Benefit or Harm Framing Influences People's Reported Threshold Probabilities for the Use of Antibiotics for Strep Throat

Robert M. Hamm, PhD, University of Oklahoma Health Sciences Center, Dept. of Family and Preventive Medicine

Threshold Probability Strategy

- Physicians can simplify some treatment decisions by treating only when the patient's disease probability exceeds a threshold probability.
  - This strategy applicable when there are
    - two states of world (e.g., disease (strep throat) or not),
    - two actions (e.g., treat (antibiotics) state 1, not state 2).
  - One can act based on a rough estimate of probability patient has disease (unless the probability is near the threshold).

What Should the Treatment Threshold Probability be?

- Treatment threshold probability is based on the utilities of the two possible mistakes:
  - to fail to treat disease (Miss) and
  - to unnecessarily treat a patient without the disease (False Alarm)

Utility of Outcomes of Decision Situation

- The utility of each can be judged

	Not Strep Throat (1 – p(strep))	**Strep Throat** p(strep)
Give No Antibiotics	Appropriate refraining from using antibiotics ***u(TN) = u(not treat non-strep)***	**Miss**, unnecessary suffering (a day or so?), possible risk of later and worse strep infection. ***u(FN) = u(miss strep)***
Give Antibiotics	**False alarm**, unnecessary treatment ***u(FP) = u(unnecessary treatment)***	Appropriate use of antibiotics ***u(TP) = u(treat strep)***

Choose to Treat if it has Higher Expected Utility

- Expected utility depends on probability of disease, as well as on utilities of the outcomes.
- EU(treat sore throat with antibiotics)
  = p(strep) * ***u(antibiotics|strep)***
  + (1-p(strep)) * ***u(antibiotics|not strep)***
- EU(don't treat sore throat with antibiotics)
  = p(strep) * ***u(no antibiotics|strep)***
  + (1-p(strep)) * ***u(no antibiotics|not strep)***
- Treatment Threshold probability =
  that p(Disease) at which EU(treat) = EU(do not treat)

Set Action EUs Equal and Solve for p(Dx)

$$p(Dx)*u(TP)+(1-p(Dx))*u(FP) = p(Dx)*u(FN)+(1-p(Dx))*u(TN)$$

$$p(Dx)*[u(TP)-u(FN)]=(1-p(Dx))*[u(TN)-u(FP)]$$

$$\frac{p(Dx)}{1-p(Dx)}=\frac{u(TN)-u(FP)}{u(TP)-u(FN)}$$

Treatment threshold probability, stated in terms of the odds, is equal to a ratio of utility differences.

Physician Use of Treatment Threshold Probability Strategy

- Do clinicians know about the treatment threshold strategy?
  - Do they have a threshold probability for each particular clinical presentation?
  - In making a treatment decision, do they judge patient's disease probability and compare it to a numerical threshold probability?
- Where do their thresholds come from?
  - Have they been taught, or read, a threshold for a particular situation?
  - Do they base thresholds on own assessments of utility?
  - Do utility-derived thresholds depend on the utility-assessment method?

How Accurately do Physicians Judge Treatment Threshold Probabilities?

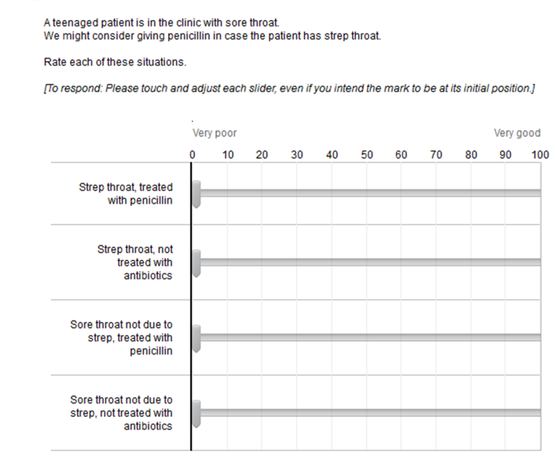
- We ask people to judge:
  - Gist (ordinal judgment):
    - Whether it is worse to "miss" than to "false alarm"
  - Precise:
    - 4 utilities
    - 2 utility differences
    - 1 utility difference ratio
    - The threshold per se.
- Are precise estimates consistent with gist?
- Are the 4 different precise estimates consistent?

Probabilities From Odds

$$Odds(Dx)=\frac{p(Dx)}{1-p(Dx)}=\frac{u(TN)-u(FP)}{u(TP)-u(FN)}$$

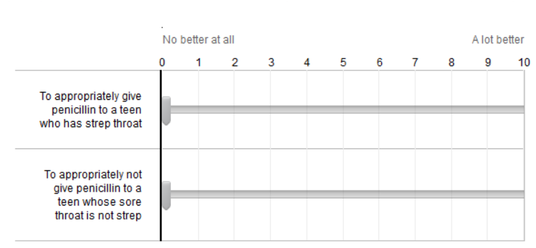
- From 4 utility judgments
$$p(Dx)=\frac{Odds(Dx)}{1+Odds(Dx)}=\frac{1}{\frac{1}{Odds(Dx)}+1}$$
- From 2 utility differences
$$p(Dx)=\frac{1}{\frac{u(TP)-u(FN)}{u(TN)-u(FP)}+1}$$
- From 1 patient misdiagnosis tradeoff
$$p(Dx)=\frac{1}{\frac{Number(FalseTreatments)}{Number(MissedTreatments)}+1}$$

Judgment of 4 Utilities: Response on a Utility Scale

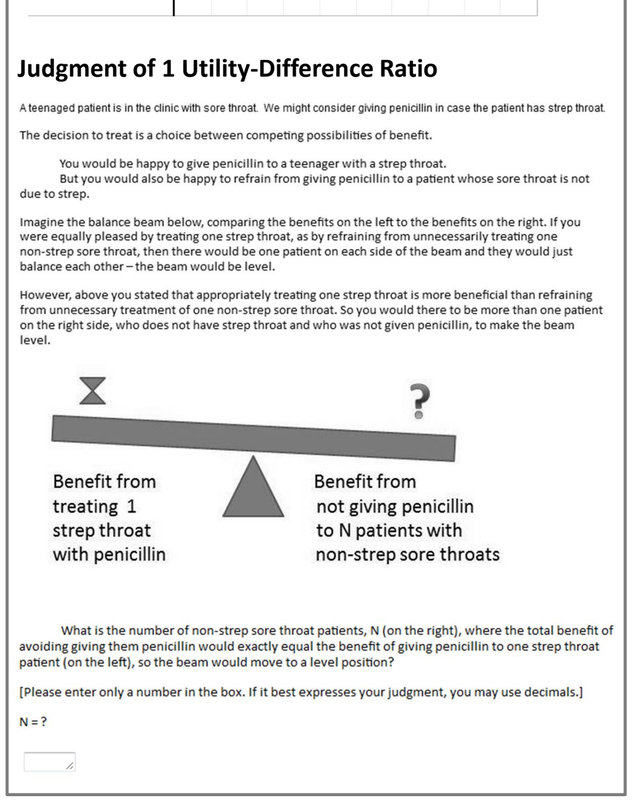


Judgment of 2 Utility Differences

- Response is on a utility difference scale.
  - Similar to Djulbegovic et al's (2014) questions, except theirs asked about "regret" with anchor at midpoint.



Judgment of 1 Utility-Difference Ratio



Method

- **Sample**. A web survey (Qualtrics) was promoted among convenience samples of primary care clinicians and residents, medical and physician assistant students, undergraduate students, and the public (patients).
- **Threshold judgment method**. Each respondent provided judgments to support four different calculations of their treatment threshold.
- **Pro/Con bias.** Bias in description of the situation was varied (presenting more detailed reasons either in favor of treating strep throat, or against overusing antibiotics).
- **Benefit or harm framing.** Descriptions of the judged outcomes and the response scales were framed either as benefits measured on a positive scale, or as harms measured on a negative scale.
- **Task order,** and order of presentation of responses within tasks, were randomly varied.

Results (1)

- **Sample**. 950 started survey and 735 (77.4%) finished
- **Threshold judgment method**.
  - Reported various degrees of awareness and use of threshold probability strategy with sore throat **(see Figure 1)**
  - Each threshold judgment method was noisy
  - The numerical thresholds were not consistent with the gist importance judgments **(see Figure 2)**
  - Many people stated threshold was > 50%, while their utility judgments implied threshold < 50% **(see Figure 3)**
- Technical quality of responses **(see Figure 4)**
  - 20% of responses on 4 utility judgment method were incoherent (e.g., error better than correct response)
  - Person tradeoff method had fewest technical problems (constrained not to allow incoherent answers)

Self Reported use of Threshold Probability Strategy

Figure 1.

	Percent endorsing this awareness or use of threshold strategy (% of 452) (multiple response possible)
Never follow treatment threshold probability strategy for sore throat	8.2
Have thought about it before but do not consciously follow it	8.6
Often do, but not with explicit probability numbers	32.7
Often, comparing specific patient probability to specific threshold probability	13.7
Often estimate patient disease probability, but do not compare to a threshold probability	20.6
Usually test and give antibiotic if positive	15.7
Other	20.6

- Participants stated which situation was more important not to make an error in.
  - If missing strep is more important, threshold should be < 50%.
  - More precise thresholds were often not consistent with the importance judgment.

Figure 2.

		47.6%		95.0%		71.9%		79.0%	
		Stated Numerical		**Person Tradeoff**		**4 Utility Judgment**		**2 Utility Difference**	
		< 0.5	**> 0.5**	**< 0.5**	**> 0.5**	**< 0.5**	**> 0.5**	**< 0.5**	**> 0.5**
Judged as More Important	**Treating versus missing strep**	**91**	**300**	**350**	**19**	**275**	**62**	**354**	**26**
	Not giving unwarranted antibiotics for nonstrep	**42**	**220**	**11**	**217**	**90**	**113**	**105**	**138**

Figure 3. **Proportion of Responses in Various Ranges for Each of Four Threshold Probability Elicitation Methods**

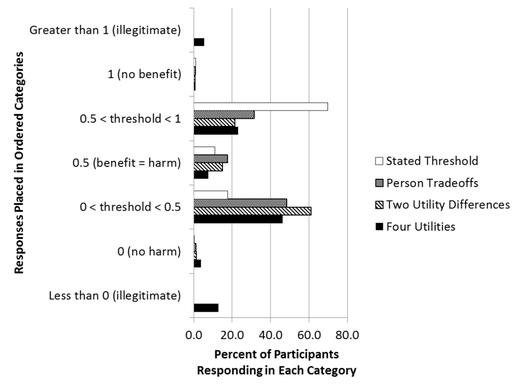
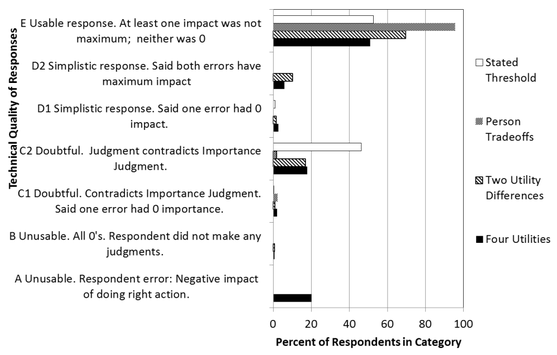


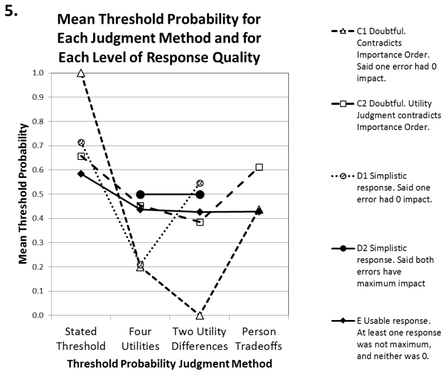
Figure 4. **Proportion of Each Method's Responses Assessed to Have Each Level of Quality**



Results (2)

- Different threshold probabilities by different methods **(see Figure 5)**
  - Highest quality responses
    - Same mean for all three utility-judgment based methods (0.43 or 0.44)
    - Stated threshold was higher (0.58)
  - Lower quality responses: means varied more
- Correlations between methods are higher with higher quality judgments **(see Figure 6)**

Figure 5.



Correlations Among Methods

Figure 6.

- Using only the highest level of quality

	4 Utilities	**2 Utility Differences**	**1 Utility Difference Ratio (Person Tradeoff)**
Stated threshold	.692* (N = 183)	.707* (N = 256)	.752* (N = 367)
4 utilities		.667* (N = 361)	.670* (N = 361)
2 utility differences			.693* (N = 484)

- Highest and 'acceptable' levels of quality

	4 Utilities	**2 Utility Differences**	**Person Tradeoff**
Stated threshold	.243* (N = 539)	.272* (N = 702)	.174* (N = 703)
4 utilities		.334* (N = 527)	.327* (N = 529)
2 utility differences			.376* (N = 689)

Summary

- Clinicians (& students) are aware of treatment threshold probability strategy
  - 46% follow its gist; only 14% actually explicitly use it
    - So we can't say "clinicians generally decide by referring to threshold probability"
- Respondents' precise threshold estimates contradict their gist utility assessments (which error is worse)
  - Different methods have different contradiction rates
  - Contradictions occur 5%, 21%, 28%, and 52% of the time.
    - Typical of behavior on an UNFAMILIAR judgment task
    - Assessment method can facilitate REASONABLE responses.
- Some respondents use scales illogically or simplistically.
  - Better quality judgments correlate better with each other.
    - Supports hypothesis that they understand and use the concept, but make errors expressing it on unfamiliar scales.
- Precise estimates of strep throat treatment threshold probabilities based on high quality utility judgments average 0.43 or 0.44 (for each of 3 methods); different than the stated threshold probabilities (0.58).
  - Supports hypothesis that people think differently about thresholds if asked to consider the underlying utilities upon which thresholds should be based.

Implications

- The noisiness and unfamiliarity of clinicians' threshold judgments suggests
  - they aren't using a well founded treatment threshold probability strategy, and
  - if we want them to do so, we should calculate it for them and provide them with aids for assessing disease probability, rather than relying on clinician judgment of either "thresholds" or "utilities."

Collaborators

National University of Ireland:
Christopher Dwyer, PhD; Pádraig MacNeela, PhD
Texas College of Osteopathic Medicine:
Frank Papa, DO, PhD
University of Oklahoma College of Medicine:
Preston Seaberg, MD; Dewey Scheid, MD, MPH; Bruna Varalli-Claypool, MHS, PA-C