Personality and physician performance pay: Evidence from a behavioral experiment in health

Mona Groß
Department of Health Care Management,
University of Cologne, Germany

Heike Hennig-Schmidt
Department of Economics,
University of Bonn, Germany
Departement of Health Economics and Health Management,
University of Oslo, Norway

Daniel Wiesen
Department of Health Care Management,
University of Cologne, Germany

UNIVERSITY OF OSLO
HEALTH ECONOMICS RESEARCH NETWORK
Working paper 2023:5

Health Economics Research Network at the University of Oslo
ISSN 1501-9071 (print version.), ISSN 1890-1735 (online), ISBN 978-82-7756-279-7

Personality and physician performance pay: Evidence from a behavioral experiment in health*

Mona Groß, Heike Hennig-Schmidt, Daniel Wiesen

August 19, 2023

Abstract

We study how the heterogeneity in responses to performance pay can be explained by personality traits. We utilize data from behavioral experiments and surveys on personality traits with physicians, medical students, and non-medical students. Performance pay is introduced at a within-subject level and complements either fee-for-service or capitation. We find that the payment system matters regarding the behavioral impact of personality traits. More conscientious and more agreeable individuals provide higher quality of care under capitation. Although performance pay further improves the quality, more conscientious and agreeable individuals respond less to capitation-based performance pay. Other personality traits are not behaviorally relevant. Under fee-for-service-based schemes, personality traits do not significantly related to individuals' behavior. Our findings inform the incentive design for physicians and the potential sorting into incentive schemes based on personality traits.

Keywords: Fee-for-service, capitation, blended pay for performance, personality traits, quality of care, heterogeneity

JEL-Classification: C91, I11

1 Introduction

Improving the quality of care is a key goal of health care policy-makers. In order to achieve this goal, performance pay has become a preferred means, for example, in the UK (e.g., Roland, 2004; Doran et al., 2006; Rosenthal et al., 2005, 2006; Kristensen et al., 2014; Roland and Campbell, 2014; Song et al., 2019), the US, and in several developing countries (e.g., Olken et al., 2014; Peabody et al., 2011; Van de Poel et al., 2016; Basinga et al., 2011; De Walque et al., 2015; Sherry et al., 2017). Paying for performance

*Groß: Department of Health Care Management, University of Cologne, Germany, e-mail: mona.gross@wiso.uni-koeln.de; Hennig-Schmidt: Department of Economics, University of Bonn, Germany, Departement of Health Economics and Health Management, University of Oslo, Norway, e-mail: hschmidt@uni-bonn.de; Wiesen: Department of Health Care Management, University of Cologne, Germany, e-mail: wiesen@wiso.uni-koeln.de (corresponding author). We are grateful for valuable comments and suggestions from conference participants at BEHnet workshop Vienna, EuHEA Oslo and GfeW Salzburg as well as from Geir Godager, Glenn Harrison, Mathias Kifmann, Jonathan Kolstad, Olga Kostopoulou, Ludwig Kuntz, and Sandra Sülz. Financial support by the German Research Foundation (BR 2346/2-1/2) and by the University of Oslo (Grant Number 296114, Research Council of Norway)is gratefully acknowledged. This study is partially based on a data set described in "Physicians' incentives, patients' characteristics, and quality of care: A systematic experimental comparison of fee-for-service, capitation, and pay for performance" (Ruhr Economic Paper No. 923).

typically rewards individuals for achieving pre-specified performance targets which speaks to healthcare providers' *extrinsic* motivation (e.g., Prendergast, 1999; Lazear, 2000). Despite the extended use of pay-for-performance (P4P), the evidence of its impact on the quality of care is inconclusive, and, if at all, moderate with substantial heterogeneity in providers' responses (e.g., Scott et al., 2011; Eijkenaar et al., 2013; Miller and Babiarz, 2014; Milstein and Schreyögg, 2016; Mendelson et al., 2017; Mathes et al., 2019; Jia et al., 2021). A deeper understanding of what drives this heterogeneity is mostly lacking.

Personality traits[1] seem to be a potential candidate as they relate to peoples' *intrinsic* motivation and have been found to explain behavior and responses to incentives in various contexts (e.g., Borghans et al., 2008; Almlund et al., 2011; Heckman et al., 2021). Personality relates to, for example, prosocial behavior (e.g., Thielmann et al., 2020), education (e.g., Heckman et al., 2006), productivity (e.g., Bowles et al., 2001; Cubel et al., 2016; Butschek et al., 2022; Opitz et al., 2022), and self-selection into the public sector (e.g., Dal Bó et al., 2013). In health economics, however, only findings from Donato et al.'s (2017) field experiment with Indian health workers suggests that P4P affects more conscientious and more neurotic providers to a lesser extent. Related lab experiments on P4P (e.g., Brosig-Koch et al., 2020, 2021; Oxholm et al., 2021) do not consider personality traits. A systematic analysis on how personality traits may explain providers' responses to financial incentives is mostly lacking.

This study contributes to filling this gap by investigating the influence of personality traits on the quality of care under the two payment schemes capitation (CAP) and fee-for-service (FFS). We also examine whether personality traits moderate the effects of P4P on care quality by using experimental and survey data from Brosig-Koch et al. (2021). The behavioral experiment allows for a causal analysis of how payment-system variations affect healthcare provision. Physicians, medical and non-medical students are randomly assigned to either CAP- or FFS-payment schemes, followed by threshold-based blended P4P-systems (CAP+P4P or FFS+P4P) at a within-subject level. The Big Five Inventory (BFI-10) elicits personality traits (Rammstedt and John, 2007; John et al., 2008).

Our results show that only the personality traits conscientiousness and agreeableness are behaviorally relevant for the quality of care. More conscientious and more agreeable subjects provide relatively higher quality of care under non-blended CAP, but not under non-blended FFS. Performance pay significantly improves the quality of care under both payment systems. Yet, more conscientious and more agreeable individuals respond significantly less under capitation-based P4P incentives. We propose an explanation of our findings that draws on different dynamics of the interplay between financial incentives and personality traits in the CAP-based and FFS-based payment systems likely to create different perceptions of acceptable provision behavior. Our findings seem informative for better incentivizing physicians and sorting them into incentive schemes.

[1] The American Psychological Association defines personality traits as "a relatively stable, consistent, and enduring internal characteristic that is inferred from a pattern of behaviors, attitudes, feelings, and habits in the individual." (American Psychological Association, nd).

2 Methods

2.1 Experiment and procedure

Participants in the role of physicians decide on medical care provision in a within-subject design, and are confronted sequentially with a non-blended and a blended payment system. In the first part of the experiment, each participant is incentivized by either baseline FFS or CAP. In the second part, P4P is introduced in addition to the respective baseline payment (FFS+P4P or CAP+P4P). Each subject is randomly assigned to only one of the two baseline payment systems.

A physician decides on the quantity of medical services for nine abstract patients, each of them suffering from one of two illnesses (characterized by high and low marginal health benefits) and one of three severities of illness (mild, intermediate, high). The patient-optimal treatment quantity q^* depends on the severity of illness. Patient characteristics are kept constant over the two parts of the experiment. The physician's profit is either a lump-sum per patient in CAP, or a fee per service and patient in FFS, minus the costs resulting from the respective service provision. Under P4P, physicians get a bonus if they reach a pre-specified quality threshold tied to the patient-optimal care. For each patient, the physician simultaneously determines her profit and the patient's health-benefit measured in monetary terms. The accumulated benefits are transferred to a charity and are exclusively used for financing surgical cataract treatment of real patients.[2] In total, EUR 2,509.40 were transferred enabling the treatment of 83 cataract patients.

In total, 107 subjects participated in our experiment: 44 medical and 43 non-medical students in lab experiments, and 20 physicians in lab-in-the-field experiments with 55 participants in CAP/CAP+P4P, 52 in FFS/FFS+P4P, 22 medical students and 10 physicians under each payment system.

The computerized experiment was programmed with z-Tree (Fischbacher, 2007). Physicians and students were presented with identical computer screens, instructions, and comprehension questions. The main difference was a 4.25 times higher payment for physicians compared to students and some minor differences in the experimental procedures. The lab experiments were conducted between 2011 and 2013 at the Essen Laboratory for Experimental Economics at the University of Duisburg-Essen. Student subjects were recruited online via ORSEE (Greiner, 2015). The lab-in-the-field experiments were conducted with the mobile lab of elfe at the Academy for Training and Education of Physicians in Bad Nauheim (Germany) in 2012 and 2013. Physicians were recruited by announcements in their training courses and voluntarily participated after their courses.

2.2 Empirical model

We apply the following estimation model for each payment system:

$$y_{ij} = \beta_0 + \beta_1 \mathrm{P4P}_j + \beta_2 \textsc{Trait}_i + \beta_3 \mathrm{P4P}_j \mathbf{x} \textsc{Trait}_i + \beta_4 \mathbf{X}_i + \beta_5 \mathbf{Z}_j + \epsilon_{ij}, \quad (1)$$

[2] For further details of the design, see the Supporting Online Information.

where y_{ij} is the *quality outcome*, namely the absolute deviation from patient-optimal care ($|q_{ij} - q_j^*|$). It is measured by the absolute number of medical services that a subject i's chosen quantity (q_{ij}) for a patient j deviates from the patient optimal quantity (q_j^*). A physician provides the highest quality of care ($y_{ij} = 0$) whenever she is delivering q_j^*. The larger the difference to the patient-optimal quantity, the lower is the quality y_{ij}. P4P_j is a dummy variable indicating the introduction of performance pay. $\textsc{Trait}_i$ is a vector of a subject's personality traits representing the score of each trait of the BFI-10. Our main interest are the interaction effects between P4P and each $\textsc{Trait}_i$.[3] X_i represents a vector of time-invariant subject characteristics, i.e., gender and medical experiences (non-medical student, medical student, or physician). Vector Z_j comprises a patient's severity of illness and the marginal health benefit varied systematically in the experiment.[4].

3 Results

3.1 Descriptive statistics

We observe a substantial degree of between-subject heterogeneity for each trait under each baseline payment system (Figure 1). On a scale from 1 to 5 (see the notes of Figure 1), our participants on average exhibit rather high levels of conscientiousness (FFS 3.452, CAP 3.691), extraversion (3.683, 3.436) and openness (3.548, 3.627), a moderate level of agreeableness (3.115, 3.000) and a rather low level of neuroticism (2.798, 2.745).[5] For subject-pool specific values see Table A.1in the Appendix. The CAP- and FFS-samples are rather similarly distributed for all personality measures; we find no significant differences ($p \geq 0.097$, Mann-Whitney-U test (MWU); $p \geq 0.382$, Kolmogorov-Smirnov test (KS)). See Table 1.

3.2 Regression analyses

We first analyze how personality traits relate to the quality of care physicians provide under CAP and FFS. Models (1) and (4) of Table 2 report the regression coefficients on the average effects on provision quality. Estimates show a significant quality improvement under CAP only for participants higher in conscientiousness and in agreeableness as the regression coefficients significantly deviate from zero; see Model (1). A negative regression coefficient indicates that the deviation from the patient-optimal care becomes smaller and thus implies a *quality improvement*. Regarding FFS, Model (4) provides no evidence that personality traits are behaviorally relevant for the quality of care. None of the coefficients is

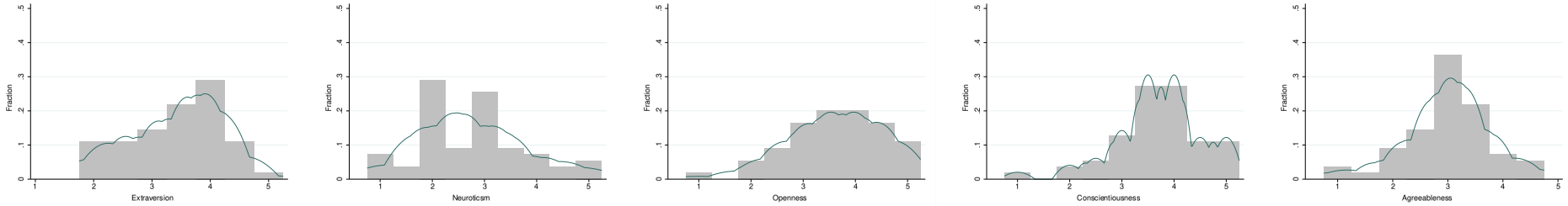
[3] We accounted for potential multicollinearity problems when including all traits in a joint model. Neither the tolerance and variance inflation matrix, nor the correlations of estimated coefficients give reason for concern. We also performed separate trait-by-trait regressions for robustness analyses for CAP and FFS, respectively. While the findings are qualitatively similar to our joint model, the adjusted R^2 is higher for our model. For the respective additional regressions see the Supporting Online Information.

[4] For a systematic comparison between CAP and FFS on the overall and patient-severity specific effects of P4P on the quality of care, see Brosig-Koch et al. (2021)

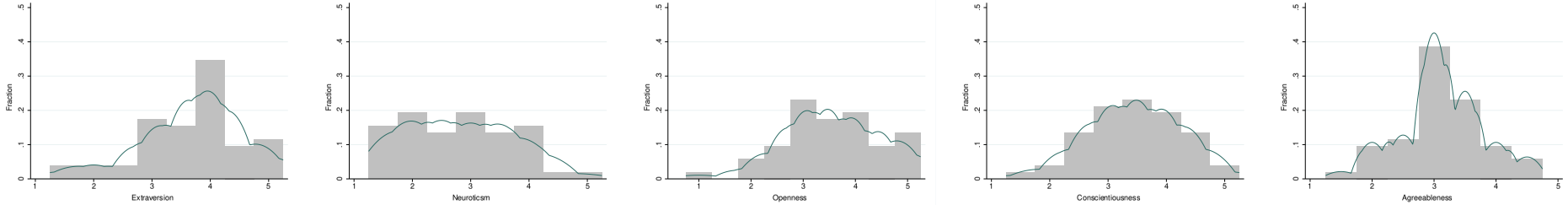
[5] Note that the measurement reliability (Cronbach alpha; α) for our data ranges only from good to fairly low; extraversion ($\alpha = 0.74$), neuroticism ($\alpha = 0.58$) openness ($\alpha = 0.42$), conscientiousness ($\alpha = 0.42$), and agreeableness ($\alpha = 0.18$). Since α coefficients depend on the number of items used, the rather low consistency is not surprising (Furnham, 2008). When using Cronbach's alpha as reliability measure for a two-item scale, the true reliability is usually underestimated (Eisinga et al., 2013). Our α's do not appear particularly striking compared to other studies using the BFI-10, reporting ranges from $\alpha = 0.03$ to $\alpha = 0.75$ (e.g., Thalmayer et al., 2011; Crede et al., 2012; Carciofo et al., 2016; Balgiu, 2018 and Hennig-Schmidt et al., 2019).

Figure 1: Distribution of personality traits among subjects by payment condition

(a) Capitation



(b) Fee-for-service



Notes. This figure shows the distributions of personality traits among subjects by payments condition. The solid line reflects the kernel density plot. Personality traits are measured by the BFI-10 (Rammstedt and John, 2007). Subjects rate their agreement to ten statements on a five-point Likert scale with 1: disagree strongly, 2: disagree a little, 3: neutral, 4: agree a little, and 5: agree strongly. Two out of the ten statements form one trait. We report the mean responses to the two statements representing one personality trait.

Table 1: Personality traits by payment system

	Mean (s.d.)			p-value	
	FFS	CAP	Difference	MWU	KS
Extraversion	3.683	3.436	0.247	0.114	0.609
	(0.863)	(0.788)			
Neuroticism	2.798	2.745	0.053	0.732	0.995
	(0.925)	(1.018)			
Openness	3.548	3.627	-0.079	0.596	0.992
	(0.930)	(0.909)			
Conscientiousness	3.452	3.691	-0.239	0.097	0.382
	(0.794)	(0.825)			
Agreeableness	3.115	3.000	0.115	0.495	1.000
	(0.669)	(0.758)			
Subjects	52	55			

Notes. For the description of how personality traits are measured, see the notes of Figure 1. Variable standard deviations are reported in parentheses. p-values of Mann-Whitney-U tests (MWU) for differences between subsamples and of two-sample Kolmogorov-Smirnov tests (KS) are reported in the second last and in the last column, respectively.

significantly different from zero; see also Figure 3.

Result 1. *Under CAP, more conscientious and more agreeable individuals provide higher quality of care. Other traits do not significantly relate to providers' behavior. We find no significant relation between personality and provision quality under FFS.*

We next analyze whether personality traits moderate the effect of P4P on the quality of care. First, we find consistent with earlier experiments of Brosig-Koch et al. (2021, 2023) or Oxholm et al. (2021) that P4P leads to a significant improvement of the quality of care under both payment systems CAP+P4P and FFS+P4P; see Models (2) and (5) in Table 2. Using Equation (1) and interacting each personality trait with CAP+P4P in Model (3) and with FFS+P4P in Model (6) in Table 2, we find that conscientiousness and agreeableness significantly moderate the P4P effect (see also Panel A of Figure 3). This implies that the positive effect of P4P on the quality of care decreases for subjects scoring higher in both traits. Thus, CAP+P4P may not be as effective to improve the quality of care for these participants, who provide already higher quality of care under CAP. This finding supports and extends the results of Donato et al. (2017) who report that providers with higher conscientiousness and neuroticism respond less to P4P.

Under FFS, personality traits do not significantly moderate the quality of care in any statistical significant way, see Model (6), Table 2, and Panel B of Figure 3.[6]. In contrast to our findings under CAP, individual responses under blended FFS are not affected by personality traits, similar to non-blended FFS. This difference suggests that participants perceive both payment systems differently, despite the mirror-image design of the payment systems. Finally, extraversion, openness, and neuroticism are not significantly related to individuals' behavior in any of the four payment scheme.

Result 2. *Performance pay significantly improves the quality of care under CAP and FFS. Yet, more con-*

[6]To check the robustness of our results, we ran regressions for several versions of our base model Equation (1), estimations based on tobit and multilevel regressions, and separate trait-by-trait analyses. The estimation results of each of these analyses are qualitatively similar to the results shown in Table 2, see the Supporting Online Information.

Table 2: Interaction effects of performance pay and personality traits on the quality of care

	CAP	CAP / CAP+P4P		FFS	FFS / FFS+P4P	
	(1)	(2)	(3)	(4)	(5)	(6)
Extraversion	-0.772	-0.430	-0.743	0.345	0.275	0.376
	(0.486)	(0.296)	(0.492)	(0.489)	(0.298)	(0.473)
Neuroticism	-0.455	-0.301	-0.553	0.339	0.274	0.367
	(0.367)	(0.226)	(0.357)	(0.440)	(0.258)	(0.400)
Openness	0.088	0.198	0.041	-0.082	-0.021	-0.074
	(0.376)	(0.252)	(0.392)	(0.498)	(0.292)	(0.488)
Conscientiousness	-1.086**	-0.690**	-1.309***	0.596	0.428	0.231
	(0.480)	(0.305)	(0.461)	(0.544)	(0.325)	(0.492)
Agreeableness	-1.159**	-0.766**	-1.162**	0.191	0.136	-0.067
	(0.484)	(0.297)	(0.496)	(0.548)	(0.315)	(0.560)
P4P		-1.117***	-1.716***		-1.199***	-1.290***
		(0.180)	(0.220)		(0.172)	(0.292)
P4P×Extraversion			0.627			-0.202
			(0.460)			(0.406)
P4P×Neuroticism			0.506			-0.186
			(0.305)			(0.339)
P4P×Openness			0.314			0.105
			(0.342)			(0.429)
P4P×Conscient.			1.238***			0.394
			(0.390)			(0.394)
P4P×Agreeable.			0.791*			0.406
			(0.439)			(0.552)
Constant	1.191***	1.440***	1.739***	3.324***	2.621***	2.667***
	(0.332)	(0.242)	(0.267)	(0.447)	(0.315)	(0.376)
Observations	495	990	990	468	936	936
Subjects	55	55	55	52	52	52
R^2	0.325	0.287	0.328	0.309	0.329	0.336

Notes. This table shows estimates from OLS-regressions on the effects of personality traits, P4P, and their interactions on the quality of care on the quality of care measured as the absolute deviation from patient-optimal care $|q_{ij} - q_j^*|$. Robust standard errors clustered for subjects in parentheses. Personality traits measured on a scale from -1 to +1. The coefficients thus reflect the effect of a one-unit change. All models control for patient health characteristics (severity of illness, marginal health benefit) and subject controls (gender, medical experience such as non-medical students, medical students, physicians). For the respective estimates, see Table A.2 in the Appendix. *$p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

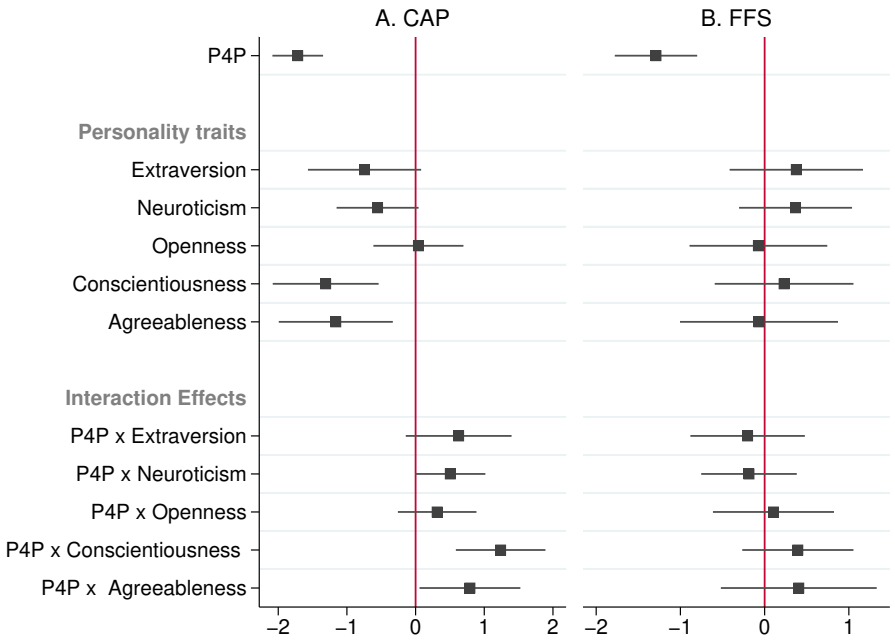
scientious and more agreeable individuals respond significantly less under CAP+P4P. Under FFS+P4P, personality traits are apparently not behaviorally relevant.

4 Discussion and conclusion

Our behavioral results evidence the importance of personality traits for understanding healthcare providers' responses to payment-scheme incentives. Under CAP, more conscientious and more agreeable subjects provide relatively higher quality of care than subjects lower in these traits. Introducing P4P enhances the quality of care. Yet, this positive effect is weaker for more conscientious and agreeable subjects. We find no such effects under FFS.

We propose an explanation of our findings that draws on different dynamics of the interplay between financial incentives and personality traits in the two payment systems likely to create different percep-

Figure 3: Regression estimates: Effect of performance pay and personality traits on the quality of care



Notes. This figure shows regression coefficients from separate regressions on the impact of P4P and personality traits on the quality of care for both baseline payment systems, CAP in Panel A and FFS in Panel B. Error bars represent 90-percent confidence intervals, with standard errors clustered at the individual subject level. Regression estimates corresponding to this figure are reported in Models (3) and (6) of Table 2. For a description of personalty-traits measurement and the additional covariates see the notes of Table 2.

tions of acceptable behavior. This approach combines two aspects. On the one hand, the difference in the monetary incentives of our two payment systems may change how medical service provision is perceived by the participants; see Gneezy et al. (2011), Gächter et al. (2022).

On the other hand, the facets of agreeableness and conscientiousness are likely to trigger a certain behavior as response to the different incentives in CAP-based and FFS-based payment schemes. CAP incentivizes underprovision as every additional service creates costs to be paid from the fixed capitation payment reducing the physician profit. Increasing their profits would require to follow a "no(low)-effort-policy", which simultaneously reduces the patients' benefits to (nearly) zero. Thus, CAP-incentives are in conflict with those characteristics of agreeableness that motivates peoples' concern and cooperation towards patients. Moreover, individuals higher in conscientiousness might be more reluctant to adopt a low-effort policy, which is likely to be perceived as selfish. Therefore, the inhibitive aspects of conscientiousness, such as moral scrupulousness and cautiousness, may become influential on behavior (Costa et al., 1991). Introducing CAP+P4P incentivizes physicians to enhance the quality of care. More conscientious and agreeable individuals are likely to respond less strongly because they have less potential for quality improvements compared to less patient-regarding physicians, as the former already provide a quality fairly close to the patient-optimum under CAP.

FFS in contrast, incentivizes overprovision because every additional service increases the physician profit, as the additional payment is larger than the additional costs. Here, the proactive aspects of conscientiousness, for example, the need for achievement and commitment to work (Costa et al., 1991), is in

line with the FFS-incentive structure, and no personality-induced corrections on behavior seem necessary.

Within the confines of our experiment, the key take-aways of this study are as follows. Certain personality traits mitigate the effectiveness of P4P in settings particularly attractive for physicians high in conscientiousness and agreeableness. Even though these personality traits may negatively relate to responses under P4P, the moderating effects are not detrimental for patients. Whether CAP-based P4P serves as a means to efficiently improve the quality of care seems, however, to depend on the shares of physicians high in contentiousness and agreeableness. Moreover, achieving a higher quality of care does not necessarily require additional performance incentives. Selecting more conscientious and agreeable physicians might lead to an improved provision quality and presents an alternative approach to introducing P4P in care settings characterized by CAP-payments like in Germany (Brosig-Koch et al., 2020).

Our study adds to the potential of personality traits within the scope of worker selection strategies to improve the performance (e.g., Ashraf and Lee., 2016). Our findings can be useful in developing targeted interventions aimed at selecting an efficient and patient-oriented medical workforce. Any policy intervention ideally rests on known behavioral and psychological mechanisms. As these mechanisms seem to be incentive-scheme specific, different payment systems such as capitation and fee-for-service might require different quality-enhancing approaches.

References

Almlund, M., Duckworth, A. L., Heckman, J., and Kautz, T. (2011). Chapter 1 - Personality psychology and economics. in Hanushek, E.A., Machin, S., and Woessmann, L. (Eds.): *Handbook of the Economics of Education*, 4:1-181. Elsevier. doi.org/10.1016/B978-0-444-53444-6.00001-8.

American Psychological Association (n.d.). *APA Dictionary of psychology.* Retrieved December 14, 2021, from `https://dictionary.apa.org/personality-trait`.

Ashraf, Nava, O. B. and Lee., S. S. (2016). Do-gooders and go-getters: Selection and performance in public service delivery. *International Growth Centre Working Paper.*

Balgiu, B. (2018). The psychometric properties of the Big Five inventory-10 (bfi-10) including correlations with subjective and psychological well-being. *Global Journal of Psychology Research: New Trends and Issues*, 8:61–69.

Basinga, P., Gertler, P. J., Binagwaho, A., Soucat, A. L., Sturdy, J., and Vermeersch, C. M. (2011). Effect on maternal and child health services in rwanda of payment to primary health-care providers for performance: an impact evaluation. *The Lancet*, 377(9775):1421–1428.

Borghans, L., Duckworth, A. L., Heckman, J. J., and ter Weel, B. (2008). The Economics and Psychology of Personality Traits. *Journal of Human Resources*, 43(4).

Bowles, S., Gintis, H., and Osborne, M. (2001). Incentive-enhancing preferences: Personality, behavior, and earnings. *American Economic Review*, 91(2):155–158.

Brosig-Koch, J., Groß, M., Hennig-Schmidt, H., Kairies-Schwarz, N., and Wiesen, D. (2021). Physicians'

incentives, patients' characteristics, and quality of care: A systematic experimental comparison of fee-for-service, capitation, and pay for performance. Ruhr Economic Papers 923. doi:10.4419/96973079.

Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N., Kokot, J., and Wiesen, D. (2020). Physician performance pay: Experimental evidence. HERO Online Working Paper Series 2020:3, University of Oslo, Health Economics Research Programme.

Brosig-Koch, J., Hennig-Schmidt, H., Kairies-Schwarz, N., Kokot, J., and Wiesen, D. (2023). A new look at physicians' responses to financial incentives: Quality of care, practice characteristics, and motivations. doi:10.2139/ssrn.4366378.

Butschek, S., Gonzàlez Amor, R., Kampkötter, P., and Sliwka, D. (2022). Motivating gig workers – evidence from a field experiment. *Labour Economics*, 75:102105.

Carciofo, R., Yang, J., Song, N., Du, F., and Zhang, K. (2016). Psychometric evaluation of Chinese-language 44-item and 10-item Big Five personality inventories, including correlations with chronotype, mindfulness and mind wandering. *PLOS ONE*, 11:1–26: doi.org/10.1371/journal.pone.0149963.

Costa, P. T., McCrae, R. R., and Dye, D. A. (1991). Facet scales for agreeableness and conscientiousness: A revision of the neo personality inventory. *Personality and Individual Differences*, 12(9):887–898.

Crede, M., Harms, P., Niehorster, S., and Gaye-Valentine, A. (2012). An evaluation of the consequences of using short measures of the Big Five personality traits. *Journal of Personality and social Psychology*, 102:874–888. doi:10.1037/a0027403.

Cubel, M., Nuevo-Chiquero, A., Sánchez-Pagés, S., and Vidal-Fernandez, M. (2016). Do personality traits affect productivity? Evidence from the laboratory. *The Economic Journal*, 126:654–681. doi: 10.1111/ecoj.12373.

Dal Bó, E., Finan, F., and Rossi, M. A. (2013). Strengthening state capabilities: The role of financial incentives in the call to public service. *The Quarterly Journal of Economics*, 128(3):1169–1218.

De Walque, D., Gertler, P. J., Bautista-Arredondo, S., Kwan, A., Vermeersch, C., de Dieu Bizimana, J., Binagwaho, A., and Condo, J. (2015). Using provider performance incentives to increase hiv testing and counseling services in rwanda. *Journal of Health Economics*, 40:1–9.

Donato, K., Miller, G., Mohanan, M., Truskinovsky, Y., and Vera-Hernández, M. (2017). Personality traits and performance contracts: Evidence from a field experiment among maternity care providers in India. *American Economic Review*, 107:506–510. doi:10.1257/aer.p20171105.

Doran, T., Fullwood, C., Gravelle, H., Reeves, D., Kontopantelis, E., Hiroeh, U., and Roland, M. (2006). Pay-for-performance programs in family practices in the United Kingdom. *New England Journal of Medicine*, 355:375–384. doi:10.1056/NEJMsa055505.

Eijkenaar, F., M. Emmert, M. Scheppach, and Oliver Schoeffski (2013). Effects of Pay for Performance in Health Care: A Systematic Review of Systematic Reviews. *Health Policy*, 110:115–130.

Eisinga, R., Grotenhuis, M., and Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health*, 58:637–642. doi:10.1007/s00038–012–0416–3.

Fischbacher, U. (2007). z-Tree: Zurich toolboox for readymade economic experiments – Experimenter's manual. *Experimental Economics*, 10:171–178.

Furnham, A. (2008). Relationship among four Big Five measures of different length. *Psychological Reports*, 102:312–316. doi:10.2466/pr0.102.1.312–316.

Gächter, S., Kölle, F., and Quercia, S. (2022). Preferences and perceptions in provision and maintenance public goods. *Games and Economic Behavior*, 135:338–355.

Gneezy, U., Meier, S., and Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives*, 25(4):191–210.

Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association*, 1:114–125.

Heckman, J., Jagelka, T., and Kautz, T. (2021). *Some contributions of economics to the study of personality. In John, O.P., Robins, R.W., and Pervin, L.A. (Eds.): Handbook of personality: Theory and research (3rd ed.), 3:853–892.* Guilford Press.

Heckman, J. J., Stixrud, J., and Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3):411–482.

Hennig-Schmidt, H., Jürges, H., and Wiesen, D. (2019). Dishonesty in health care practice: A behavioral experiment on upcoding in neonatology. *Health Economics*, 28:319–338. doi:10.1002/hec.3842.

Jia, L., Meng, Q., Scott, A., Yuan, B., and Zhang, L. (2021). Payment methods for healthcare providers working in outpatient healthcare settings. *Cochrane Database of Systematic Reviews*, 1:CD011865. doi:10.1002/14651858.CD011865.pub2.

John, O., Naumann, L., and Soto, C. (2008). *Paradigm shift to the integrative Big Five taxonomy. In John, O.P., Robins, R.W., and Pervin, L.A. (Eds.): Handbook of personality: Theory and research (3rd ed.), 3:114-158.* Guilford Press.

Kristensen, S. R., Meacock, R., Turner, A. J., Boaden, R., McDonald, R., Roland, M., and Sutton, M. (2014). Long-term Effect of Hospital Pay for Performance on Mortality in England. *New England Journal of Medicine*, 371(6):540–548.

Lazear, E. P. (2000). Performance Pay and Productivity. *American Economic Review*, 90(5):1346–1361.

Mathes, T., Pieper, D., Morche, J., Polus, S., Jaschinski, T., and M., E. (2019). Pay for performance for hospitals. *Cochrane Database of Systematic Reviews*, 7. doi:10.1002/14651858.CD011156.pub2.

Mendelson, A., Kondo, K., Damberg, C., Low, A., Motuapuaka, M., Freeman, M., O'Neil, M., Relevo,

R., and Kansagara, D. (2017). The Effects of Pay-for-Performance Programs on Health, Health Care Use, and Processes of Care: A Systematic Review. *Annals of Internal Medicine*, 166(5):341–353.

Miller, G. and Babiarz, K. S. (2014). Pay-for-Performance Incentives in Low- and Middle-Income Country Health Programs. In Anthony J. Culyer, editor, *Encyclopedia of Health Economics*, pages 457 – 466. Elsevier.

Milstein, R. and Schreyögg, J. (2016). Pay for performance in the inpatient sector: A review of 34 P4P programs in 14 OECD countries. *Health Policy*, 120:1125–1140. doi:10.1016/j.healthpol.2016.08.009.

Olken, B. A., Onishi, J., and Wong, S. (2014). Should aid reward performance? evidence from a field experiment on health and education in indonesia. *American Economic Journal: Applied Economics*, 6(4):1–34.

Opitz, S., Sliwka, D., Vogelsang, T., and Zimmermann, T. (2022). The targeted assignment of incentive schemes. ECONtribute Discussion Paper, No 187, University of Bonn and University of Cologne, Reinhard Selten Institute (RSI), Bonn and Cologne.

Oxholm, A.-S., Di Guida, S., and Gyrd-Hansen, D. (2021). Allocation of health care under pay for performance: Winners and losers. *Social Science & Medicine*, 278:113939. doi:10.1016/j.socscimed.2021.113939.

Peabody, J., Shimkhada, R., Quimbo, S., Florentino, J., Bacate, M., McCulloch, C. E., and Solon, O. (2011). Financial incentives and measurement improved physicians' quality of care in the philippines. *Health Affairs*, 30(4):773–781.

Prendergast, C. (1999). The Provision of Incentives in Firms. *Journal of Economic Literature*, 37(1):7–63.

Rammstedt, B. and John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, 41:203–212. doi:10.1016/j.jrp.2006.02.001.

Roland, M. (2004). Linking physicians' pay to the quality of care - A major experiment in the United Kingdom. *New England Journal of Medicine*, 351:1448–1454. doi:10.1056/NEJMhpr041294.

Roland, M. and Campbell, S. (2014). Successes and Failures of Pay for Performance in the United Kingdom. *New England Journal of Medicine*, 370(20):1944–1949.

Rosenthal, M. B., Frank, R. G., Li, Z., and Epstein, A. M. (2005). Early Experience with Pay-for-Performance: From Concept to Practice. *Journal of the American Medical Association*, 294(14):1788–1793.

Rosenthal, M. B., Landon, B. E., Normand, S.-L. T., Frank, R. G., and Epstein, A. M. (2006). Pay for performance in commercial HMOs. *New England Journal of Medicine*, 355:1895–1902. doi:10.1056/NEJMsa063682.

Scott, A., Sivey, P., Ouakrim, D. A., Willenberg, L., Naccarella, L., Furler, J., and Young, D. (2011).

The Effect of Financial Incentives on the Quality of Health Care Provided by Primary Care Physicians. *Cochrane Database of Systematic Reviews*, page https://doi.org/10.1002/14651858.CD008451.pub2.

Sherry, T. B., Bauhoff, S., and Mohanan, M. (2017). Multitasking and heterogeneous treatment effects in pay-for-performance in health care: evidence from rwanda. *American Journal of Health Economics*, 3(2):192–226.

Song, Z., Ji, Y., Safran, D. G., and Chernew, M. E. (2019). Health care spending, utilization, and quality 8 years into global payment. *New England Journal of Medicine*, 381:252–263. doi:10.1056/NEJMsa1813621.

Thalmayer, A., Saucier, G., and Eigenhuis, A. (2011). Comparative validity of brief to medium-length big five and big six personality questionnaires. *Psychological Assessment*, 23:995–1009. doi:10.1037/a0024165.

Thielmann, I., Spadaro, G., and Balliet, D. (2020). Personality and prosocial behavior: A theoretical framework and meta-analysis. *Psychol. Bull.*, 146(1):30—90.

Van de Poel, E., Flores, G., Ir, P., and O'Donnell, O. (2016). Impact of performance-based financing in a low-resource setting: a decade of experience in cambodia. *Health Economics*, 25(6):688–705.

Appendix: Additional information and analyses

Table A.1: Sample characteristics

	Sample			
	Full sample	Medical students	Non-medical students	Physicians
A. All payment systems				
Main characteristics				
Male	40.2%	27.3%	53.5%	40.0%
Age (Mean, s.d.)	28.6 (9.8)	25.2 (7.0)	24.3 (3.4)	45.3 (6.7)
Personality traits (Mean, s.d.)				
Extraversion	3.6 (0.83)	3.7 (0.84)	3.5 (0.90)	3.5 (0.56)
Neuroticism	2.8 (0.97)	2.6 (0.91)	3.0 (0.92)	2.6 (1.10)
Openness	3.6 (0.92)	3.8 (0.88)	3.3 (0.97)	3.6 (0.82)
Conscientiousness	3.6 (0.81)	3.6 (0.69)	3.2 (0.81)	4.3 (0.53)
Agreeableness	3.1 (0.71)	3.3 (0.63)	2.8 (0.73)	3.1 (0.64)
N	107	44	43	20
B. FFS				
Main characteristics				
Male	44.2%	40.9%	50.0%	40.0%
Age (Mean, s.d.)	28.4 (9.9)	24.3 (4.5)	24.1 (4.0)	45.9 (7.2)
Personality traits (Mean, s.d.)				
Extraversion	3.7 (0.86)	3.8 (0.89)	3.7 (0.92)	3.4 (0.66)
Neuroticism	2.8 (0.92)	2.4 (0.82)	3.1 (0.85)	3.0 (0.08)
Openness	3.5 (0.93)	3.9 (0.85)	3.3 (1.00)	3.5 (0.80)
Conscientiousness	3.5 (0.79)	3.5 (0.66)	3.1 (0.82)	4.1 (0.61)
Agreeableness	3.1 (0.67)	3.3 (0.66)	2.9 (0.61)	3.2 (0.63)
N	52	22	20	10
C. CAP				
Main characteristics				
Male	36.4%	13.6%	56.5%	40.0%
Age (Mean, s.d.)	28.8 (9.9)	26.1 (8.8)	24.6 (2.9)	44.6 (8.2)
Personality traits (Mean, s.d.)				
Extraversion	3.4 (0.79)	3.5 (0.79)	3.4 (0.88)	3.4 (0.58)
Neuroticism	2.7 (1.02)	2.8 (0.99)	2.9 (0.99)	2.3 (1.11)
Openness	3.6 (0.91)	3.7 (0.92)	3.4 (0.95)	4.0 (0.69)
Conscientiousness	3.7 (0.82)	3.8 (0.70)	3.3 (0.81)	4.5 (0.44)
Agreeableness	3.0 (0.76)	3.2 (0.61)	2.9 (0.83)	3.0 (0.90)
N	55	22	23	10

Notes. This table corresponds to Table A.1 in Brosig-Koch et al. (2021) presenting summary statistics of subjects' characteristics for the full sample of our experiment, medical and non-medical students, and physicians differentiated by payment system.

Table A.2: Regressions on different versions of our base model

	A. CAP					B. FFS				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
P4P	-1.117***	-1.716***	-1.716***	-1.716***	-1.716***	-1.199***	-1.290***	-1.290***	-1.290***	-1.290***
	(0.180)	(0.220)	(0.220)	(0.220)	(0.220)	(0.172)	(0.291)	(0.291)	(0.291)	(0.292)
InterSEV	0.436***		0.436***		0.436***	-0.529***		-0.529***		-0.529***
	(0.074)		(0.074)		(0.074)	(0.086)		(0.086)		(0.086)
HighSEV	0.958***		0.958***		0.958***	-0.997***		-0.997***		-0.997***
	(0.134)		(0.134)		(0.134)	(0.141)		(0.141)		(0.142)
HighMHB	-0.115**		-0.115**		-0.115**	-0.054		-0.054		-0.054
	(0.044)		(0.044)		(0.044)	(0.051)		(0.051)		(0.051)
Medical students	0.192			0.192	0.192	-0.388			-0.388	-0.388
	(0.280)			(0.281)	(0.281)	(0.299)			(0.299)	(0.300)
Physician	-0.438			-0.438	-0.438	-1.520***			-1.520***	-1.520***
	(0.322)			(0.323)	(0.323)	(0.266)			(0.267)	(0.267)
Male	0.385			0.385	0.385	0.050			0.050	0.050
	(0.263)			(0.263)	(0.263)	(0.222)			(0.222)	(0.223)
Extraversion	-0.430	-0.649	-0.649	-0.743	-0.743	0.275	0.441	0.441	0.376	0.376
	(0.296)	(0.522)	(0.523)	(0.492)	(0.492)	(0.298)	(0.493)	(0.494)	(0.472)	(0.473)
Neuroticism	-0.301	-0.617	-0.617	-0.553	-0.553	0.274	0.313	0.313	0.367	0.367
	(0.226)	(0.374)	(0.374)	(0.356)	(0.357)	(0.258)	(0.422)	(0.423)	(0.400)	(0.400)
Openness	0.198	-0.009	-0.009	0.041	0.041	-0.021	0.022	0.022	-0.074	-0.074
	(0.252)	(0.425)	(0.425)	(0.391)	(0.392)	(0.292)	(0.504)	(0.504)	(0.487)	(0.488)
Conscientiousness	-0.690**	-1.573***	-1.573***	-1.309***	-1.309***	0.428	-0.438	-0.438	0.231	0.231
	(0.305)	(0.446)	(0.447)	(0.460)	(0.461)	(0.325)	(0.502)	(0.503)	(0.491)	(0.492)
Agreeableness	-0.766**	-1.099**	-1.099**	-1.162**	-1.162**	0.136	-0.402	-0.402	-0.067	-0.067
	(0.297)	(0.539)	(0.540)	(0.496)	(0.496)	(0.315)	(0.668)	(0.669)	(0.559)	(0.560)
P4P×Extraversion		0.627	0.627	0.627	0.627		-0.202	-0.202	-0.202	-0.202
		(0.459)	(0.460)	(0.460)	(0.460)		(0.404)	(0.405)	(0.405)	(0.406)
P4P×Neuroticism		0.506	0.506	0.506	0.506		-0.186	-0.186	-0.186	-0.186
		(0.304)	(0.304)	(0.304)	(0.305)		(0.338)	(0.338)	(0.338)	(0.339)
P4P×Openness		0.314	0.314	0.314	0.314		0.105	0.105	0.105	0.105
		(0.341)	(0.341)	(0.341)	(0.342)		(0.428)	(0.428)	(0.428)	(0.429)
P4P×Conscientiousness		1.238***	1.238***	1.238***	1.238***		0.394	0.394	0.394	0.394
		(0.389)	(0.389)	(0.389)	(0.390)		(0.393)	(0.393)	(0.393)	(0.394)
P4P×Agreeableness		0.791*	0.791*	0.791*	0.791*		0.406	0.406	0.406	0.406
		(0.437)	(0.438)	(0.438)	(0.439)		(0.550)	(0.551)	(0.551)	(0.552)
Constant	1.440***	2.381***	1.955***	2.165***	1.739***	2.621***	1.822***	2.348***	2.140***	2.667***
	(0.242)	(0.251)	(0.220)	(0.286)	(0.267)	(0.315)	(0.365)	(0.401)	(0.349)	(0.376)
Observations	990	990	990	990	990	936	936	936	936	936
Subjects	55	55	55	55	55	52	52	52	52	52
R^2	0.287	0.248	0.308	0.269	0.328	0.329	0.173	0.242	0.267	0.336

Notes.
This table shows estimates from OLS-regressions on the effects of P4P and personality traits on the quality of care under CAP and FFS (Panels A and B, respectively) for different versions of Equation (1). Robust standard errors clustered for subjects in parentheses. Patient controls: Severity of illness (intermediate InterSEV, High HighSEV); reference category: mild. Marginal health benefit MHB; reference category: low. Subject controls: gender, reference category: female. Medical experience: medical students, physicians; reference category: non-medical students. For a description of personalty-traits measurement and the personality traits see the notes of Table 2. * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.