Peer Feedback in Clinical High-Volume Activities: How Ranking Designs and Ability Impact Physician Effort

K. Huesmann
U of Munster

Y. Ndiaye
U of Cologne

C. Waibel
ETH

D. Wiesen
U of Cologne

EEA 2023
Barcelona

Motivation: How to ensure high quality of care?

- Peer comparison through public performance feedback may motivate physicians to increase efforts
- Kolstad (2013): Performance feedback may motivate physicians even stronger than financial incentives
- Rankings which assign physicians to outcomes based on ranks are one way to provide feedback
- Still little known about the relevance of ranking design, i.e. setting of thresholds for ranks
- Physicians may react differently to rankings if they differ with respect to which outcomes they are able to reach

Does providing peer rankings improve physician effort?
How does the effect depend on ranking design and ability?

Example Colonoscopy: High-volume activities with measurable performance indicators

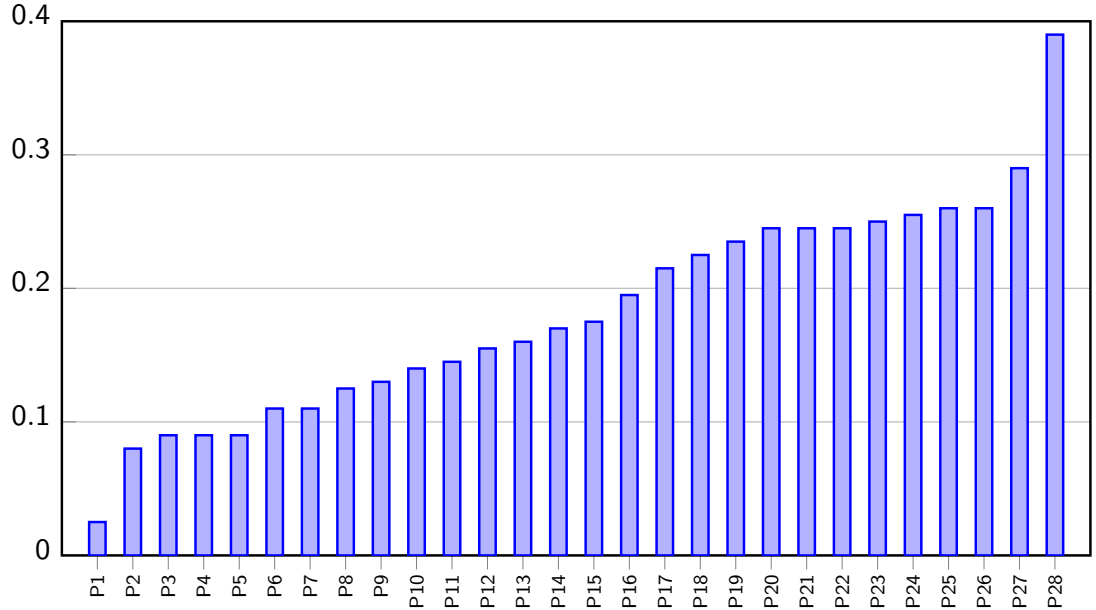


Figure: Adenoma detection rate for individual physicians

- Currently feedback is often provided in a rather pragmatic way
- Many options of designing a ranking based on the outcomes
- Non-anonymous feedback advocated by medical societies

Method: Theory and Lab-in-the-field experiment based on the conjecture that relative rank matters

Central conjecture: Physicians experience status utility which

- Positively depends on physicians ranked below
- Negatively depends on physicians ranked above

Theory: Transfer clinical setting to a model of status concerns (building on Dubey & Geanokoplos, 2010)

- Physicians exert costly effort with stochastic patient outcomes
- Physicians are heteregoneous with respect to their ability, i.e., their achievable patient outcomes

Experiment: Lab-in-the-field experiment with 124 medical students

Results: Rankings design and abilities matter

1. Peer feedback via non-anonymous rankings impacts on effort
   - The effect depends on the ranking design
   - The effect is mainly positive, but can be negative as well
2. The impact of the ranking design depends on abilities
   - Thresholds that can be crossed by increasing effort motivate
   - A threshold that cannot be crossed demotivates

A model of status concerns: Overview

- N physicians
- Each physicians i chooses effort level e_i at a cost $c(e_i)$ with $c' > 0$
- Stochastic patient benefit $x_i(e_i) \geq 0$ distributed according to $F_i(\cdot|e_i)$ with $F_i(x_i|e_i)$ decreasing in e_i
- $\mathcal{R}$: all possible rankings of N, with ties allowed (*pooling*)
- $\gamma : X^N \to \mathcal{R}$ is an absolute ranking scheme
  - Equal outcomes imply the same rank
  - Higher outcomes lead to weakly higher ranks

A model of status concerns: Status Utility

- Physician's i utility under a ranking scheme γ for effort e_i and others' outcome x_{-i}:

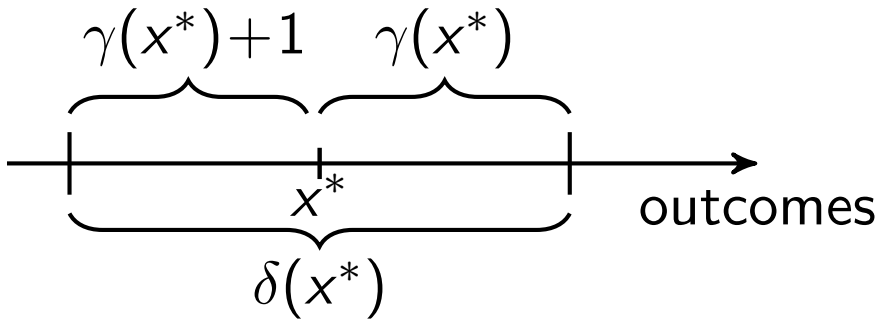
$$u_i^{\gamma}(e_i, x_{-i}) = \underbrace{\alpha_i x_i(e_i)}_{\text{altruistic utility}} - \underbrace{c(e_i)}_{\text{effort cost}} + \underbrace{\lambda_i S^{\gamma}(x_i(e_i), x_{-i})}_{\text{status utility}}$$

- $S^{\gamma}(x_i, x_{-i}) = S(U^{\gamma}(x_i, x_{-i}), O^{\gamma}(x_i, x_{-i}))$
  - U^{γ} = set of physicians ranked below under γ
  - O^{γ} = set of physicians ranked above under γ
  - S is increasing in U^{γ} and decreasing in O^{γ}
- Optimal effort for ranking scheme γ and others' outcome x_{-i}

$$e_i^*(\gamma) = \arg\max \mathbb{E} u_i^{\gamma}(e_i, x_{-i})$$

What is the impact of ranking design on ptimal effort?

Let γ offer an additional threshold at x^* compared to δ



What is the effect on status utility?

- Increase for outcomes above x^*: separation from those below
- Decrease for outcomes below x^*: no pooling with those above

Proposition: Optimal effort increases if the prob. of reaching $\gamma(x^*)$ increases and the prob. of reaching $\gamma(x^*)+1$ decreases in effort.

Hypotheses

H I Ranking Design matters: Effort varies depending on the placement of thresholds for the ranks

H II Abilities matter: thresholds within the range of achievable outcomes increase effort while they decrease effort outside.

Experimental design: Set-up & decision situation

Set-up

- 124 medical students take physicians' role
- Each physician is assigned to either the role of a low-ability physician or a high-ability physician
- Patients not present but patient benefit is transferred to real patients outside the lab

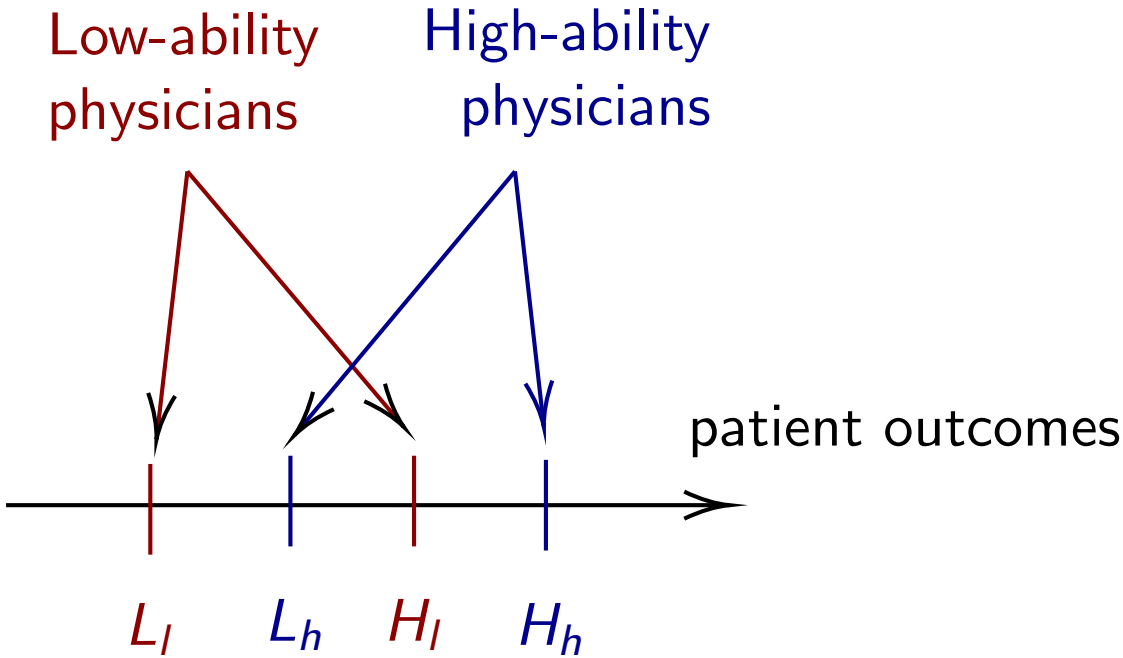
Decision situation

- Physicians choose costly effort $e_i \in \{1, 2, ..., 10\}$.
- Patients: None (passive)

Experimental design: Payoffs

(Monetary) Payoffs

- $\pi^{Physician}$ = Fixed wage - effort costs $c(e_i)$
- $\pi^{Patient}$ = Stochastic patient outcome $x_i(e_i) \in \{L_i, H_i\}$ with $x_l \in \{0, 20\}$ and $x_h \in \{5, 25\}$



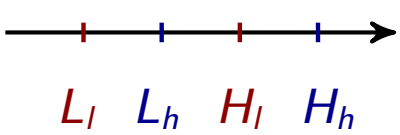
Experimental timing: 3 parts

Participants randomly assigned to seat at tables for 4 persons.

1. Benchmark effort decision without ranking (strategy method)
   - Each physician decides about effort as a low-ability physician
   - Each physician decides about effort as a high-ability physician
2. Type assignment and introduction
   - Incentivised real effort task: Entrance test for medical studies
   - Ability assignment (private, 2 low- and 2 high-abil. per group)
   - Stand up and tell names within groups
3. Effort decisions with rankings
   - Physicians decide simultaneously how much effort to exert for each ranking condition of 5 rankings
   - One randomly chosen ranking published within group

Experimental variation of rankings within individuals

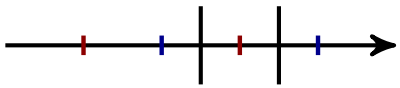
Baseline



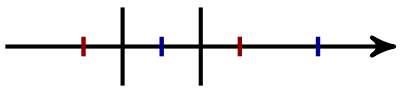
No ranking (*Base*)

Rankings

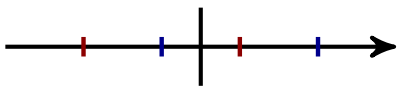
Top threshold (T)



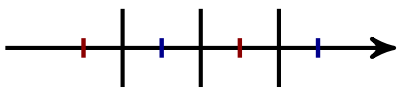
Top and middle threshold (TM)



Middle and bottom threshold (MB)



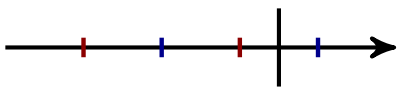
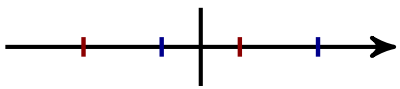
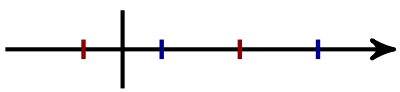
Middle threshold (M)



Top, middle and bottom threshold (TMB)

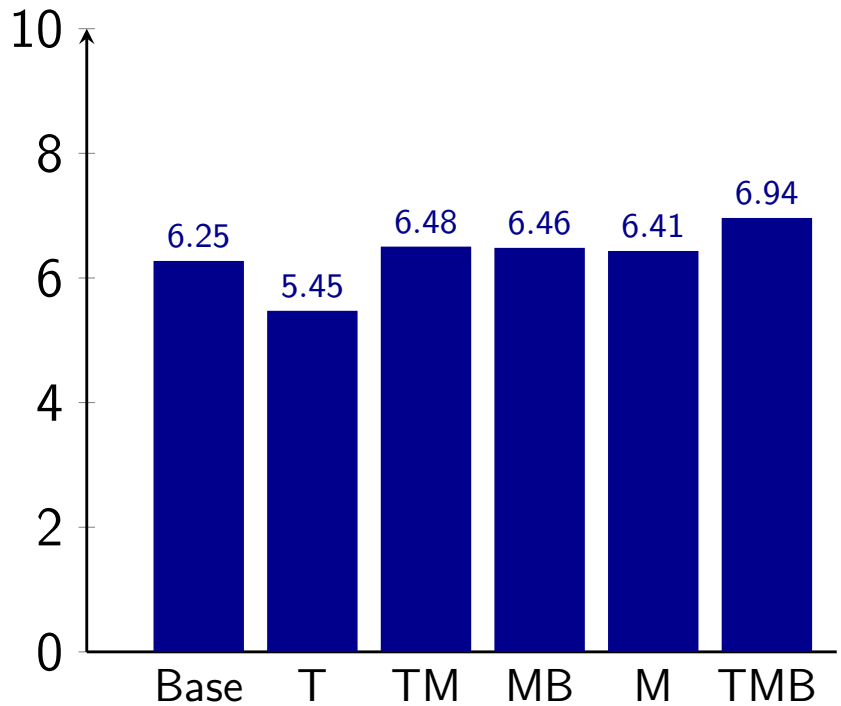
Expected (de-)motivational effects of thresholds

Hypotheses II: Thresholds within the range of achievable outcomes increase effort while they decrease effort outside.

	Effect on effort for	
	High-ability	Low-ability
 Top threshold	+	-
 Middle threshold	+	+
 Bottom threshold	-	+

Result I: Ranking design matters

Average Effort

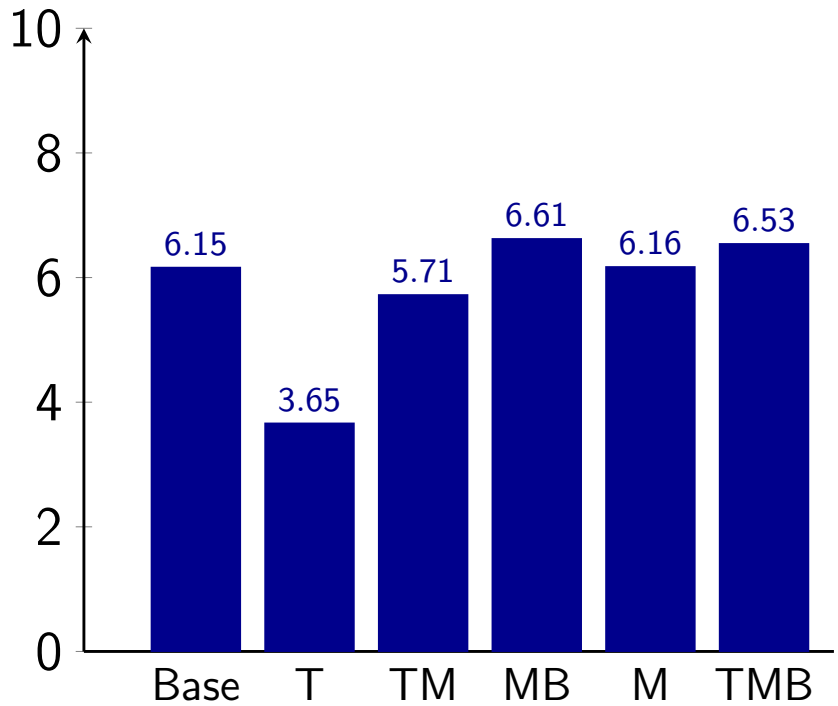


Result I: Effort choices depend on the ranking design

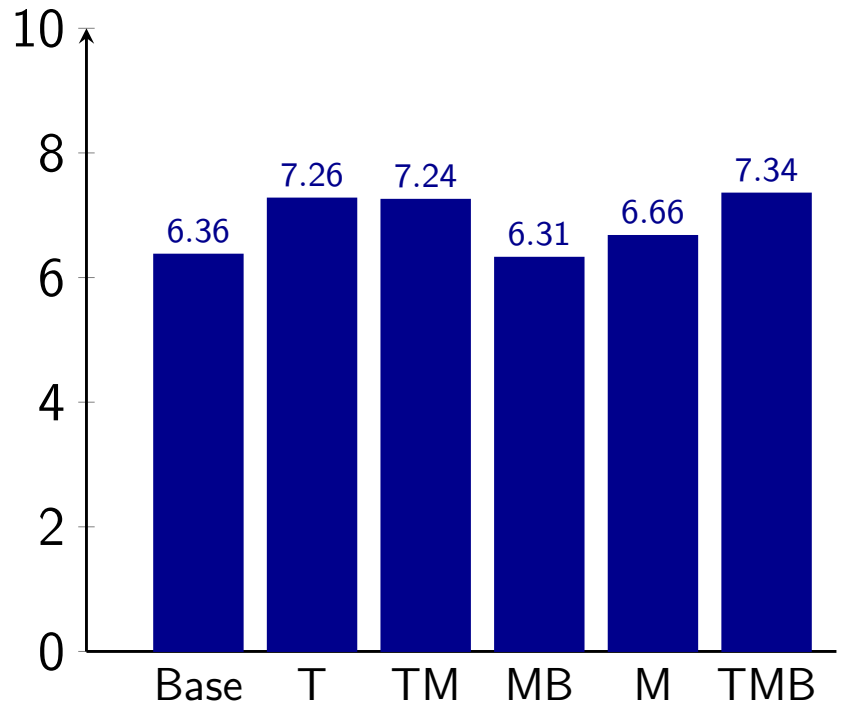
- The ranking with BMT thresholds motivates most ($p < 0.01$)
- The ranking with only a T threshold demotivates ($p < 0.01$)

Result II: Abilities matter

Effort low-ability type



Effort high-ability type



Result II: Impact of ranking design depends on abilities

- Low-ability physicians: highest effort with MB thresholds ($p < 0.1$), lowest effort for only T threshold ($p < 0.01$)
- High-ability physicians: highest efforts if top threshold is in place (T, TM, TMB: $p < 0.01$)

Conclusion: Careful design of rankings can improve quality in health care

- Peer feedback in from of rankings can motivate but also demotivate
- Effects depend on ability in terms of reachable outcomes
- Thresholds within the range of reachable outcome rather motivates while outside it may demotivate
- Demotivating effect less severe if also motivating thresholds are present
- Understanding of how ranking design and abilities interact can help to tailor rankings to group specifications
- Results are a base for a broader research agenda

Appendix

Our contribution: complement the sparse evidence on the impact of feedback in a health care context

- Social comparisons seem to work (e.g., Roels and Su 2014, Dranove et al. 2003, Cutler et al. 2004, Kolstad 2013)
- Only few studies about feedback in health care context (e.g., Song et al. 2018, Navathe et al. 2020)
- Impact of non-anonymous feedback stronger than anonymous feedback (e.g., Tafkov 2013)
- Usually: granular rankings (e.g., Tafkov 2013, Kuziemko et al. 2014) or honoring the top (e.g., Ashraf et al. 2014, Bradler et al. 2016)
- Ability matters (e.g., Kuziemko et al.2014, Ashraf et al. 2014)
- Scarce evidence on ranking design, settings differ (see Hannan et al. 2018)

Example Colonoscopy: High-Volume activities with measurable performance Indicators

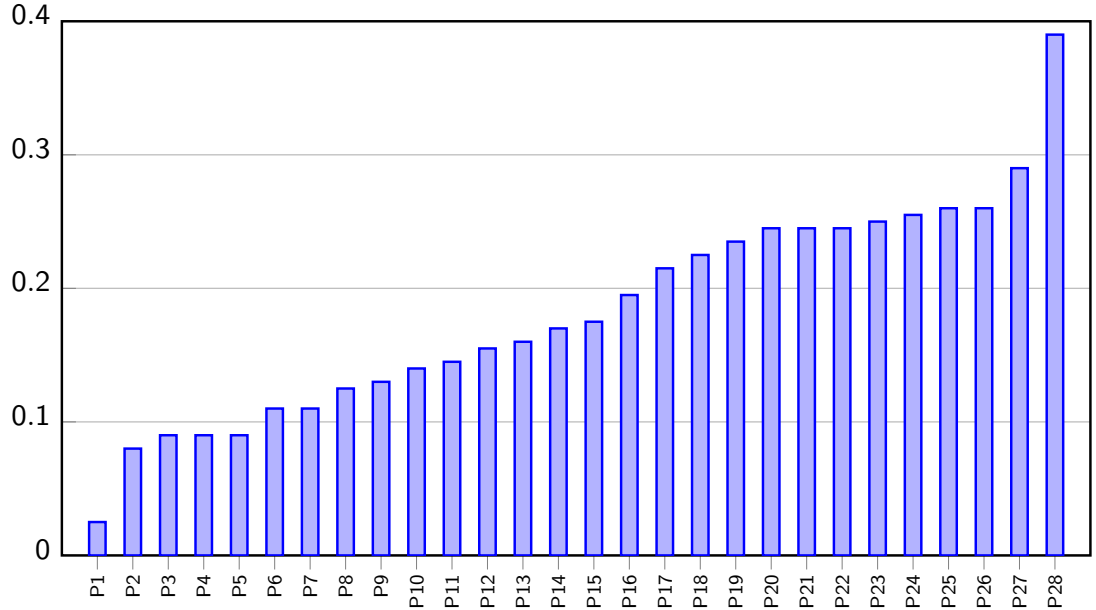


Figure: Adenoma detection rate for individual physicians

- Currently feedback is often provided in a rather pragmatic way
- Many options of designing a ranking based on the outcomes
- Non-anonymous feedback advocated by medical societies

Experimental Design: Distribution of Health Outcomes Depending on Effort

Effort-level e_i	Cost of effort $c(e_i)$	Physician with low ability		Physician with high ability	
		Probability $x = 0$	Probability $x = 20$	Probability $x = 5$	Probability $x = 25$
0	0	100%	0%	100%	0%
1	2	90%	10%	90%	10%
2	4	80%	20%	80%	20%
3	6	70%	30%	70%	30%
4	8	60%	40%	60%	40%
5	10	50%	50%	50%	50%
6	12	40%	60%	40%	60%
7	14	30%	70%	30%	70%
8	16	20%	80%	20%	80%
9	18	10%	90%	10%	90%
10	20	0%	100%	0%	100%

Discussion/Limitations

- Experimental results provide qualitative insights
  - Qualitative effects rather transferable than quantitative ones
  - Status concerns expected to be stronger in real-world settings
  - Experiment helps to understand underlying mechanisms
- Sample size tailored to within-subject comparisons
  - Comparison of sym. and asym. rankings not meaningful
  - Two ranking conditions were necessary to reduce complexity
- Effects were clearly larger for high-ability physicians
  - Might be driven by demotivational effects of learning abilities
  - Might be a self-selecting effect