American Evaluation Association
Orlando, Florida November 13 2009

Methods & Tools for Comparative Effectiveness Research

J. Michael Menke

Psychology, Evaluation Group for Analysis of Data, Pharmacoeconomics

University of Arizona

menke@email.arizona.edu

Foreground

menke@email.arizona.edu

The Demise of US Health Policy Agencies

- OTA's health program, 1975
- National Center for HC Technologies, 1978
- IOM's Council on HC Tech Assessment, 1984
- AHCPR, 1989

All dead by 2009

menke@email.arizona.edu

déjà vu?
US Health Policy Agency in 2009?

- American Recovery and Reinvestment Act
- $1.1 billion
- Conrad – Baucus Senate Bill
- HC Comparative Effectiveness Research Institute
- How can it / will the CER work?
- (Re)Inspired by a paper by Gail Wilensky (2006)

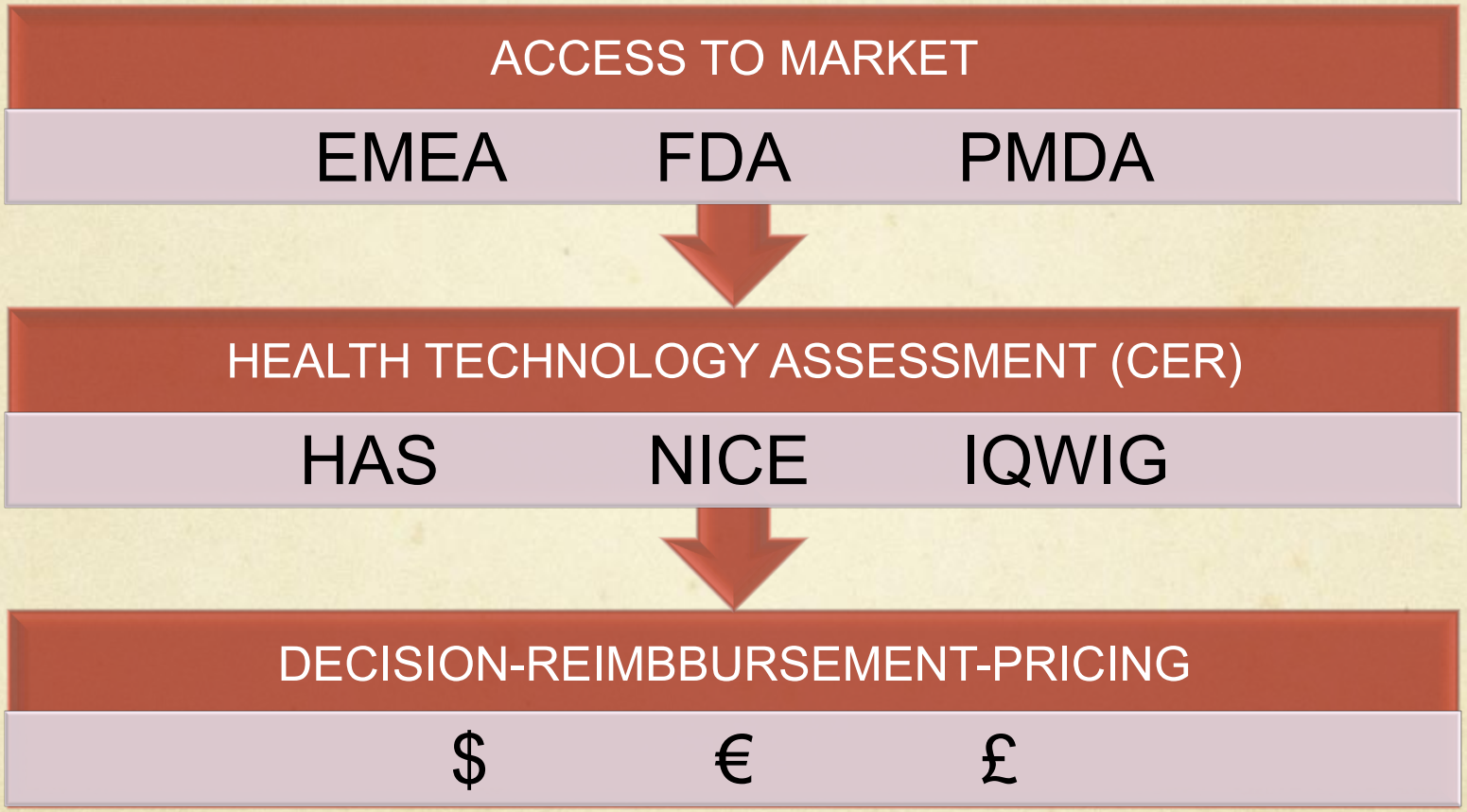
menke@email.arizona.edu

CER Well-established in Other Countries

- UK: NICE (National Institute for Health and Clinical Excellence)
- France: HAS (Haute Autorite de Sante)
- Germany: IQWiG (Institute fur Qualtat und Wirtschaftlichkeit im Gesundhetiswesen)
- Australia: (Pharmaceutical Benefits Scheme)
- Some Canadian provinces

menke@email.arizona.edu

Innovation



menke@email.arizona.edu

France: Comparative Effectiveness

1	• Efficacy to effectiveness • Phase 2 to 3
2	• Good enough to be reimbursed?
3	• Better than other treatments?

menke@email.arizona.edu

A Hot Potato

- Free markets do not apply to health care
  - “Let market forces work” objection to US CER
  - The guy who orders it does not pay for it
- “If medical care had been any other industry it would have failed years ago.”
  - Health care lacks transparency
  - Medical system uncertainty
- Health care system structure is highly resistant to change
  - Even though US medical innovation is so renowned for innovation

menke@email.arizona.edu

A Hotter Potato

- Valuing human life in $
  - QALY & cost/QALY
- Rule of rescue
- Americans not coping well with mortality
- Forgotten meaning / purpose of insurance
  - History of health / life insurance
- Blending of risk with disease (Aronoff)

menke@email.arizona.edu

Purpose of CER

- To tell us what works, when, and for whom? (well… maybe)
- To aid in making informed clinical and health policy decisions

menke@email.arizona.edu

Map of Presentation

Question >	Evidence Synthesis >	Comparison & Decide >	Implement or Clean-up
PIV	Smoking cessation	Bayesian mixtures	EVPI
		TreeAge decision modeling	
	Low back pain	Effect size to probability Different kinds of distributions Measurement CEA – units CUA – QALY's Economics	

menke@email.arizona.edu

Innovative Systems and System Innovations to Improve Lives

- Must evaluate
  - Perform as intended
  - Are they worth the cost?
- Evaluators need to capture data to
  - Inform policy
  - Inform service-level decision-making
- Continue or terminate (summative evaluation)
- Steer (formative evaluation)

menke@email.arizona.edu

An Innovation may Work in a Complex Manner

- Some helped some not
- Some helped and some hurt
- Which ones patients?
  - Does an innovation cost more – or less?
  - Too often, the comparison is to its previous state or to a control state.

menke@email.arizona.edu

A Standard of Comparison

- Effect sizes comparing treatments to control are insufficient to decide among treatments
- A new treatment must be compared to something
  - Progression of science
  - Perhaps every intervention does better than nothing
- For health care innovation, the best comparison is the current standard of care.

menke@email.arizona.edu

Questions Remaining after Evaluation

- Has the uncertainty been reduced enough to make a decision after evaluation?
- Or do we need to know more for policy implementation?
- Can we help clinical decision making?
- Will a specific particular client benefit?

Health technology assessment methods can help answer these questions

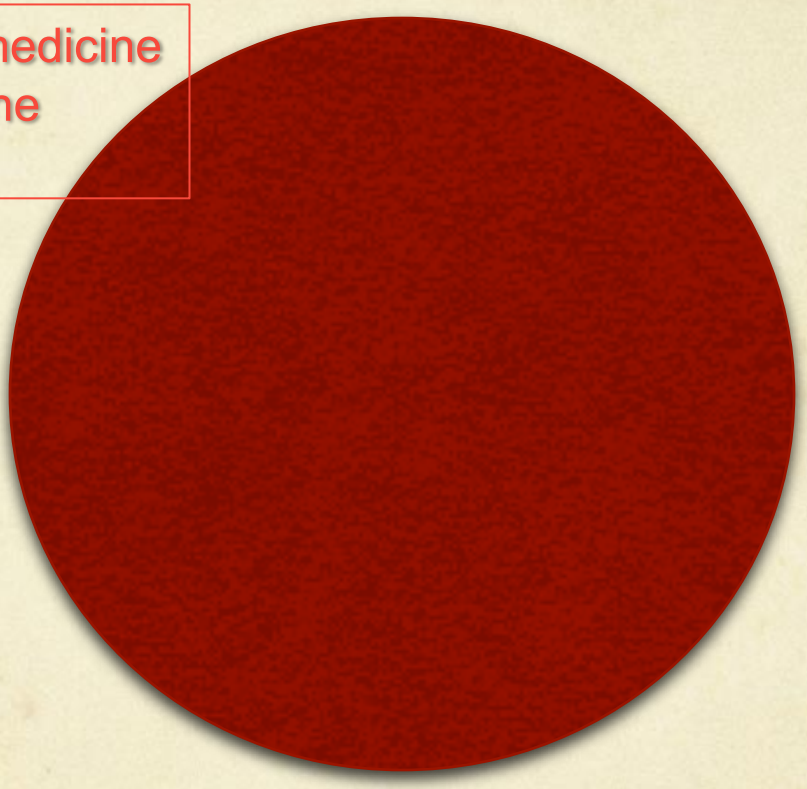
menke@email.arizona.edu

Evaluating Medical Innovations

menke@email.arizona.edu

The Challenge for Alternative Medicine

The effect size of alternative medicine per a single condition, within the larger health care system

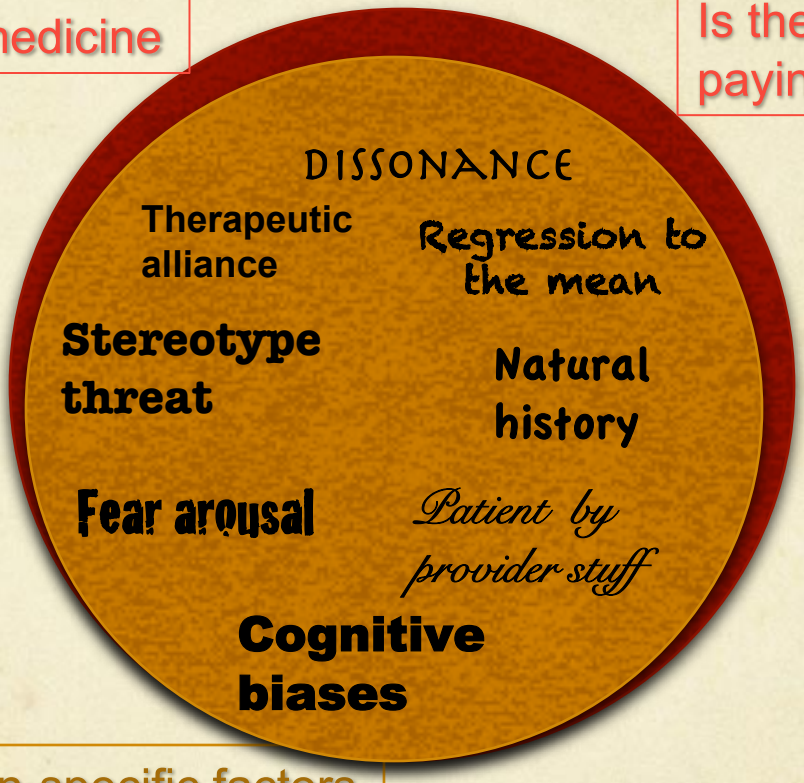


menke@email.arizona.edu

The Challenge for Alternative Medicine: How Big Is the Margin?

The effect size of alternative medicine

Is the margin worth paying for?



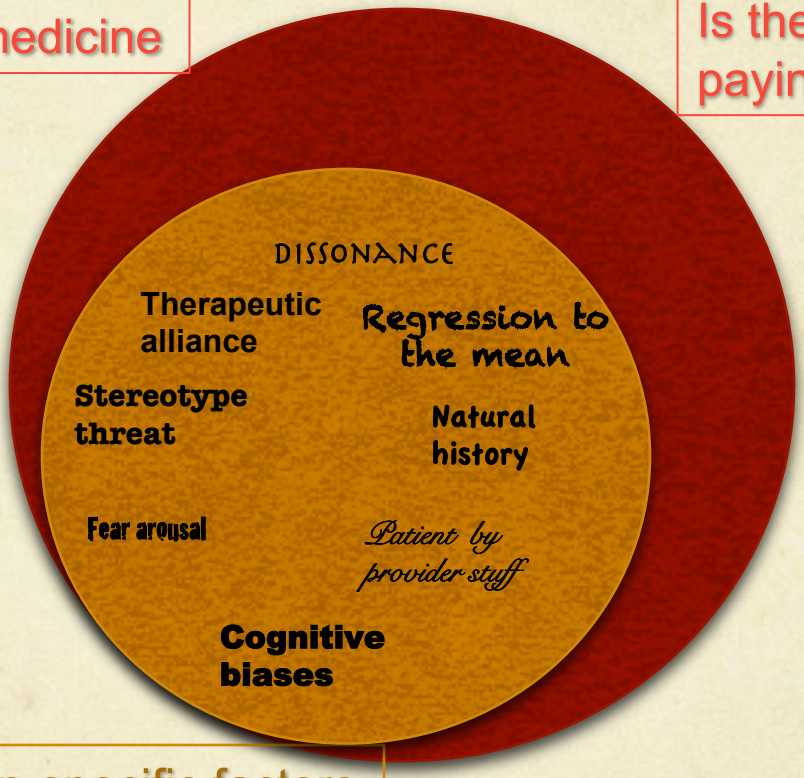
The effect size of non-specific factors

menke@email.arizona.edu

The Challenge for Alternative Medicine: How Big Is the Margin?

The effect size of alternative medicine

Is the margin worth paying for?



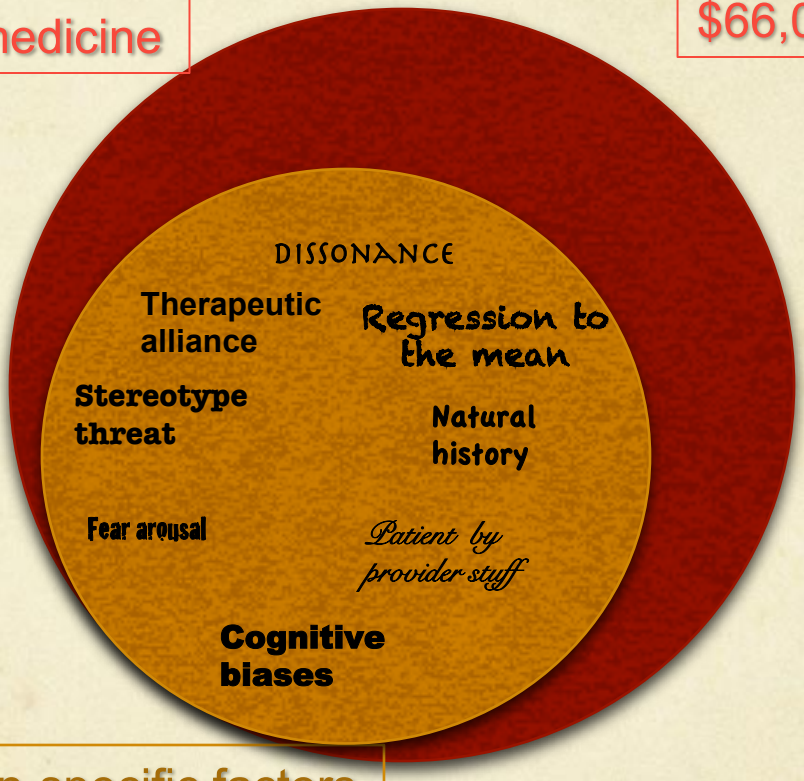
The effect size of non-specific factors

menke@email.arizona.edu

How Much Does the Margin Cost?

The effect size of alternative medicine

$66,000 / QALY?



The effect size of non-specific factors

menke@email.arizona.edu

But

- Reducing demand for expensive medical care is a win!
- But still, a “winner” can be
  - More directed,
  - More wisely referred to,
  - Improved, and
  - Replaced with better, cheaper alternatives

menke@email.arizona.edu

Risks are Becoming Diseases

- High blood pressure
- Cancer survivors
- Practitioner provided prevention
  - “Not so fast!”
  - Unaffordable, even when cheap

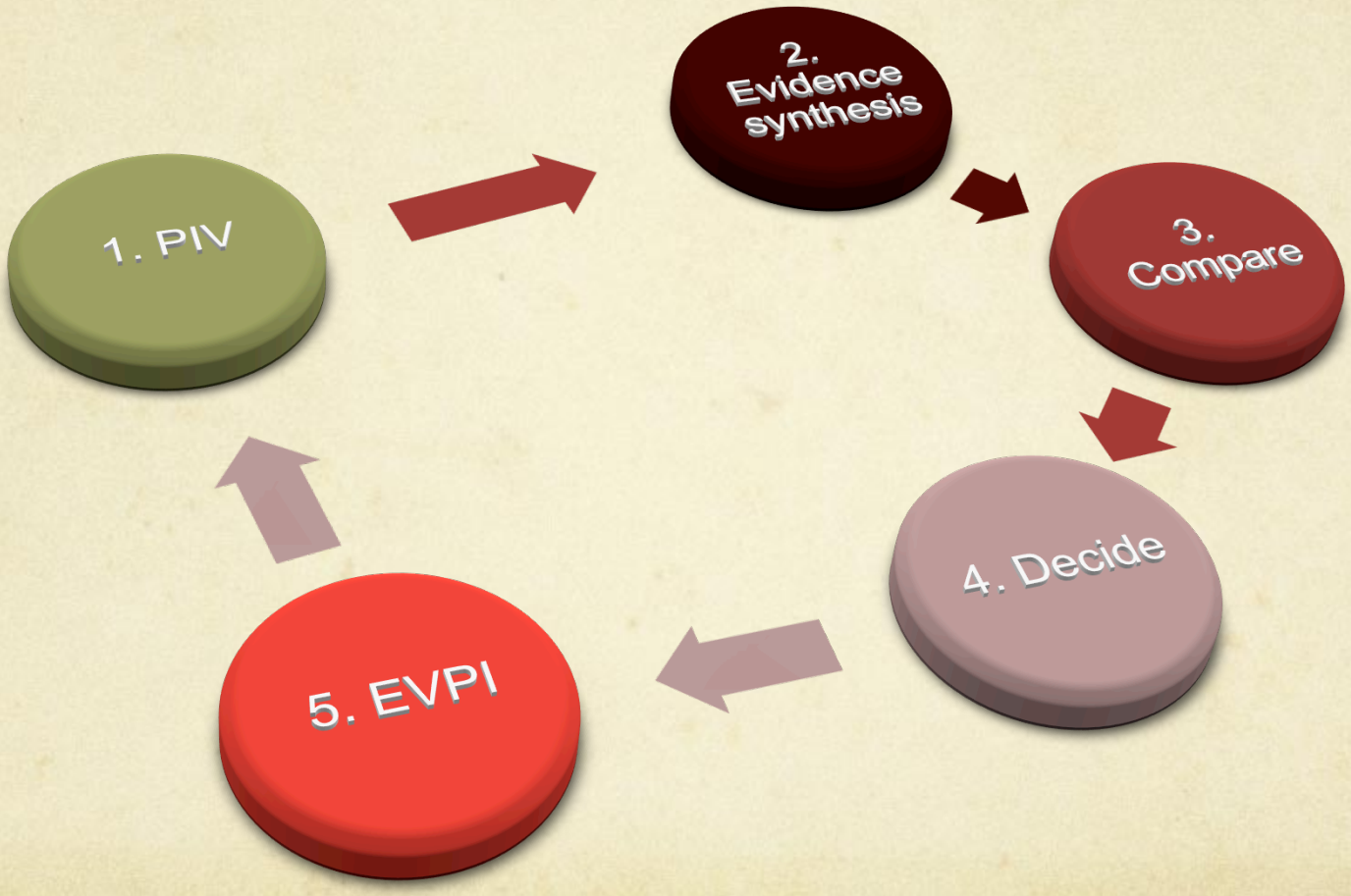
menke@email.arizona.edu

CER Guiding Principles

UK / NICE	US – imputed cynically
Robust (for what?)	Yep, for efficacy, but not for informing policy and clinical decision-making
Inclusive	Divisive
Transparent	Opaque
Independent of financial interests	Industry sponsored
Timely	Working on it! C-Path for fast-tracking

menke@email.arizona.edu

CER Steps



menke@email.arizona.edu

STEP 1. Prior Information Value, *ex ante Valuation*

- Is the problem important enough to warrant reimbursement in by a public or semi-public scheme?
  - A *guaranteed* treatment for athlete's foot.
  - Cost: $1 million per patient
- What is the PIV (prior information value)?
  - Larson RC, Kaplan EH. Decision-oriented approaches to program evaluation. *New Directions for Program Evaluation: Evaluation of Complex Systems. 1981(10):49-68.*
- How valuable is the solution a particular health problem?

menke@email.arizona.edu

STEP 2. Evidence Synthesis

- Synthesized evidence must be usable
- The result must assist & not delay and obfuscate decision-making
- In effect, the evidence should reduce system uncertainty

menke@email.arizona.edu

Synthesizing Research By

- Meta-analyses
- Systematic reviews
- Cochrane collaboration
- Bayesian mixture-method

menke@email.arizona.edu

Cochrane Collaboration

- Archie Cochrane's call for systematic, up-to-date reviews of all relevant RCTs of health care
- Originally for reviews of controlled trials in pregnancy and childbirth
- To support the UK National Health Service
- Cochrane Centre' opened in Oxford, England in October 1992
- October 1993 – first Cochrane Colloquium - 77 people from eleven countries co-founded 'The Cochrane Collaboration'
- Currently > 5,000 health care researchers, providers, policy makers, managers, consumers and educators

menke@email.arizona.edu

Cochrane Collaboration

- Meta-analyses & systematic reviews
- Based on synthesized NHST research, so has many of the weaknesses of NHST
  - Avoids Type I error
  - More likely to make Type II error
    - (more likely to rule out an effective program or treatment)
  - Comparison groups vary (WLC, placebo, no-tx)
- Conclusions are often not informative, or do not address which innovation is better.

menke@email.arizona.edu

Cochrane Sample Summaries

Telephone: **"Our review of trials found telephone counseling to be effective; multiple sessions are likely to be most helpful."**

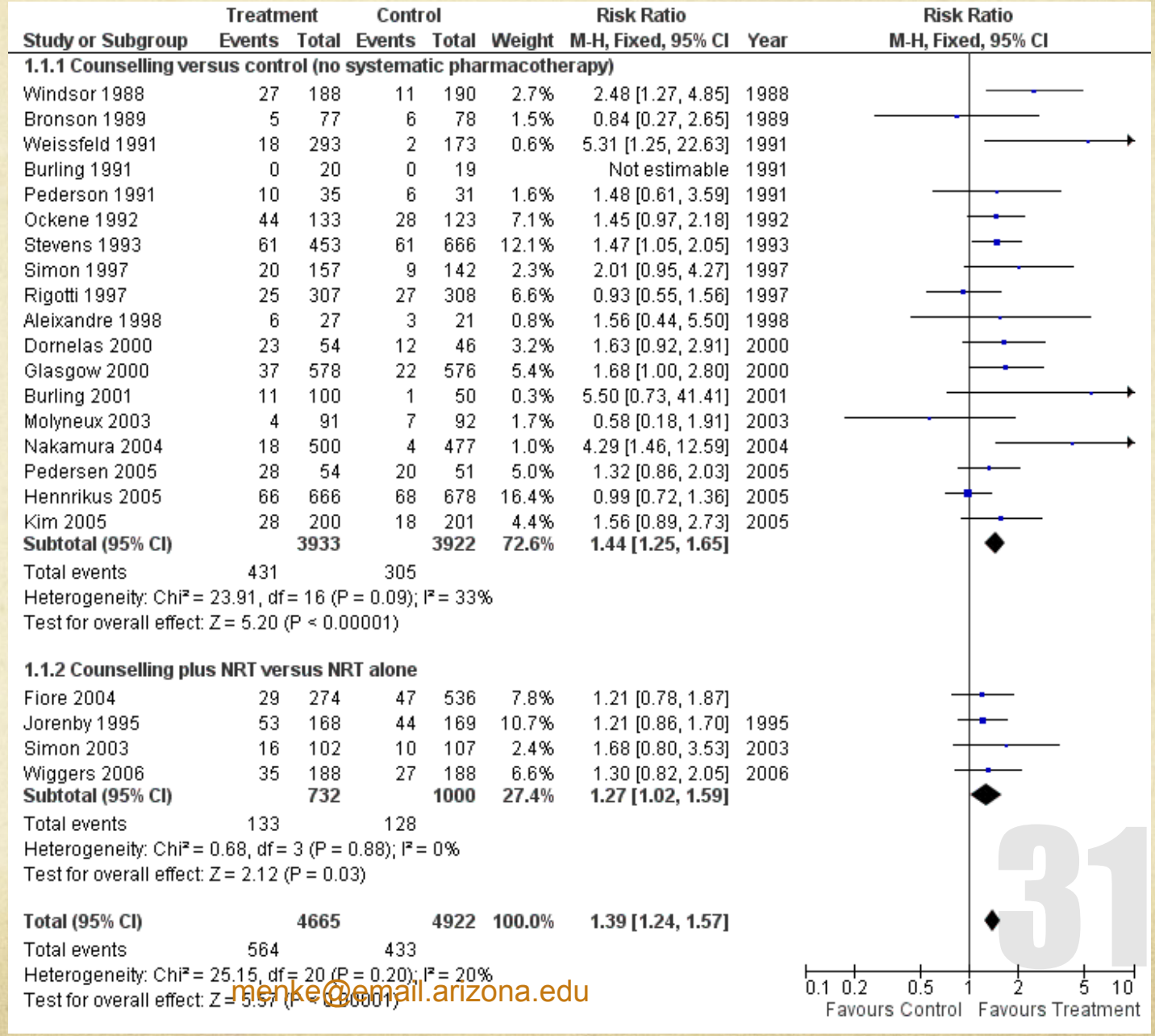
Physician: **"when doctors provide brief simple advice about quitting smoking this increases the likelihood that someone who smokes will successfully quit and remain a nonsmoker 12 months later. More intensive advice may result in slightly higher rates of quitting." [p. 2]**

Individual counseling: **"The review found that individual counseling could help smokers quit, but there was not enough evidence about whether more intensive counseling was better."**

menke@email.arizona.edu

Forest Plot

Figure 2. Forest plot of comparison: 1 Individual counselling compared to minimal contact control, outcome: 1.1 Smoking cessation at longest follow-up.



Study or Subgroup	Treatment Events	Total	Control Events	Total	Weight	Risk Ratio M-H, Fixed, 95% CI	Year
1.1.1 Counselling versus control (no systematic pharmacotherapy)							
Windsor 1988	27	188	11	190	2.7%	2.48 [1.27, 4.85]	1988
Bronson 1989	5	77	6	78	1.5%	0.84 [0.27, 2.65]	1989
Weissfeld 1991	18	293	2	173	0.6%	5.31 [1.25, 22.63]	1991
Burling 1991	0	20	0	19		Not estimable	1991
Pederson 1991	10	35	6	31	1.6%	1.48 [0.61, 3.59]	1991
Ockene 1992	44	133	28	123	7.1%	1.45 [0.97, 2.18]	1992
Stevens 1993	61	453	61	666	12.1%	1.47 [1.05, 2.05]	1993
Simon 1997	20	157	9	142	2.3%	2.01 [0.95, 4.27]	1997
Rigotti 1997	25	307	27	308	6.6%	0.93 [0.55, 1.56]	1997
Aleixandre 1998	6	27	3	21	0.8%	1.56 [0.44, 5.50]	1998
Dornelas 2000	23	54	12	46	3.2%	1.63 [0.92, 2.91]	2000
Glasgow 2000	37	578	22	576	5.4%	1.68 [1.00, 2.80]	2000
Burling 2001	11	100	1	50	0.3%	5.50 [0.73, 41.41]	2001
Molyneux 2003	4	91	7	92	1.7%	0.58 [0.18, 1.91]	2003
Nakamura 2004	18	500	4	477	1.0%	4.29 [1.46, 12.59]	2004
Pedersen 2005	28	54	20	51	5.0%	1.32 [0.86, 2.03]	2005
Hennrikus 2005	66	666	68	678	16.4%	0.99 [0.72, 1.36]	2005
Kim 2005	28	200	18	201	4.4%	1.56 [0.89, 2.73]	2005
Subtotal (95% CI)		**3933**		**3922**	**72.6%**	**1.44 [1.25, 1.65]**	
Total events	431		305				
Heterogeneity: $Chi^2 = 23.91$, df = 16 (P = 0.09); $I^2 = 33\%$							
Test for overall effect: Z = 5.20 (P < 0.00001)							
1.1.2 Counselling plus NRT versus NRT alone							
Fiore 2004	29	274	47	536	7.8%	1.21 [0.78, 1.87]	
Jorenby 1995	53	168	44	169	10.7%	1.21 [0.86, 1.70]	1995
Simon 2003	16	102	10	107	2.4%	1.68 [0.80, 3.53]	2003
Wiggers 2006	35	188	27	188	6.6%	1.30 [0.82, 2.05]	2006
Subtotal (95% CI)		**732**		**1000**	**27.4%**	**1.27 [1.02, 1.59]**	
Total events	133		128				
Heterogeneity: $Chi^2 = 0.68$, df = 3 (P = 0.88); $I^2 = 0\%$							
Test for overall effect: Z = 2.12 (P = 0.03)							
Total (95% CI)		**4665**		**4922**	**100.0%**	**1.39 [1.24, 1.57]**	
Total events	564		433				
Heterogeneity: $Chi^2 = 25.15$, df = 20 (P = 0.20); $I^2 = 20\%$							
Test for overall effect: Z = 5.57 (P < 0.00001)							

Risk Ratio M-H, Fixed, 95% CI

0.1 0.2 0.5 1 2 5 10

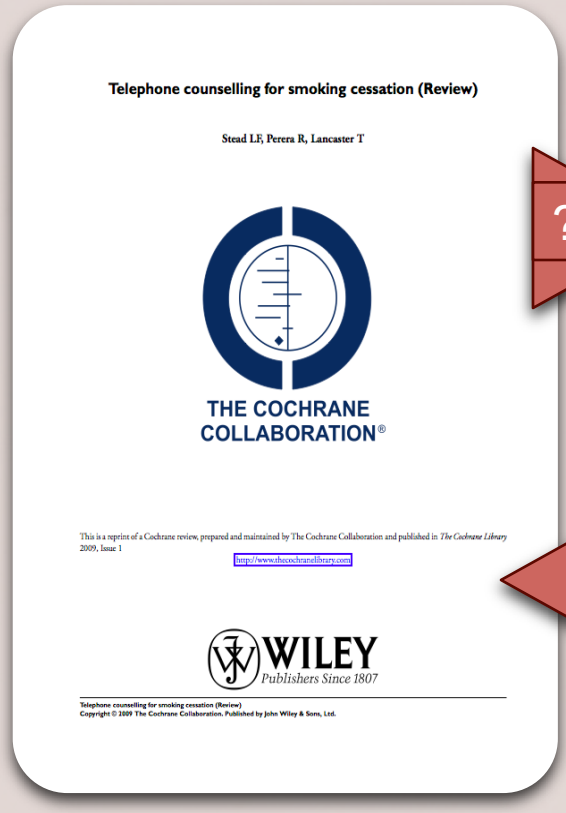
Favours Control Favours Treatment

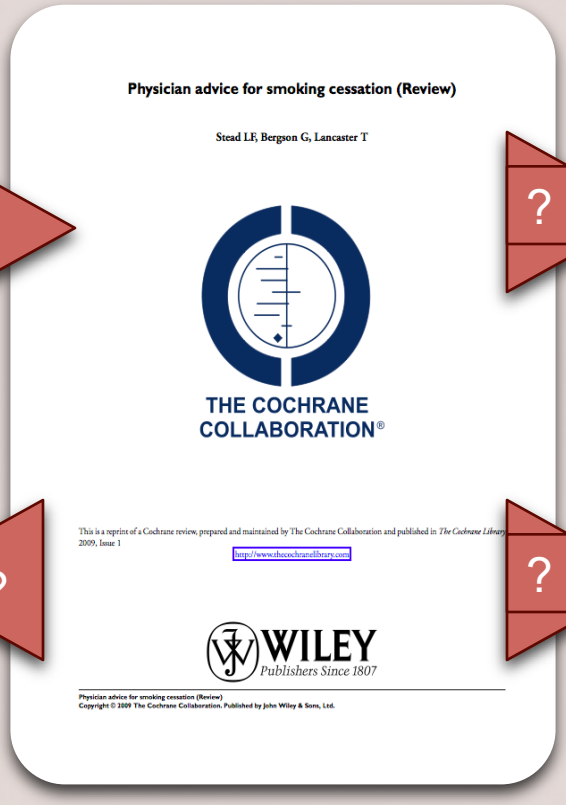
Weaknesses of Cochrane

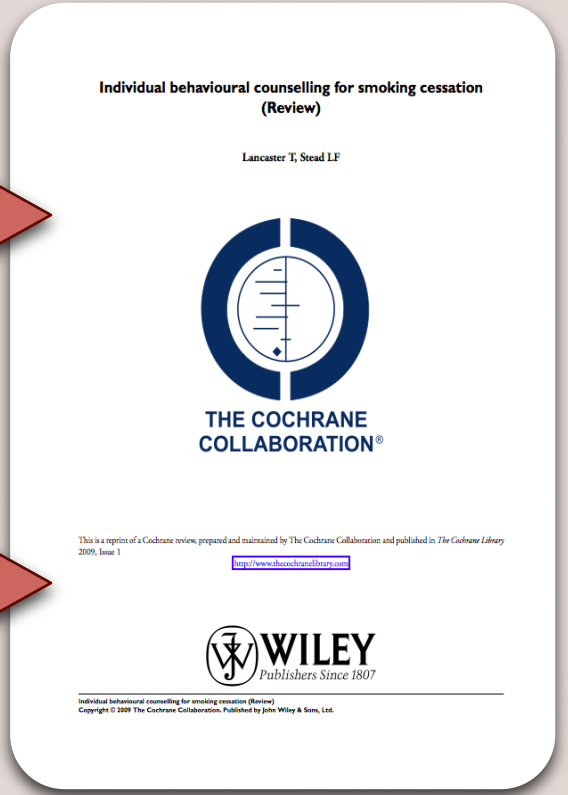
- Products are health care related research only
- Depends on lots of published research
- Resource intensive
  - Experts and time required
- Not amenable to cross-comparing systems
  - Outcomes are usually against a control group
  - Great science, but not for decision-making
- Uncertainty not managed well

menke@email.arizona.edu

Cochrane: But Which Smoking Cessation Intervention is Best?







?

?

?

?

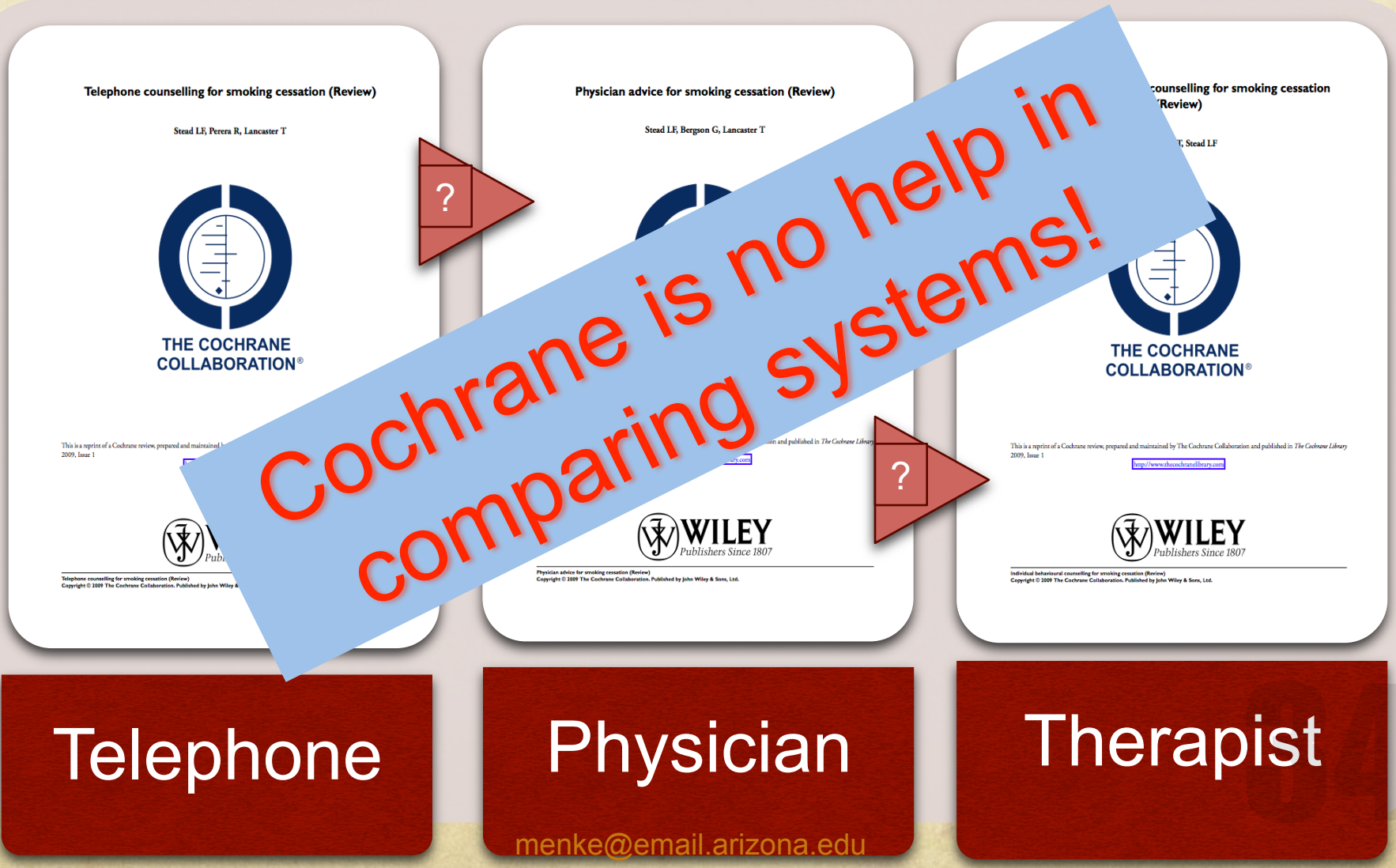
Telephone

Physician

Therapist

menke@email.arizona.edu

Cochrane: But Which Smoking Cessation Intervention is Best?



menke@email.arizona.edu

Bayesian Evidence Synthesis

- Decision-oriented
- Robust to deviations from normal distributions
- May track effects as compared to groups
- Gives relative effect sizes in comparison to a standard
- Can rank treatments – which include the various comparisons, including control groups.

menke@email.arizona.edu

Choosing Outcomes

menke@email.arizona.edu

QALYs as Effects

- Morbidity
- Mortality
- Exchangeable in Cost Effectiveness Analysis and Cost Benefit Analysis
- Willingness to pay (WTP)

menke@email.arizona.edu

Cost (per QALY)

- A year of life adjusted for its quality or its value. A year in perfect health is considered equal to 1.0 QALY. The value of a year in ill health would be discounted. For example, a year bedridden might have a value equal to 0.5 QALY. [medicineNet.com]

menke@email.arizona.edu

QALY Example: CRC

Search Results (Back) Article/Ratios

Your search returned 60 results

Pick Columns to Display(Sort by)

Article ID	Weight	Health State	Publication Year	Reference
2007-01-03159	0.536	Post-op colectomy	2007	Hayes 2007 ANZ J Surg
2007-01-03047	0.6	Colorectal cancer patient with following disease progression	2007	Tappenden 2007 Eur J Cancer
2007-01-03047	0.8	Colorectal cancer patient prior to disease progression	2007	Tappenden 2007 Eur J Cancer
2007-01-03030	0.83	Disease recurrence	2007	de Verteuil 2007 Int J Technol Assess Health Care
2007-01-03030	0.83	Nonoperative management	2007	de Verteuil 2007 Int J Technol Assess Health Care
2007-01-03030	0.83	Initial operation	2007	de Verteuil 2007 Int J Technol Assess Health Care
2007-01-02323	0.25	Colorectal Cancer, Dukes' D	2007	Tappenden 2007 Gut
2007-01-02323	0.5	Colorectal Cancer, Dukes' C	2007	Tappenden 2007 Gut
2007-01-02323	0.7	Colorectal Cancer, Dukes' B	2007	Tappenden 2007 Gut
2007-01-02323	0.74	Colorectal Cancer, Dukes' A	2007	Tappenden 2007 Gut
1 2 3 4 5 6				

CEA Registry
Center for the Evaluation of Value and Risk in Health
menke@email.arizona.edu
Tufts Medical Center
Institute for Clinical Research & Health Policy Studies

Problems with Effect Size

- A 20% effect size means?
- 20% get all better?
- Everyone gets 20% better?
- Some combination?
- 40% get better, 20% die
- Milton Friedman: “Who wants to wade across a river which averages 4 feet deep?”

menke@email.arizona.edu

We Need Some Estimate of the Demand on Resources

- Full evaluations require some sense of cost
- Costs are determined by perspective
  - Payer (reduce payments)
  - Society (improve productivity)
  - Patient (pain relief)
- Opportunity costs
- Indirect costs
- Externalities

menke@email.arizona.edu

A One-slide Course in Health Economics

- Strictly comparative to a current standard (no placebos, please!)
- Welfare economic theory
  - Pareto optimization
    - At least one helped, no one hurt
  - Cost-benefit analysis (consequences)
  - Willingness to pay, contingent valuation
- Operations research and management science
  - Constraint maximization
  - Social decision-making under finite resources
  - Cost-effectiveness method

menke@email.arizona.edu

CEA versus CBA

- Cost benefit analysis (CBA) born out of social welfare theory.
  - Need measure combining morbidity and mortality
  - QALY
- Cost effectiveness analysis (CEA) born out of management science and operations research
  - Original units

menke@email.arizona.edu

Analytic Perspective

- Depending on perspective ~
  - Patient / consumer
    - Costs of care
    - Externalities / indirect costs
    - Opportunity costs
  - Health care system
  - Payers
  - Societal – includes loss of life and productivity

menke@email.arizona.edu

Conceptual Structure for Bayesian Indirect Comparisons

	Treatment A	Treatment B	Control A	Control B
Study # 1	X	X		
Study # 2	X		X	
Study # 3		X		X
Study # 4	X	X	X	
Study # 5	X			X
Study # 6	X			X
Study # 7	X	X	X	X

menke@email.arizona.edu

Weight by Study Quality

	Treatment A	Treatment B	Control A	Control B	Study Quality
Study # 1	X	X			10
Study # 2	X		X		4
Study # 3		X		X	9
Study # 4	X	X	X		6
Study # 5	X			X	8
Study # 6	X			X	1
Study # 7	X	X	X	X	7

menke@email.arizona.edu

Compare Treatments B to C

	Treatment A	Treatment B	Treatment C	Control	Study Quality
Study # 1	X	X			10
Study # 2	X		X		4
Study # 3		X		X	9
Study # 4	X		X		6
Study # 5	X			X	8
Study # 6	X			X	1
Study # 7	X	X		X	7

menke@email.arizona.edu

Bayesian Evidence Synthesis Demonstration

- [Run Demo]
- Show organization of studies
- Code
- Data
- Trace
- Convergence
- Distributions

menke@email.arizona.edu

WinBUGS Results

Node statistics

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
rk[1]	4.445	0.5528	0.00278	3.0	4.0	5.0	10001	50000
rk[2]	3.232	0.7578	0.004848	2.0	3.0	4.0	10001	50000
rk[3]	2.309	0.7888	0.005909	1.0	2.0	4.0	10001	50000
rk[4]	1.992	0.9466	0.007296	1.0	2.0	4.0	10001	50000
rk[5]	3.021	1.992	0.009304	1.0	5.0	5.0	10001	50000

Random effects model for Smoking Cessation data

==========

49 trials (47 + two 3-arm-trials),

96 data points,

5 treatments (var = tx)

1 = baseline - control

2 = quitline

3 = physician minimal

4 = physician intensive

5 = individual counselling

==========

menke@email.arizona.edu

After 50,000 samples, Relative Effectiveness is…

label	Mean rank	SD	Error (X 10^{-3})	2.5%	median	97.5%	Distribution of Rankings
Baseline/ control	4.4	0.55	2.8	3	4	5	
Telephone	3.2	0.76	4.8	2	3	4	
Physician - min	2.3	0.78	5.9	1	2	4	
Physician - intense	2.0	0.95	7.2	1	2	4	
Therapist	3.0	2.0	9.3	1	5	5	

menke@email.arizona.edu

Effectiveness or Cost Effectiveness?

- WinBUGS can take different outcomes

menke@email.arizona.edu

Step 3. Decision Analysis

- “A systematic approach to decision making under conditions of imperfect knowledge; a practical application of probability theory. Used to calculate the optimal strategy from among a series of alternative strategies.”

menke@email.arizona.edu

Incremental Cost Effectiveness Ratio (ICER)

- Plain cost-effectiveness can mislead
- Something cheap and ineffective can be cost effective
- Also, the ICER method does exactly what we want in Comparative Effectiveness Research: it compares a novel system to a current standard.

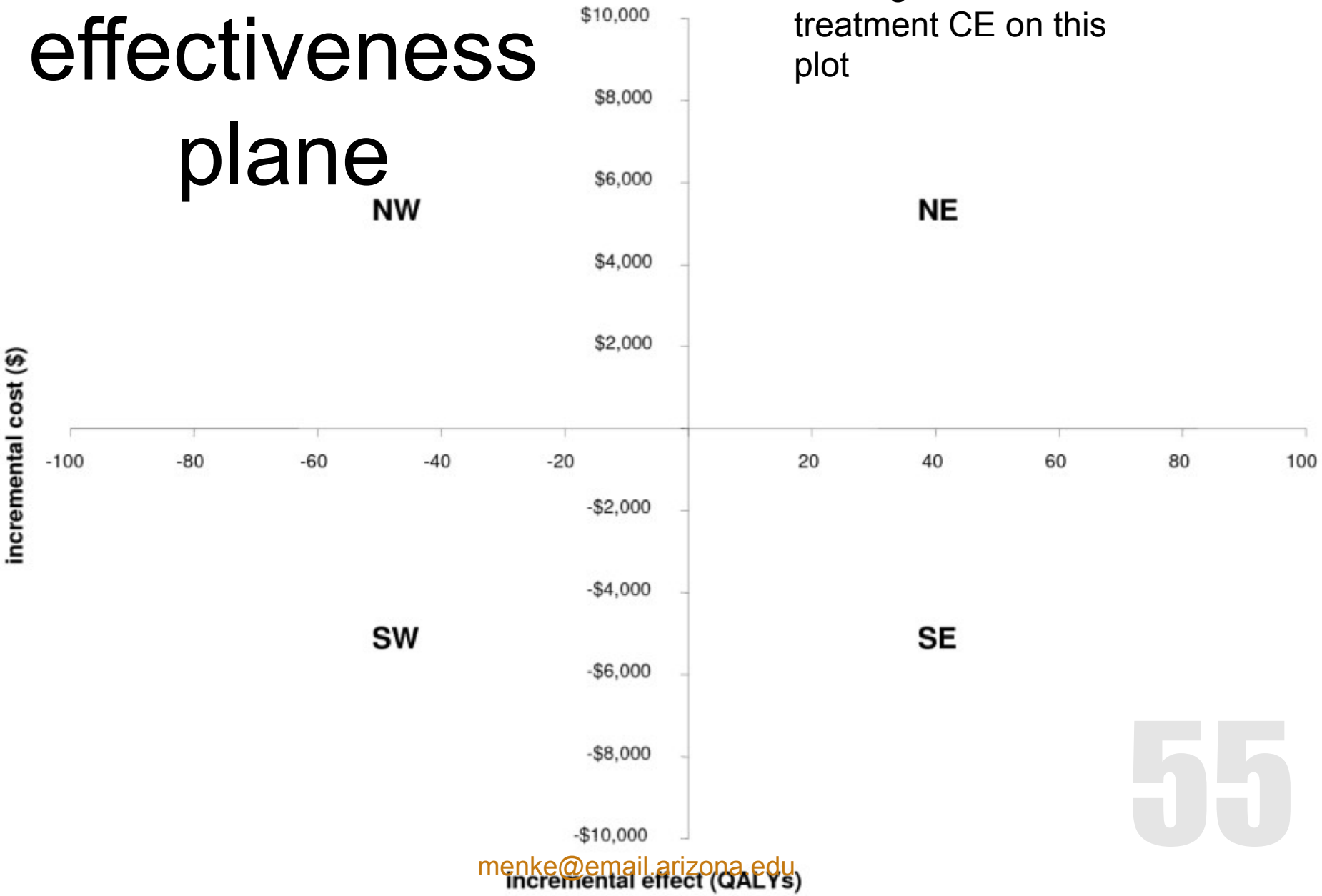
menke@email.arizona.edu

The cost effectiveness plane

menke@email.arizona.edu

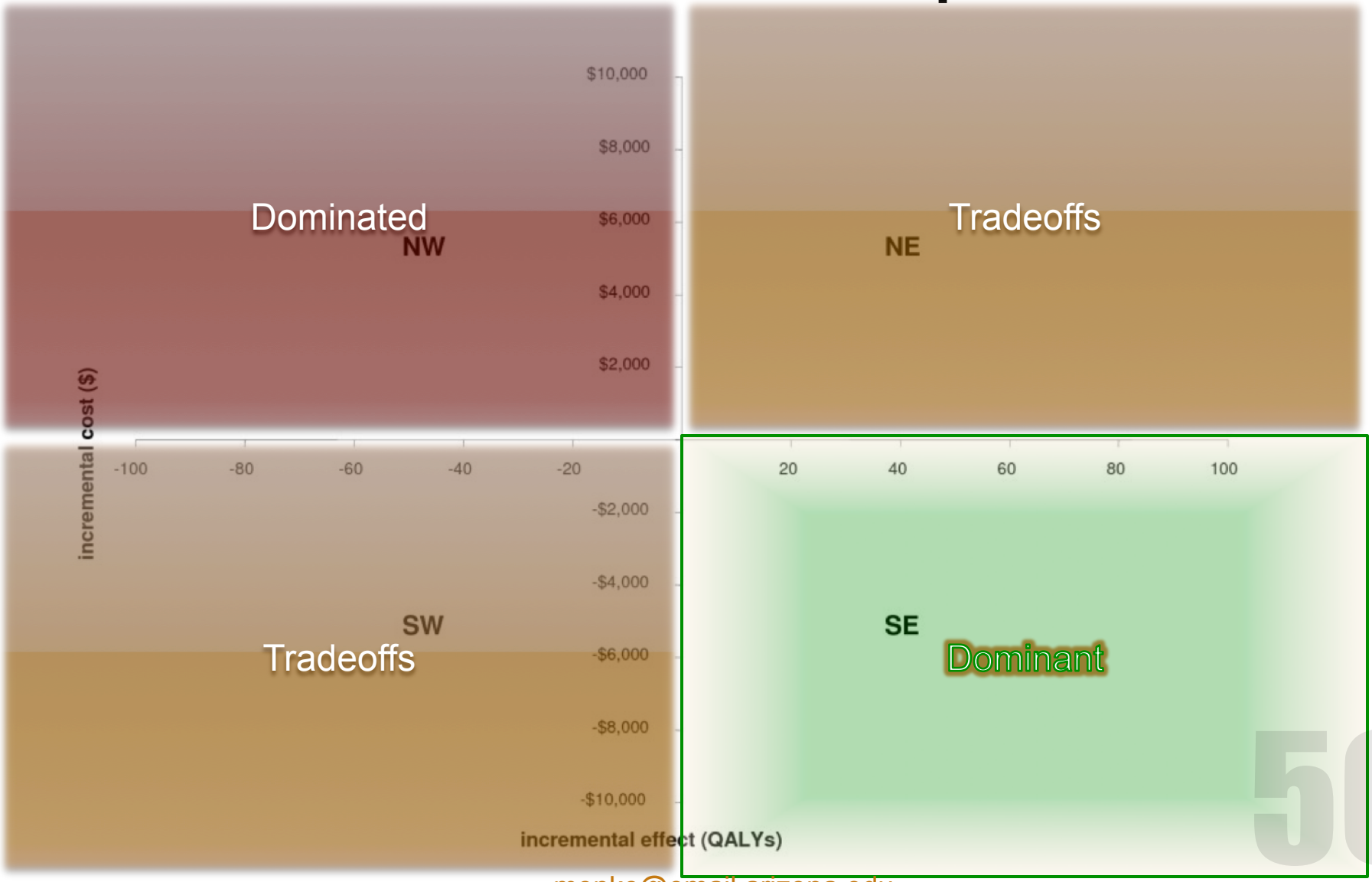
Cost effectiveness plane

Plotting new treatment CE on this plot



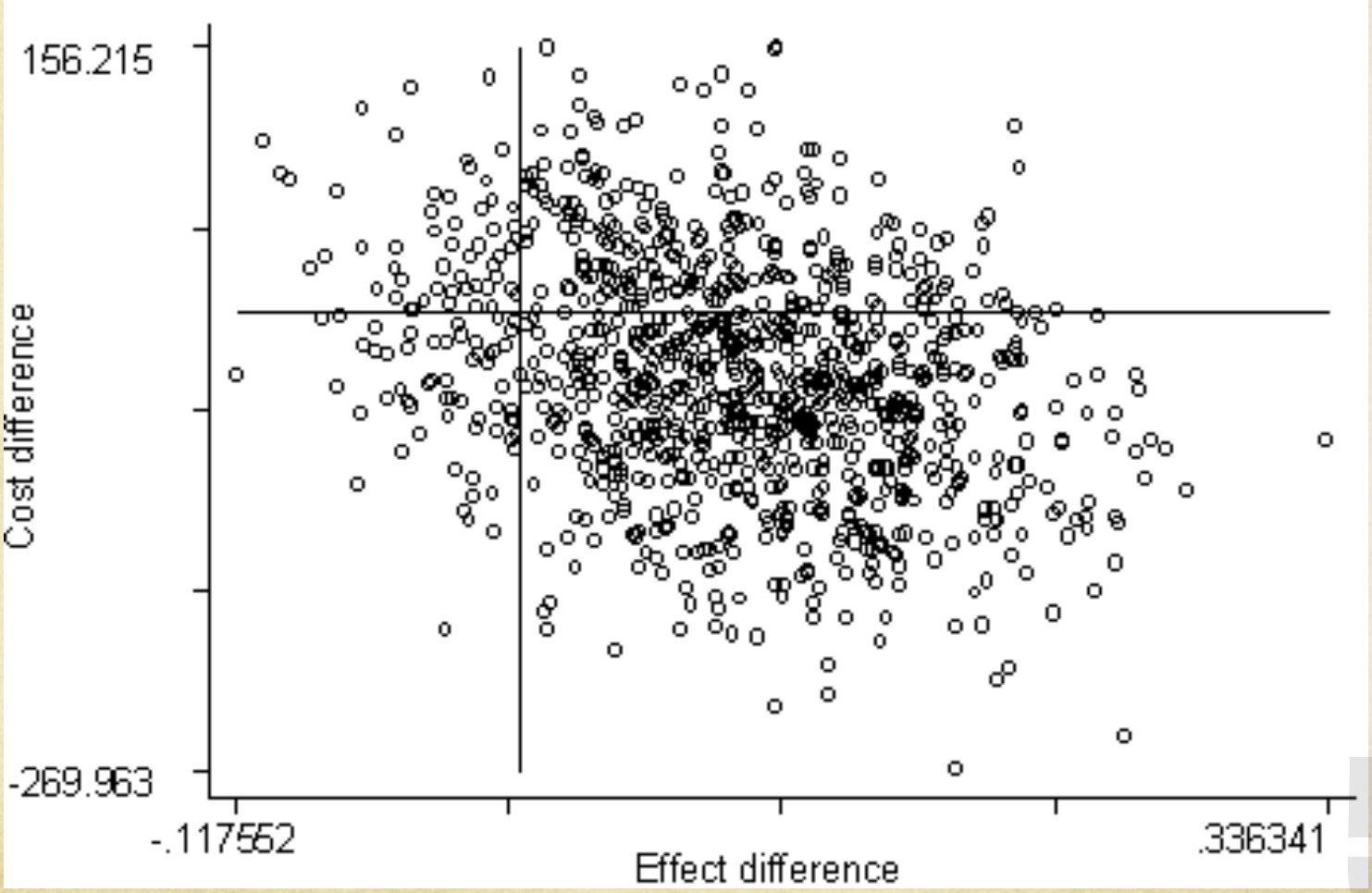
menke@email.arizona.edu

Cost effectiveness plane



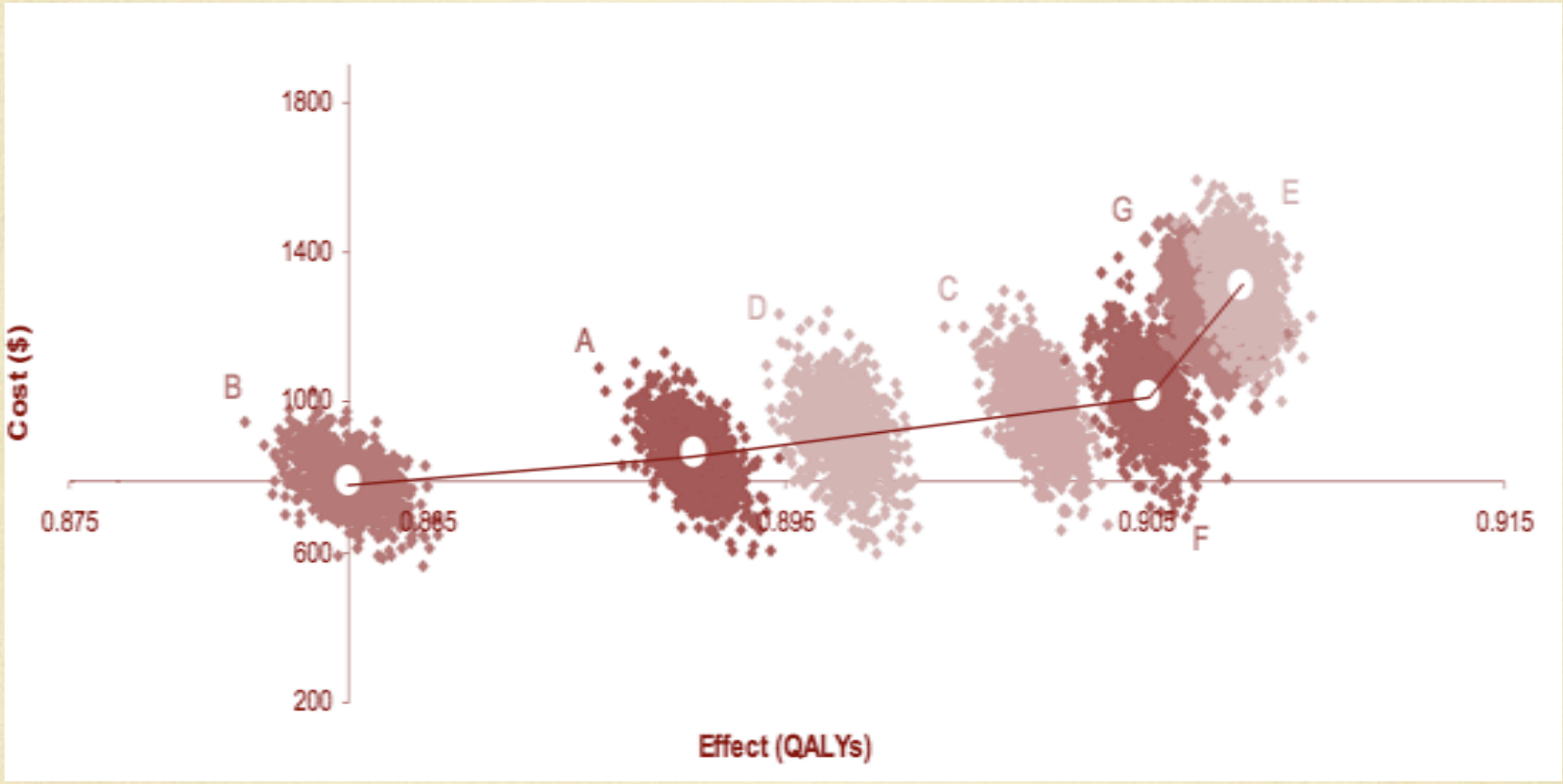
menke@email.arizona.edu

Incremental CE Ratio plot



menke@email.arizona.edu

Multiple ICER's



From: Barton, Briggs, and Fenwick, 2005

menke@email.arizona.edu

Incremental cost effectiveness Ratio

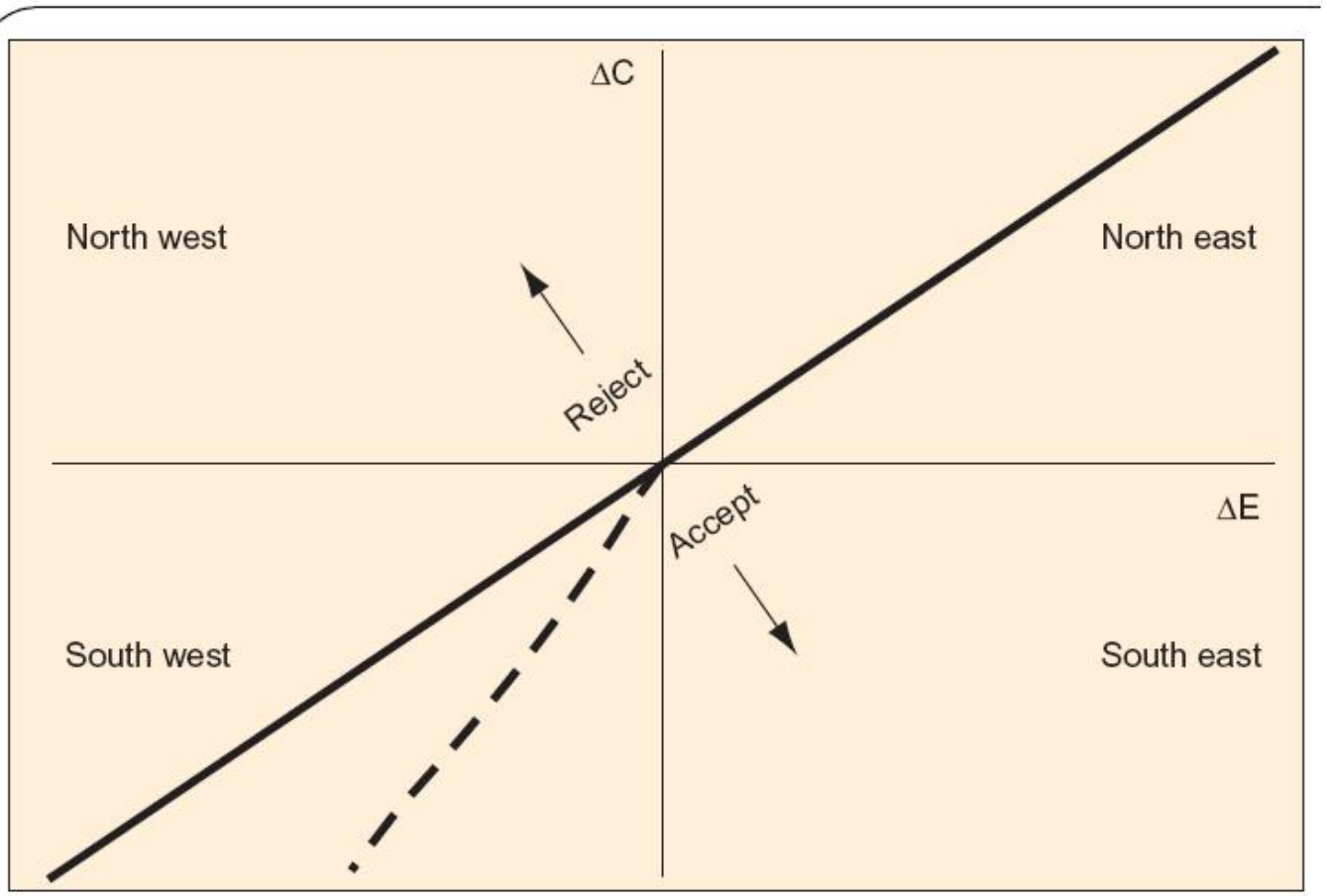


Figure 1. Incremental costs (ΔC) and effects (ΔE) of a new technology over an old one and the maximum acceptable incremental cost–effectiveness ratio without a kink (solid line) and with a kink (dotted line).

menke@email.arizona.edu

ICER strength

- Shows CE relative to a current standard of therapy or care
  - A direct comparison of two programs or interventions
  - CE's then compare-able
  - Whereas CE ratios are not directly compare-able
  - That which is barely effective and cheap could be just as cost effective as something very effective and very expensive

menke@email.arizona.edu

STEP 4. *Ex post* VOI

- Cleaning up the analysis - what just happened?
  - Is there enough of an effect to continue looking for evidence?
- Value of information (VOI) analysis
  - Estimates degree of uncertainty
  - Affixes monetary value of reducing uncertainty

menke@email.arizona.edu

Expected value of perfect information

menke@email.arizona.edu

Expected value of perfect information

- “The expected costs of uncertainty can be interpreted as the expected value of perfect information.”
  - Claxton 2006

menke@email.arizona.edu

Expected Value of Perfect Information (EVPI)

- Assume you could “buy” information that perfectly predicts a future outcome

- The expected value of perfect information (EVPI) is the difference between expectation of the maximum benefit and expected net benefit:

menke@email.arizona.edu

Personalized Example

- The US economy
  - Credit uncertainty
  - Stock market volatility

- How much would *you* pay to reduce uncertainty?
  - There would be an upper limit.
  - Probably not more than what you are “worth”

menke@email.arizona.edu

Briggs Sculpher & Claxton, 2006

- The expected cost of uncertainty is determined jointly by the probability that
  - 1) a decision based on existing information will be wrong, probability of error, and
  - 2) the consequences of a wrong decision (expected opportunity loss)
- This is variously called “expected cost of uncertainty” or “expected opportunity loss surrounding decisions”

menke@email.arizona.edu

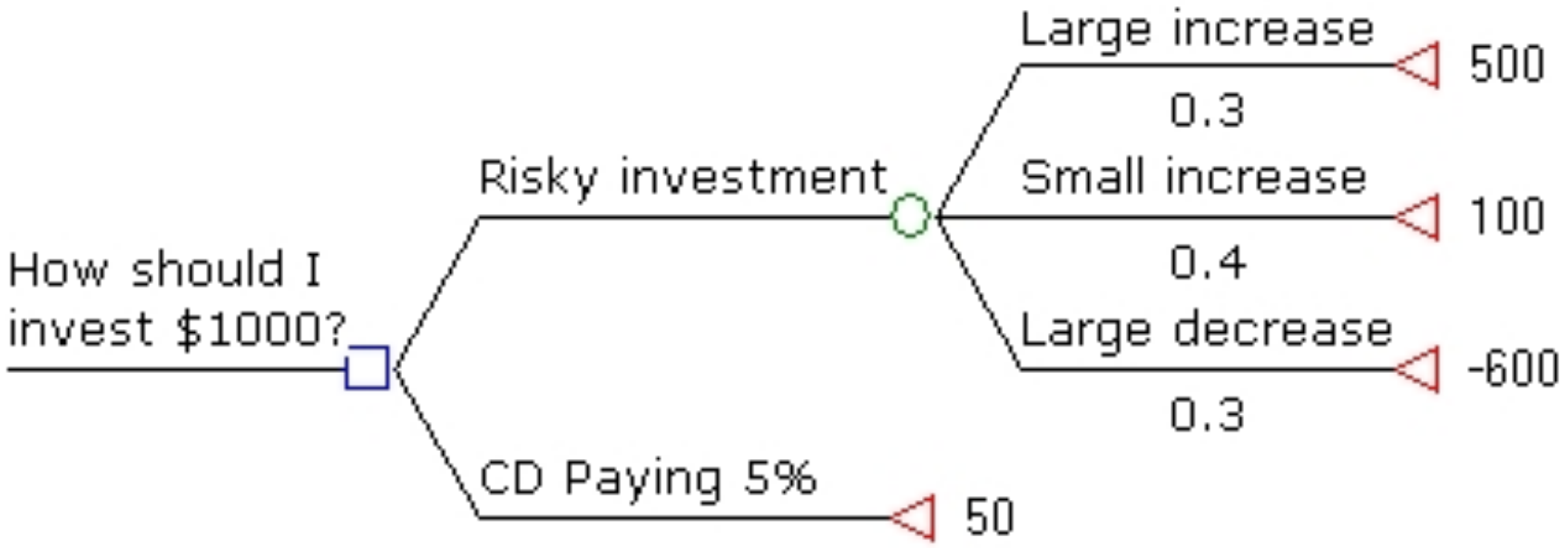
EVPI: Three core tasks

1. Decision analytic model to represent the problem
2. Probabilistic analysis (PSA)
3. Establish the value of additional information

- *EVPI estimates are for the individual patient or client!*

menke@email.arizona.edu

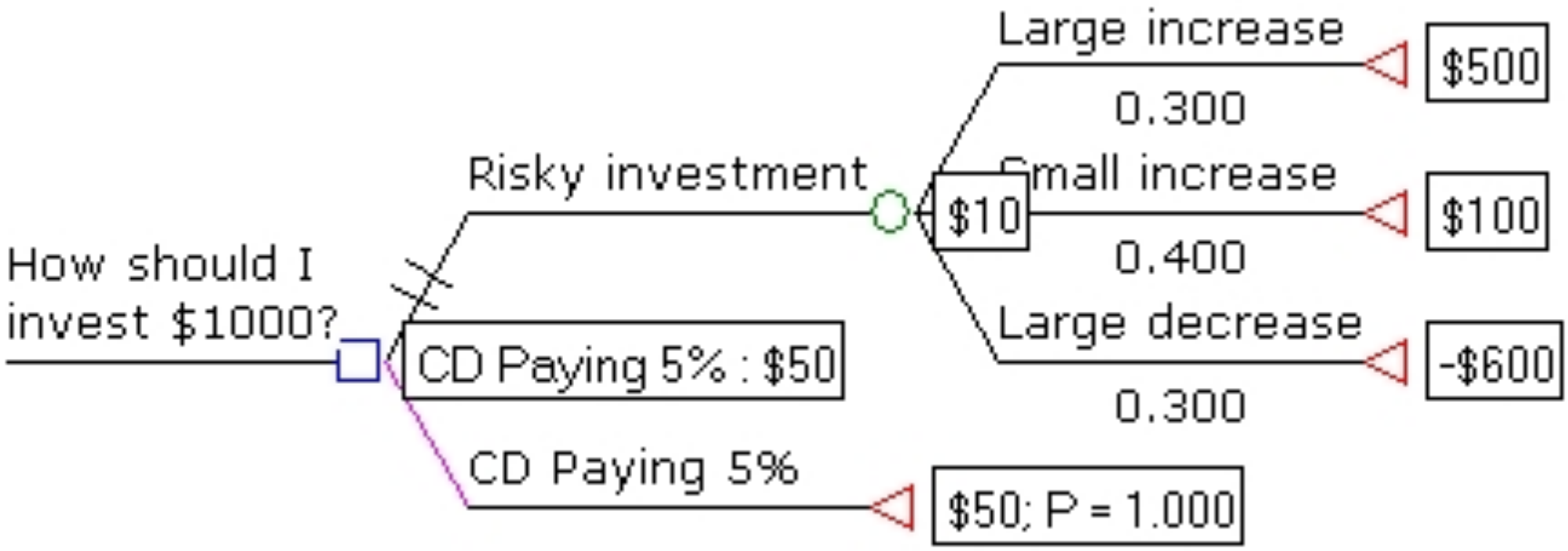
Calculation of EVPI – An example



Source: Tree Age Pro 2009 Users Manual

menke@email.arizona.edu

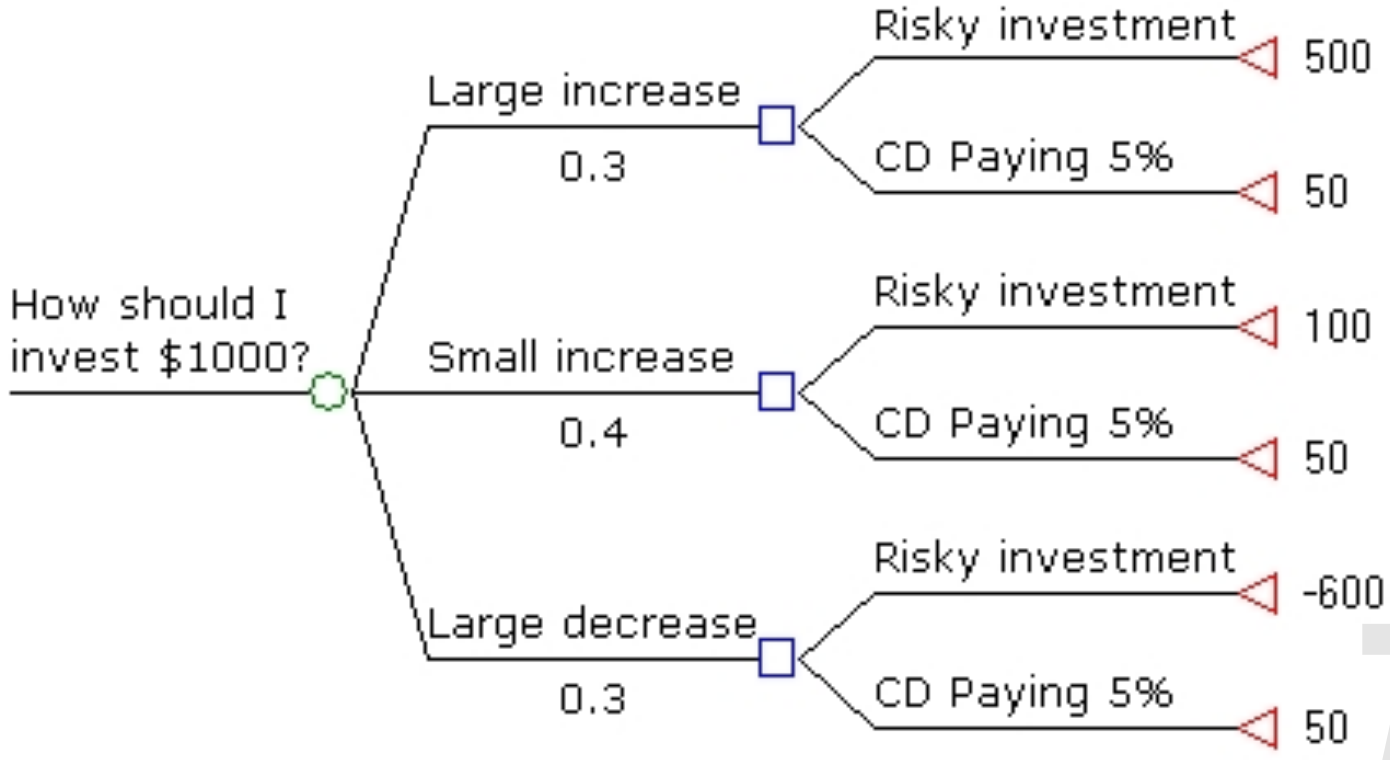
Calculation of EVPI – Rollback of Stock Tree



Source: Tree Age Pro 2009 Users Manual

menke@email.arizona.edu

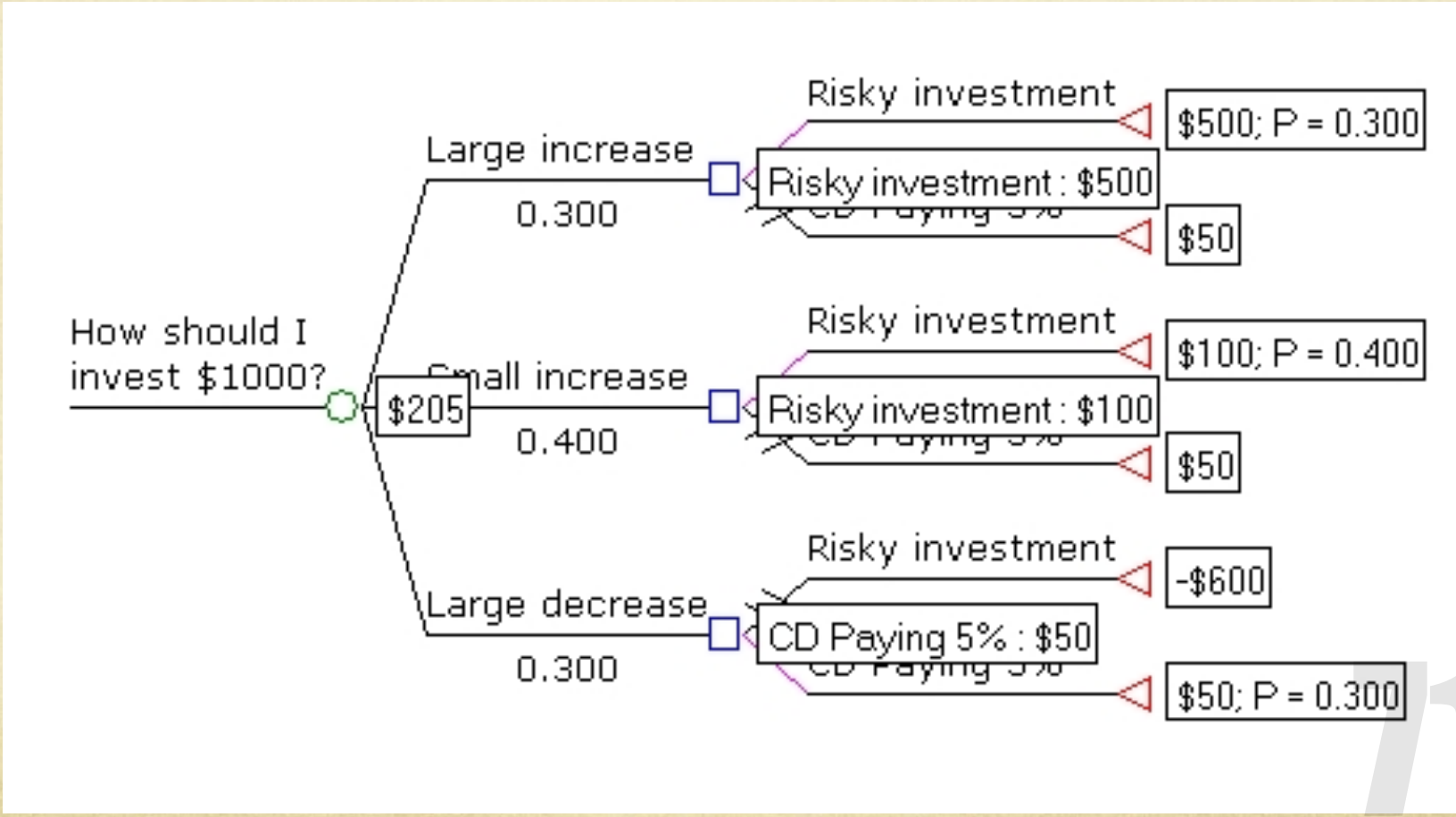
Rearranging the Tree for Best Possible Outcome



Source: Tree Age Pro 2009 Users Manual

menke@email.arizona.edu

Rollback of Best Possible Outcome



Source: Tree Age Pro 2009 Users Manual

menke@email.arizona.edu

EVPI for Stock Investment

- $205 - $50 = $155
- It makes sense to pay *up to* $155 for market information that would allow you to predict the outcome

menke@email.arizona.edu

EVPI Conceptually

	Drug A	Drug B	Optimal choice	Maximum net benefit	Opportunity loss if choose "B"
Iteration 1	9	12	B	12	0
Iteration 2	12	10	A	12	2
Iteration 3	14	20	B	20	0
Iteration 4	11	10	A	11	1
Iteration 5	14	13	A	14	1
Expectation	12	13		13.8	0.8

Current Information: 13
Perfect Information: 13.8
EVPI = 13.8 – 13 = 0.8

Source: Claxton, K – University of York

menke@email.arizona.edu

EVPI

- EVPI (Expected Value of Perfect Information) – the theoretical maximum worth to the decision maker of additional information about uncertain states of nature that is absolutely unerring.

menke@email.arizona.edu

What EVPI Means

- If EVPI > Decision threshold then collecting more information is worthwhile
- Reflects the amount of uncertainty in the data that is present
- One should delay adoption of technologies when the EVPI is large

menke@email.arizona.edu

Situations where EVPI may be useful

- Expensive technologies that have marginal benefits
- Concerns about TX safety – it may be worthwhile to delay adoption because the value of additional information exceeds the value gained from immediate adoption
- Setting research priorities for:
  - Health insurance plans
  - Pharmaceutical manufacturers
  - NIH and other government agencies

menke@email.arizona.edu

Conclusion

- Comparative effectiveness research can be accomplished in 4 general steps
  - Establishing prior information value
  - Evidence synthesis
  - Decision analysis
  - Value of information analysis
- There is absolutely no reason why CER cannot be carried out to improve health care policy and decisions.

menke@email.arizona.edu