# Facilitating Due Diligence:
## Specifications to Improve Usability of External Systematic Reviews and Guidelines

Jeffrey S. Harris MD, MPH, MBA
Methodologist,
Kaiser Permanente
National Guideline Program

care management | institute

KAISER PERMANENTE®

# Background, Objectives, Methods

- Background
  - Large internal guideline development effort
  - Intent to be more efficient; focus on implementation
  - Questions about using external resources
- Objectives of this presentation
  - To share experiences in exploring the adoption of outside Systematic Review / Meta-Analyses (SR/MAs) and guidelines (GL)
  - To suggest solutions for discussion and possible action
- Methods
  - Qualitative observations from attempted use, adaptation
  - Collected suggestions, process improvements

# Position, Action, Benefits

- Position
  - Variable Clinical Questions, grading systems, appraisal processes, and reporting/formats precludes interoperability
  - Variance is generally viewed as a sign of lower quality
  - The ability to easily assess or verify the validity of SRs and guidelines allows confident use of “trusted sources.”
- Action
  - Propose standards for evidence identification and appraisal, grading/levels, guideline processes and reporting/format
- Benefits
  - Reduced collective expense and time
  - Reduced variance in recommendations, patient information

# The Quest for Improved Efficiency for Guideline Developers and the Guideline community…

## Goals of GL Programs

- To avoid duplication of effort and expense
- To improve efficiency and consistency
- To find trusted sources for SRs, MAs and guidelines
- To ensure that this process and output are always of high quality

## Current Situation

- Many are doing their own SR/MAs and GLs
- Long /intensive development, differing recommendations
- Many are trying to define “trusted source” by (subjective) checklist
- Analyzing Clinical Questions, search, analysis, recommendations, usability
- Seeking trusted sources

# The Ideal Trusted Source

- Reliability and consistency of process
- Consistency of output from process
- Internal validity of conclusions
- Transparency of process, logic, conflicts
- Explicitness of recommendations
- **Easy to verify quality and acceptability**
  - **Initial, spot check**
  - Found in a repository of assessed SRs or GLs

# The Search for the Trusted Source

- Trust but verify *
  - How were conclusions reached?
  - Is there enough information to verify the logic?
  - How much effort is this ?
    - Using PRISMA and AMSTAR for SRs/MAs
    - Using AGREE II for guidelines
    - Going beyond checklists to granular quality control
  - "It comes from a good organization"
    - Beware the return of eminence-based medicine

*Courtesy of the IAEA; audit and quality control principle

# Due Diligence Requires Granularity
## High level checklists can be subjective…

- Assessment of the completeness and quality of systematic reviews
  - Clinical Questions
  - Search terms, inclusion/exclusion, process, outcome
  - Risk of bias of studies and bodies of evidence
  - Clinical logic of conclusions
- Assessment of processes and outcomes of individual recommendations, collected guidelines
  - Benefits vs harms
  - Preferences/values, cost/tradeoffs

# High-level Qualitative Observations
## from assessments for possible use

- The quality of SRs, MAs and guidelines varies
  - Design and execution
  - Presentation, accessibility
- It requires a ferret to find the methodology, clinical questions, evidence grades, rationale etc.
  - No standard format
    - Masses of basic text
    - Separate documents and sites…
    - Incomplete information
- It took more time to analyze existing materials than it would to do it over from the beginning

# Specific Observations

- Methodology
  - Often inaccessible, vague or obtuse
    - In separate documents, on web sites - some sites password protected
    - Inadequately described, boilerplate, reprint of others methods
- Clinical Questions
  - Often not present or not well-framed
    - Some PICOS elements missing
- Evidence searches incomplete
  - Many studies not listed in PubMed
    - Single database searches
  - Hand searches absent
    - E.g., various recent low back guidelines have from 200 to 1800 references
  - Some abstracts and PubMed incorrectly label studies
    - E.g., “RCTs” and “SRs” are not as advertised

# Specific Observations

- Evidence tables absent in some NGC listings
  - Varying formats and data elements
  - Wide variation in insightfulness
- Summary paragraphs often descriptive rather than analytic
  - Restatement of facts
  - Absent in some NGC listed materials
- Variable evidence ranking schemes
  - Ranking by study design mistaken for quality
  - Inclusion of unanalyzed guidelines, industry material, textbooks, review articles, case reports (MS, pain GLs)

# Specific Observations

- Rating the body of evidence
  - Various schemes in use: the good, the bad and the ugly
    - USPSTF, GRADE de facto standards but different
    - Non-standard “levels” based on both RCTs and observation
    - Aggregations of lower quality studies upgraded
    - Heterogeneous studies blended
  - Not done in some cases
- Many SRs and MAs done by non-clinicians
  - Clinical logic or significance may be missed
- Variable recommendation levels
  - Expert consensus, general acceptance as Level I (ACC, ACEP)

# Observations

- Indirect evidence mistaken for direct evidence
  - Screening without standardized treatment
    - Prostate, breast, colon cancer screening literature
  - Class effects (statins, anti-hypertensives)
  - Inference from basic science (pain management)
  - Intermediate variables (lipid levels, blood pressure)
  - Unrelated variables (pain vs function)
  - Shorter term studies (low back procedures, opioids)
- Recommendations linked to evidence grade or not
  - Separating them can be more explicit  about their basis
- Logic for recommendations not always well explained
- Conflicts not always disclosed or acted on
- Absent or incomplete use, disclosure of external review

# Observations

- Panel process not always clearly described
  - Training, facilitation
  - Recommendation formulation and approval
  - Dissenting views
  - Potential conflicts (e.g. headaches and opioids)
- Single discipline guidelines may be myopic
  - Consensus based on referred populations
  - Population harms, preferences may not be considered
- Many guidelines consider only efficacy, not
  - Harms
  - Effectiveness
  - Patient, population and organizational values
  - Costs and alternative uses of resources

# High Level Solutions

- Transparency
  - Process steps
  - Values of practitioners, patients, organization
  - Potential conflicts of all participants
- Explicitness
  - Data (evidence tables)
  - Process (GRADE-like throughput)
- Specificity
  - Direct evidence, focused clinical questions
- Standardization
  - Variance = low quality

# Empirical Recommendations
## to facilitate due diligence

- Easily accessible methodology
  - Explicit, standard methodology  for SR/MAs (Cochrane)
  - Explicit, standard methodology, recommendation levels for GL (TBD)
    - Include AGREE II elements but without the need to search for them
- Easily found clinical questions
  - Use PICOS format, focused questions, explicit population
  - Single intervention vs natural history; head to head comparisons
- Evidence searches
  - List and diagram search terms and yield
    - Hand searches mandatory
  - Reasonable, standard inclusion, exclusion criteria
  - Search multiple dbs, e.g., EMBASE, CENTRAL, CINAHL, PEDRO…
  - Review abstract for design

# Empirical Recommendations
## to facilitate due diligence

- Evidence appraisal
  - Evidence tables with standard data elements
    - e.g. N, blinding, interventions, dropout, stats, risks of bias
  - Analytic summary paragraphs (elements TBD)
  - Explicit risk of bias evaluations
    - Cochrane RoB with added elements, granularity, levels
    - Automate?
  - Explicit analysis of heterogeneity
    - Do not do MAs if present
  - Standard scheme for assessment of the body of evidence
    - e.g., GRADE - like
      - Resolve nature of harms data vs. levels of evidence
      - Separate strength of recommendation from strength of evidence

# Empirical Recommendations
## to facilitate due diligence

- Net benefit
  - Consider harms vs benefits, from all perspectives 
    - Stakeholder value review or involvement
  - Consider population benefit vs alternatives
  - Resolve issue of low quality of harm studies
- Recommendations
  - Independent
  - Standard levels consistent with evidence (TBD)
    - Expert consensus or general acceptance labeled as such, not Level 1
  - Explain logic concisely but clearly
  - Compare with similar guidelines
    - Explain differences
  - Make recommendations actionable (see Usability below)
  - Make recommendations sequential (see Values appendix)

# Empirical Recommendations
## to improve quality and facilitate due diligence

- Standardized staff and panel training
  - Manuals, case study method, mentoring; describe explicitly
- External protocol and pre-publication review
  - Similar to Cochrane process
  - Standard tools with more granularity re: risk of bias assessment
    - PRISMA/AMSTAR/Cochrane Back Group criteria
    - AGREE II
- Disclosures
  - Staff, authors and reviewers
  - Affiliations and conflicts of interest
- Central source of objectively appraised SRs and MAs
    - E.g., DARE + objective rating
- Central source of appraised guidelines

# Incorporate Values…

- Inconvenience, anxiety, chasing false positives
- Complications
  - What
  - Likelihood
  - Pain and suffering
- Net benefit
  - NNT, NNH
  - Level of clinical benefit
  - Population benefit
- Resource stewardship
  - Best use of resources, screening for population

# Usability Issues

- Sequence/prioritize recommendations
  - Maximize clinical utility, yield, efficacy
  - Improve efficiency, resource stewardship
  - Reduce variance in practice (no “options”)
- Include useful clinical information
  - Class effect generalizability
  - Preference order
  - Dose, duration
- Place extraneous material in separate appendices
  - Didactic material - basic science; current practices

# Summary

- Standardize to improve quality and efficiency
  - Some standards need development
  - Standards adoption…
- Develop repository of appraised guidelines
- Trust but verify
  - Granularity, objective criteria, explanations are good
  - Time and effort vs *de novo* effort
  - If it takes longer to verify than do, it's not efficient
- Beware the return of eminence-based medicine
  - Spot check trusted sources…

# Appendix*

## Values to Inform Recommendations and Implementation

*Do not remove

# Values to Inform Recommendations:
## Diagnostic Testing

- Tests should be performed only if results will affect the course of treatment
- Imaging or testing should generally be done to confirm a clinical impression prior to surgery or other major, invasive treatment
  - No fishing…
    - Do you have a fracture?
      - Go fish…

# Values to Inform Recommendations:
## Relative Effectiveness

- Treatments should improve on the natural history of the disorder
  - In many cases the natural history is recovery without treatment
- When there are options for testing or treatment available, choose the option supported by [the best] clinical and statistical significance

# Values to Inform Recommendations:
## Invasive treatment

- Invasive treatment should, in almost all cases, be preceded by adequate conservative treatment
- Invasive treatment should be performed
  - If conservative treatment does not improve the health problem
  - If there is evidence of net effectiveness for a specific diagnosis, indication, and situation
- The more invasive and permanent the invasive tests or treatments
  - The more caution should be exercised
  - The stronger the evidence of net efficacy should be

# Values to Inform Recommendations:
## Cost Effectiveness

- The more costly the test or intervention
  - The more caution should be generally exercised prior to ordering the test or treatment
  - The stronger the evidence of efficacy should be
- When two testing or treatment methods appear equivalent, the most cost-effective method is preferred

# Values for Implementation:
## Shared Decision Making

- Testing and treatment decisions should be a collaboration between clinician and patient with full disclosure of benefits and risks.
  - The best treatment strategy should be recommended.
  - In cases where the patient cedes that judgment to the clinician, the clinician's judgment as to the best treatment strategy should be implemented.

# Values for Implementation:
## Care Management

- Management
  - Treatment should have specific, objective goals
    - Monitored for achievement of those goals within a reasonable time
    - Failure to achieve a goal does not change the risk/benefit calculation for a subsequent test or treatment
- Disability management
  - Treatment should not create dependence or functional disability