# Physician performance pay: Experimental evidence

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

iHEA Session at ASSA Meeting, Philadelphia

January 5-7th, 2018

# Why do we care?

- Mounting health care costs in many Western countries (e.g., Baicker and Goldman, 2011 JEP; Chandra and Skinner, 2012, JEL)
- Understanding how physicians respond to incentives is a important for policy-makers and researchers alike
- Traditional payment system: fee-for-service incentivizes "too many" services (e.g., Ellis and McGuire 1986, JHE)
- A prominent attempt to control costs: lump-sum capitation (CAP) payments (e.g., in managed care); CAP incentivizes underprovision of medical services (e.g., Cutler 1995, ECMA)
- P4P-programs are frequently suggested to improve the quality of health care (e.g., UK, USA)

# Empirical literature on P4P

- Inconclusive evidence on the effect of performance pay on physicians' behavior (e.g., Sutton, 2012, NEJM, Witter et al. 2012, Cochrane Rev., Eijkenaar et al. 2013, EJHE)
- (If at all) rather moderate effects of P4P (e.g., Mullen et al., 2010, RAND, Li et al., 2014, HE)
- Health outcomes might be biased due to measurement errors (e.g., Campbell et al., 2009, NEJM, Gravelle et al. 2010 EJ)
- Often performance pay is introduced with other interventions (e.g., public reporting of performance)

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

# Other-regarding motivations in public services

- Other-regarding motivations are a fundamental determinant of public service provision (e.g., Besley and Ghatak 2005, AER; Delfgauuw and Dur 2010, JPubE; Prendergast 2007, AER)
- Financial incentives might lead to crowding-out other regarding motivations (e.g., Deci 1971; Frey et al. 1996, JPE; Frey 1997, EJ)
- Some experimental evidence for motivation crowding-out in work effort (e.g., Gneezy and Rustichini 2000, QJE; Huffman and Bognanno, 2017, MS)
- Level of incentive may be an important driver of a behavioral change (Ariely et al. 2009, REStud)

▷ No empirical evidence on whether P4P crowds-out physicians' altruistic (patient-regrading) motivation

# This paper

- Artefactual field experiments (in the sense of Harrison and List, 2004, JEL) with physicians from a representative sample of resident physicians in Germany
- 'Clean' performance measure tied to 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
- First field experiment studying the causal effect of introducing performance pay on physicians' behavior
- Link of behavioral data to physicians' individual characteristics

# Outline

# Related behavioral experiments in health

- Health care markets: Kessler and Roth (2014 AER, 2014 JPubE)
- Non-monetary incentives: Kesternich et al. (2015, JPubE), Godager et al. (2016 JEBO)
- Fee-for-service and capitation: Hennig-Schmidt et al. (2011, JHE), Hennig-Schmidt and Wiesen (2014, SSM), Lagarde and Blauuw (2017, SSM), Brosig-Koch et al. (2016, JEBO)
- Mixed payment systems: Brosig-Koch et al. (2017, HE)
- P4P: Lagarde and Blauuw (2016, WP), Cox et al. (2016, JEBO)

# Research questions

1. How does performance pay affect physicians' behavior?

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

3. Does performance pay crowd-out physicians' altruistic motivation?

4. How do physicians' characteristics relate to crowding-out of motivation?

# Some background

- Overall, 104 primary care physicians (PCPs) participated in our artefactual field experiment
- Sub-sample (∼10%) of PCPs enrolled in the Zi-Praxis-Panel (ZiPP) of the *Zentralinstitut der Kassenärtzlichen Bundesvereinigung* 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
- About 72 million people (∼88% of German population) are enrolled in a statutory health insurance scheme

# PCPs: Some sample characteristics

- Average age: 54 years (ZiPP: 54, German PCPs: ∼53 years)
- Share of female PCPs: 35% (ZiPP: 39% German PCPs: ∼44%)
- Distribution of locations similar to ZiPP
  - City: Our sample: ∼37%; ZiPP: ∼34%
  - Outer conurbation: Our sample: ∼44%; ZiPP: ∼37%
  - Rural: Our sample: ∼19%; ZiPP: ∼29%

# General experimental design

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

| Treatment | Payment system, part *I* | Performance-pay systems, part *II* | # Sub. (# pat.) |
|---|---|---|---|
| 20% bonus | CAP | CAP+P4P-20% | 51 (918) |
| 5% bonus | CAP | CAP+P4P-5% | 53 (954) |

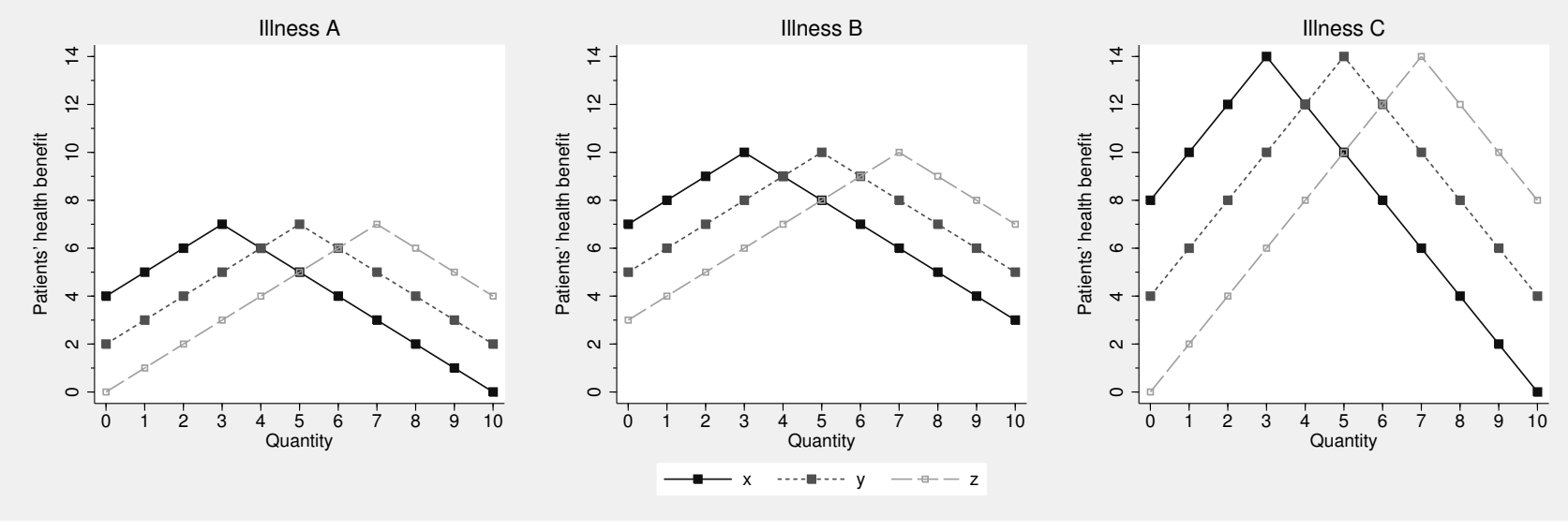
- Bewteen-subject comparison for performance-pay systems

# Decision situation

- Framed physician decision-making experiment
- Subjects decide (in the role of physicians) 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 $x, y, z$ (i.e., low, intermediate, high)




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

# Payment systems

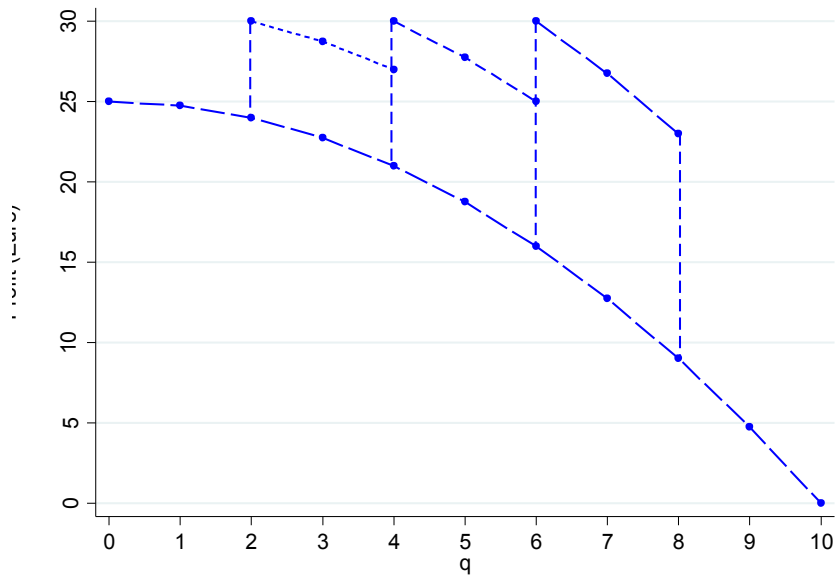
- Performance pay linked to patients' benefit (health outcome) and adjusted for severities of illness
- Bonus is granted if quality threshold is reached $|q - q^*| \leq 1$
- Different rates for patients' severity of illness can be interpreted as of risk adjustment (e.g., Glazer and McGuire, 2000, AER)
- CAP: lump-sum payment of 25 EUR for physicians
- Reflects asymmetric information between payer and provider

| Severity | CAP+P4P-20% | CAP+P4P-5% |
|---|---|---|
| $x(q^* = 3)$ | 6 | 2.25 |
| $y(q^* = 5)$ | 9 | 5.25 |
| $z(q^* = 7)$ | 14 | 10.25 |

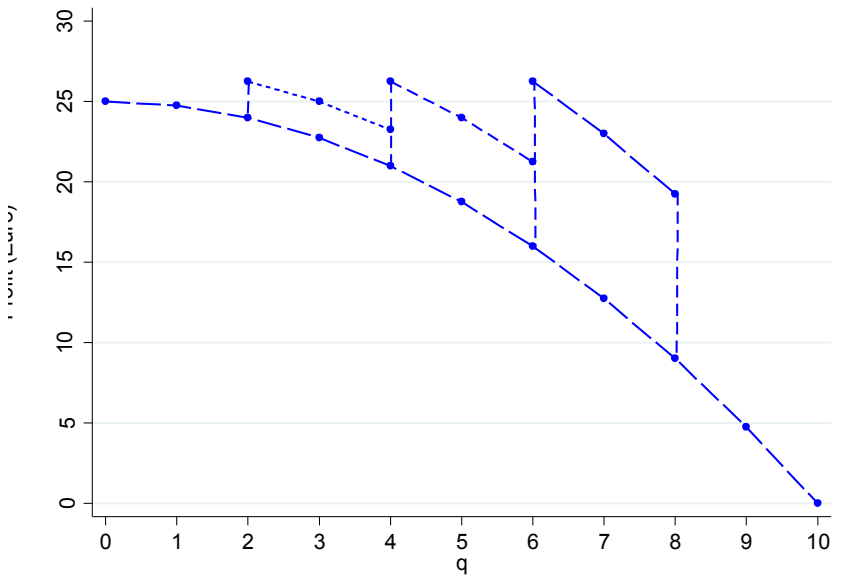
- Cost are convex $c(q) = q^2/4$

# Parameters: Profits in CAP+P4P-20 and CAP+P4P-5

CAP-P4P-20%




CAP-P4P-5%



# Sample decision screen

SoPHIE

**Runde 1: Patient 1**

Link zu Instruktionen

| Anzahl medizinischer Leistungen | Ihre pauschale Vergütung (in Euro) | Ihre Bonusvergütung (in Euro) | Ihre Kosten (in Euro) | Ihr Verdienst = Vergütung + Bonus − Kosten (in Euro) | Nutzen des Patienten mit Erkrankung B Schweregrad x (in Euro) |
|---|---|---|---|---|---|
| **0** | 25.00 | 0.00 | 0.00 | 25.00 | 17.50 |
| **1** | 25.00 | 0.00 | 0.25 | 24.75 | 20.00 |
| **2** | 25.00 | 6.00 | 1.00 | 30.00 | 22.50 |
| **3** | 25.00 | 6.00 | 2.25 | 28.75 | 25.00 |
| **4** | 25.00 | 6.00 | 4.00 | 27.00 | 22.50 |
| **5** | 25.00 | 0.00 | 6.25 | 18.75 | 20.00 |
| **6** | 25.00 | 0.00 | 9.00 | 16.00 | 17.50 |
| **7** | 25.00 | 0.00 | 12.25 | 12.75 | 15.00 |
| **8** | 25.00 | 0.00 | 16.00 | 9.00 | 12.50 |
| **9** | 25.00 | 0.00 | 20.25 | 4.75 | 10.00 |
| **10** | 25.00 | 0.00 | 25.00 | 0.00 | 7.50 |

Welche Anzahl medizinischer Leistungen möchten Sie erbringen?

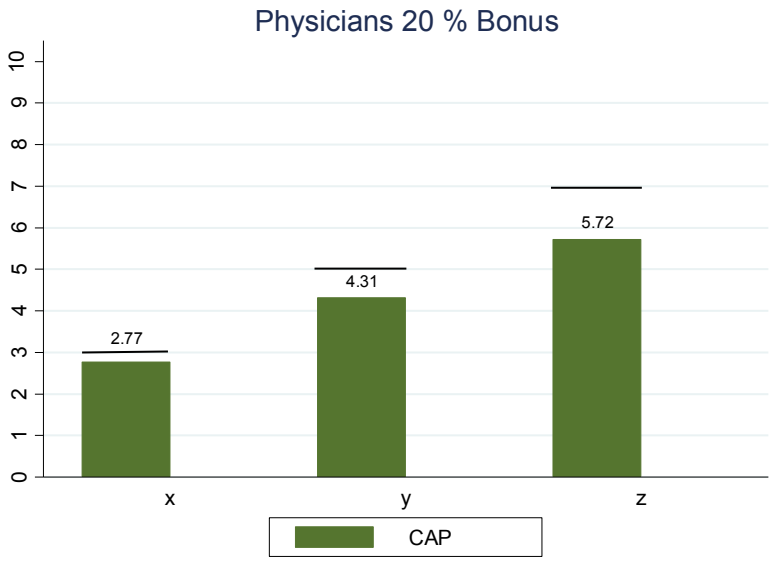
Absenden ...

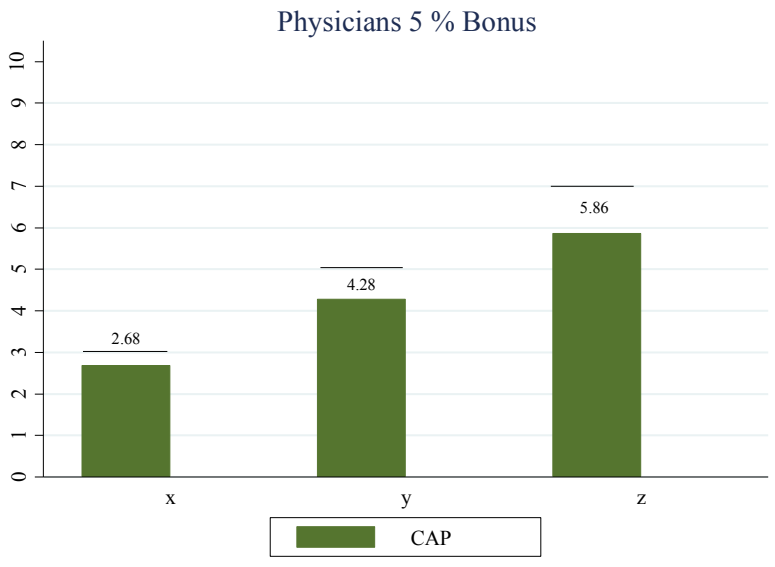
# Experimental protocol

- Field experiments were run in March 2016; average duration of about 25 minutes
- Double-blind procedure: experiment facilitated by trust office (usually running the ZiPP) ensured anonymity of subjects, payment procedure via notary office
- Random payment technique: One decision is randomly selected for payment from each part
- Average payment per subject: 45.93 EUR (total: 4,823 EUR)
- Average patient benefit: 47.64 EUR (total: 5,002.50 EUR)

# Physicians' medical service provision in CAP

## Average quantities by severity of illness

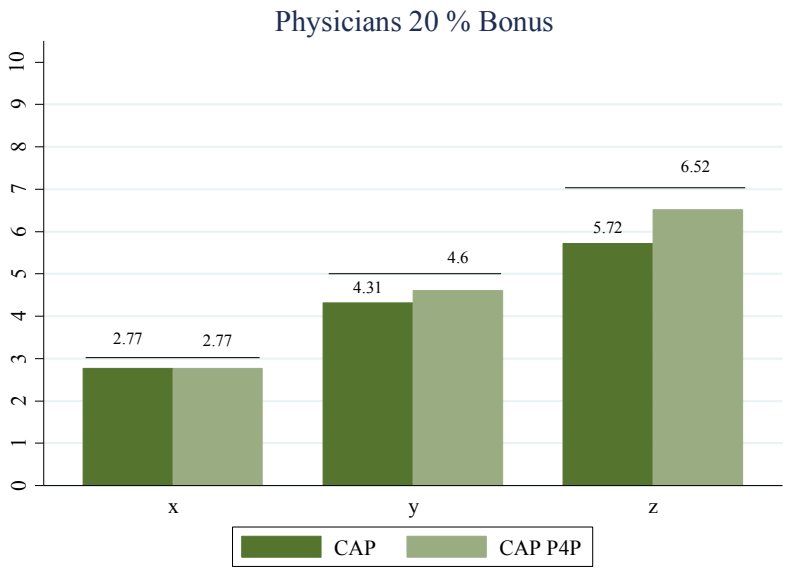


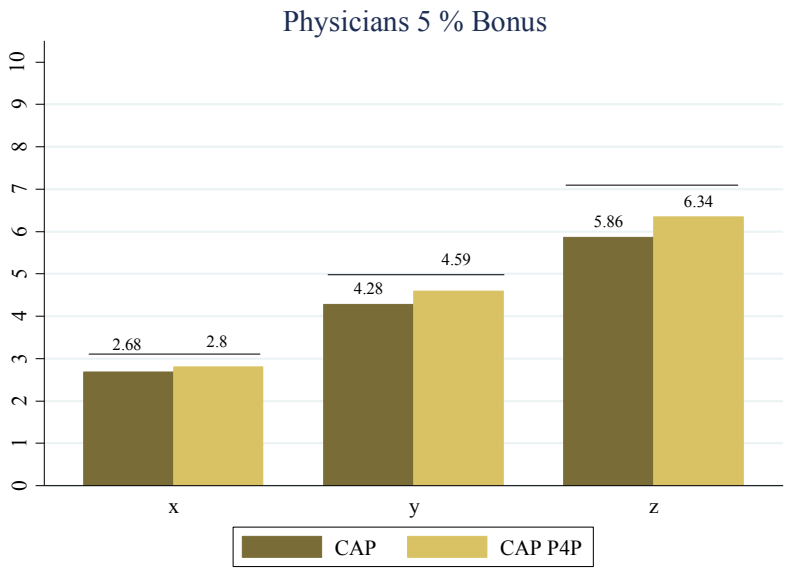




- Physicians significantly underprovide medical services in CAP ($p \leq 0.007$, Wilcoxon signed-rank test)
- Severity of illness significantly affects quantity choices
- Consistent with findings in the empirical and experimental literature (e.g., Cutler, 1995, ECMA Hennig-Schmidt et al. 2011, JHE)

# How performance pay affects physicians' behavior

## Average medical services by payment system and severity



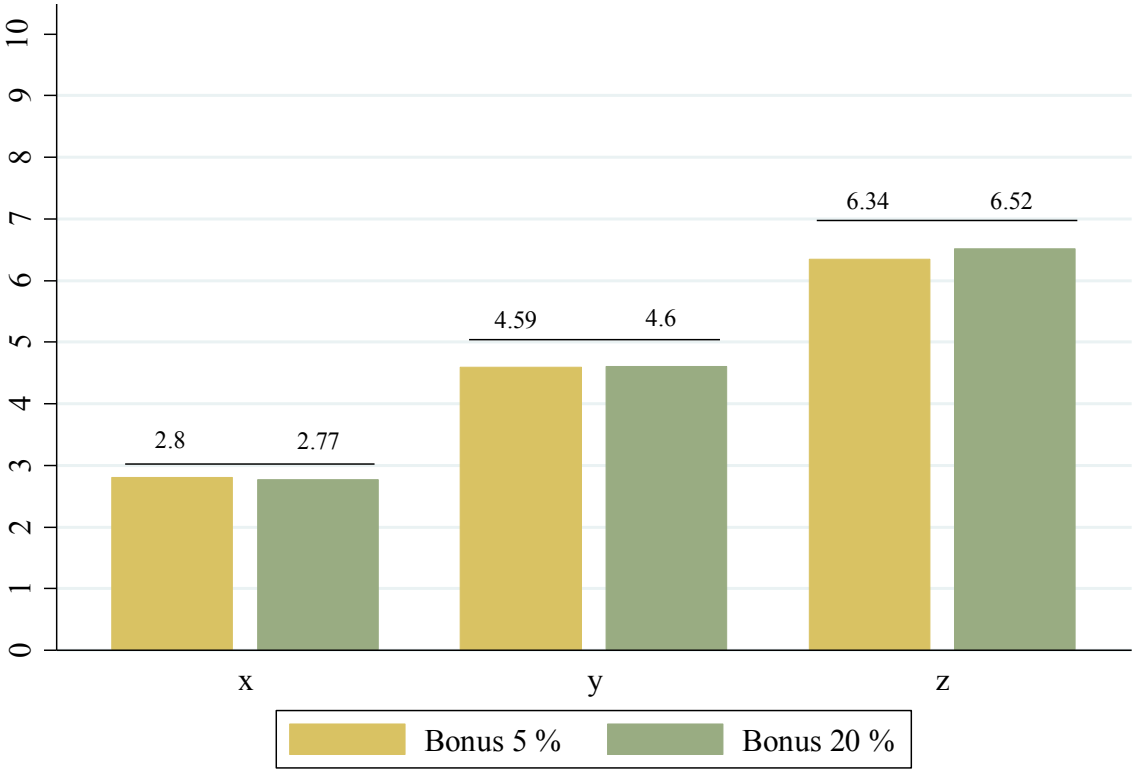
# Within-subject comparisons

- Underprovision is significantly reduced for intermediately ($y$) and severely ill ($z$) patients in CAP+P4P-20% and CAP+P4P-5% ($p \leq 0.0014$, Wilcoxon signed-rank test)
- For low severity patients ($x$), the reduction in underprovision is weakly significant in CAP+P4P-20% ($p \leq 0.0823$)
- For CAP+P4P-5%, the reduction for low severity patients ($x$) is not significant ($p \geq 0.6284$)

▷ Performance pay reduces underprovision of medical services inherent in CAP for intermediately and severely ill patients under both P4P schemes.

# Physicians' behavior under high and low performance pay

## Average medical services by payment system and severity




- Between-subjects: No significant effect of bonus level

▷ A reduction in the bonus level does not significantly affect behavior.

# Crowding-out of altruistic motivation

- 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: 1.5%; BM: 54%; TO: 41%; Other: 2.5%
  - Part *II* (CAP+P4P): PM: 30%; BM: 64%; TO: 0%; Other: 4%
- Transitions:
  - Crowding out: BM $\longrightarrow$PM: 7% ($\sim$ 14% of BM); TO $\longrightarrow$ PM: 24%
  - Crowding in: PM $\longrightarrow$ BM: 1%; TO $\longrightarrow$ BM: 17%

# Crowding out and physicians' characteristics

Logit regression on crowding out of altruistic behavior

| | (1) | (2) |
|---|---|---|
| Condition: 20%-Bonus | 0.0089 | 0.0105 |
| | (0.0094) | (0.0092) |
| Low severity (= 1 if $I = x$) | 0.0376 | 0.0364 |
| | (0.0274) | (0.0264) |
| Interm. severity (= 1 if $I = y$) | 0.0203 | 0.0195 |
| | (0.0207) | (0.0200) |
| Marginal health benefit | -0.0300** | -0.0289** |
| | (0.016) | (0.0142) |
| Age | -0.0005 | -0.0034 |
| | (0.0012) | (0.0023) |
| Gender (= 1 if male) | -0.0392* | -0.0430** |
| | (0.0220) | (0.0202) |
| City | -0.0447*** | -0.4581*** |
| | (0.1738) | (0.0162) |
| Outer conurbation | -0.0416** | -0.0432** |
| | (0.1915) | (0.1846) |
| Years in practice controls | No | Yes |
| Other characteristics | Yes | Yes |
| Observations | 936 | 936 |
| Subjects | 104 | 104 |

Dependent variable: Crowding-out (1 if $q_I^* \rightarrow \hat{q}_{II}$; 0 otherwise)
Logit reg.; marginal effects; ref. category: 'high severity $z$' and 'rural'
clustered for subjects; robust SE; ***$p < .01$, **$p < .05$, *$p < .1$

# Crowding out of altruistic behavior

- Some evidence for crowding out of physicians' patient-regarding/ altruistic behavior
- Likelihood for crowding out decreases in the patients' marginal health benefit and is not significantly affected by the severity
- Likelihood for crowding out is significantly higher for physicians in rural areas than for physicians in cities and towns
- Likelihood for crowding-out is lower for male physicians

# Concluding remarks

- First controlled (artefactual) field experiment to analyze effect of performance pay on physicians' provision behavior
- Underprovision in CAP is significantly reduced in P4P-systems
- Severity of illness significantly affect physicians' behavior
- Level of bonus does *not* significantly affect physicians' behavior
- Non-negligible evidence for crowding-out of patient-regarding/ altruistic behaviors
- Physicians' gender and location significantly relate to crowding out of altruistic behavior