



Weighing Value and Quality of Evidence: The Knowns and Unknowns

***Dan Ollendorf, MPH
Chief Review Officer
Institute for Clinical and Economic Review***

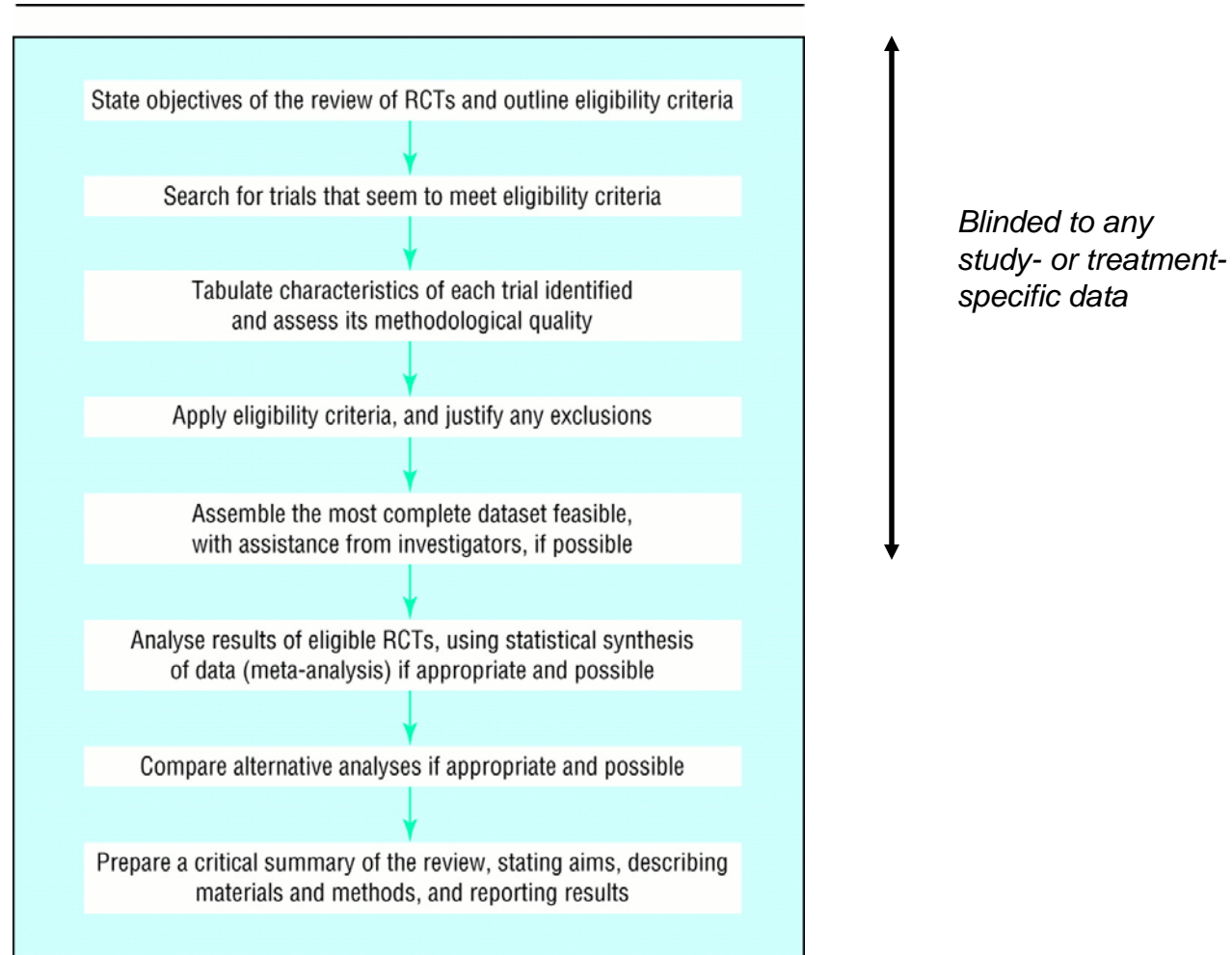
Comparative Effectiveness Research

- Goal → judgments on benefits/harms of multiple treatment alternatives
- Reality → explicit comparisons of alternatives largely nonexistent
- Compromise → detailed and thorough analysis of what can be gleaned from current evidence

Evidence Synthesis

- Most clinical reviews conducted in “journalistic style”:
 - Data chosen *post hoc* and selectively to support answers to key questions
- Systematic reviews:
 - Based on *a priori* set of key questions
 - Systematic and reproducible selection process
 - Data collected drive answers to key questions

Systematic Review Flowchart



Greenhalgh, BMJ 1997

Benefits of Systematic Review*

- Bias-limiting methods for identifying and rejecting studies
- Quick assimilation of large amounts of information
- Formal evaluation of evidence base for generalizability and consistency
- Quantitative techniques to assess heterogeneity

*Chalmers I, Altman DG, Eds. *Systematic Reviews*. London: BMJ Publishing Group, 1995.

Sources of Review Data

- Electronic literature databases (e.g., MEDLINE, EMBASE)
- Cochrane registers
- “Grey literature”
- Manual review of references
- Other unpublished sources and/or data
- Foreign-language literature

Rating Evidence Quality

- Multiple rating systems developed (GRADE, STROBE, etc.)
- Key considerations common to all:
 - Study design
 - Sample selection and methods
 - Precision of estimates
 - Consistency across studies
 - Generalizability

Qualitative Evidence Assessment

- Structure and formal methods add to credibility of systematic reviews
- Qualitative concerns must still be addressed:
 - Measures/outcomes of primary interest to audience
 - “Most important” studies → based on design, findings, generalizability, etc.
 - Judgments on whether between-treatment differences are clinically meaningful

Quantitative Evidence Assessment

- Meta-analysis an important tool for evidence assessment:
 - Improved statistical power in series of small studies
 - Improved precision of effect measures
 - Comparison of multiple treatment alternatives
 - Examination of statistical heterogeneity

Meta-Analysis

- Subject to over-use:
 - Unrepresentative or biased set of studies
 - Methods, measurement of outcomes too diverse
 - May not be any better than 1 large well-designed study!
- Subject to under-use*:
 - Fear of heterogeneity
 - Methods to explore/adjust for inconsistencies under-utilized

Decision Analysis

- Use of simulation models as complement to evidence synthesis:
 - Extend data on short-term or surrogate endpoints to potential long-term effects
 - Evaluate costs and effects for multiple populations
 - Assess potential impact of variations in treatment approach
 - Multiple methods to present uncertainty in estimates

Key Clinical Model Outputs for Technology Appraisals

- Intermediate outcomes (e.g., triglycerides)
- Long-term outcomes (e.g., CHD, diabetes, stroke)
- Changes in diagnostic/treatment pathway
- Mortality
- Life expectancy
- Quality of life/utility

Measures of Economic Impact for Technology Appraisals

- Impact on service use
 - Tests, hospitalizations
- Cost to reduce adverse outcomes
- Cost to achieve same clinical goal
- Cost per life year gained
- Cost per QALY
- Budget impact per 1,000
- System issues, e.g. manpower, acceptability

Value-of-Information Analysis

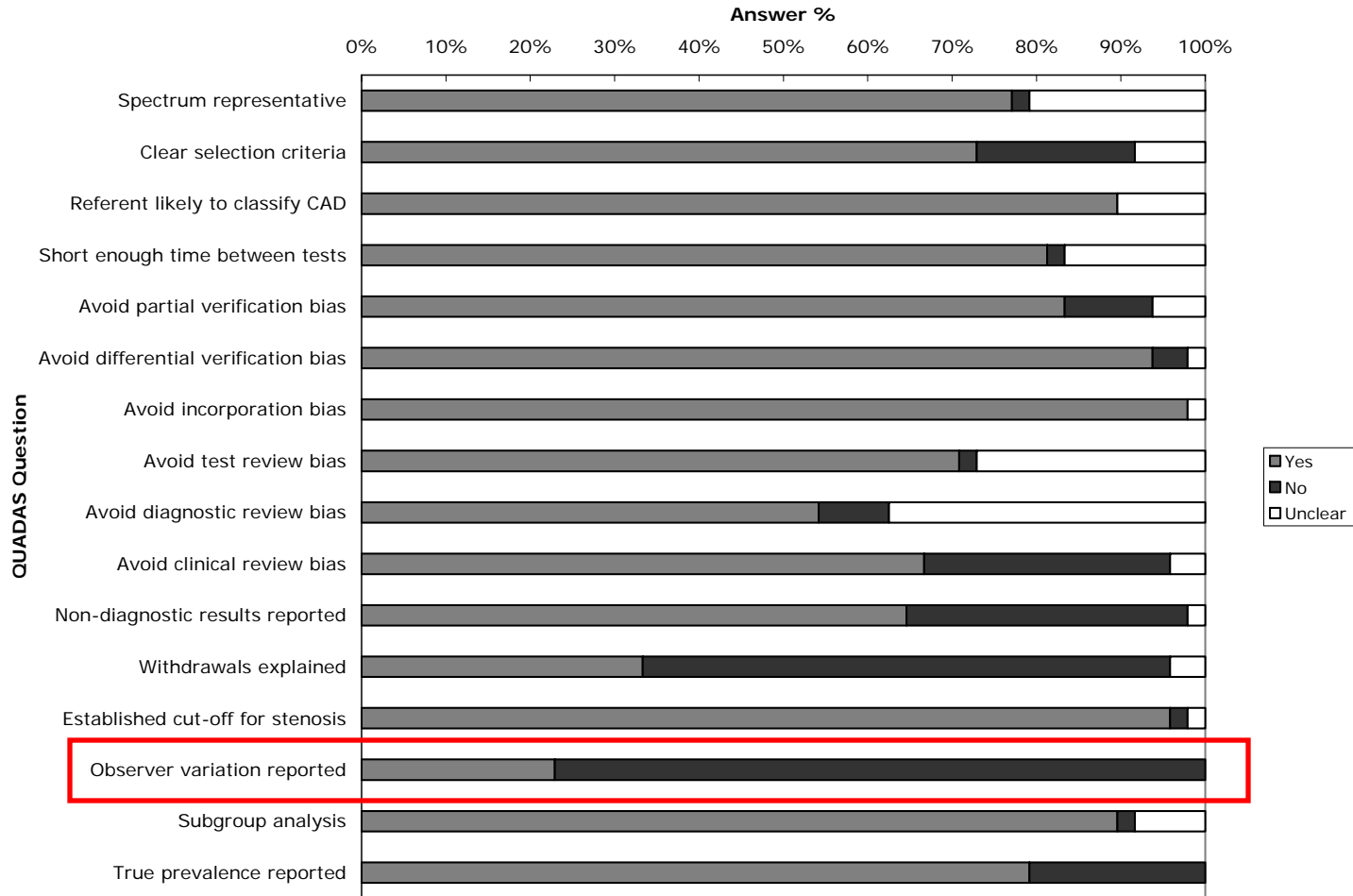
- Quantitative assessment of costs and benefits of resolving uncertainty
 - Combined measure of probability of error and its consequences
 - EVPI: maximum value of resolving all uncertainty for individual patient
 - Can be adjusted to reflect population size, timeframe for use of new evidence, benefits if uncertainty resolved for certain parameters, etc.

Case Study: Coronary CT Angiography (CCTA)

- Widespread growth of this non-invasive test for diagnosis of coronary artery disease
- Questions regarding impact on testing rates, candidate populations, downstream effects
- Also, issues common to all diagnostic tests*
 - Rapid evolution
 - Interpreter variability
 - Spectrum bias
 - “Best hands”

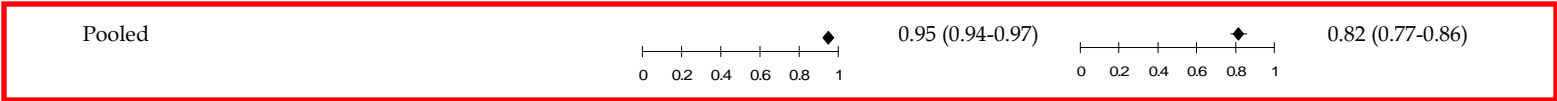
*Pearson SD, Health Affairs, 2008

Evidence Quality

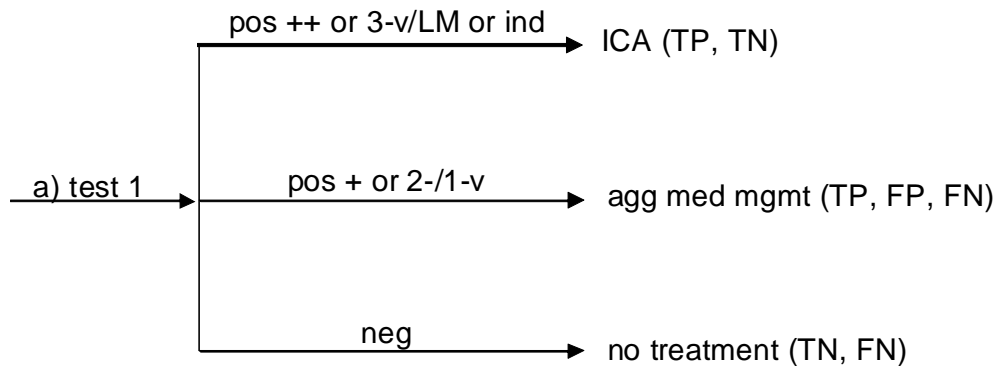


Evidence Synthesis

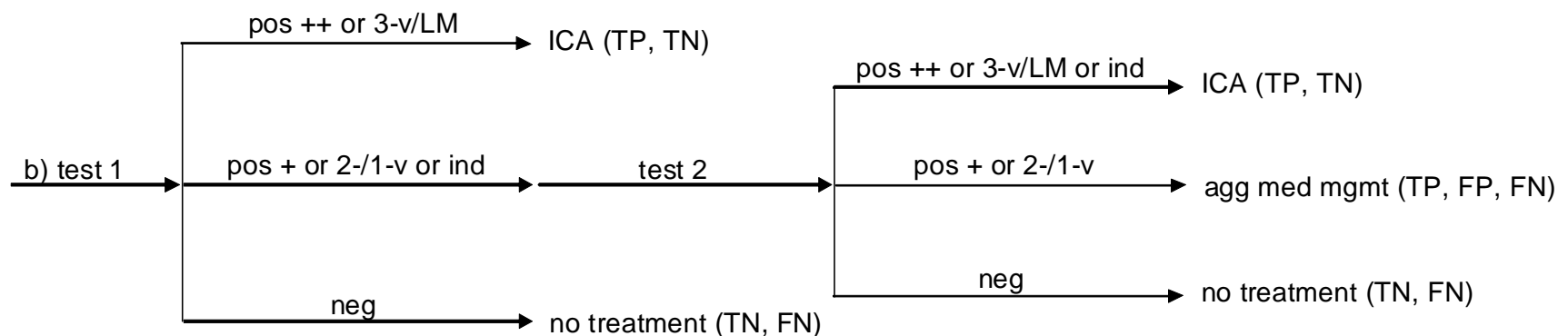
Author, Year	TP	FP	FN	TN	Sensitivity	Sensitivity	Specificity	Specificity
Achenbach (a) 2008	34	27	7	35		0.83 (0.68-0.93)		0.56 (0.43-0.69)
Achenbach (b) 2008	39	9	1	51		0.98 (0.87-1.00)		0.85 (0.73-0.93)
Bayrak 2008	64	4	0	32		1.00 (0.94-1.00)		0.89 (0.74-0.97)
Budoff 2008	52	30	3	142		0.95 (0.85-0.99)		0.83 (0.76-0.88)
Cademartiri (1) 2008	82	32	2	29		0.98 (0.92-1.00)		0.48 (0.35-0.61)
Cademartiri (2) 2007	20	1	0	51		1.00 (0.83-1.00)		0.98 (0.90-1.00)
Ehara 2006	59	3	1	6		0.98 (0.91-1.00)		0.67 (0.30-0.93)
Fine 2006	35	5	2	24		0.95 (0.82-0.99)		0.83 (0.64-0.94)
Ghostine 2006	28	2	1	35		0.97 (0.82-1.00)		0.95 (0.82-0.99)
Hacker 2007	11	7	2	10		0.85 (0.55-0.98)		0.59 (0.33-0.82)
Husmann 2008	16	20	2	119		0.89 (0.65-0.99)		0.86 (0.79-0.91)
Johnson (2) 2007	17	2	0	16		1.00 (0.80-1.00)		0.89 (0.65-0.99)
Leber 2005	22	7	3	17		0.88 (0.69-0.97)		0.71 (0.49-0.87)
Leber 2007	20	9	1	60		0.95 (0.76-1.00)		0.87 (0.77-0.94)
Leschka 2005	47	0	0	20		1.00 (0.92-1.00)		1.00 (0.83-1.00)
Meijboom (2) 2007	69	7	3	37		0.96 (0.88-0.99)		0.84 (0.70-0.93)
Meijboom (3) 2007	28	1	0	4		1.00 (0.88-1.00)		0.80 (0.28-0.99)
Miller 2008	136	16	28	111		0.83 (0.76-0.88)		0.87 (0.80-0.93)
Mollet 2005	38	2	0	12		1.00 (0.91-1.00)		0.86 (0.57-0.98)
Muhlenbruch 2007	44	3	1	3		0.98 (0.88-1.00)		0.50 (0.12-0.88)
Nikolaou 2006	38	10	1	23		0.97 (0.87-1.00)		0.70 (0.51-0.84)
Oncel 2007	62	0	0	18		1.00 (0.94-1.00)		1.00 (0.81-1.00)
Plass 2006	40	1	0	9		1.00 (0.91-1.00)		0.90 (0.55-1.00)
Pugliese 2006	25	1	0	9		1.00 (0.86-1.00)		0.90 (0.55-1.00)
Pugliese 2008	38	0	0	13		1.00 (0.91-1.00)		1.00 (0.75-1.00)
Pundziute 2008	53	7	1	42		0.98 (0.90-1.00)		0.86 (0.73-0.94)
Raff 2005	38	3	2	27		0.95 (0.83-0.99)		0.90 (0.73-0.98)
Ropers D 2006	25	8	1	50		0.96 (0.80-1.00)		0.86 (0.75-0.94)
Ropers U 2007	41	11	1	47		0.98 (0.87-1.00)		0.81 (0.69-0.90)
Rubinshtein (1) 2007	26	6	1	70		0.96 (0.81-1.00)		0.92 (0.84-0.97)
Scheffel 2006	14	0	1	15		0.93 (0.68-1.00)		1.00 (0.78-1.00)
Schuijf (3) 2006	27	6	0	25		1.00 (0.87-1.00)		0.81 (0.63-0.93)
Shabestiri 2007	104	15	4	20		0.96 (0.91-0.99)		0.57 (0.39-0.74)
Shapiro 2007	28	3	1	5		0.97 (0.82-1.00)		0.63 (0.24-0.91)



Decision-Analytic Model: Outpatient Setting



Extend data on test characteristics to downstream effects



Model Results: 30% CAD Prevalence

Strategy	Effectiveness (QALY)	Costs (\$)
SECHO-SPECT	15.140	7,576
CCTA-SECHO	15.146	7,605
SECHO-CCTA	15.151	7,343
CCTA-SPECT	15.154	7,911.
SPECT-CCTA	15.157	8,077
SECHO	15.167	7,998
SPECT	15.172	9,051
CCTA	15.183	8,207

C/E vs. Stress-ECHO → ~\$13,000/QALY

Model Results: 10% CAD Prevalence

Strategy	Effectiveness (QALY)	Costs (\$)
SECHO	16.012	4,543
CCTA-SECHO	16.014	3,962
SECHO-SPECT	16.014	4,068
SECHO-CCTA	16.015	3,831
CCTA-SPECT	16.017	4,175
CCTA	16.018	4,645
SPECT-CCTA	16.024	4,450
SPECT	16.030	5,633

C/E vs. Stress-ECHO → ~\$17,000/QALY

Issues/Challenges with Evidence Synthesis

- Weighing evidence from RCTs vs. observational studies:
 - Significant variation in approach by region/setting
- Tension between responsible summary of evidence and applying judgment
 - When is a difference a difference?
 - Is the difference important?
- Presentation of results in useful format for variety of decision-makers

Integrated Evidence Rating: CCTA vs. Alternative Strategies for Stable Chest Pain, Outpatient Presentation

<i>Comparative Clinical Effectiveness</i>	Superior: A	Aa	Ab	Ac
	Incremental: B	Ba	Bb	Bc
	Comparable: C	Ca	Cb	Cc
	Unproven/Potential: U/P	Ua	CCTA=Ub	Uc
	Insufficient: I	I	I	I
		a High	b Reasonable/Comp	c Low
		<i>Comparative Value</i>		

What's Next?

- Systematic review and modeling useful for characterizing the “knowns” as well as the “unknowns” about *current* evidence
- Remainder of day will be spent learning about methods for generating *new* evidence
- Techniques vary in terms of intensity, duration, cost, and control of bias:
 - As with appraisals of current evidence, key questions will help drive study design

Questions?

Recommended Readings

- **Systematic Review**

- Cook DJ, Mulrow CD, Haynes RB. Systematic reviews: synthesis of best evidence for clinical decisions. *Ann Intern Med* 1997;126:376-80.
- Slutsky J, Atkins D, Chang S, et al. Comparing medical interventions: AHRQ and the Effective Health Care Program. In: Agency for Healthcare Research and Quality. Methods Guide for Comparative Effectiveness Reviews [posted November 2008]. Rockville, MD.

- **Grading Evidence**

- The GRADE Working Group. Grading quality of evidence and strength of recommendations. *BMJ* 2004;doi:10.1136/bmj.328.7454.1490.
- Atkins D, Eccles M, Flottorp S, et al. Systems for grading the quality of evidence and the strength of recommendations: critical appraisal of existing approaches. *BMC Health Services Research* 2004;doi:10.1186/1472.6963.4.38

- **Value of Information Analysis**

- Felli JC, Hazen GB. Sensitivity analysis and the expected value of perfect information. *Med Decis Making* 1998;18:95-109.
- Claxton K. Exploring uncertainty in cost-effectiveness analysis. *Pharmacoeconomics* 2008;26:781-98.

- **Case Study**

- Institute for Clinical and Economic Review. CCTA Final Appraisal Report. Access at: <http://www.icer-review.org/index.php/ccta.html>