSICIAN AND PATIENT BLEEDING THRESHOLDS FOR WARFARIN



Slide courtesy: G. Guyatt, Study: Devereaux et. al., 2001

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."

- Understanding values & preferences necessary to trade-off benefits and downsides
- Values and preferences should ideally be informed by systematic reviews, but evidence is often sparse

9

Example recommendation

ACCP AT9 recommendation:

In patients undergoing major orthopedic surgery (e.g., total hip replacement), we suggest the use of LMWH in preference to the other agents.

Patients who place a high value on avoiding bleeding complications and a low value on its inconvenience are likely to choose a compression device (IPCD) over the drug options.

10

4 determinants of the strength of recommendation

Factors that can weaken the strength of a recommendation	Explanation
❑ Lower quality evidence	The higher the quality of evidence, the more likely is a strong recommendation.
❑ Uncertainty about the balance of benefits versus harms and burdens	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 is a weak recommendation warranted.
❑ Uncertainty or differences in patients' values	The greater the variability in values and preferences, or uncertainty in values and preferences, the more likely weak recommendation warranted.
❑ Uncertainty about whether the net benefits are worth the costs	The higher the costs of an intervention – that is, the more resources consumed – the less likely is a strong recommendation warranted.

11

Balance of benefits & harms

Benefits clearly outweighs the downsides	Benefits probably still outweighs the downsides	Harms probably outweighs the benefits	Harms clearly outweighs the benefits
⇩	⇩	⇩	⇩
Strong recommendation for an action	Conditional/weak recommendation for an action	Conditional/weak recommendation against an action	Strong recommendation against an action

The GRADE grid

	Strong	Weak/ conditional	Neutral	Weak/ conditional	Strong
Assessors view of the balance	Desirable **clearly** outweigh undesirable	Desirable **probably** outweigh undesirable	Trade-offs equally balanced or uncertain	Undesirable **probably** outweigh desirable	Undesirable **clearly** outweigh desirable
Recommen-dation	"We recommend doing…"	"We suggest doing…"	None. Research gap!	"We suggest against doing…"	"We recommend against doing…"
PICO 1					
PICO 2					
PICO 3					

Implications of a *strong* recommendation

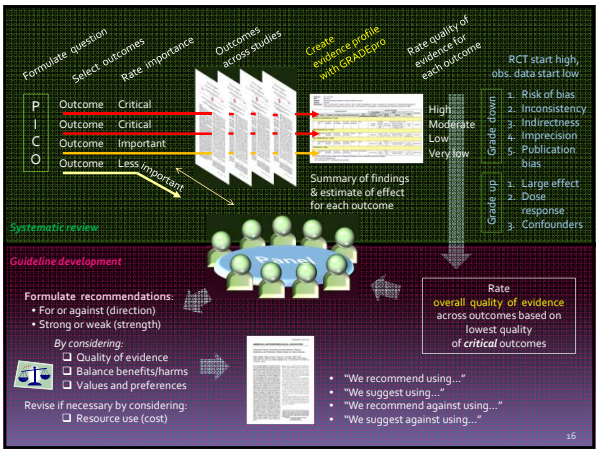
- Population: Most people in this situation would want the recommended course of action and only a small proportion would not
- Health care workers: Most people should receive the recommended course of action
- Policy makers: The recommendation can be adapted as a policy in most situations

14

Implications of a *conditional* recommendation

- Population: The majority of people in this situation would want the recommended course of action, but many would not
- Health care workers: Be prepared to help people 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

15



Additional practical points

- Some recommendations should not be graded
- Make recommendations actionable
- Use of "There is insufficient data…"
- Recommending against an intervention

17

Some recommendations should not be graded

- Motherhood statements:
  - Too vague, not actionable, leave out: "Take a thorough history and physical"
  - Helpful, but should not be graded: "Pregnant women should be offered evidence-based information and support to enable them to make informed decisions regarding their care"
  - Should probably be graded: "Perform the A1C test at least two times a year in patients who are meeting treatment goals "

Make recommendation actionable

- Avoid statement of facts
  - "AZA or 6-MP is effective for maintenance of remission in patients with CD regardless of disease distribution (Grade A)." (AGA 2006)
- Use of "there is insufficient evidence"
  - "The currently available evidence is insufficient to support the use of methotrexate for the induction or maintenance of remission in patients with active UC (Grade B) (AGA 2006)"
  - Often overused
  - Most of the time not helpful for clinicians
  - Instead recommend as part of study
- Keep the actual recommendation concise

19

Recommending against an intervention

- Avoid: "we do not recommend…"
- Instead use: "we recommend against…"
- Always include the comparator in the recommendation ("use X rather than Y…"), unless self explanatory
- Decide whether to recommend for the intervention of against the alternative

20

Avoid vague terms

- "Although there is evidence-based data to support the use of corticosteroids…"
- "Long-term treatment with corticosteroids is undesirable…"
- "Ileal-release preparations of budesonide (Entocort) are indicated for the treatment of patients with…"

21

Summary

- The quality of evidence is not the only factor that drives the strength of recommendations
- Balance between desirable and undesirable effects is equally important
- Keep patient's values & preferences in mind
- GRADE helps to make your judgments transparent

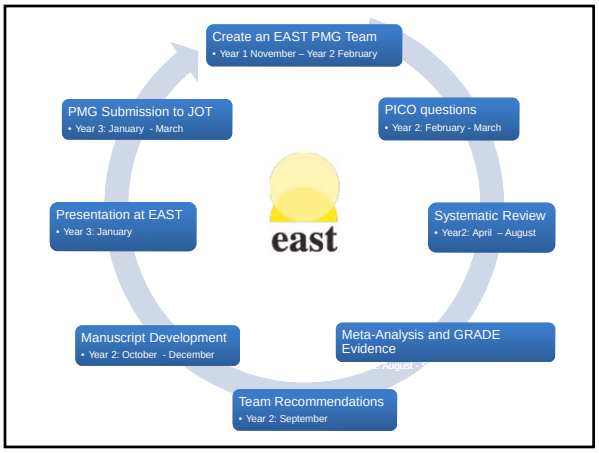
22

Defining a PICO question

John J. Como, MD, MPH, FACS
MetroHealth Medical Center

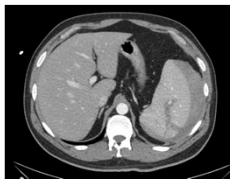
Guidelines

- Thousands of healthcare guidelines now exist
  - 2700 guidelines in the Agency for Healthcare Research and Quality's National Guideline Clearinghouse
  - At least 6800 guidelines in the database of the Guidelines International Network
- GRADE methodology is very commonly used
  - Methodology endorsed by EAST
- One of the first steps in generating guidelines using the GRADE methodology is the formulation of PICO questions



PICO principle

- Finding answers to specific questions in the medical literature is often a challenge
  - Which blunt trauma patients should receive splenic angioembolization?



PICO principle

- The PICO principle
  - Involves dissecting a question into its component parts to facilitate finding its answer
  - Is an essential first step in Evidence Based Medicine
- We want the PICO question to generate an answer which will be in the form of an action statement
- Divides questions into 4 parts

PICO questions - components

- P - Population
- I - Intervention or Indicator
  - Planned Action
- C - Comparator or Control
  - Alternative Action
- O - Outcome

PICO question

- Population
  - Who are the relevant patients?
  - What is the problem?

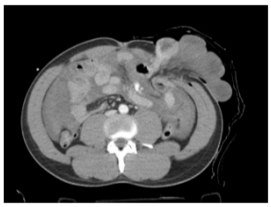
PICO question

- Intervention or Indicator
  - Management strategy
  - Diagnostic test
  - Exposure

- Could be:
  - Surgical procedure
  - Medication
  - Other clinical therapies
  - Diagnostic test
  - Screening for disease
  - Lifestyle changes
    - Diet
    - Exercise
  - Exposure
  - Social activities

PICO question

- Comparator or Control



- Alternative
  - Treatment strategy
  - Test
  - Exposure

PICO question

- Outcome
  - What are the patient relevant consequences of the intervention?

PICO question

P	Patient Population Problem	Describe your patient group?
I	Intervention	Which main intervention, prognostic factor or exposure is being considered?
C	Comparison	What is the main alternative treatment to compare to I (intervention)?
O	Outcome	What outcome is relevant to you and your patient?

PICO(TS) question

- Time frame
  - Optional
- Setting
  - Optional

PICO(TS) questions

Table 7. The PICOTS Format for Asking Answerable Clinical Questions

Component	Comment	Diagnosis Question Example[a]	Treatment/Harm Question Example[b]	Prognosis Question Example[c]
Population	Patient, population, or problem to which the question applies	Adults with acute upper respiratory infection	Adults with acute bacterial sinusitis	Adults with acute bacterial sinusitis
Intervention	Service, planned action, prognostic factor, or cause of interest	History, physical examination, or diagnostic test	Antibiotic therapy for 7 to 10 days	Prognostic factors, including age, illness severity, comordid conditions (eg, allergic rhinitis)
Comparator (optional)	When applicable, an alternative intervention or comparison	None	Placebo or no therapy	None
Outcome(s)	Measurements to determine the impact of the intervention and comparator	Distinguish bacterial vs viral sinusitis	Clinical improvement of presenting signs and symptoms; harms and adverse events	Identify patients who are likely to benefit most from antibiotic therapy
Time frame (optional)	Timing or time frame of interest	Within the first 3 weeks of illness	During and after treatment	During and after treatment
Setting (optional)	Clinical care or other setting of interest	Any setting	Any setting	Any setting

Framing the question

PICO format

Question:

- In stable blunt splenic trauma patients with contrast blush noted on CT of the abdomen (P)
- Should angioembolization be performed (I)
- Versus observation without embolization (C)
- To improve splenic salvage? (O)

Framing the question

- If formulated properly, the PICO question will generate an answer in the form of an action statement

Key action statements

- An ideal key action statement or recommendation describes:
  - WHO
  - *May*, *should*, or *must* do WHAT
  - To WHOM
  - UNDER WHAT CIRCUMSTANCES
  - HOW
  - WHY
- Prescriptions of a specific behavior of a provider
- Should be unambiguous and precise
- Should not be vague or underspecified
- Should not be statements of fact

Examples

- Action statements
  - An admission EKG should be performed on all patients in whom blunt cardiac activity is suspected
  - A single preoperative dose of prophylactic antibiotics with broad spectrum coverage should be administered to all patients sustaining penetrating abdominal wounds
- Not action statements
  - Computed tomography with axial collimation is superior to plain films in screening for TLS fractures
    - Statement of fact
  - MTBI is defined as an acute alteration in brain function caused by a blunt external force and is characterized by a GCS score of 13-15, loss of consciousness <= 30 minutes, and duration of posttraumatic amnesia of 24 hours
    - Definition

PICO example

- PICO question
  - In stable patients with blunt splenic trauma with contrast blush noted on CT of the abdomen (P)
  - Should angioembolization of the spleen be performed (I)
  - Versus observation without embolization (C)
  - To improve splenic salvage? (O)
- Action statement generated
  - In stable patients with blunt splenic trauma with contrast blush noted on CT of the abdomen (P)
  - Angioembolization of the spleen should be performed (I)
  - Versus observation without embolization (C)
  - To improve splenic salvage. (O)

PICO example

- PICO question
  - In stable penetrating abdominal trauma patients without peritonitis (P)
  - Should selective nonoperative management be employed (I)
  - Versus mandatory laparotomy (C)
  - To decrease the incidence of non-therapeutic laparotomy? (O)
- Action statement generated
  - In stable penetrating abdominal trauma patients without peritonitis (P)
  - Selective nonoperative management should be employed (I)
  - Versus mandatory laparotomy (C)
  - To decrease the incidence of non-therapeutic laparotomy. (O)

Conclusion

- First step in guideline development, once the topic has been decided, is the development of the PICO question
- Dissects a question into its component parts to facilitate finding its answer
- Will generate an answer which will be in the form of an action statement
- We want our guidelines to be *actionable* statements

MetroHealth
Saving lives is only the beginning.

Conducting a Literature Search and a Systematic Review

Catherine G. Velopulos, MD, MHS
Assistant Professor, Johns Hopkins
Department of Surgery

JOHNS HOPKINS MEDICINE

Practice Management Guideline

- New format modeled off the Cochrane systematic review format – rigorous and protocolized
  - Modified for PMG goals to produce recommendations rather than an overall conclusion
  - Incorporate GRADE methodology to rate quality
  - Recommendations are based on *quality* of the studies and *magnitude of effect*
- Manual and example document on EAST PMG website.

Before You Start

- Identify and information specialist who is trained in systematic review
  - Most university-affiliated libraries have people with this expertise
  - Chapter 6 in Cochrane Handbook
- Have a plan (protocol) to document search strategy and sources, and how the review will be conducted
  - Time/date
  - Database of origin
  - Background and Methods

Before You Start

- Software you will need:
  - A reference manager: e.g. EndNote (not free) or RefWorks (free)
    - Usually available through institution
  - Review Manager - RevMan (Cochrane)
    - Free, online. Data cannot be shared by users unless working on an actual Cochrane review.
  - GRADEpro
    - Free, online.

JOHNS HOPKINS MEDICINE

Types of Sources

- Database
  - Major: Pubmed, EMBASE, Cochrane, LILACS, etc.
  - Subject-specific: PsychINFO, etc.
  - Citation: Web of Science, Scopus
- Hand-searching
  - Bibliography/References
  - Table of contents in journals
  - Meeting proceedings/Abstracts ("Gray literature")

JOHNS HOPKINS MEDICINE

Table 1 Articles available on selected injury prevention and safety promotion topics in four literature databases, 2005.

Topic	Database	Articles available	% of total	Unique articles
Bicycle-related brain injuries: 51 (7 (13.7%))*	EMBASE	31	60.8	8
	MEDLINE	33	64.7	3
	PsycINFO	16	34.4	2
	WoS	27	52.9	5
Ethanol-impaired driving: 234 (39 (16.7%))*	EMBASE	96	41.0	27
	MEDLINE	167	71.4	34
	PsycINFO	80	34.2	9
	WoS	141	60.3	28
House fires: 54 (3 (5.6%))*	EMBASE	12	31.5	2
	MEDLINE	26	48.1	3
	PsycINFO	9	16.7	3
	WoS	44	81.5	24
Road rage: 40 (3 (7.5%))*	EMBASE	10	25.0	4
	MEDLINE	16	40.5	7
	PsycINFO	17	42.5	2
	WoS	24	60.0	10
Suicide among adolescents: 966 (66 (6.8%))*	EMBASE	346	35.8	77
	MEDLINE	667	69.1	217
	PsycINFO	277	28.7	38
	WoS	427	44.2	111

*Total number of articles (number common to all databases (%)).
WoS, World of Science.

Source: Lawrence, DW. What is lost when searching only one literature database for articles relevant to injury prevention and safety promotion? Inj Prev, 14(6), 401-404

JOHNS HOPKINS MEDICINE

Table 2. Type of electronic database searched

No. studies	Comparison	Type of search	Median no. trials	HS retrieval	ES retrieval	% Difference
31 studies	HS versus MEDLINE	Complex/HSSS or Simple	209 (IQR 75-392)	0.95 (95% CI 0.94-0.95)	0.55 (95% CI 0.54-0.56)	40%
7 studies	HS versus databases other than MEDLINE	Complex/HSSS or Simple	37 (IQR 23-259)	0.99 (95% CI 0.99-1.00)	0.50 (95% CI 0.48-0.52)	49%
4 studies	HS versus EMBASE	Complex/HSSS or Simple	149 (IQR 38-1332)	0.99 (95% CI 0.99-1.00)	0.49 (95% CI 0.47-0.52)	51%
4 studies	HS versus PsycINFO	Complex/HSSS or Simple	37 (IQR 26-39)	0.99 (95% CI 0.99-1.00)	0.67 (95% CI 0.58-0.76)	32%
1 study	HS versus Lilacs & CENTRAL	Complex/HSSS or Simple	9	1.00	0.11 (95% CI 0.01-0.15)	89%

Source: Hopewell S, et al. Handsearching versus electronic searching to identify reports of randomized trials. Cochrane Database Syst Rev. 2007 Apr 18;(2):MR000001

JOHNS HOPKINS MEDICINE

Constructing the Search

- Controlled Vocabulary
  - Alternative spellings/plurals
  - Related terms and synonyms
- Keywords
  - Truncate if possible
  - MeSH terms in PubMed – Medical Subject Headings
  - Start with broader terms, narrower are automatically included

JOHNS HOPKINS MEDICINE

Constructing the Search

- Use PICO question as a guide
  - (**P**opulation OR Synonym OR . . .) AND (**I**ntervention OR Synonym OR . . .) AND (**C**omparator OR Synonym OR . . .) AND (**O**utcome)
  - Keep simple at first, use initial search to explore MeSH
  - Be careful to use Boolean operators correctly
- Filter by study type
- Try to include languages other than English – should be able to find a translator for most things

JOHNS HOPKINS MEDICINE

Exploring for MeSH Terms

subaxial spine injuries with different indication and risk profiles. In case of incomplete neurological injury there is significant improvement. Operative treatment certainly decreases the complications related to prolonged immobilisation in recovery phase by making the patient mobile early.
PMID:
23785912
[PubMed - indexed for MEDLINE]

MeSH Terms
MeSH Terms
- Adolescent
- Adult
- Bone Plates
- Bone Screws
- Cervical Vertebrae/injuries*
- Cervical Vertebrae/radiography
- Cervical Vertebrae/surgery*
- Female
- Follow-Up Studies
- Fracture Fixation, Internal/methods*
- Humans
- Male
- Middle Aged
- Spinal Fractures/complications
- Spinal Fractures/radiography
- Spinal Fractures/surgery*
- Spinal Fusion
- Treatment Outcome
- Young Adult

JOHNS HOPKINS MEDICINE

Constructing the Search

- Use PICO question as a guide
  - (**P**opulation OR Synonym OR . . .) AND (**I**ntervention OR Synonym OR . . .) AND (**C**omparator OR Synonym OR . . .) AND (**O**utcome)
  - Keep simple at first, use initial search to explore MeSH
  - Be careful to use Boolean operators correctly
- Filter by study type
- Try to include languages other than English – should be able to find a translator for most things

JOHNS HOPKINS MEDICINE

Documenting the Search

- Timing
  - Year of publication
  - Date accessed
- Search strategy for each database
- Sources other than database

JOHNS HOPKINS MEDICINE

Using the Search

- Deduplication
- Title and Abstract Review – 2 independent reviewers
  - Yes
  - Maybe
  - No
- All "Yes" and "Maybe" by either person pulled for full-text review

JOHNS HOPKINS MEDICINE

Using the Search

- Full-text review for suitability based on PICO-based inclusion/exclusion criteria
  - Yes
  - Maybe
  - No
- All "Yes" included; "Maybe" requires resolution by consensus
- Would like at least 20-50 articles for full-text review, 15-30 for systematic review

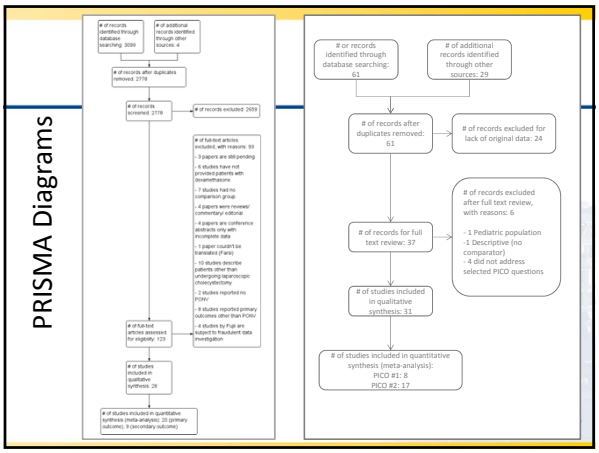
JOHNS HOPKINS MEDICINE

Using the Search

- Data extraction form created by team
- Double data extraction performed independently, then compared – recommend Excel, set up with columns similar to RevMan
- Decision for articles to include for qualitative synthesis and quantitative synthesis.

JOHNS HOPKINS MEDICINE

PRISMA Diagrams



Parts of the Systematic Review

- Qualitative Synthesis/Analysis
  - Overview of the results of the studies
  - Includes all outcomes evaluated
  - Summary of outcomes that are not discretely measureable
  - ALWAYS included: a systematic review *may* include a quantitative analysis (meta-analysis) as well, but not required

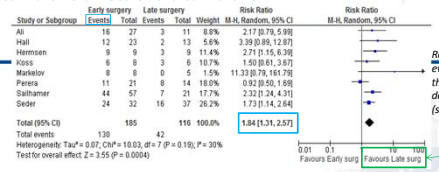
JOHNS HOPKINS MEDICINE

Parts of the Systematic Review

- Quantitative Synthesis/Meta-analysis
  - Need a CRITICAL and MEASURABLE outcome
  - Decide measure of association: varies, but in general use RR for high-prevalence outcomes (>20%).
  - Plug data from extraction tables into RevMan to create Forest Plots
  - For event rates, keep in mind that RevMan records "event" as the worse outcome when generating the Forest plot

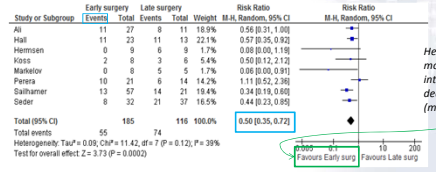
JOHNS HOPKINS MEDICINE

Survival is Event – Relative Risk of Survival is 1.84



RevMan interprets events as "bad," thus the intervention that decreases the event (survival) is favorable

Mortality is Event – Relative Risk of Mortality is 0.50



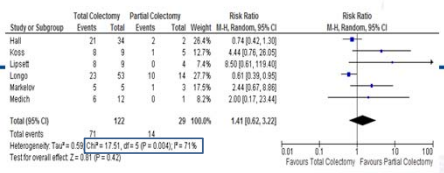
Here, the event is mortality, thus the intervention that decreases the event (mortality) is favorable

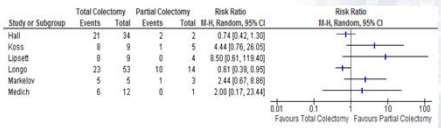
JOHNS HOPKINS MEDICINE

To Pool or Not To Pool

- Appropriate to pool the data into one measure of effect only if I^2 is acceptable
  - I^2 is the % of the variability in the effect estimate that is due to heterogeneity (differences between the studies) rather than random error
  - 100% means that all of the variability is because the studies are too different to combine; Chi-square test for heterogeneity associated with significant p-value
  - 0% means that there is essentially no difference between the studies; p-value would be non-significant

JOHNS HOPKINS MEDICINE





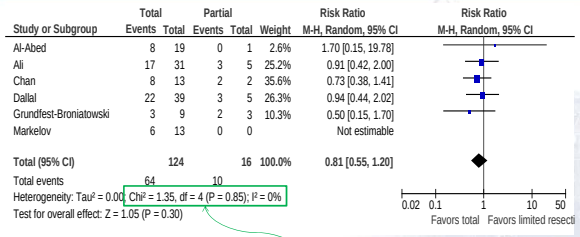
JOHNS HOPKINS MEDICINE

To Pool or Not To Pool

- Heterogeneity is a primarily clinical and methodological determination
  - Statistical measures of heterogeneity, (as all statistical measures are), remain sensitive to population size and event rate
  - Must use both to determine appropriateness of combining into one effect measure
  - Will dictate strength of recommendation

JOHNS HOPKINS MEDICINE

Fooled by the I^2



Studies are very different, but imprecision reflected in wide confidence intervals (due to low study numbers) masks this heterogeneity

JOHNS HOPKINS MEDICINE

It's Not Just Quantity, It's Quality!

- Data from RevMan is now ready for importation into GRADEpro*
- Recommendations in GRADE are based on quality of the studies, and magnitude of effect

*If there is no quantitative analysis, quality profile will be based on risk of bias assessment (reference Cochrane)

JOHNS HOPKINS MEDICINE

GRADE: ASSESSING THE QUALITY OF EVIDENCE

east || 27th Annual Scientific Assembly

So you want to write a practice management guideline?

Naples, FL, February 8, 2014

Philipp Dahm, MD, MHSc, FACS
Professor of Urology & Residency Program Director, University of Florida & Malcom Randall VA Medical Center, Gainesville, Florida
Coordinating Editor, Cochrane Prostatic Diseases and Urological Cancers Group

1

Where are We?

- Defined our key questions in PICO format
- Identified important outcomes
  - Benefits, Harms (and resource utilization)
- Rated the outcomes as critical, important or not important
- Appreciate the importance of
  - Systematic reviews ("totality of evidence")
  - Need to rate the quality of evidence beyond study design

Learning Objectives

- Rate the quality of evidence for a body of evidence by outcome using GRADE
  - Study limitations
  - Inconsistency
  - Imprecison
  - Indirectness
  - Publication bias
- Rate the overall quality of evidence

Quality of Evidence

The quality of evidence (QoE) reflects the extent to which we have confidence that an estimate of the effect is adequate to support recommendations

The Filet of GRADE

5

GRADE – Quality of Evidence

Quality of Evidence	Study Design	Lower if	Higher if
High	Randomized trial	**Study quality:** -1 Serious limitations -2 Very serious limitations -1 Important **inconsistency** **Directness:** -1 Some uncertainty -2 Major uncertainty -1 **Sparse or imprecise data** -1 High probability of **reporting bias**	**Strong association:** +1 Strong, no plausible confounders +2 Very strong, no major threats to validity +1 Evidence of **a dose response gradient** +1 All plausible **confounders** would have reduced the effect
Moderate			
Low	Observational study		
Very low	Any other evidence		

Safeguards Against Bias

- Randomization
- Concealed allocation
- Blinding
- Completeness of follow-up
- Intention-to-treat analysis

Empirical Evidence of Bias

Inadequate RCT Methodology
Associated with Exaggerated Treatment Effects

Allocation concealment
- inadequate +41%
- unclear +30%

Blinding
- no "double-blinding" +17%

→ Inadequate RCT methodology associated with bias

Schulz, KF et al, JAMA (1995)

Randomization

Definition

- Achieves balance for both known and unknown prognostic variables
- Observed differences not greater than might be expected due to chance

Allocation Concealment

Definition

- Investigators should not be able to determine the allocation of the next patient to be entered into a trial
- Decision to accept/reject participant should be made in ignorance of next assignment
- ≠ Blinding

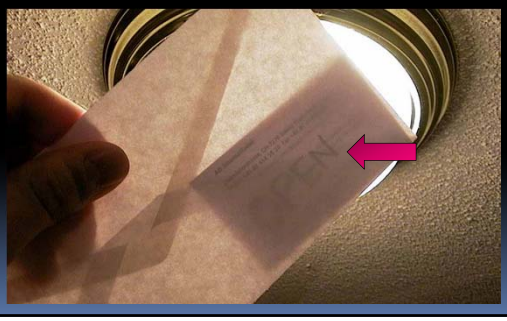
What Methods are Concealed?

Use Hospital Chart Numbers to Randomize Patients	No
Place Patient Treatment Allocations into Envelopes	Maybe
Use Telephone Randomization System	Yes

When Concealment is LOST

- Australian investigators undertook a randomized trial of open versus laparoscopic appendectomy. The trial ran smoothly during the day.
- But "not so" smoothly at night…

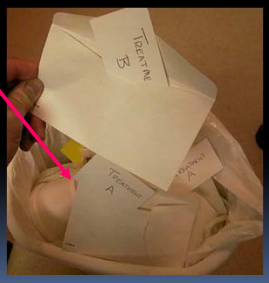
Was Randomization Concealed?



Was Randomization Concealed?

Garbage Can Randomization

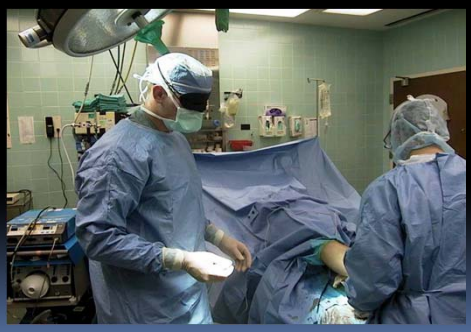
"If you open enough envelopes, you'll eventually get your treatment of choice!"



Blinding

- Prevents study participants' and study personnel's knowledge of study group assignment
- Protects trial from:
  - 1) treatment differences between the groups other than the randomized treatments
  - 2) biased assessment of outcomes (least important for death as outcome)
- Need to assure that neither participants nor research teams can identify treatment assignment
  (e.g. prostate cancer prevention trial and PSA levels)

Can Surgeons be Blinded?



Can Surgeons be Blinded?

Courtesy of Mo Bhandari and Rudolph Poolman, McMaster University

Whom Can we Blind in a Surgical RCT?

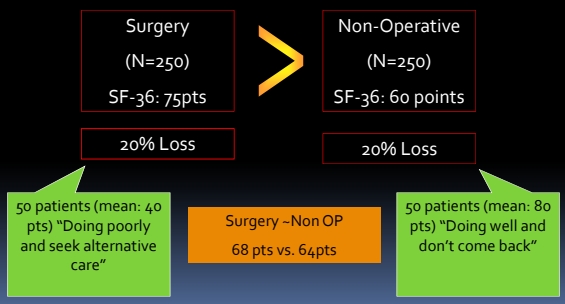
- Patients
- Caregivers
- Collectors of outcome data
- Adjudicators of outcomes
- Data analysts
- [Manuscript writers]

Completeness of Follow-up

- Protects trial from non-random loss of participants
- Guards against selection bias

Completeness of Follow-up

Worst Case Scenario Method



Analysis of an RCT

Per Protocol versus Intention-to-treat

Intention-to-Treat: Every participant analyzed as randomized (regardless whether pt actually received treatment)

Advantage: Guards against other bias
Disadvantage: May underestimate the full treatment effect

Per Protocol: Analyses limited to those participants in both groups who actually received the treatment

Disadvantage: Pts who adhere to study treatment may be different than the drop-outs
e.g. analysis limited to pts with metastatic bladder cancer who received 3 cycles of MVAC …

Inconsistency/Heterogeneity

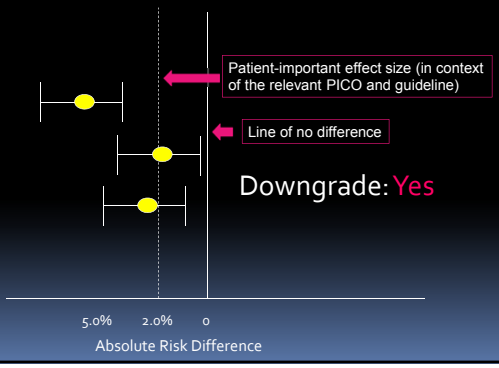
- Definition: Unexplained differences of results between individual trials
- Common explanations relate to
  - Clinical differences (patients, interventions, outcomes)
  - Methods (risk of bias)
  - Chance
- Hypotheses to explain inconsistency should be defined at the get-go (*a priori*)

Inconsistency/Heterogeneity

- Visual representation in Forest plot
- Eye ball test
  - Variation in effect size
  - Overlap of confidence intervals

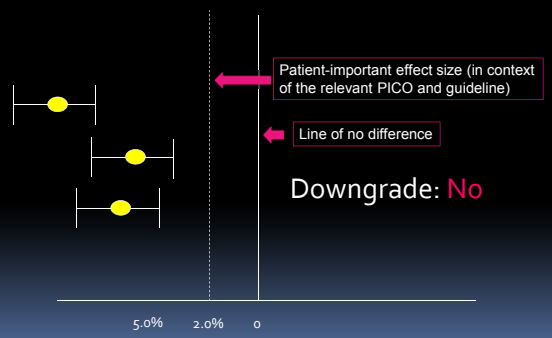
- Statistical tests for heterogeneity
  - Chi-square
  - I^2 (preferred)
    - < 40% - low
    - 30% - 60% moderate
    - 50% - 90% substantial
    - 75% - 100% considerable

Judgments about Inconsistency



Downgrade: Yes

Judgments about Inconsistency



Downgrade: No

Indirectness of Evidence

- Differences in:
  - Population/patients
    - i.e. old versus young
  - Interventions
    - i.e. intravenous versus oral drug administration
  - Outcomes
    - i.e. patient important versus surrogates
- Indirect comparisons
  - A > B; B > C; A > C ??

Imprecision

- Uncertainty about "true" effect size as reflected by:
  - Low event rate
  - Wide confidence intervals
- Downgrade for imprecision when:
  - Confidence in estimate of effect not adequate to support decision

Imprecision: Neoadjuvant Chemotherapy in Colon CA

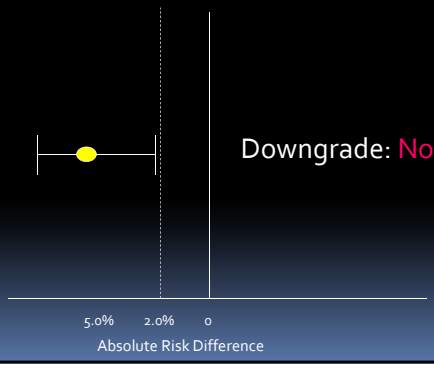
- SR of supports impact on overall survival
- HR 0.86 (95% CI: 0.77 – 0.95)
- Corresponds to 5% absolute survival advantage
- Failure to report harm; however considerable morbidity and some treatment related mortality likely

Imprecision: Neoadjuvant Chemotherapy in Colon CA

- What degree of uncertainty surrounding the 5% improvement in overall survival are you willing to accept to recommend neoadjuvant chemotherapy?

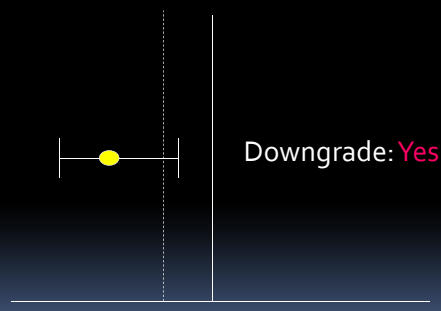
Absolute risk reduction: 5%
95% CI: 9% - 2%

Judgments about Imprecision



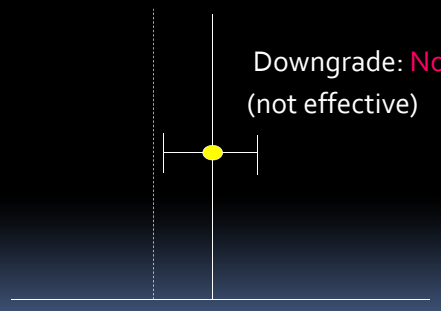
Downgrade: No

Judgments about Imprecision



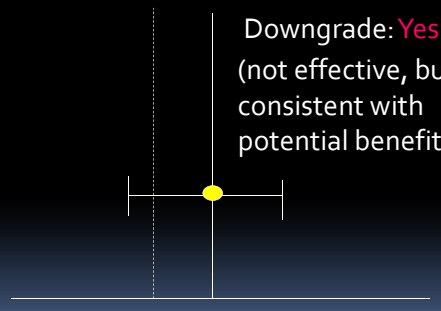
Downgrade: Yes

Judgments about Imprecision



Downgrade: No
(not effective)

Judgments about Imprecision

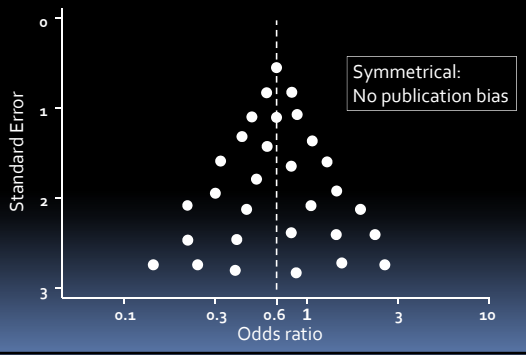


Downgrade: Yes
(not effective, but consistent with potential benefit)

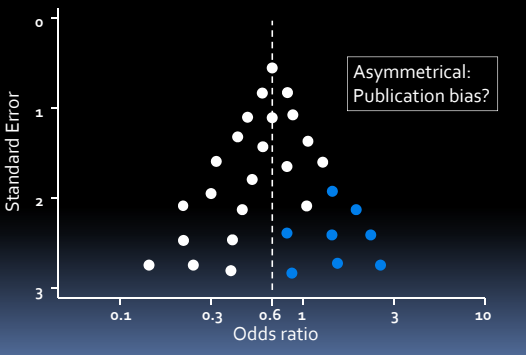
Publication Bias

- Faster and multiple publication of "positive" trials
- Slower and fewer publication of "negative" trials
- SR needs to search "grey literature" for unpublished studies
- Evaluation using "Funnel plot" and other methods

Funnel Plot



Funnel plot



What can raise quality?

Large effect size

- 1 level: Large magnitude of effect (RRR 50%)
- 2 levels: Very large magnitude of effect (RRR 80%)
- Common criteria:
  - Everyone used to do badly
  - Almost everyone does well
- Example: Functional improvement after hip replacement

Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials

Gordon C S Smith, Jill P Pell

What is already known about this topic

Parachutes are widely used to prevent death and major injury after gravitational challenge

Parachute use is associated with adverse effects due to failure of the intervention and iatrogenic injury

Studies of free fall do not show 100% mortality

What this study adds

No randomised controlled trials of parachute use have been undertaken

The basis for parachute use is purely observational, and its apparent efficacy could potentially be explained by a "healthy cohort" effect

Individuals who insist that all interventions need to be validated by a randomised controlled trial need to come down to earth with a bump

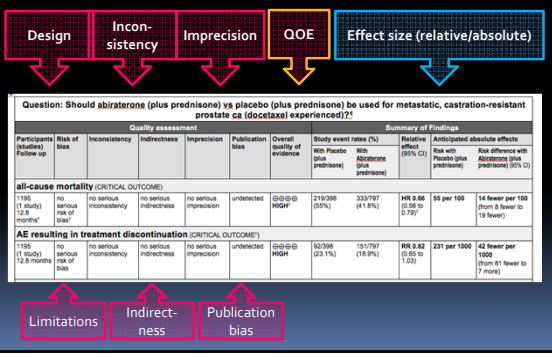
Parachutes reduce the risk of injury after gravitational challenge, but their effectiveness has not been proved with randomised controlled trials

Smith GCS and Pell JP, BMJ (2003)

How can Quality of evidence differ by outcome?

- In a systematic review, not all studies contribute to all outcomes
  - i.e. may have different time horizons, 30 days versus 1 year follow-up
- Risk of bias may vary by outcome
  - i.e. blinding of outcome assessors less important for overall versus disease-specific survival

GRADE Evidence Profile



Question: Should abiraterone (plus prednisone) vs placebo (plus prednisone) be used for metastatic, castration-resistant prostate ca (docetaxel experienced)?[1]

Quality assessment							Summary of Findings				
Participants (studies) Follow up	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	Overall quality of evidence	Study event rates (%) With Placebo (plus prednisone)	Study event rates (%) With Abiraterone (plus prednisone)	Relative effect (95% CI)	Anticipated absolute effects Risk with Placebo (plus prednisone)	Anticipated absolute effects Risk difference with Abiraterone (plus prednisone) (95% CI)
all-cause mortality (CRITICAL OUTCOME)											
1195 (1 study) 12.8 months[a]	no serious risk of bias[c]	no serious inconsistency	no serious indirectness	no serious imprecision	undetected	⊕⊕⊕⊕ HIGH[f]	219/398 (55%)	333/797 (41.8%)	HR 0.66 (0.56 to 0.79)[f]	55 per 100	14 fewer per 100 (from 8 fewer to 19 fewer)
AE resulting in treatment discontinuation (CRITICAL OUTCOME[b])											
1195 (1 study) 12.8 months	no serious risk of bias	no serious inconsistency	no serious indirectness	no serious imprecision	undetected	⊕⊕⊕⊕ HIGH	92/398 (23.1%)	151/797 (18.9%)	RR 0.82 (0.65 to 1.03)	231 per 1000	42 fewer per 1000 (from 81 fewer to 7 more)

GRADE – Quality of Evidence

Observational studies	*Quality of evidence*	*Randomized trials*
Extremely strong association and no major threats to validity	High	No serious flaws in study quality
Strong, consistent association and no plausible confounders	Moderate	Serious flaws in design or execution or quasi-randomized trials
No serious flaws in study quality	Low	Very serious flaws in design or execution
Serious flaws in design and execution	Very low	Very serious flaws and at least one other serious threat to validity

Overall Quality of Evidence

- Based on lowest quality evidence for critical outcomes
- Important to establish *a priori* which outcomes are critical

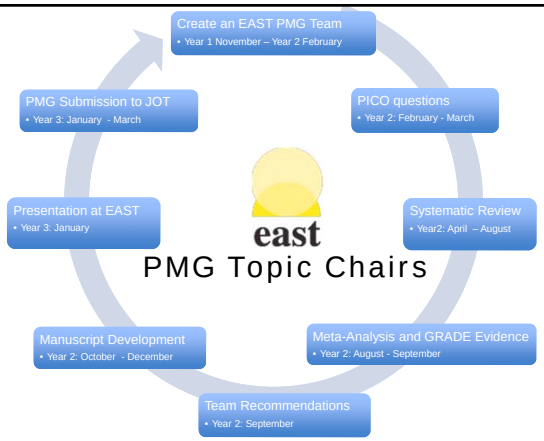
Take Home Messages

- GRADE quality of evidence ratings
  - Based on systematic review rather than individual studies
  - Outcome-specific
  - Considers additional dimensions beyond study design and limitations
  - Judgments to downgrade/upgrade are context-specific within the guideline
  - Guideline developers: rate overall evidence based on critical outcomes

Developing a Timeline

27th EAST Annual Scientific Assembly
Workshop: So You Want to Write a Practice Management Guideline
Mayur B. Patel, MD,MPH,FACS

If all difficulties were known at the outset of a long journey, most of us would never start out at all. – Dan Rather



EAST Practice Management Guideline or Evidenced-Based Review Timeline

- Circle of Time Crosses Three Calendar Years
- 12 o'clock to 6 o'clock: New Territory
  - Creating a relevant Team
  - PICO creation
  - Systematic Review
  - Meta-Analysis and GRADE evidence
- 6 o'clock to 12 o'clock: Familiar
  - Make an Actionable Recommendation, Write the Manuscript, Presentation, Submit the Manuscript

Submit a Proposal & Create a Team
November – February

- Read
  - Literature
  - Prior Guidelines
  - Develop Informal Question but Think PICO
- Talk to
  - Prior Guideline Participants
  - Prior Guideline Leaders
  - Current Guidelines Committee Members
  - Understand the Commitment

Submit a Proposal & Create a Team
November – February

- http://www.east.org/resources/treatment-guidelines/guidelines-under-development

Guidelines Under Development

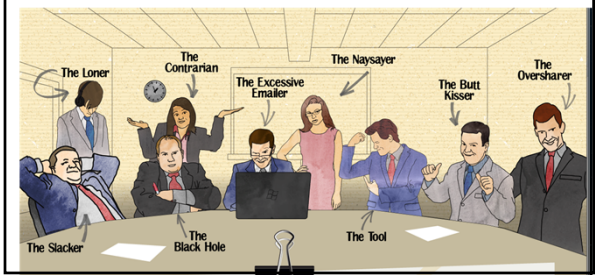
Visit this page often for a listing of the EAST Practice Management Guidelines (PMGs) and Evidence Based Reviews (EBRs) that are currently under development. To volunteer as a team member, click the volunteer button for the PMG or EBR of interest and an email willl be sent to the EAST Administrative Offices. Your volunteer request will be reviewed by the Chair of the Guidelines Committee and confirmed by email notification.

If you have a recommendation for a new PMG or EBR or to update existing ones, submit a proposal using the new online proposal form. Proposals will be reviewed by the EAST Guidelines Committee. Notification of proposal approval will be sent to the proposer via email.

Submit a proposal

Submit a Proposal & Create a Team
November – February

- Be Aware of How You Work
  - Doer, Thinker, Creator, Helper, Persuader, Organizer



Submit a Proposal & Create a Team
November –February

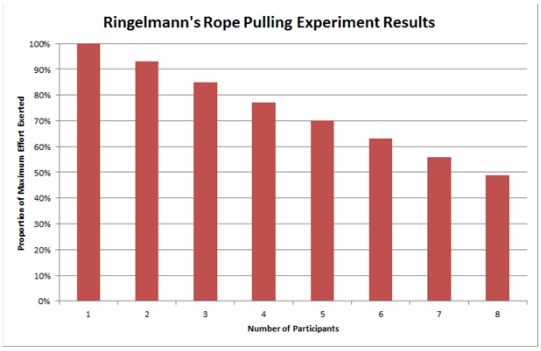
- Create a Team ≤ 10
- Experts:
  - Content: Not Only Trauma & Acute Care Surgery
  - Systematic Review Methodology
  - GRADE Methodology
  - EAST Guideline Committee Liaison
  - EAST Senior
- Energy:
  - Local
  - Accountable
- Get to Know the Rest of Your Team

AVOID the Ringlemann Effect

Avoid the Ringlemann Effect

- "When employing men, or draught animals, better use is achieved when the source of motive power works alone: as soon as one couples two or several such sources to the same load, the work performed by each of them, at the same level of fatigue, decreases as a result of the lack of simultaneity of their efforts …"
- Ringlemann Effect:
  - Social Loafing aka Motivation Loss
  - Loss of Coordination

AVOID the Ringlemann Effect



PICO Questions
February - March

- 1-2 PICOs are Reasonable
- Share With:
  - Local Colleagues
  - Guideline Group
  - Guideline Committee
  - EAST Community
- This is the fundamental and most critical step

PICO Questions
February - March

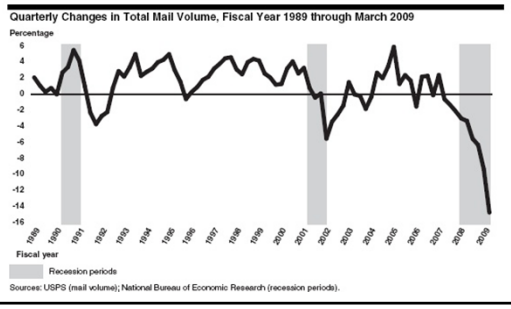
- Engage and Set Stage with your Team
  - Set a time
  - Set a pattern
  - Pick a method
    - ~~Telepathy~~
    - Electronic Chat
    - Conference Call
    - GoTo Meeting

Systematic Review
April - August

- April
  - Protocol development using PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines
    - http://www.prisma-statement.org
- May
  - Register the Protocol with PROSPERO
    - http://www.crd.york.ac.uk/prospero/
  - Begin Literature Search Process with Librarian
  - Develop data extraction forms
- June
  - Start Title/Abstract screening (<1200)
  - Advance to Full-text review (<100)
- July
  - Data Extraction

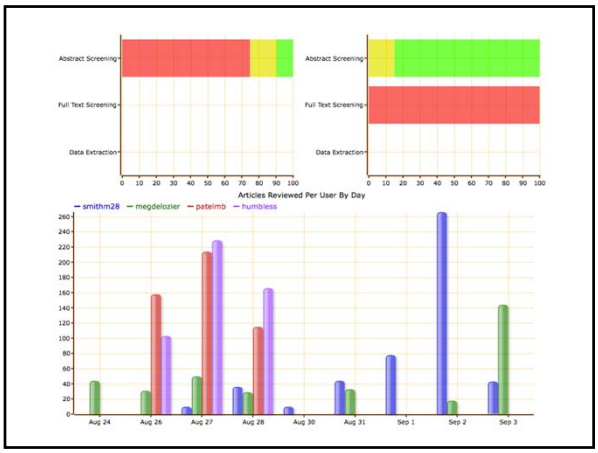
Systematic Review
April - August

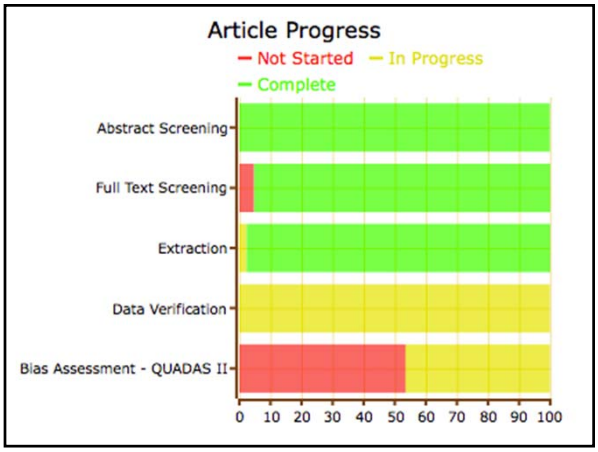
- Print Papers, Envelopes, Stamps



Systematic Review
April - August

- ~~Print Papers, Envelopes, Stamps~~
- MS Word and/or Excel
- Email .pdf's
- Dropbox or Cloud platform equivalent
- RevMan
  - http://ims.cochrane.org/revman/download
- Distiller SR
  - http://systematic-review.net





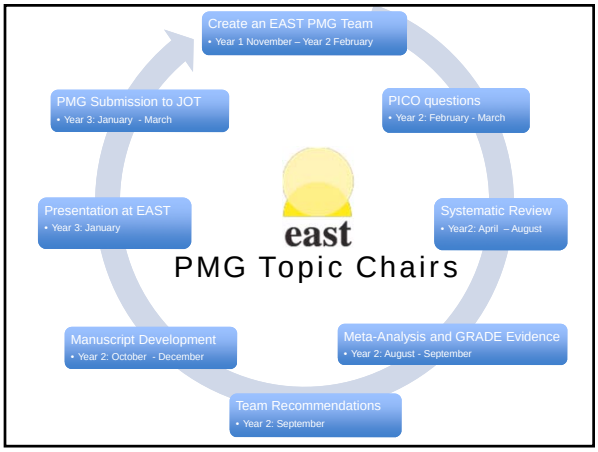
Meta-Analysis and GRADE Evidence

August - September

- Perform meta-analysis with methodology expert and Bias assessment (i.e. RevMan)
- Consider GRADE Course
- Use free GRADEPro software to create evidence profile
  - http://ims.cochrane.org/revman/gradepro
- Rate quality of evidence for each PICO across all outcomes

It's a Marathon, Not a Sprint

- Be Aware of How You Work
- Sharing a Common Vision with your Team
- Clearly Define Roles and Responsibilities
- Be Proactive and Provide Feedback
- Acknowledge and Reward
- Always Celebrate Success
- Slow and Steady Wins the Race



FOUNDATIONS OF EVIDENCE-BASED MEDICINE

east || 27th Annual Scientific Assembly

So you want to write a practice management guideline?

Naples, FL, February 8, 2014

Shahnaz Sultan, MD, MHSc
Assistant Professor , Division of Gastroenterology, University of Florida & Malcom Randall VA Medical Center, Gainesville, Florida

1

What is Evidence Based Medicine?

"The conscientious, explicit and judicious use of the **current best evidence** in making decisions about the care of individual patients".

"The practice of evidence based medicine means **integrating individual clinical expertise** with the best available external evidence from clinical research".

EBM – Historical Perspective

Three movements (~1990):

- *Evidence Based Medicine* and the need for critical appraisal (Canada)
- *Systematic reviews* and the need for research synthesis - Cochrane Collaboration (UK)
- *Clinical Practice Guidelines* and the need to standardize the practice (US)

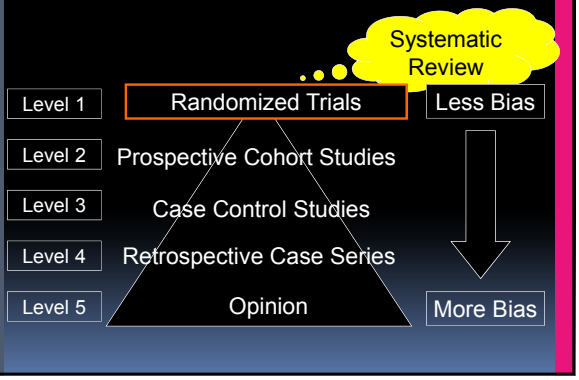
Courtesy of B. Djulbegovic

Principles of EBM

- "Hierarchy of Evidence"
  - Some studies are more likely than others to describe "the truth"
- "Evidence alone is never enough"
  - Need to integrate the evidence with the patient's values and preferences

Evidence Based Decision-Making

Hierarchy of Evidence (Therapy)

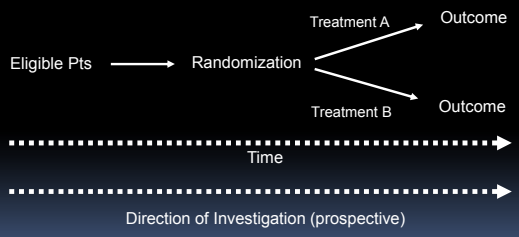


GRADE – Quality of Evidence

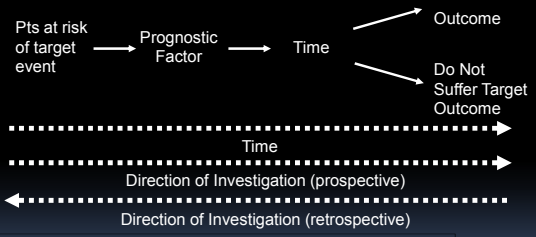
Quality of Evidence	Study Design	Lower if	Higher if
High	Randomized trial	**Study quality:** -1 Serious limitations -2 Very serious limitations -1 Important **inconsistency** **Directness:** -1 Some uncertainty -2 Major uncertainty -1 **Sparse or imprecise data** -1 High probability of **reporting bias**	**Strong association:** +1 Strong, no plausible confounders +2 Very strong, no major threats to validity +1 Evidence of **a dose response gradient** +1 All plausible **confounders** would have reduced the effect
Moderate			
Low	Observational study		
Very low	Any other evidence		

Study Designs

Randomized Controlled Trial



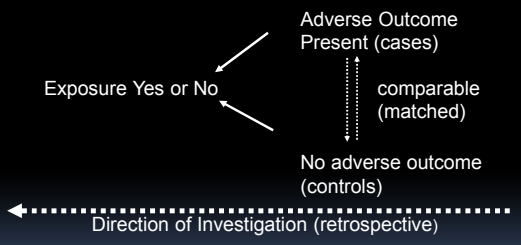
Study Designs



Examples:
- Stress urinary incontinence in women following childbirth
- Smoking and lung cancer

Study Designs

Cohort Study



Examples:
- DES exposure of mother and vaginal clear cell carcinoma
- Agent orange exposure and prostate cancer

EBM – Systematic Reviews

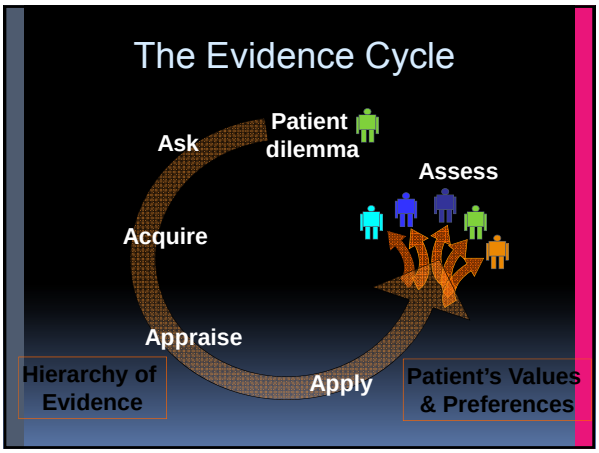
- Need to summarize the entire body of evidence for a question
- Sir Archie Cochrane – Founding of the Cochrane Collaboration
- Pioneered methodology of systematic reviews; important EBM resource

THE COCHRANE COLLABORATION®

Systematic Review

A study that uses a predefined, systematic and transparent approach to identify, select, appraise and summarize primary studies addressing a focused clinical question using methods to reduce the likelihood of bias

The Evidence Cycle



Ask

A Focused Clinical Question

Focused Clinical Question

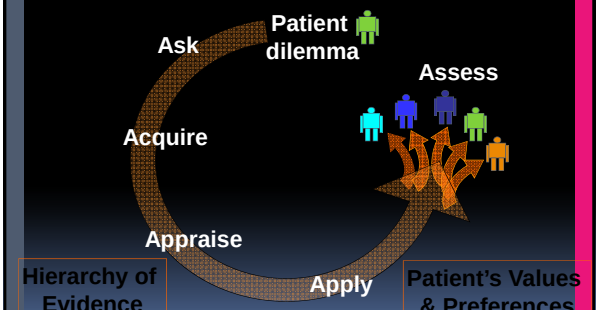
- **P** Population
- **I** Intervention
- **C** Comparison
- **O** Outcome
- **S** Study design

Importance of outcomes

- Consider all outcomes of importance to patients
- Define before formal evidence review (*a priori)*
- Distinguish between outcomes that are critical, important and non-important for decision-making: GRADE rates relative importance of outcomes to on a scale from 1-9

Guyatt GH et al, BMJ (2008)

The Evidence Cycle



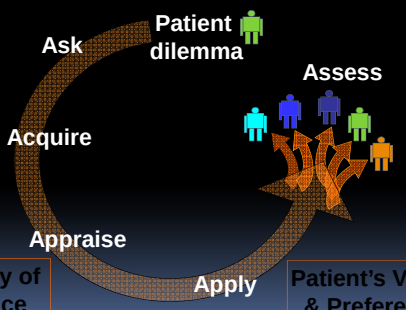
Acquire

The (Body of) Evidence

Finding the Evidence

- Systematic reviews of randomized controlled trials (RCTs)
- Individual RCTs
- Observational studies
- Economic studies

The Evidence Cycle

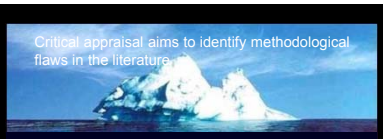


Hierarchy of Evidence

Patient's Values & Preferences

Appraise

The (Body of) Evidence



Beware of Initial Appearances

Looking at the surface is not enough!

Safeguards Against Bias

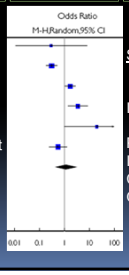
- Randomization
- Concealed allocation
- Blinding
- Completeness of follow-up
- Intention-to-treat analysis

Critical appraisal aims to identify methodological flaws in the literature

Quality of Evidence

Definition: The extent to which our confidence in an estimate of the treatment effect is adequate to support a particular recommendation

Methodological limitations	Inconsistency of results	Indirectness of evidence	Imprecision of results	Publication bias

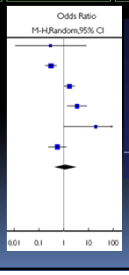


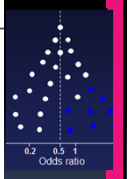
Risk of bias:

Allocation concealment
Blinding
Intention-to-treat
Follow-up
Stopped early

Sources of indirectness:

Indirect comparisons
Patients
Interventions
Comparators
Outcomes





Clinical Practice Guidelines

- "Systematically developed statements to help practitioner and patient decisions about appropriate health care for specific clinical circumstances."

Shaneyfelt, TM JAMA (1999)

Clinical Practice Guidelines

Take Home Messages

- Guiding principles of EBM:
  - Hierarchy of evidence
  - Evidence alone is never enough
- Practice of EBM represented by 5A's
- "Best evidence" derived from systematic reviews
  - Rate quality of evidence (GRADE)
  - Focus on patient-important outcomes
- Guideline recommendations important approach for promoting evidence-based clinical practice

GRADE Handout: **EAST Workshop 2014**

Exercise

- Work in small groups
- Decide whether you prefer to use GRADEpro (software) or to do the exercise on paper
- Select someone to report back to the whole group
- Watch the time

1. Familiarize yourself with the systematic review/study that you have been given (if you are not yet familiar with it): read the abstract.

2. Identify the clinical question in the PICO (Population, Intervention, Comparison, and Outcome) format. Work in your small group.

P: _____

I: _____

C: _____

O: _____

3. Select up to **7** important outcomes for this comparison (consider the following suggestions)
Suggestions

a) Generate a list of relevant outcomes (see **worksheet 1**)
   - Discuss in the group which outcomes would be relevant (think of all relevant outcomes, not only those that are in the review, but you think might be important to someone making a decision; make sure to include both benefits and downsides, e.g., adverse effects and costs, if relevant)
   - Find consensus within your small group about which outcomes are important enough to be included in the GRADE Evidence Profile

b) From this list choose up to 7 outcomes that you think are most important to a guideline panel or others making recommendations and should be included in the evidence profile; transfer them to a blank evidence profile (see **worksheet 3**) or use GRADEpro.

4. Assess the quality of evidence for this outcome according to the GRADE approach
Suggestions

- Fill in **worksheet 2** (or use GRADEpro) to assess the quality of the evidence for each outcome
  - Consult the table "GRADE quality assessment criteria" (or use help in GRADEpro).
- Fill in the Quality of the Evidence column in the Evidence Profile.

5. Move from evidence to recommendations using the evidence profile
Suggestions

- Use worksheet 4 to help decide on strong or weak recommendations
- See also "Definitions for strong and weak/conditional recommendations" and
- "Implications of strong and weak/conditional recommendations"

Worksheet 1: List of outcomes from the systematic review

Title of the systematic review: _____

List all outcomes below. When you have completed the listing, choose **up to 7** most important outcomes to be included in the GRADE evidence table.
Rate the relative importance for each outcome on a 9 point scale ranging from 1 (less important) to 9 (critically important for decision making). You can use the same rating multiple times.

1 – 3 less important and not included in the GRADE Evidence Profile
4 – 6 important but not critical for making a decision (inclusion in the Evidence Profile may depend on how many other important outcomes there are)
7 – 9 critical for making a decision and should definitely be included in the Evidence Profile

Transfer the selected outcomes into the blank GRADE evidence profile (see **worksheet 3** of this handout). If you are using GRADEpro, begin inserting the outcomes in GRADEpro.

Outcome	**Importance**	**Include in GRADE evidence profile?**	
1.		Yes	No
2.		Yes	No
3.		Yes	No
4.		Yes	No
5.		Yes	No
6.		Yes	No
7.		Yes	No
8.		Yes	No
9.		Yes	No
10.		Yes	No
11.		Yes	No
12.		Yes	No

<u>Worksheet 2</u>: Assessing the quality of evidence across studies for an outcome (see criteria and definitions in the following sections of this handout.) Footnotes should be included for any up- or downgrading.

Quality criteria	**Rating** (circle one for each criterion)	**Footnotes** (explain reasons for up- or downgrading)	**Quality of the evidence** (Circle one per outcome)
Outcome # 1:			
Risk of bias	No serious (-1) very serious (-2)		⊕⊕⊕⊕ High ⊕⊕⊕○ Moderate ⊕⊕○○ Low ⊕○○○ Very Low
Inconsistency	No serious (-1) very serious (-2)		
Indirectness	No serious (-1) very serious (-2)		
Imprecision	No serious (-1) very serious (-2)		
Publication Bias	Unlikely likely (-1) very likely (-2)		
Large effect	Large (+1) Very large (+2)		
Dose-response gradient	No Yes (+1)		
Plausible confounding would change the effect	No Yes (+1)		

Quality criteria	Rating (circle one for each criterion)	Footnotes (explain reasons for up- or downgrading)	Quality of the evidence (Circle one per outcome)
Outcome #2:			
Risk of bias	No serious (-1) very serious (-2)		⊕⊕⊕⊕ High ⊕⊕⊕◯ Moderate ⊕⊕◯◯ Low ⊕◯◯◯ Very Low
Inconsistency	No serious (-1) very serious (-2)		
Indirectness	No serious (-1) very serious (-2)		
Imprecision	No serious (-1) very serious (-2)		
Publication Bias	Unlikely likely (-1) very likely (-2)		
Large effect	Large (+1) Very large (+2)		
Dose-response gradient	No Yes (+1)		
Plausible confounding would change the effect	No Yes (+1)		

GRADE quality assessment criteria

Study design	*Quality of evidence*	*Lower if **	*Higher if **
Randomized trials → **Observational studies →**	**High** **Moderate** **Low** **Very low**	**Study limitations** -1 Serious -2 Very serious **Inconsistency** -1 Serious -2 Very serious **Indirectness** -1 Serious -2 Very serious **Imprecision** -1 Serious -2 Very serious **Publication bias** -1 Likely -2 Very likely	**Large effect** + 1 Large + 2 Very large **Dose response** + 1 Evidence of a gradient **All plausible confounding** + 1 Would reduce a demonstrated effect, or + 1 Would suggest a spurious effect when results show no effect

* 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)

Conceptualization: Quality of evidence across studies for the outcome (both versions are valid)

⊕⊕⊕⊕ 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 on our confidence in the estimate of effect and may change the estimate.

⊕⊕◯◯ Low = 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 = Any estimate of effect is very uncertain.

⊕⊕⊕⊕ High = We are very confident that the true effect lies close to that of the estimate of the effect.

⊕⊕⊕◯ Moderate = We are moderately confident in the estimate of effect: The true effect is likely to be close to the estimate of effect, but possibility to be substantially different.

⊕⊕◯◯ Low = Our confidence in the effect is limited: The true effect may be substantially different from the estimate of the effect.

⊕◯◯◯ Very low = We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect.

EXPLANATIONS FOR DOWNGRADING

Limitations of design:

- lack of allocation concealment
- lack of blinding (particularly if outcomes are subjective and their assessment highly susceptible to bias)
- large loss to follow-up
- failure to adhere to an analysis according to intention-to-treat principle
- stopping a trial early for benefit
- selective reporting of events: investigators neglect to report outcomes that they have measured (typically those for which they observed no effect).

Inconsistency:

Widely differing estimates of the treatment effect (i.e. heterogeneity or variability in results) across studies suggest true differences in underlying treatment effect. When heterogeneity exists, but investigators fail to identify a plausible explanation, the quality of evidence should be downgraded by one or two levels, depending on the magnitude of the inconsistency in the results.

Inconsistency may arise from differences in:

- populations (e.g., drugs may have larger relative effects in sicker populations)
- interventions (e.g., larger effects with higher drug doses)
- outcomes (e.g., diminishing treatment effect with time).

Indirectness:

There are two types of indirectness.

1. Indirect comparison – occurs when a comparisons of intervention A versus B is not available, but A was compared with C and B was compared with C. Such trials allow indirect comparisons of the magnitude of effect of A versus B. Such evidence is of lower quality than head-to-head comparisons of A and B would provide.
2. Indirect population, intervention, comparator, or outcome – the question being addressed by the authors of a systematic review is different from the available evidence regarding the population, intervention, comparator, or an outcome.

Imprecision:

Results are imprecise when studies include relatively few patients and few events and thus have wide confidence intervals around the estimate of the effect.

1. For dichotomous outcomes

- total (cumulative) sample size is lower than the calculated optimal information size (OIS, comparable to a sample size calculation in a single trial)
- total number of events is less than 300 – a “rule of thumb” based on simulations and dependent on the baseline risk and effect sizes
- 95% confidence interval (or alternative estimate of precision) around the pooled or best estimate of effect includes both negligible effect and appreciable benefit or appreciable harm. GRADE suggests that threshold for "appreciable benefit" or "appreciable harm" that warrants downgrading is a relative risk reduction (RRR) or relative risk increase (RRI) greater than 25%.

Exception

When event rates are very low, 95% confidence intervals around relative effects can be very wide, but 95% confidence intervals around absolute effects may be narrow. Under such circumstances one may not downgrade the quality of evidence for imprecision.

2. For continuous outcomes

- 95% confidence interval includes no effect and the upper or lower confidence limit crosses the minimal important difference (MID), either for benefit of harm.
- if the MID is not known or use of different outcomes measures required calculation of an effect size (ES), we suggest downgrading if the upper or lower confidence limit crosses an effect size of 0.5 in either direction.

Publication Bias:

Publication bias is a systematic underestimate or an overestimate of the underlying beneficial or harmful effect due to the selective publication of studies (publication bias). That is, investigators fail to report studies they have undertaken (typically those that show no effect) or journals are less likely to accept studies that show no effect for publication.

EXPLANATIONS FOR UPGRADING

Strong Association:

When methodologically strong observational studies yield large or very large and consistent estimates of the magnitude of a treatment or exposure effect, we may be confident about the results. In those situations, the weak study design is unlikely to explain all of the apparent benefit or harm, even though observational studies are likely to provide an overestimate of the true effect.

The larger the magnitude of effect, the stronger becomes the evidence.

Magnitude of effect: Effect measure large RR >2 or <0.5
(based on consistent evidence from at least 2 studies, with no plausible confounders): upgrade 1 level
very large RR >5 or <0.2 (based on direct evidence with no major threats to validity): upgrade 2 levels

Effects of all Plausible Confounding:

On occasion, all plausible confounding from observational studies or randomized trials may be working to reduce the demonstrated effect or increase the effect if no effect was observed.
For example, if only sicker patients receive an experimental intervention or exposure, yet they still fare better, it is likely that the actual intervention or exposure effect is larger than the data suggest.

Example 1
A rigorous systematic review of observational studies including a total of 38 million patients demonstrated higher death rates in private for-profit versus private not-for-profit hospitals (Devereaux 2004). One possible bias relates to different disease severity in patients in the two hospital types. It is likely, however, that patients in the not-for-profit hospitals were sicker than those in the for-profit hospitals. Thus, to the extent that residual confounding existed, it would bias results against the not-for-profit hospitals. The second likely bias was the possibility that higher numbers of patients with excellent private insurance coverage could lead to a hospital having more resources and a spill-over effect that would benefit those without such coverage. Since for-profit hospitals are likely to admit a larger proportion of such well-insured patients than not-for-profit hospitals, the bias is once again against the not-for-profit hospitals. Because the plausible biases would all diminish the demonstrated intervention effect, one might consider the evidence from these observational studies as moderate rather than low quality.

Example 2
A parallel situation exists when observational studies have failed to demonstrate an association but all plausible biases would have increased an intervention effect. This situation will usually arise in the exploration of apparent harmful effects. For example, because the hypoglycemic drug phenformin causes lactic acidosis, the related agent metformin is under suspicion for the same toxicity. Nevertheless, very large observational studies have failed to demonstrate an association (Salpeter S, Greyber E, Pasternak G, Salpeter E. Risk of fatal and nonfatal lactic acidosis with metformin use in type 2 diabetes mellitus. Cochrane Database of Systematic Reviews 2007, Issue 4. Art No: CD002967.). Given the likelihood that clinicians would be more alert to lactic acidosis in the presence of the agent and over report its occurrence, one might consider this moderate, or even high quality evidence refuting a causal relationship between typical therapeutic doses of metformin and lactic acidosis.

Only a body of evidence with no important threats to validity should be upgraded.

Dose response relation:

The presence of a dose-response gradient may increase our confidence in the findings of observational studies and thereby increase the quality of evidence.

Only studies with no threats to validity (not downgraded for any reason) can be upgraded.

Example
The observation that, in patients receiving anticoagulation with warfarin, there is a dose response gradient between higher levels of the international normalized ratio (INR), an indicator of the degree of anticoagulation, and an increased risk of bleeding increases our confidence that supratherapeutic anticoagulation levels increase bleeding risk.

Worksheet 3: Evidence Profile
PICO:

Author(s): **Date:**
Question: **Settings:** **Bibliography:**

Quality assessment							Summary of findings					Importance
							No of patients		Effect		Quality	
No of studies	Design	Limitations	Inconsistency	Indirectness	Imprecision	Other considerations	Intervention	Control	Relative (95% CI)	Absolute		

EXAMPLE: Complete response of tumor to chemotherapy											
5	randomized trial	no serious limitations	serious[1]	no serious indirectness	no serious imprecision	Publication bias[4]	216/344	211/344	RR 1.0 (0.92 to 1.1)	0 fewer per 1000 (from 49 fewer to 61 more)	⊕⊕◯◯ LOW	CRITICAL

Mention footnotes here

Worksheet 4: **Draft recommendation for consideration by the guideline panel**

⇨ **Draft recommendation**

⇨ **Values and preferences associated with this recommendation (assume a set of values for each outcome that you considered)**

Example: Values and Preferences
For this recommendation we placed a high value on the prevention of death in an illness with a high case fatality. It places relatively low values on adverse reactions, the development of resistance and costs of treatment.

Overall quality of evidence across all critical outcomes	

⇨ **Judgments about the strength of a recommendation**

Make a judgment using the table below. Add an explanation for your judgment.

Factors that can weaken the strength of a recommendation [Instructions]	**Decision**	**Explanation**
Lower quality evidence [The higher the quality of evidence, the more likely is a strong recommendation.]	☐ **Yes** ☐ **No**	
Uncertainty about the balance of benefits versus harms and burdens [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 is a weak/conditional recommendation warranted.]	☐ **Yes** ☐ **No**	
Uncertainty or differences in values [The greater the variability in values and preferences, or uncertainty in values and preferences, the more likely weak/conditional recommendation warranted.]	☐ **Yes** ☐ **No**	
Uncertainty about whether the net benefits are worth the costs [The higher the costs of an intervention – that is, the more resources consumed – the more likely is a weak/conditional recommendation warranted]	☐ **Yes** ☐ **No**	

Frequent "yes" answers will increase the likelihood of a weak/conditional recommendation

If consensus is not reached by discussion, this table below allows the panel making a recommendation to record their views (votes) about the recommendation related to a specific interventions, based on their analysis of the available evidence, the benefits and downsides, values and preferences and cost. This assessment is then mapped to the strength of recommendation for the use, or non-use, of each intervention.
Insert the number of votes for the recommendation in each category

The GRADE grid for solving disagreement:

GRADE strength	**Strong**	**Weak/ conditional**		**Weak/ conditional**	**Strong**
Assessors view of the balance of desirable and undesirable consequences of the intervention	Desirable consequences clearly outweigh undesirable consequences	Desirable consequences probably outweigh undesirable consequences	Exception	Undesirable consequences probably outweigh desirable consequences	Undesirable consequences clearly outweigh desirable consequences
Recommendation	We recommend to "do something"	We suggest/conditionally recommend to "do something"		We suggest/conditionally recommend to "not do something"	We recommend to "not do something"

Strength of the recommendation	

⇨ **Final recommendation**

⇨ **Remarks**

Definition: Strong vs. weak recommendations

☐ **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.

Implications of strong and weak recommendations

The 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; request discussion if the intervention is not offered
- For clinicians—most patients should receive the recommended course of action
- For policy makers—the recommendation can be adopted as a policy in most situations

The implications of a weak recommendation are:

- For patients—most people in your situation would want the recommended course of action, but many would not
- For clinicians—you should recognize that different choices will be appropriate for different patients and that you must help each patient to arrive at a management decision consistent with her or his values and preferences
- For policy makers—policy making will require substantial debate and involvement of many stakeholders