Physician performance pay: Experimental evidence

Jeannette Brosig-Koch, Heike Hennig-Schmidt,
Nadja Kairies-Schwarz, Johanna Kokot, Daniel Wiesen

Empirical Management Conference 2019

December 12, 2019

supported by: German Research Foundation—DFG and Central research institution for ambulatory health care in Germany—Zi

Why do we care?

- Understanding how physicians respond to incentives is important for policy-makers and researchers alike
- The traditional payment system: fee-for-service may incentivize "too many" services; overtreatment (e.g., Ellis and McGuire 1986, JHE)
- A prominent attempt to control costs: lump-sum capitation (CAP) payments (e.g., in managed care); CAP may lead to underprovision of medical services (e.g., Cutler 1995, ECMA)
- Pay for performance (P4P) programs are frequently suggested to improve the quality of health care (e.g., UK, USA)
- Ongoing health policy debate on the introduction and design of P4P

Mixed empirical evidence

- Inconclusive evidence on the effect of performance pay on physicians' behavior (e.g., Epstein, 2012, NEJM, Witter et al. 2012, Cochrane Rev., Eijkenaar et al. 2013, EJHE, Milstein and Schreyögg, 2016, HP)

- (If at all) rather moderate effects of P4P (e.g., Mullen et al., 2010, RAND, Li et al., 2014, HE)

- Possible reasons:
  - Health outcomes might be biased or difficult to observe (e.g., Campbell et al., 2009, NEJM, Gravelle et al., 2010, EJ; Roland and Olesen, 2016, BMJ)
  - P4P is introduced oftentimes with other interventions (e.g., public reporting of performance)
  - Heterogeneity in physicians' responses typically not considered

▷ Causal effect of performance pay on physicians' behavior and the quality of health care is difficult to infer using field data

Size of bonus and crowding-out of altruistic behavior

- Not well understood how the level of a performance bonus affects physicians' behavior
- Adverse effects like a crowding-out of physicians' altruistic (patient-regarding) behavior and motivation might occur
- Other-regarding motivations are fundamental in public service provision (e.g., Besley and Ghatak 2005, AER; Prendergast 2007, AER; Delfgaauw and Dur 2008, EJ) particularly in health (Arrow, 1963, AER)
- Financial incentives might lead to crowding-out of intrinsic motivation (e.g., Deci 1971; Frey et al. 1996, JPE; Frey 1997, EJ)
- Some experimental evidence for motivation crowding-out (e.g., Gneezy and Rustichini 2000, QJE; Ariely et al. 2009, REStud)

▷ No causal evidence on the effect of bonus levels and on whether P4P crowds-out physicians' altruistic (patient-regarding) behavior

This paper

- Artefactual field experiment (in the sense of Harrison and List, 2004, JEL) with primary care physicians from a representative sample of resident physicians in Germany
- 'Clean' performance measure tied to the patient-optimal quality of medical care
- Within-subjects: Exogenous variation of the payment system from CAP to CAP + performance pay
- Between-subjects comparison of different bonus levels
- Link of behavioral data to physicians' practice characteristics such as location and annual profit (Levitt and List 2007, JEP; Camerer 2015)

Research questions

1. How does performance pay affect physicians' behavior?

2. Does the bonus level affect physicians' behavior (Low bonus of 5% vs. High bonus of 20% on top of CAP payment)?

3. How do physicians' real-world characteristics relate to their behavior in the experiment?

4. Does performance pay crowd-out physicians' altruistic (patient-regarding) behavior?

Our physician sample: Some background

- Overall, 104 primary care physicians (PCPs) participated in our artefactual field experiment
- Sub-sample ($\sim$10%) of PCPs enrolled in the practice panel of the *Central research institution for ambulatory health care in Germany* (ZiPP) which is a representative sample of resident physicians in Germany
- ZiPP is run annually with about 5,000 resident physicians
- In Germany, around 33,000 resident PCPs contract with the statutory health insurance (GKV), about 1,000 PCPs participate in the ZiPP

Sample characteristics

- Average age: 56 years (ZiPP: 54, German PCPs: ∼53 years)
- Share of female PCPs: 35% (ZiPP: 39% German PCPs: ∼44%)
- Distribution of locations similar to ZiPP
  - City: ∼30%; ZiPP: ∼34%
  - Outer conurbation: ∼36%; ZiPP: ∼37%
  - Rural: ∼34%; ZiPP: ∼29%
- Annual profit: ∅150,383 EUR (ZiPP: ∅158,733 EUR)
- Our sample not significantly different from non-participating PCPs of the ZiPP

General experimental design

- Within-subject design: introduction of performance pay at two different levels

Treatment	Payment system part *I*	Payment system part *II*	# Sub. (# pat.)
Low bonus (5%)	CAP	CAP+P4P-5%	53 (954)
High bonus (20%)	CAP	CAP+P4P-20%	51 (918)

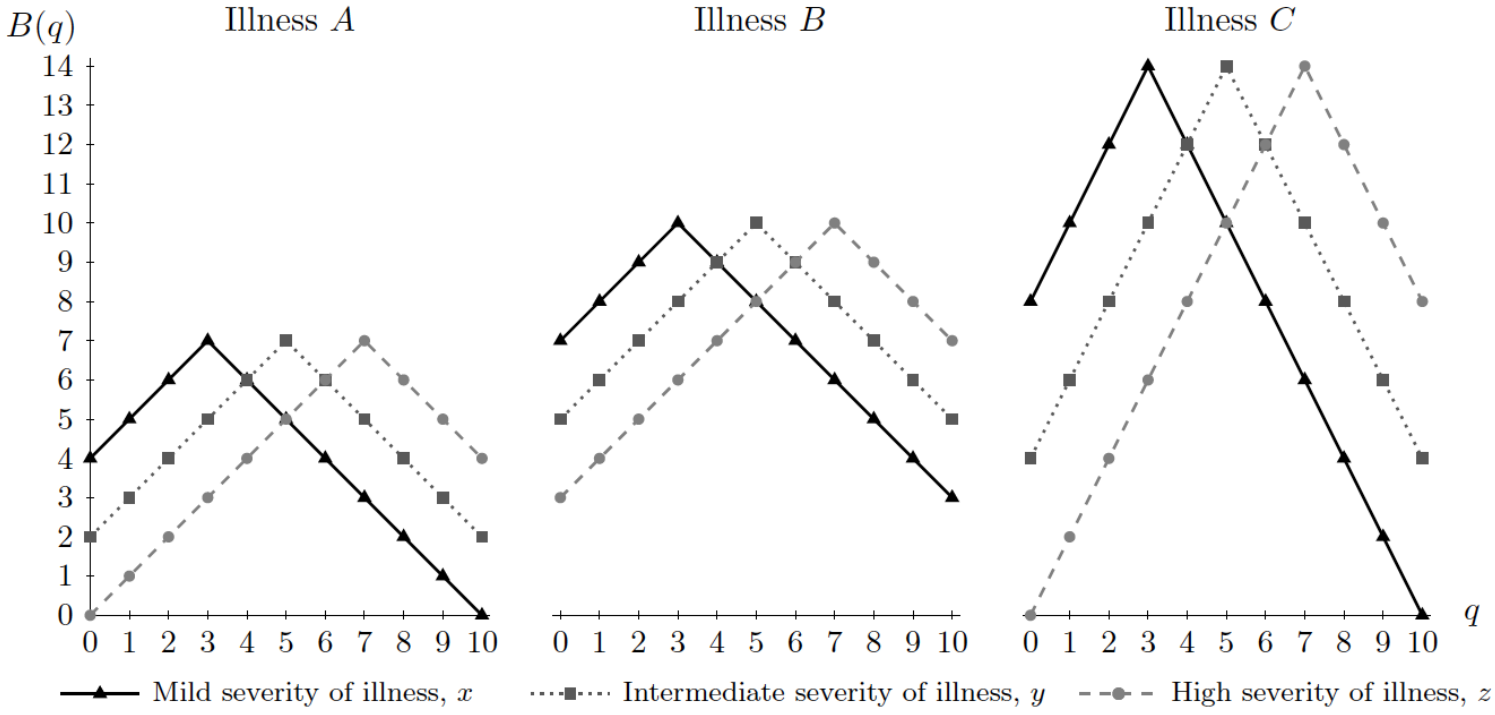
- Between-subject comparison for performance-pay systems
- Control treatments with medical students

Decision situation

- Framed physician decision-making experiment
- Physicians decide on the quantity of medical services q
- Individual decisions on $q \in \{0, 1, \ldots 10\}$ for 9 abstract patients
- Subjects simultaneously determine profit and the patient's health benefit (measured in monetary terms)
- Framing and setting are the same for all payment systems

Patients' health benefit

- Systematic variation of patients' health benefit; constant for all payment systems
- Illnesses A, B, C with three severities x (mild), y (interm.), z (high)



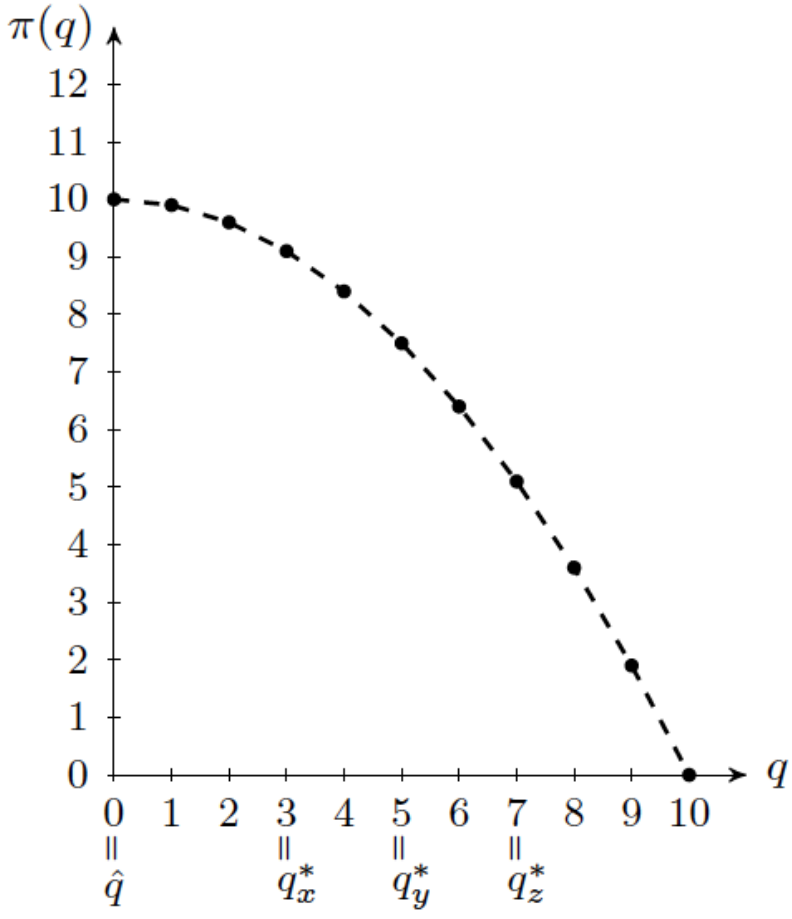
- Salient incentive: Patients' health benefit measured in monetary terms, benefits real patients' health outside the lab

Payment systems

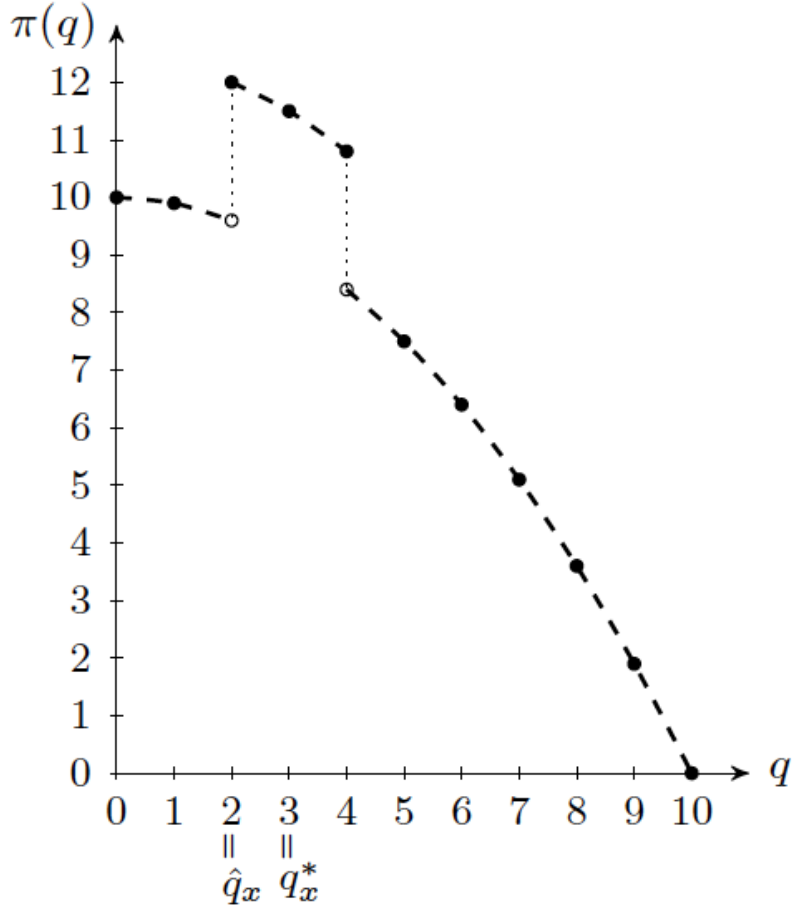
- CAP: lump-sum payment of 25 EUR for physicians
- Performance pay linked to patients' benefit (health outcome) and adjusted for severities of illness
- Discrete bonus is granted if quality threshold is reached $|q - q^*| \leq 1$
- Reflects asymmetric information between payer and physician
- Different bonus rates for patients' severity of illness; risk adjusted (e.g., Glazer and McGuire 2000, AER)
- Cost are convex $c(q) = q^2/10$

Parameters: Illustration of physicians' profits

CAP



CAP+P4P



Sample decision screen

Patient with illness B, mild severity (x)

Round 1: Patient 1 **Link to instructions**

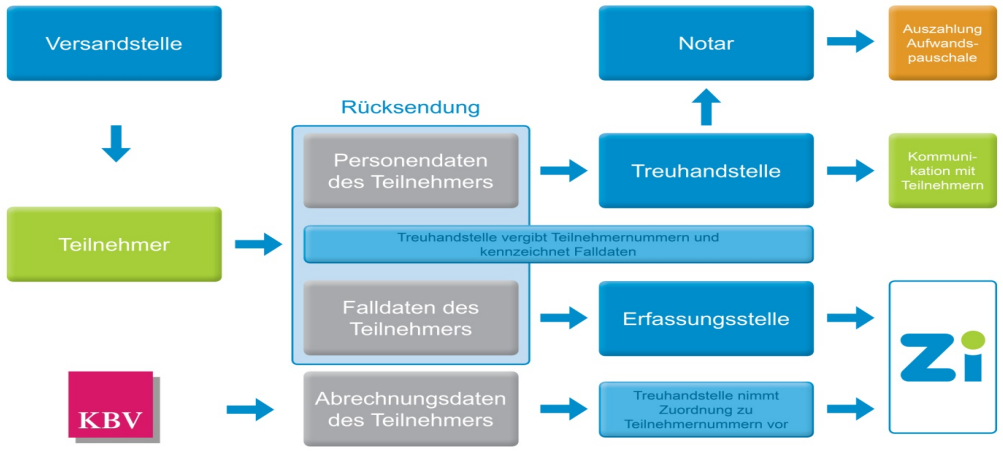
Quantity of medical services	Your lump-sum remuneration (in Euro)	Your bonus payment (in Euro)	Your costs (in Euro)	Your payoff = remuneration + bonus – costs (in Euro)	Benefit of the patient with illness B and severity x (in Euro)
0	25	0.00	0.00	25.00	17.5
1	25	0.00	0.25	24.75	20.0
2	25	2.25	1.00	26.25	22.5
3	25	2.25	2.25	25.00	25.0
4	25	2.25	4.00	23.25	22.5
5	25	0.00	6.25	18.75	20.0
6	25	0.00	9.00	16.00	17.5
7	25	0.00	12.25	12.75	15.0
8	25	0.00	16.00	9.00	12.5
9	25	0.00	20.25	4.75	10.0
10	25	0.00	25.00	0.00	7.5

Which quantity of medical services do you want to provide?

send...

Facilitation of the artefactual field experiment

ZiPP: data collection procedure



- Double-blind procedure
- Anonymity of subjects ensured
- Experiment followed the data security guidelines of the ZiPP
- Payment procedure via notary office

Experimental protocol

- Experiments with physicians were run in March 2016; average duration: about 30 minutes
- Post experimental questionnaire (e.g., risk attitudes, altruism)
- Random payment technique: One decision is randomly selected for payment in each part
- Average payment per subject: 45.93 EUR (total: 4,823 EUR)
- Average payment per patient: 47.64 EUR (total: 5,003 EUR)
- Behavioral data linkage: Administrative data on, for example, profit, location, practice characteristics are provided by Zi

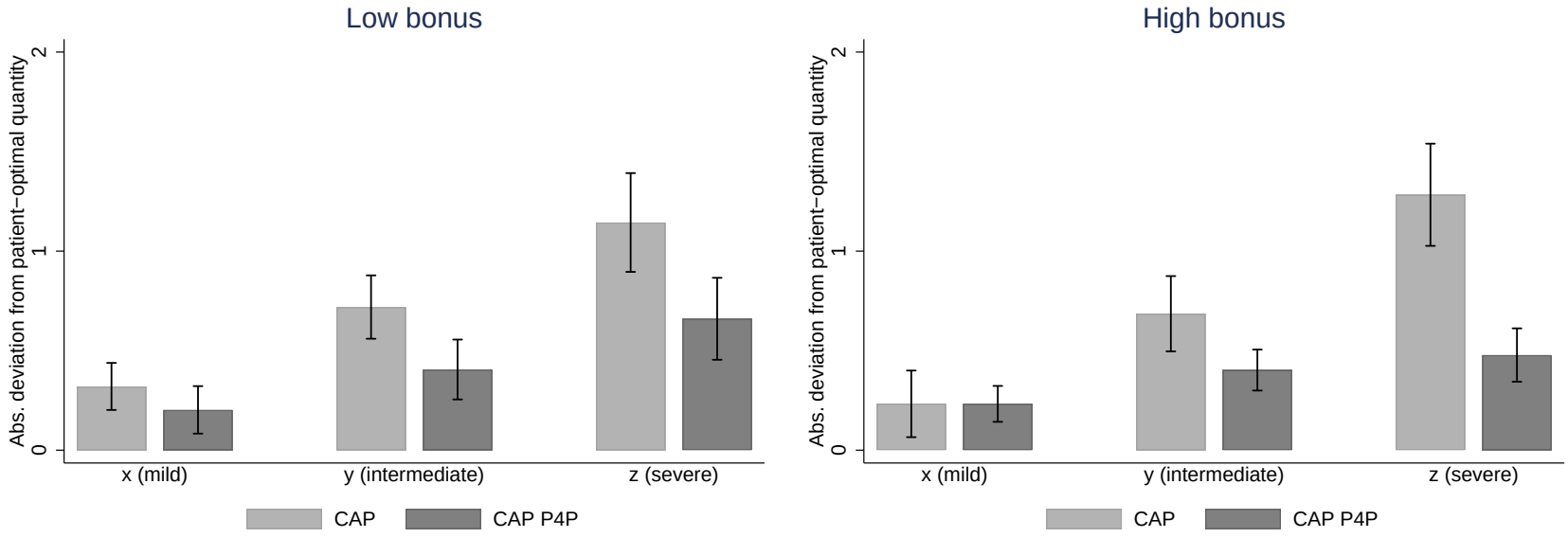
Behavioral results

Physicians' medical service provision in CAP (part *I* of the experiment)

- Physicians significantly underprovide medical services in CAP for patients with intermediate and high severity of illness ($p \leq 0.014$, Wilcoxon signed-rank test; comparison with q^* for all illnesses)
- Underprovision increases in patients' severity of illness, patients' marginal benefit does not significantly affect behavior
- Consistent with findings in the experimental literature (e.g., Hennig-Schmidt et al. 2011, JHE; Brosig-Koch et al., 2017, HE)

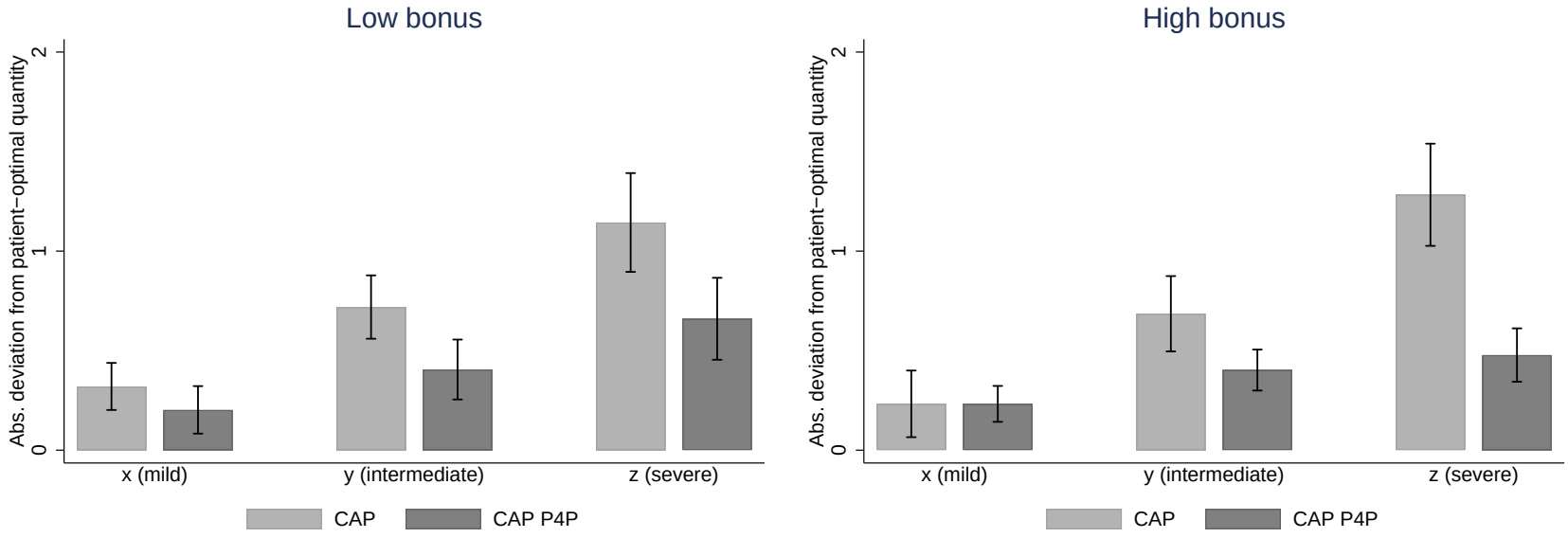
How performance pay affects physicians' behavior

Abs. deviation from the patient-optimal quantity



How performance pay affects physicians' behavior

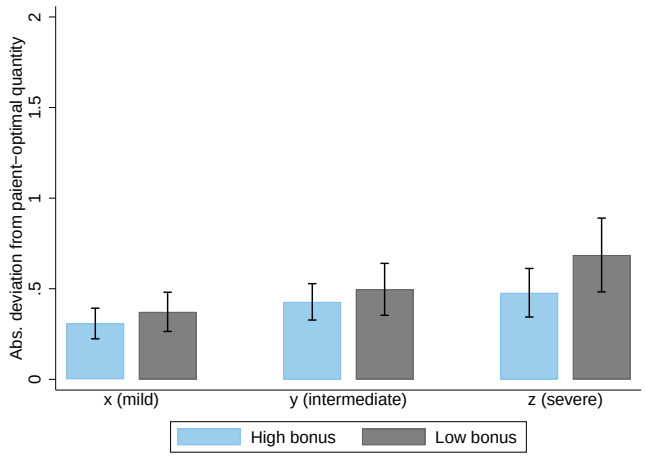
Abs. deviation from the patient-optimal quantity



- Underprovision is significantly reduced for intermediately (y) and severely ill (z) patients in CAP+P4P-20% and CAP+P4P-5% ($p \leq 0.094$, Wilcoxon signed-rank test)
- For mild severity patients (x), the reduction in underprovision is not significant ($p > 0.162$)

Does the size of the bonus affect behavior?

Absolute deviation from the patient-optimal quantity (part *II* of the experiment)



- Very similar behavioral responses for the two different bonus levels
- No statistically significant differences ($p > 0.4964$, Mann-Whitney U-Test)

▷ The bonus level does not significantly affect physicians' behavior.

Data linkage: Behavior and physicians' characteristics

Dependent variable:	Abs. deviation from patient-optimal care $\lvert q - q^* \rvert$
Performance pay	-0.116**
	(0.656)
Performance pay × Interm. severity	-0.253***
	(0.080)
Performance pay × High severity	-0.567***
	(0.113)
High annual profit (= 1 if above 147k EUR)	0.304**
	(0.130)
City	0.369**
	(0.185)
Outer conurbation	0.173
	(0.157)
Other experimental controls	Yes
Other physicians' characteristics	Yes
Constant	1.100
	(0.739)
Observations	1,800
Subjects	100
R^2	0.163

OLS; ref. cat.: CAP; robust SE clustered for sub.; 'Other experimental controls are dummies for High bonus, severities y, z, and illness C. 'Other physicians' characteristics' comprise 'Share of SHI-patients', 'time per SHI-patient', 'profit share from SHI', '# of physicians in practice', and self-reported attitudes. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

Quality of care and physicians' characteristics

So far...

- Performance pay significantly increases the quality of care (i.e., reduces the abs. deviation to the patient-optimal quantity)
- Performance pay effect depends on patients' severity of illness
- ▷ Deviations from the patient-optimal quantity is higher for physicians with high annual profit...

  ...and are significantly higher for physicians practicing in cities compared to rural areas

Crowding-out of altruistic (patient-regarding) behavior

Descriptive analysis

- Analysis is based on how (104x9) individual patients are treated in both parts
- Treatment types:
  - Profit maximization (PM)
  - Benefit maximization (BM)
  - Trade-off (TO)
- Treatment types by part of the experiment:
  - Part *I* (CAP): PM: 1%; BM: 54%; TO: 42%; Other: 3%
  - Part *II* (CAP+P4P): PM: 30%; BM: 64%; TO: 0%; Other: 6%
- Transitions:
  - Crowding out: BM $\longrightarrow$PM: 7% ($\sim$ 14% of BM); TO $\longrightarrow$ PM: 22%
  - Crowding in: PM $\longrightarrow$ BM: 1%; TO $\longrightarrow$ BM: 17%

Policy implications

Costs of P4P from a payer's perspective:

- Low bonus: 22.6% increase of remuneration
- High bonus: 37.4% increase of remuneration

Increase of health benefit:

- Low bonus: 8%
- High bonus: 7.5%

Arc-elasticity of patient benefit $\frac{(H_2-H_1)/(H_2+H_1)}{(R_2-R_1)/(R_2+R_1)}$

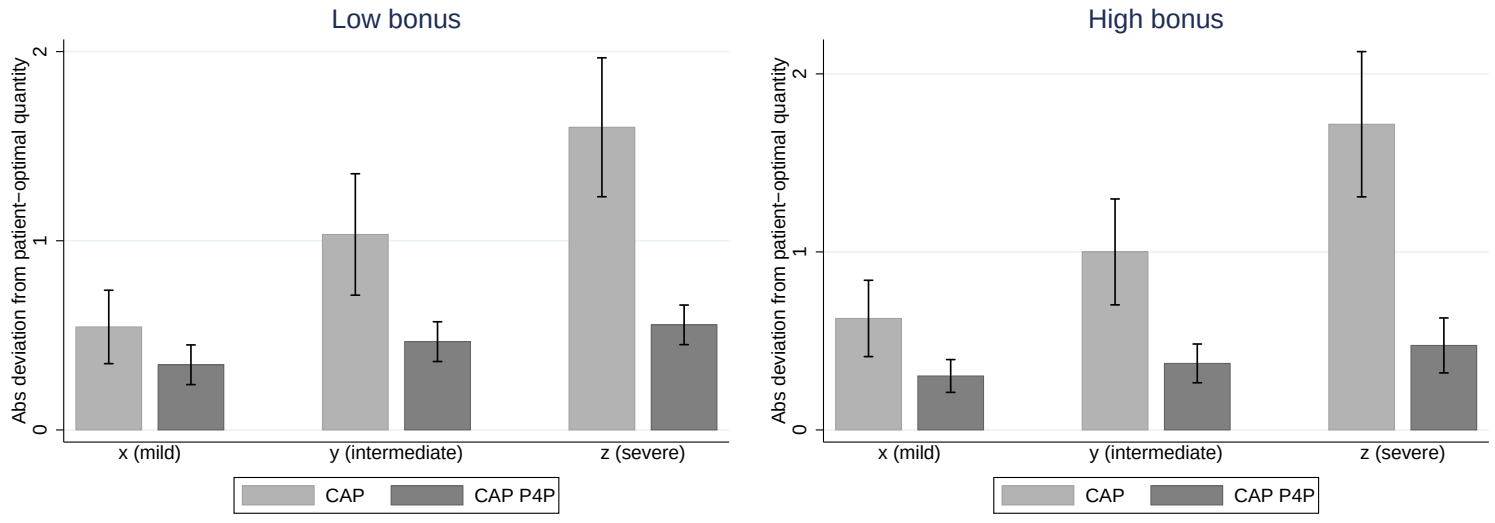
- Low bonus: 0.18
- High bonus: 0.08

Concluding remarks

- Controlled artefactual field experiments to test the effect of introducing performance pay on physicians' behavior
- Underprovision in CAP is significantly reduced under performance pay
- Patients' severity of illness affect physicians' behavior
- Surprisingly, the level of the bonus pay does *not* significantly affect physicians' behavior
- Physicians with higher profits are less concerned about the quality of care
- Non-negligible evidence for crowding-out of patient-regarding altruistic behavior

THANK YOU FOR YOUR ATTENTION!!!

Does the behavior of physicians and med. students differ?



- Within-subjects: Underprovision in CAP is significantly reduced under performance pay
- Level of bonus pay does not significantly affect students either

▷ Performance pay affects students' behavior very similarly.

Robustness of results: Evidence from control treatments with medical students

- "Taking performance pay away" (reverse order) does not affect medical students behavior in a significant way compared to introducing performance pay
- No significant differences under constant maximum incentives (increased capitation)

▷ Findings are robust across subject pools and towards order of payment systems as well as levels of incentives.

Why an experiment? A complementary approach

1. Lab and artefactual field experiments are well suited to testing explicit predictions of simple theoretical models under controlled conditions.

2. No patients are harmed

3. Experiments often provide unique opportunities to study behavior that is hidden or prohibited in the field.

4. Experimental data, combined with field studies and social surveys, can help us understand sources of heterogeneity in behaviors.

5. Experiments are highly replicable.