The Power of Data: Assessing Primary Care Performance Using Routinely Collected Emergency Department Data

(Authors' names blinded for peer review)

This online appendix presents additional investigations and robustness checks on the analysis presented in the main paper.

Contents

1. List of ACS conditions and publicaly available data bases used in this study

Table 1 List of ACS conditions

Acute	Chronic	Vaccine-preventable
Cellulitis	Angina	Influenza
Dehydration	Asthma	Pneumonia
Dental conditions	Chronic obstructive pulmonary disease	Tuberculosis
Ear, nose, and throat infections	Congestive heart failure	Other vaccine-preventable conditions
Gangrene	Convulsions and epilepsy	
Gastroenteritis	Diabetic complications	
Nutritional deficiencies	Hypertension	
Pelvic inflammatory disease	Iron deficiency	
Perforated or bleeding ulcer		
Urinary tract infections or pyelonephritis		

Table 2 Publicly available data bases used in this study

Publisher, Data set	Extracted Information	Link
NHS Digital, Epraccur	Name and addresses of PCPs and when they ended operating (time of closure)	https://digital.nhs.uk/services/organisation-data-service/data-downloads/gp-and-gp-practice-related-data
NHS Digital, Ebranches	PCP branches and when they started operating	https://digital.nhs.uk/services/organisation-data-service/data-downloads/gp-and-gp-practice-related-data
NHS Digital, Patients Registered at a GP Practice	Number and demographics of patients registered at PCP	https://digital.nhs.uk/data-and-information/publications/statistical/patients-registered-at-a-gp-practice
NHS Digital, Payments to General Practice	PCP patient case-mix	https://digital.nhs.uk/data-and-information/publications/statistical/nhs-payments-to-general-practice
NHS Digital, Etrust	Location of hospital sites	https://digital.nhs.uk/services/organisation-data-service/data-downloads/other-nhs-organisations
NHS Digital, General and Personal Medical Services	FTE physicians working at PCPs	https://webarchive.nationalarchives.gov.uk/20180328140045/http://digital.nhs.uk/catalogue/PUB20503
NHS Digital, Quality & Outcomes Framework	QOF achievements per PCP	https://digital.nhs.uk/data-and-information/publications/statistical/quality-and-outcomes-framework-achievement
NHS England, GP Patient Survey	Patient perception of PCPs	https://gp-patient.co.uk/about
Free Map Tools, UK Postcode Database	Latitude and longitude coordinates of UK postcodes	https://www.freemaptools.com/contact.htm
Ministry of Housing, Communities & Local Government	English indices of deprivation at small area level	https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015
Office for National Statistics	Postcode to Output Area	https://geoportal.statistics.gov.uk/datasets/e7824b1475604212a2325cd373946235/about

2. Descriptive statistics for the PCPs

Table 3 Descriptive statistics for the 80 PCPs.

		2013	2014	2015	2016	2017
Proportion of ACS ED visits	Mean	0.171	0.173	0.166	0.167	0.165
	SD	0.037	0.036	0.035	0.041	0.029
Proportion of PCPs closest to EDs	Mean	0.148	0.148	0.148	0.140	0.140
	SD	0.357	0.357	0.357	0.349	0.349
Depriviation rank (in 1,000)	Mean	9.031	9.031	9.031	9.159	9.159
	SD	5.220	5.220	5.220	5.211	5.211
PCP scale (in 1,000 patients)	Mean	8.455	8.757	9.253	8.998	9.447
	SD	3.872	3.936	3.898	3.653	3.794
Proportion of female patients	Mean	0.505	0.495	0.499	0.495	0.495
	SD	0.043	0.043	0.021	0.042	0.041
Proportion of elderly patients	Mean	0.035	0.034	0.035	0.034	0.033
	SD	0.013	0.012	0.012	0.013	0.013
Case-mix Index	Mean	0.998	0.984	0.983	0.971	0.967
	SD	0.069	0.067	0.065	0.063	0.063

3. Alternative model specifications

For the model presented in (1),

$$A_{gkt} = \alpha_0 + \alpha_C C_{gkt} + u_g + \epsilon_{gkt} \tag{1}$$

we normalize the number of ACS attendances by dividing with the total number of attendances. This normalization controls for heterogeneity that affects patients with ACS and Non-ACS conditions equally. In addition, we explicitly control for observed PCP and ED heterogeneity that might affect ACS patients differently than Non-ACS patients. In this section we investigate three alternative modeling choices.

3.1. Adaptive centering approach

In this section we employ we employ Raudenbush's adaptive centering approach (Raudenbush 2009). This method involves transforming time-varying covariates by centering them around cluster means across multiple dimensions (in our case, PCP and hospital). This transformation helps to separate within-cluster effects from between-cluster effects, potentially mitigating bias from time-invariant confounding. We estimate an alternative model using these transformed variables for time-varying PCP covariates (results reported in Table 4, Column (2)) and compared the resulting $\widehat{u}_g$ with those from the primary model (for comparison, Table 4, Column (1) repeats the results of the main paper). The high correlation (0.920, $p<0.001$) between these estimates suggests robustness to time-invariant confounding from observable PCP covariates. As shown in Figure 1, the histograms and kernel density estimates for $\widehat{u}_g$ from both models suggest that there

are no significant differences between the distributions generated by these two methods. This is further supported by a Kolmogorov-Smirnov test for equality of distributions, where the p-value is 1.000. This suggests $\widehat{u}_g$ is robust to time-invariant confounding from observable PCP covariances. However, we acknowledge that this approach does not directly test whether the true u_g is uncorrelated with the observable PCP covariates and cannot eliminate the possibility of confounding from unobservable PCP characteristics. Therefore, we also estimate a fixed-effects model in the next section.

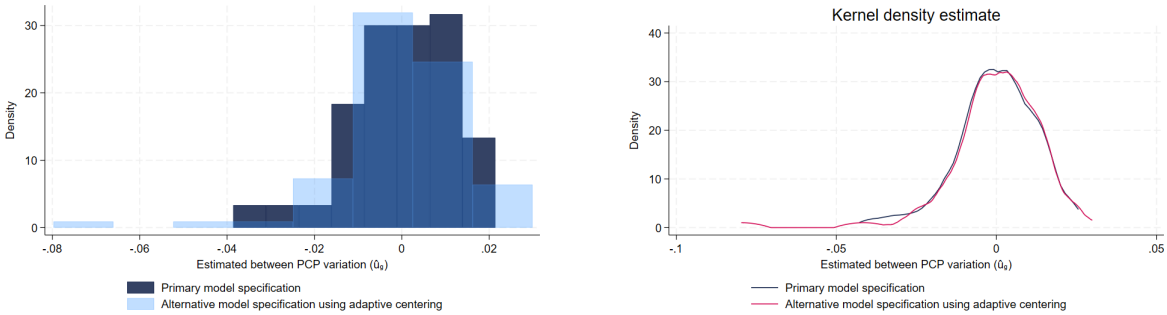

Figure 1 $\widehat{u}_g$ **comparision primary model vs. adaptive centering**

3.2. Fixed-effect specification

In this section we estimate model (1) by treating the unobservable heterogeneity between PCPs u_g as a fixed effect. In contrast to the random effects model estimated in the main paper, this specification does not place any parametric assumptions on PCP heterogeneity (the random-effects model assumed Normally distributed heterogeneity) and does not assume that PCP heterogeneity is orthogonal to any of the time varying controls included in the model. We note that this model specification cannot include the distance variable $closest_{gk}$ or the deprivation index D_g as they are time-invariant and are therefore collinear with the PCP fixed effects. We estimate the fixed-effect model using the Stata command *xtreg, fe* and provide the results in Table (4) Column (4). In the fixed-effect specification 77.5% of the variation is attributed to systematic differences between PCPs compared to the 51.9% in the random-effect specification (Table (4) Column (1)). The larger between-variation in the fixed-effect specification may be the result of absorbing time-invariant heterogeneity between PCPs that is explicitly controlled for in the random-effect specification.

We argued in the main paper that due to the low number of observations per PCP the fixed-effect specification will estimate the unobservable heterogeneity between PCPs with more uncertainty than the random-effect specification. This is illustrated in Figure 2. In fact, the variance of $\widehat{u}_g$ is 0.00054122 in the fixed-effect specification, which is 214% higher than the variance of the random-effect specification (0.0001722).

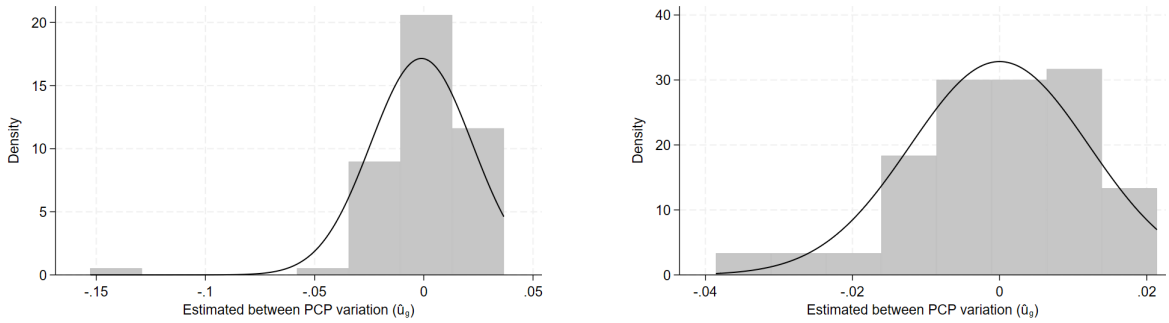

Figure 2 Variation between PCPs: Fixed-effect specification (left) and Random-effect specification (right)

Nevertheless, the PCP-effect estimated by the fixed effect model is highly correlated to that estimated by the random effect model (correlation=0.751, p<0.001), which indicates that the two specifications overlap in identifying PCP that perform better (or worse) than average.

For validation purposes, we re-estimate the different models outlined in §4 of the main paper, i.e. first a model using the results of the patient survey:

$$P^i_{gt} = \beta_0 + \beta_U \widehat{u}_g + \beta_\epsilon \widehat{\epsilon}_{gt} + \beta_C C_{gt} + u^P_g + \epsilon^P_{gt}. \tag{2}$$

second, a model using the QOF scores:

$$\begin{aligned} HighQ^*_{gt} &= \gamma_0 + \gamma_U \widehat{u}_g + \gamma_\epsilon \widehat{\epsilon}_{gt} + \gamma_C C_{gt} + u^Q_g + \epsilon^Q_{gt}, \\ HighQ_{gt} &= 1[HighQ^*_{gt} > 0]. \end{aligned} \tag{3}$$

and third, a model using the patient-to-staff ratio as a performance antecedent:

$$\widehat{u}_g = \delta_0 + \delta_P PpFTE_g + \delta_C C_g + \epsilon^U_g. \tag{4}$$

The results of estimating models (2) – (4) with $\widehat{u}_g$ obtained through the fixed-effect specification appear in Table 5 – 7 and are similar to the results reported in the main paper. We find that PCPs with a higher $\widehat{u}_g$ are also the ones patients are more likely dissatisfied with, PCPs with a higher $\widehat{u}_g$ are associated with a lower probability of obtaining QOF scores of at least 90%, and we find positive association between patient-to-staffing ratio and $\widehat{u}_g$. Hence, using a fixed-effect specification does not alter the paper's conclusions.

3.3. Testing non-linear scale effects

One of the controls included in the model of (1) is PCP scale. This is the number of patients registered with a PCP. In this section we test whether scale might affect PCP performance in a non-linear fashion by adding $Scale_{gt} \times Scale_{gt}$ to the model. For comparison, Column (1) in Table 4 repeats the results of the main paper and Column (3) shows the results of the model with $Scale_{gt} \times Scale_{gt}$. There is no evidence of non-linear scale effects as the coefficients of $Scale_{gt}$ and $Scale_{gt} \times Scale_{gt}$ are close to zero (and not statistically significant).

Table 4 Decomposing variation in PCP performance: Alternative model specifications

	(1) A_{gkt}	(2) A_{gkt}	(3) A_{gkt}	(4) A_{gkt}
Closest hospital	-0.002	-0.000	-0.002	
	(0.003)	(0.003)	(0.003)	
Deprivation rank	-0.001***	-0.001***	-0.001***	
	(0.000)	(0.000)	(0.000)	
Scale	0.000	0.001	-0.002	0.001
	(0.000)	(0.001)	(0.002)	(0.001)
Scale × Scale			0.000	
			(0.000)	
Female	0.137	-0.233	0.156	-0.233
	(0.101)	(0.143)	(0.103)	(0.145)
Elderly	-0.170	0.330	-0.184	0.330
	(0.161)	(0.375)	(0.159)	(0.381)
CMI	0.061*	0.040	0.056*	0.040
	(0.033)	(0.051)	(0.034)	(0.051)
Constant	Yes	Yes	Yes	Yes
ED × Year FE	Yes	Yes	Yes	Yes
PCP Effect	Random	Random	Random	Fixed
τ^2	0.0001722	0.0002492	0.0001657	0.00054122
τ^2 95% CI	[.0001038; .0002856]	[.0001083; .0005735]	[.0000981; .0002801]	n/a
ICC	51.9%	62.10%	50.90%	77.50%
ICC 95% CI	[37.8%; 65.7%]	[41.0%; 79.5%]	[.3632278; .6523892]	n/a
Wald χ^2 (F-statistic)	916.5	915.2	1,008.4	33.8
Prop $> \chi^2$ (F)	0.000	0.000	0.000	0.000
Log Pseudolikelihood	1,182.1	1,177.9	1,182.8	n/a
Observations	426	426	426	426
Number of groups	80	80	80	80

PCP-Clustered standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.
Note. # F-statistic applies to column (4). na: statistics not provided in the fixed-effects estimation.

Table 5 Estimating model (2) using $\widehat{u}_g$ obtained from a fixed-effect specification

	(1) Not rec	(2) No access
$\widehat{u}_g$	0.914***	0.886***
	(0.334)	(0.289)
$\widehat{\epsilon}_{gt}$	-0.053	0.501**
	(0.222)	(0.246)
Closest hospital	-0.008	-0.005
	(0.015)	(0.014)
Deprivation rank	0.001	-0.002***
	(0.001)	(0.001)
Scale	-0.001	0.001
	(0.001)	(0.001)
Female	-0.414**	0.001
	(0.190)	(0.183)
Elderly	1.865***	1.699***
	(0.527)	(0.496)
CMI	-0.246**	-0.314***
	(0.108)	(0.087)
Constant	Yes	Yes
Year FE	Yes	Yes
PCP effect	Random	Random
Wald χ^2	21.0	59.7
Prop $> \chi^2$	0.051	0.000
Log Pseudolikelihood	614.7	607.4
Observations	387	387
Number of groups	80	80

PCP-Clustered standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1
Not rec: Proportion of patients refraining to recommend their PCP. No access: Proportion of patients experiencing access problems at their PCP.

Table 6 **Estimating model (3) using $\widehat{u}_g$ obtained from a fixed-effect specification**

	(1) $P(QOF_{gt} \geq 90\%)$	(2) $P(QOF_{gt} \geq 95\%)$
$\widehat{u}_g$	-21.025*	6.380
	(11.344)	(13.849)
$\widehat{\epsilon}_{gt}$	11.589	15.960*
	(12.074)	(8.723)
Closest hospital	0.899	0.795
	(0.687)	(0.550)
Deprivation rank	-0.056	-0.010
	(0.043)	(0.049)
Scale	-0.038	-0.027
	(0.054)	(0.067)
Female	12.350*	4.465
	(6.321)	(7.459)
Elderly	-28.097	14.372
	(19.494)	(19.254)
CMI	-1.128	-5.220
	(3.840)	(3.867)
Constant	Yes	Yes
Year FE	Yes	Yes
PCP effect	Random	Random
Wald χ^2	25.6	24.7
Prop $> \chi^2$	0.012	0.016
Log Pseudolikelihood	-115.6	-202.7
Observations	389	389
Number of groups	80	80

PCP-Clustered standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 7 **Estimating model (4) using $\widehat{u}_g$ obtained from a fixed-effect specification**

	(1) $\widehat{u}_g$
PpFTE	0.009***
	(0.002)
Closest hospital	-0.002
	(0.003)
Deprivation rank	-0.001***
	(0.000)
Scale	-0.001**
	(0.000)
Female	0.397***
	(0.061)
Elderly	-0.667***
	(0.165)
CMI	0.025
	(0.036)
Constant	Yes
Observations	79
R^2	0.692

Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

4. Robustness checks

In this section we test the robustness of several sample-defining choices.

4.1. Alternative thresholds to define relevant PCPs

The primary analysis focuses on the top PCPs accounting for 50% of attendances at Hospital 1 and 65% at Hospital 2, resulting in 81 distinct PCPs. In this section we test the robustness of our findings with respect to these thresholds. More specifically, we re-estimate models (1) – (4) for

- the top PCPs accounting for 55% of attendances at Hospital 1 and 60% at Hospital 2, resulting in 80 distinct PCPs.
- the top PCPs accounting for 55% of attendances at Hospital 1 and 70% at Hospital 2, resulting in 91 distinct PCPs.
- the top PCPs accounting for 60% of attendances at Hospital 1 and 60% at Hospital 2, resulting in 90 distinct PCPs.
- the top PCPs accounting for 60% of attendances at Hospital 1 and 70% at Hospital 2, resulting in 101 distinct PCPs.

All alternative samples yield similar results to those presented in the main paper (see Tables 8 – 13). We therefore conclude that the findings are not driven by the discretionary thresholds.

4.2. Excluding PCP closures and expanding PCPs

In the primary analysis we include data from PCPs that expanded to new branches up until one year before the expansion. To alleviate concerns that the results may be driven by expansions, we now exclude expanding (N=1) PCPs from the entire study period and re-estimate models (1)–(4). The results are presented in Tables 8 – 13 and are quantitatively comparable to the primary analysis presented in the paper.

Table 8 Estimating model (1) on different samples

	Top 80 PCPs	Top 91 PCPs	Top 90 PCPs	Top 101 PCPs	Excluding expanding PCPs
	(1)	(2)	(3)	(4)	(5)
	A_{gkt}	A_{gkt}	A_{gkt}	A_{gkt}	A_{gkt}
Closest hospital	0.003	0.004	0.002	0.003	-0.000
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Deprivation rank	-0.001**	-0.001**	-0.001***	-0.001**	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Scale	-0.000	-0.000	-0.000	-0.000	0.000
	(0.001)	(0.000)	(0.001)	(0.000)	(0.000)
Female	0.125	0.122	0.128	0.119	0.138
	(0.108)	(0.105)	(0.113)	(0.111)	(0.100)
Elderly	0.011	-0.007	-0.111	-0.098	-0.166
	(0.174)	(0.136)	(0.145)	(0.131)	(0.162)
CMI	0.025	0.045	0.030	0.044	0.060*
	(0.039)	(0.033)	(0.040)	(0.035)	(0.033)
Constant	Yes	Yes	Yes	Yes	Yes
ED × Year FE	Yes	Yes	Yes	Yes	Yes
τ^2	0.0002358	0.0002041	0.0002334	0.0002192	0.0001736
τ^2 95% CI	[.0001445; .0003846]	[.0001353; .0003079]	[.0001519; .0003587]	[.0001497; .0003209]	[.0001046; .0002878]
ICC	56.6%	51.5%	49.2%	47.9%	51.80%
ICC 95% CI	[41.8%; 70.3%]	[39.7%; 63.2%]	[36.0%; 62.5%]	[36.7%; 59.4%]	[37.7%; 65.7%]
Wald χ^2	502.9	743.0	560.4	881.2	918.6
Prop $> \chi^2$	0.000	0.000	0.000	0.000	0.000
Log Pseudolikelihood	1,154.8	1,422.3	1,282.7	1,551.2	1,163.3
Observations	428	528	497	597	420
Number of groups	80	91	90	101	79

PCP-Clustered standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Table 9 Estimating model (2) on different samples

	Top 80 PCPs	Top 91 PCPs	Top 90 PCPs	Top 101 PCPs	Excluding expanding PCPs
	(1)	(2)	(3)	(4)	(5)
	Not rec	Not rec	Not rec	Not rec	Not rec
$\widehat{u}_g$	1.021**	0.991**	1.119***	1.036***	1.241***
	(0.406)	(0.398)	(0.358)	(0.339)	(0.463)
$\widehat{\epsilon}_{gt}$	0.084	0.036	0.088	0.045	-0.008
	(0.203)	(0.189)	(0.160)	(0.153)	(0.216)
Closest hospital	-0.004	-0.008	-0.005	-0.009	-0.009
	(0.016)	(0.016)	(0.015)	(0.015)	(0.017)
Deprivation rank	0.000	0.000	-0.000	-0.000	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Scale	-0.003**	-0.003**	-0.003**	-0.003**	-0.002*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female	-0.073	-0.094	-0.072	-0.095	-0.147
	(0.119)	(0.113)	(0.110)	(0.106)	(0.123)
Elderly	1.291**	1.124**	0.664	0.661*	1.659***
	(0.559)	(0.462)	(0.423)	(0.363)	(0.503)
CMI	-0.196	-0.189*	-0.147	-0.147	-0.242**
	(0.120)	(0.107)	(0.113)	(0.102)	(0.107)
Constant	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
PCP effect	Random	Random	Random	Random	Random
Wald χ^2	24.7	24.9	23.4	24.0	21.3
Prop $> \chi^2$	0.016	0.015	0.023	0.020	0.046
Log Pseudolikelihood	598.2	687.9	675.9	766.4	610.3
Observations	384	439	433	488	384
Number of groups	80	91	90	101	79

PCP-Clustered standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.
Not rec: Proportion of patients refraining to recommend their PCP.

Table 10 Estimating model (2) on different samples

	Top 80 PCPs	Top 91 PCPs	Top 90 PCPs	Top 101 PCPs	Excluding expanding PCPs
	(1)	(2)	(3)	(4)	(5)
	No access	No access	No access	No access	No access
$\widehat{u}_g$	0.955***	0.889***	1.369***	1.238***	1.040***
	(0.336)	(0.343)	(0.323)	(0.323)	(0.369)
$\widehat{\epsilon}_{gt}$	0.508**	0.362*	0.349**	0.248	0.520**
	(0.211)	(0.202)	(0.165)	(0.163)	(0.240)
Closest hospital	0.002	-0.005	0.002	-0.003	-0.006
	(0.014)	(0.014)	(0.013)	(0.013)	(0.015)
Deprivation rank	-0.003**	-0.002**	-0.002**	-0.002**	-0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Scale	-0.001	-0.001	0.000	-0.000	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female	0.318***	0.300***	0.224**	0.217**	0.305***
	(0.106)	(0.103)	(0.112)	(0.109)	(0.115)
Elderly	1.272***	0.973**	0.827**	0.673**	1.328***
	(0.449)	(0.406)	(0.331)	(0.306)	(0.452)
CMI	-0.308***	-0.258***	-0.228**	-0.194**	-0.301***
	(0.090)	(0.083)	(0.091)	(0.083)	(0.088)
Constant	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
PCP effect	Random	Random	Random	Random	Random
Wald χ^2	64.5	57.2	61.0	54.7	64.0
Prop $> \chi^2$	0.000	0.000	0.000	0.000	0.000
Log Pseudolikelihood	605.2	682.8	680.5	757.9	602.5
Observations	384	439	433	488	384
Number of groups	80	91	90	101	79

PCP-Clustered standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.
No access: Proportion of patients experiencing access problems at their PCP.

Table 11 Estimating model (3) on different samples

	Top 80 PCPs	Top 91 PCPs	Top 90 PCPs	Top 101 PCPs	Excluding expanding PCPs
	(1) $P(QOF_{gt} \geq 90\%)$	(2) $P(QOF_{gt} \geq 90\%)$	(3) $P(QOF_{gt} \geq 90\%)$	(4) $P(QOF_{gt} \geq 90\%)$	(5) $P(QOF_{gt} \geq 90\%)$
$\widehat{u}_g$	-18.184	-19.552	-6.053	-2.895	-39.599**
	(12.681)	(13.570)	(15.470)	(15.649)	(15.693)
$\widehat{\epsilon}_{gt}$	16.933	14.448	5.300	3.797	12.967
	(11.160)	(10.681)	(8.229)	(8.087)	(11.467)
Closest hospital	0.880	0.894	0.944	1.007	0.780
	(0.716)	(0.706)	(0.726)	(0.716)	(0.696)
Deprivation rank	-0.009	-0.021	-0.009	-0.021	-0.026
	(0.040)	(0.038)	(0.035)	(0.033)	(0.038)
Scale	0.010	-0.026	0.050	0.019	-0.023
	(0.060)	(0.057)	(0.062)	(0.058)	(0.055)
Female	3.961	4.513	3.854	4.336	6.576*
	(3.920)	(3.827)	(3.480)	(3.287)	(3.819)
Elderly	-13.933	-10.853	-25.477*	-20.872	-25.717
	(18.693)	(16.766)	(14.726)	(14.490)	(19.241)
CMI	-1.854	-2.457	-1.669	-2.405	-0.819
	(4.012)	(3.710)	(3.722)	(3.500)	(3.864)
Constant	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
PCP effect	Random	Random	Random	Random	Random
Wald χ^2	21.8	23.1	25.3	26.4	26.4
Prop $> \chi^2$	0.040	0.027	0.013	0.009	0.009
Log Pseudolikelihood	-113.4	-128.0	-134.8	-150.0	-114.4
Observations	385	440	434	489	386
Number of groups	80	91	90	101	79

PCP-Clustered standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Table 12 Estimating model (3) on different samples

	Top 80 PCPs	Top 91 PCPs	Top 90 PCPs	Top 101 PCPs	Excluding expanding PCPs
	(1)	(2)	(3)	(4)	(5)
	$P(QOF_{gt} \geq 95\%)$	$P(QOF_{gt} \geq 95\%)$	$P(QOF_{gt} \geq 95\%)$	$P(QOF_{gt} \geq 95\%)$	$P(QOF_{gt} \geq 95\%)$
$\widehat{u}_g$	-8.701	-8.431	-3.980	-0.678	-10.643
	(12.758)	(13.453)	(13.821)	(13.742)	(17.380)
$\widehat{\epsilon}_{gt}$	19.986**	16.067**	12.731*	9.768	19.067**
	(8.297)	(7.880)	(7.080)	(6.791)	(8.722)
Closest hospital	0.668	0.724	0.698	0.778	0.640
	(0.548)	(0.534)	(0.545)	(0.531)	(0.555)
Deprivation rank	-0.009	-0.020	-0.013	-0.023	-0.013
	(0.036)	(0.035)	(0.034)	(0.033)	(0.042)
Scale	-0.016	-0.037	-0.001	-0.018	-0.044
	(0.062)	(0.059)	(0.058)	(0.056)	(0.066)
Female	5.530	6.505	6.491	7.395	7.558
	(5.843)	(5.526)	(5.438)	(5.083)	(5.737)
Elderly	10.506	15.850	0.625	6.635	5.099
	(16.197)	(15.205)	(13.401)	(13.540)	(18.118)
CMI	-4.085	-5.703	-4.801	-6.251*	-4.505
	(3.806)	(3.560)	(3.754)	(3.542)	(3.826)
Constant	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes
PCP effect	Random	Random	Random	Random	Random
Wald χ^2	28.3	31.2	21.0	25.0	25.3
Prop $> \chi^2$	0.005	0.002	0.050	0.015	0.014
Log Pseudolikelihood	-204.5	-227.5	-228.4	-251.2	-201.1
Observations	385	440	434	489	386
Number of groups	80	91	90	101	79

PCP-Clustered standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Table 13 Estimating model (4) on different samples

	Top 80 PCPs	Top 91 PCPs	Top 90 PCPs	Top 101 PCPs	Excluding expanding PCPs
	(1)	(2)	(3)	(4)	(5)
	$\widehat{u}_g$	$\widehat{u}_g$	$\widehat{u}_g$	$\widehat{u}_g$	$\widehat{u}_g$
PpFTE	0.009***	0.008***	0.008***	0.007***	0.008***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Closest hospital	-0.006*	-0.005*	-0.006*	-0.005	-0.002
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Deprivation rank	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Scale	0.000	-0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Female	0.003	0.006	-0.021	-0.009	0.027
	(0.071)	(0.068)	(0.083)	(0.080)	(0.048)
Elderly	-0.194	-0.099	-0.166	-0.127	-0.174
	(0.157)	(0.143)	(0.106)	(0.113)	(0.137)
CMI	0.019	0.014	0.039	0.032	0.012
	(0.037)	(0.030)	(0.034)	(0.030)	(0.030)
Constant	Yes	Yes	Yes	Yes	Yes
Observations	79	90	88	99	78
R^2	0.217	0.175	0.185	0.145	0.249

Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.
As in the main paper, we excluded PCPs with implausible staffing numbers (one PCP excluded in (1), (2) and (5), two PCPs excluded in (3)-(4)).

5. Single hospital study

In this section we outline that it is feasible to apply the methodology with data from a single hospital. Following the steps outlined in §3 of the paper, we estimate model (1) independently for Hospital 1 and Hospital 2. Table 14 presents the results and Figure 3 shows the $\hat{u}_g$ distribution for both hospitals. The results indicate greater uncertainty in estimating between-PCP variability when analyzing each hospital independently compared to the combined analysis, leading to wider 95% CIs. Specifically, the ICC estimated based on Hospital 1 is 52.6% (95% CI [29.2%; 75.1%]), and for Hospital 2 is 33.9% (95% CI [22.6%; 47.6%]). Crucially, despite this increased uncertainty, the individual estimates of PCP burden on EDs ($\hat{u}_g$) are highly consistent between the single-hospital and combined analyses. Figure 4 illustrates this consistency, showing a strong positive correlation between the $\hat{u}_g$ estimates from the combined data and those from each hospital individually (correlation of 0.886 (p<0.001) for Hospital 1, and 0.769 (p<0.001) for Hospital 2).

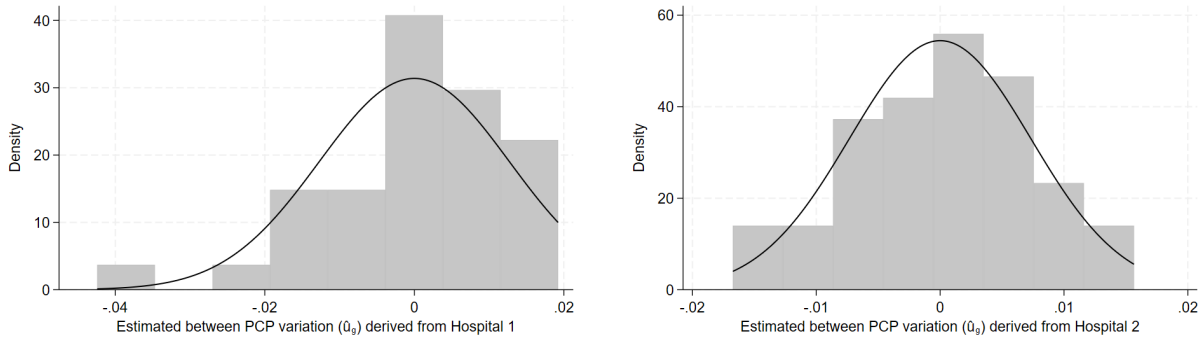

Figure 3 Variation between PCPs.

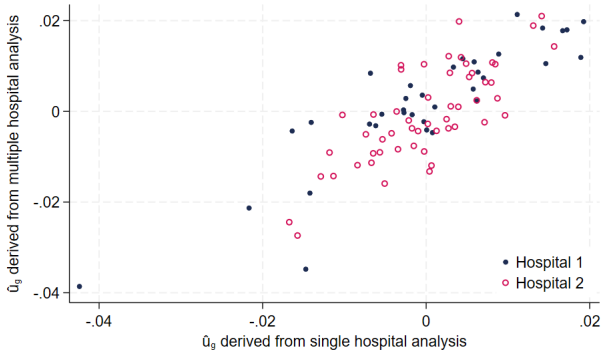

Figure 4 Correlation between the U_g

Subsequently, we replicate models (2) – (4) to demonstrate that the measure $\widehat{u}_g$ estimated using single hospital data is correlated with PCP surveys, QOF scores, and patient-to-staff ratios. Not surprisingly, since the single hospital analysis uses less data, some of the results are estimated with

more uncertainty compared to the analysis using both hospitals. Specifically, we find that PCPs with a higher $\widehat{u}_g$ are also the ones patients are more likely dissatisfied with. Table 15 indicates $\beta_U > 0$ for all models, albeit in Column (1) not distinguishable from zero. Regarding the QOF score, most of the coefficients γ_U are negative as expected but they are estimated with very large standard errors leading to results that are not statistically significant (Table 16). For the performance antecedents we find, as expected, a positive association between patient-to-staffing ratio and $\widehat{u}_g$ for both hospitals (Table 17). Taken together the results provide similar indications of validity as the analysis conducted on the joint data.

Table 14 Estimating model (1) for both hospitals separately

	Hospital 1 (1)	Hospital 2 (2)
Closest hospital	-0.016***	0.000
	(0.006)	(0.006)
Deprivation rank	-0.002**	-0.001**
	(0.001)	(0.000)
Scale	0.001	0.000
	(0.001)	(0.000)
Female	0.172***	-0.213***
	(0.063)	(0.066)
Elderly	-0.151	-0.406***
	(0.351)	(0.150)
CMI	0.143*	0.065**
	(0.082)	(0.025)
Constant	Yes	Yes
Year FE	Yes	Yes
τ^2	0.0001875	0.0000736
τ^2 95% CI	[.0000782; .0004494]	[.0000441; .000123]
ICC	52.7%	33.9%
ICC 95% CI	[29.2%; 75.1%]	[22.6%; 47.6%]
Wald χ^2	136.1	179.4
Prop $> \chi^2$	0.000	0.000
Log Pseudolikelihood	450.8	751.3
Observations	165	261
Number of groups	35	53

PCP-Clustered standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Table 15 Estimating model (2) for both hospitals separately

	Hospital 1		Hospital 2	
	(1)	(2)	(3)	(4)
	Not rec	No access	Not rec	No access
$\widehat{u}_g$	1.024	0.921**	1.915**	2.037***
	(0.670)	(0.449)	(0.772)	(0.727)
$\widehat{\epsilon}_{gt}$	0.142	0.776**	-0.096	0.197
	(0.355)	(0.316)	(0.196)	(0.281)
Closest hospital	0.020	0.022	-0.035	-0.025
	(0.019)	(0.016)	(0.021)	(0.019)
Deprivation rank	-0.001	-0.004**	-0.000	-0.003***
	(0.002)	(0.002)	(0.001)	(0.001)
Scale	0.001	0.003**	-0.003**	-0.002
	(0.002)	(0.001)	(0.001)	(0.001)
Female	-0.183	0.299**	-0.640**	-0.254
	(0.115)	(0.124)	(0.309)	(0.260)
Elderly	2.902***	2.561***	0.934*	0.992*
	(1.106)	(0.699)	(0.482)	(0.598)
CMI	-0.449**	-0.620***	-0.271**	-0.277***
	(0.212)	(0.147)	(0.125)	(0.105)
Constant	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
PCP effect	Random	Random	Random	Random
Wald χ^2	17.9	113.4	37.5	30.6
Prop $> \chi^2$	0.118	0.000	0.000	0.002
Log Pseudolikelihood	266.6	257.0	418.0	421.1
Observations	162	162	261	261
Number of groups	35	35	53	53

PCP-Clustered standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Table 16 Estimating model (3) for both hospitals separately

	Hospital 1		Hospital 2	
	(1)	(2)	(3)	(4)
	$P(QOF_{gt} \geq 90\%)$	$P(QOF_{gt} \geq 95\%)$	$P(QOF_{gt} \geq 90\%)$	$P(QOF_{gt} \geq 95\%)$
$\widehat{u}_g$	-30.024	9.945	-99.535***	-23.309
	(23.629)	(21.886)	(30.325)	(37.208)
$\widehat{\epsilon}_{gt}$	8.994	25.058**	15.249	4.214
	(17.450)	(10.381)	(14.449)	(11.900)
Closest hospital	0.638	1.446*	0.661	-0.516
	(0.922)	(0.773)	(1.034)	(0.913)
Deprivation rank	0.034	-0.060	-0.048	-0.027
	(0.077)	(0.064)	(0.045)	(0.060)
Scale	-0.005	0.104	-0.028	-0.173*
	(0.082)	(0.070)	(0.061)	(0.091)
Female	6.712*	2.201	1.645	7.847
	(3.863)	(5.475)	(13.295)	(15.283)
Elderly	-44.601	10.765	-20.077	-3.489
	(29.225)	(26.732)	(21.060)	(25.136)
CMI	5.439	-5.087	-0.158	-5.056
	(7.284)	(6.196)	(4.424)	(5.065)
Constant	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
PCP effect	Random	Random	Random	Random
Wald χ^2	39.5	26.8	29.3	20.7
Prop $> \chi^2$	0.000	0.005	0.004	0.055
Log Pseudolikelihood	-35.9	-83.2	-81.4	-129.6
Observations	135	164	261	261
Number of groups	35	35	53	53

PCP-Clustered standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.
The number of observations differ between (1) and (2) due to perfect prediction.

Table 17 Estimating model (4) for both hospitals separately

	Hospital 1	Hospital 2
	(1)	(2)
	$\widehat{u}_g$	$\widehat{u}_g$
PpFTE	0.010***	0.005***
	(0.003)	(0.001)
Closest hospital	-0.005	0.002
	(0.005)	(0.004)
Deprivation rank	0.000	-0.000
	(0.001)	(0.000)
Scale	-0.000	0.000
	(0.000)	(0.000)
Female	0.016	0.123*
	(0.039)	(0.072)
Elderly	-0.293	-0.027
	(0.317)	(0.112)
CMI	0.040	0.002
	(0.076)	(0.020)
Constant	Yes	Yes
Observations	35	52
R^2	0.291	0.242

Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.
As in the main paper, we excluded the one PCP with implausible staffing numbers, leading to N=52 for Hospital 2.

6. Bootstrapping the effect size

In §6 of the main paper we estimate the impact on ACS attendances and costs of a counterfactual scenario where PCPs that performed worse than the 25% could improve their performance to the 25% percentile. These results were estimated using the following bootstrapping algorithm:

1. We draw a random sample of 80 PCPs with replacement, ensuring that the resampling respects the natural grouping of the data, specifically the proportion of ACS attendances A_{gkt} within each PCP g. We chose 80 PCPs because this matches the number of PCP clusters used in the primary analysis. Since PCP clusters are sampled with replacement, some clusters may appear more than once, while others may not appear at all in a given bootstrap sample. As a result, the number of observations in the bootstrap sample may differ from the original, depending on how many times individual clusters are selected and their size.

2. We estimate model (1) based on this new dataset.

3. We assume that for each PCP-hospital-year the number of Non-ACS attendances remains as observed, and calculate the predicted number of ACS attendances for each PCP-year, using $ACS = \frac{\widehat{A}_{gkt}}{1-\widehat{A}_{gkt}} NonACS$.

4. We aggregate these predictions to the ED level to obtain the status quo predictions.

5. We determine the 25% percentile of $\widehat{u}_g$, denoted as $\widehat{u}_{25}$.

6. For PCPs that perform worse than the 25% percentile (i.e. $\widehat{u}_g > \widehat{u}_{25}$), we set $\widehat{u}_g$ equal to $\widehat{u}_{25}$. PCPs that perform better than the 25% percentile are left unchanged.

7. We repeat steps 3 to 4 to obtain the predicted number of ACS attendances for the counterfactual scenario.

8. We calculate the difference in ACS attendances compared to the status quo.

9. We repeat steps 1 to 8 10,000 times to estimate a 95% confidence interval for the difference in ACS attendances based on the empirical distribution.

7. Determining the impact of different resource allocation policies on ACS ED attendances

In §6 of the main paper we estimate the impact of four different resource allocation policies on ACS ED attendances. Each allocation policy uses a different criterion to identify PCPs that are selected to receive an additional 0.5 FTE physicians:

- Policy 1 allocates resources to PCPs with the highest proportion of ACS attendances (as measured by $\widehat{u}_g$).
- Policy 2 selects PCPs with the highest proportion of patients reporting access problems (averaged over the duration of study).
- Policy 3 allocates resources according to the PCP's proportion of ACS admissions (as measured according to §4.4 of the paper).

- Policy 4 is a random policy in which PCPs are selected randomly with equal probability.

For each policy the results were estimated according to the steps below. Note that for the random policy 4 we repeat the steps 1,000 times to derive the mean of the counterfactual predictions.

1. We calculate the impact at PCP-level averaged across EDs and years: First, we use the model estimates from (1) to compute the PCP's adjusted proportion of ACS visits: $\widehat{A}_{gkt} = \widehat{\alpha}_0 + \widehat{\alpha}_C C_{gkt}$, average this over the duration of study, and denote this as $\tilde{A}_g$. $\tilde{A}_g$ is the averaged part of PCP performance that is not related to staffing and therefore not affected by the allocation policies.

2. We assume that $X = \{1, \ldots, 15\}$ FTE can be allocated in increments of 0.5 FTE across different PCPs. This means that $N = X/0.5$ PCPs will see an increase in staffing.

3. We rank PCPs in descending order of the selection criterion and select the first N PCPs to see an increase in staffing.

4. We compute counterfactual patient-to-staff ratios $PpFTE_g^c$. Specifically, for PCPs that are ranked $\leq N$, we set their staffing $PpFTE_g^c$ equal to the average number of patients per physician FTE and add 0.5 FTE to the denominator. For PCPs that are ranked $> N$, their staffing levels are left unchanged, i.e. $PpFTE_g^c = PpFTE_g$.

5. Subsequently, we use the model estimates from (4) to compute the counterfactual random effect $\tilde{u}_g = \widehat{\delta}_0 + \widehat{\delta}_P PpFTE_g^c + \widehat{\delta}_C C_g$.

6. We assume that for each PCP the average number of Non-ACS attendances remains as observed, and calculate the predicted number of ACS attendances for each PCP, using $ACS = \frac{(\tilde{A}_g + \tilde{u}_g)}{1-(\tilde{A}_g + \tilde{u}_g)} NonACS$.

7. We aggregate these predictions to the ED level to obtain the counterfactual predictions at annual ED level.

The results of this exercise are presented in §6 of the main paper.

References

Raudenbush SW (2009) Adaptive centering with random effects: An alternative to the fixed effects model for studying time-varying treatments in school settings. *Education Finance and Policy* 4(4):468–491.