Appendix

Did the Acute Frailty Network improve outcomes for older people living with frailty? A staggered difference-in-difference panel event study

Andrew Street[1], Laia Maynou[1,2,3] and Simon Conroy[4]

[1]Department of Health Policy, London School of Economics and Political Science, Houghton Street, London, WC2A 2AE, UK

[2]Department of Economics, Econometrics and Applied Economics, Universitat de Barcelona, Avinguda Diagonal, 690, 08034, Barcelona, Spain

[3]Center for Research in Health and Economics (CRES), Universitat Pompeu Fabra, Ramon Trias Fargas 25-27, 08005 Barcelona, Spain

[4]MRC Unit for Lifelong Health and Ageing, University College London, 5th Floor, 1-19 Torrington Place, London, WC1E 7HB, UK

A Methods

Event study design

Hospital sites joined the AFN in six sequential cohorts, the first starting in January 2015, the sixth in May 2018 (Table 2). Two other cohorts joined later, after the period covered by our data, so were not included in the analysis. We accounted for this differential phasing of entry into the AFN by employing a staggered difference-in-difference (DiD) panel event study approach [1] and by implementing the Callaway and Sant'Anna estimator [2]. The event study involved estimating the following equation:

$$Y_{ijt} = \alpha + \sum_{l=2}^{L} \beta_l \text{Lag}l_{ijt} + \sum_{k=1}^{K} \beta_k \text{Lead}k_{ijt} + \gamma^{'} X_{ijt} + T_t + \mu_j + e_{ijt} \quad (1)$$

where Y_{ijt} refers to one of the following outcomes: LoS, in-hospital mortality, institutionalisation and readmission within 30 days. The analysis was carried out at patient level, hence the subscripts refer to a patient i treated at site j in month t. $\text{Lag}l_{ijt}$ is a set of binary variables denoting each month prior to AFN membership (the intervention), with L indicating the number or pre-intervention months and the first lag, $l = 1$, omitted to avoid multi-collinearity. $\text{Lead}k_{ijt}$ is a set of binary variables with K indicating the number of post-intervention months in addition to the first month (denoted $k = 0$) when the site became an AFN member. To identify patients in the control group, the Lag$l$ and Lead$k$ dummy variables are assigned a missing value. X_{ijt} is a set of covariates capturing patient characteristics, T_t denotes time fixed-effects, μ_j denotes site fixed-effects and e_{ijt} is a classical error term.

The standard event study methodology assumes that the impact of the intervention is common across all cohorts. But this assumption may not hold, particularly for interventions that evolve over time. This might be the case for the AFN membership model, for two reasons. First, there may have been selection effects, if organisational characteristics influenced when sites decided to join the network. For instance, those that joined early might have had pre-existing features that were more closely aligned with those of the AFN and may have been more enthusiastic about the network's aims than those that joined later. Second, there may have been evolutionary effects, if the way that the AFN operated and worked with its members evolved over time. In recognition of these possibilities we apply the Callaway and Sant'Anna estimator [2] which relaxes the assumption of common cohort effects and involves calculating the effect of the intervention for each cohort. The Callaway and Sant'Anna estimator is termed the "group-time average treatment effect", ATT_{gt}, defined as:

$$ATT_{gt} = E\left[\left(\frac{G_g}{E(G_g)} - \frac{\frac{p_g(X)C}{1-p_g(X)}}{E[\frac{p_g(X)C}{1-p_g(X)}]}\right)(Y_t - Y_{g-1})\right] \quad (2)$$

where G_g equals 1 if site j is in the first cohort that became AFN members at time t, and

zero otherwise, and C takes a value of 1 for sites that never become AFN members (termed "never-treated" sites). Patients from the never-treated sites are used as controls, and there is an option also to include as controls those patients cared for in sites before they become AFN members (termed "not-yet-treated" sites).

The estimator allows for the possibility that patients subject to the intervention might differ from those who do not. To ensure that characteristics are balanced between patients in the cohort and control groups, propensity score weighting is used when calculating the ATT_{gt}. The generalised propensity score is defined as:

$$p_g(X) = P(G_g = 1|X, G_g + C = 1) \tag{3}$$

which is the probability that an individual is treated in a site that is an AFN member, conditional on (i) having characteristics X and (ii) being a member of cohort g or the control group C. This ATT_{gt} correction works by up-weighting patients from the control group that have characteristics similar to those frequently found in group g and down-weighting patients from the control group that have characteristics rarely seen in group g [2; 3]. The weighting is designed to establish parallel trends in outcomes prior to AFN membership for the control and cohort groups.

The ATT_{gt} is estimated for each cohort so, to obtain unbiased standard errors for the ATT_{gt}, the Callaway and Sant'Anna estimator uses wild bootstrapping and clustering at the site level. The ATT_{gt} estimates are reported as average marginal effects (AMEs). For the LoS equation, the AMEs can be interpreted a change in the number of days in hospital due to being cared for in an AFN site. For the other three equations, the AMEs indicate the difference in the probability of dying, being institutionalised or being readmitted for those in AFN sites compared to those in control sites.

In our main analysis, we made the following decisions. First, we applied the same length of pre-intervention period to all AFN cohorts, with $L = 36$. Second, we applied the same post-intervention period to all cohorts, covering eleven months, such that $K + 1 = 11$ (noting that the estimator assigns $k = 0$ to the first intervention month). This ensured that the post-intervention period was the same across cohorts, the last month of available data being eleven months from enrolment for those sites in cohort six. Third, patients in the control group were drawn from both "never-treated" sites that never joined the AFN and from "not-yet-treated" sites that subsequently enrolled.

Robustness checks

We conducted three types of robustness check to assess the sensitivity of our results to our analytical choices. The first choice concerned the length of intervention period, enabling us to identify longer-term effects. We ran one analysis in which we allowed the intervention

period to run to the last month for which data were available, March 2019. This meant that the intervention period varied by cohort, covering 39 months for sites in cohort one but just eleven months for those in cohort six. In another analysis, the intervention period was set to twelve months ($K+1=12$) rather than eleven months, though this meant data were missing for the twelfth month for cohort six sites.

Our second robustness check concerned the selection of controls, restricting these to so-called "never-treated" patients cared for in hospital sites that never joined the AFN. In essence, this restriction meant that patients cared for in sites before they became AFN members played no part for the purposes of this analysis.

Our third robustness check applied a tighter definition of the AFN intervention, focusing on sites thought to have engaged most actively, as defined according to adoption and implementation of the AFN best practice principles. In theory, some sites may have already adopted all of the principles prior to joining the network; others may have joined the network, but subsequently failed to adopt any of the principles. In either case, joining the AFN would have made no difference to the adoption of best practice principles. In reality, sites lay between these extremes, with variation both in how many of the principles were in place prior to enrolment and in how many principles were adopted after joining. If the adoption of these principles did indeed drive improved outcomes, the impact of AFN membership should be most evident for those sites that adopted more of the principles after they joined the AFN. Sites from cohort two through to cohort six were surveyed about which of the ten principles were in place prior to joining and within twelve months of joining. On average, joining the AFN was associated with the adoption of an additional four of the principles. Using this information we defined two forms of active engagement. The first restricted the intervention group to those AFN sites that adopted four or more of the principles subsequent to joining. Implementation of this form of the robustness check meant that the other AFN sites (including all those from cohort one, which were not surveyed) were dropped from this specific analysis. The second form of this robustness check considered adoption of each best practice principle in turn. In this set of analyses, only those sites that adopted the specific principle were considered to have been intervention sites. This meant dropping from the analyses those AFN sites that had already adopted the principle prior to joining the AFN and those sites that did not adopt the principle upon joining.

B Data

We analysed anonymised patient-level data from the Hospital Episode Statistics (HES) which contains details of all admissions to NHS hospitals. The HES data dictionary provides information about the variables in this dataset [4], those used for the purposes of analysis indicated below using square brackets. The unit of observation in HES is a consultant "episode", which captures the time a patient spends under the care of a senior doctor, who may discharge the patient or transfer responsibility to another doctor, the latter case generating a new episode. We identify all episodes for each patient during their hospital stay allowing us to analyse information recorded during their entire spell in hospital.

The AFN focused on the acute care needs of older people living with frailty so we identified all patients in HES aged 75+ with a high Hospital Frailty Risk Score (HFRS) [5; 6] who had an emergency hospital admission to departments of Acute Internal Medicine [MainSpef=326, general [MainSpef=300] or geriatric medicine [MainSpef=430] between 1 January 2012 to 31 March 2019.

The HFRS is calculated by combining a weighted set of 109 3-character International Statistical Classification of Diseases and Related Health Problems, tenth revision (ICD-10) diagnostic codes recorded across all the episodes associated with their current hospital spell and across all the episodes associated with their two previous emergency admissions occurring in the prior two years [6]. The HFRS takes values from 0 to 173.2 and patients are categorised as having low (HFRS score <5), intermediate (HFRS score 5-15) or high (HFRS score >15) frailty.

We constructed four outcome variables for each patient.

- Length of stay (LoS) was calculated as the difference between the patient's discharge and admission dates.
- We determined whether or not the patient died in hospital from the discharge method [DisMeth] variable which records the circumstances under which a patient left hospital. If the patient died in hospital, DisMeth takes a value of 4.
- We defined a patient as being institutionalised following their hospital stay if they were admitted from their own home but discharged to a facility providing long term care. This was determined by comparing Admission Source [AdmiSorc], which records where the patient was immediately prior to admission, with Destination on Discharge [DisDest, which identifies where patients went after leaving hospital. Institutionalisation took a value of 1 for patients admitted from their own home [AdmiSorc=19] and but discharged to an NHS run care home [DisDest=54, Local Authority residential care [DisDest=65], a non-NHS run care home [DisDest=85] or a non-NHS run hospice [DisDest=88]; 0 otherwise.

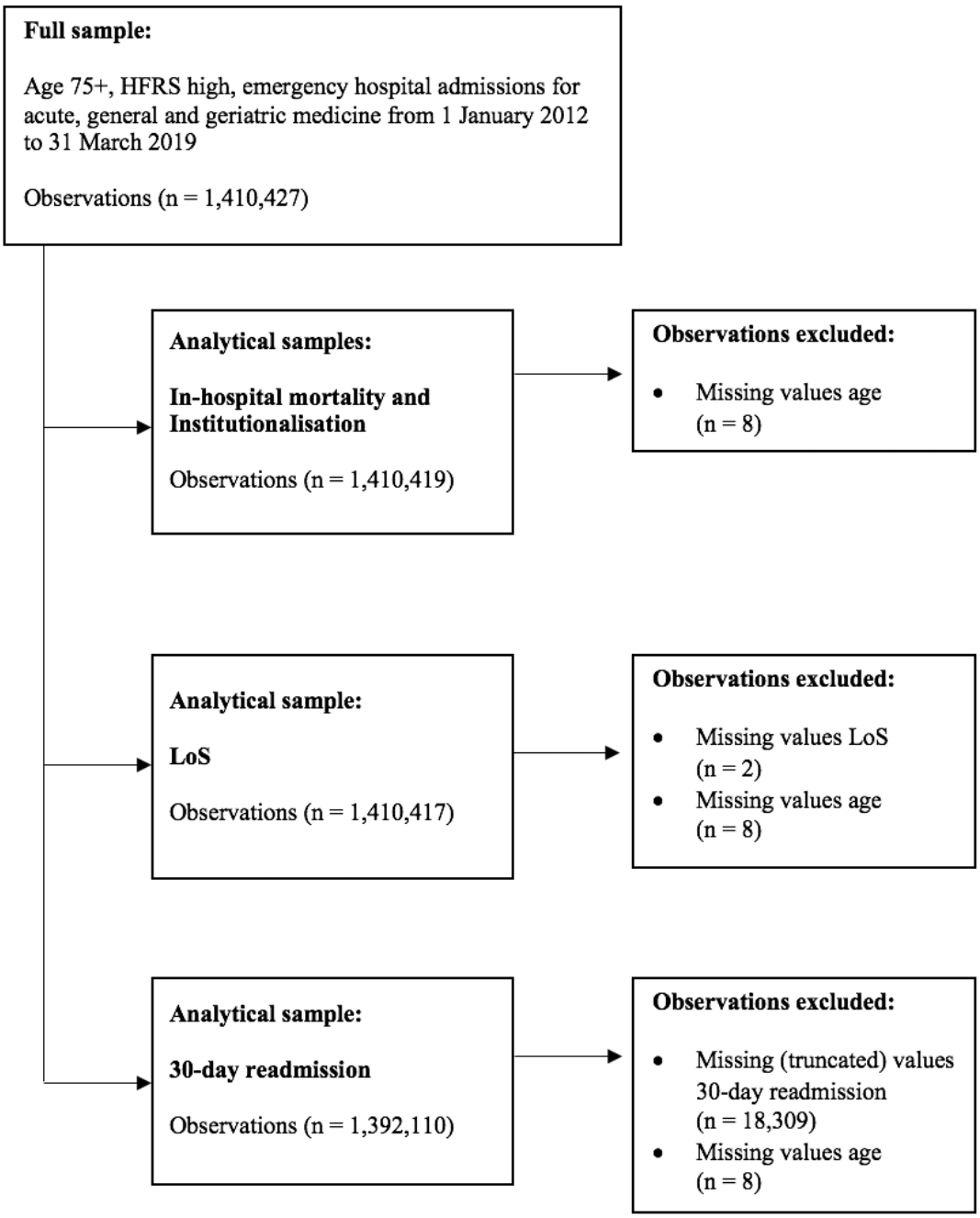
- Unplanned readmission was defined as an emergency readmission to any hospital within 30 days of being discharged after the index admission.

The analyses controlled for patient age, categorised into 5-year age bands, and gender. The socioeconomic conditions of where patients lived was accounted for using the deciles of Index of Multiple Deprivation [IMD04Decile] [7]. This variable identifies the area in which the patient lives, allocated to one of ten groups ordered from most deprived areas [IMD04Decile=1] through to least deprived areas [IMD04Decile=10]. The IMD variables were not included in the final specifications to ensure that all models converged and featured the same set of variables.

While all patients were categorised as having high frailty risk, we accounted for their actual HFRS score [5; 6]. Clinical complexity was captured using the Charlson comorbidity index [8; 9], which uses age and ICD-10 indicators of comorbidity to estimate mortality risk. The index takes values from 0 to 17 but was categorised for the analysis into four groups (0,1,2 and 3+) [10]. We also included counts of the number of diagnosis codes and of the number of emergency admissions in the previous year. In the analyses of LoS, institutionalisation and readmission, we controlled for whether the patient died in hospital.

Starting and analytical samples

Figure A1: Analytical Samples



C Robustness check - Long term effects

The rationale for this robustness check is that effects may take longer than 12 months to be emerge. This possibility is examined by exploiting the full longitudinal potential of the dataset, which ran until 31 March 2019.

Implementing this check meant that the post-intervention period varied by cohort, from 39 months for sites in cohort one to 11 months for those in cohort six. Obviously, this had implications for the number of patients considered to have been exposed to the intervention, which increased from 105,292 in the main analysis to 307,503 in this analysis (the number of controls is unchanged). While the number of patients in cohort six was unaffected, there was an increase for all other cohorts, notably for cohort one, where the number increased from 10,966 to 71,326 as shown in Table A1.

Table A2 provides descriptive statistics for the newly-defined patients subject to the intervention, compared to the controls. As in the main analysis, those subject to the intervention had shorter LoS, lower proportions died in hospital or changed institutional status, while a higher proportion was re-admitted within 30 days of discharge. They also differed with respect to socio-demographic and clinical characteristics.

The top row (labelled "ATT") of Table A3 reports the overall estimates of the long-term average treatment effect of AFN membership on each outcome, with A2 presenting the monthly overall effects and confidence intervals. Note that confidence intervals become wider the later the post-intervention month. This is because there are progressively fewer observations on which to base these long term estimates.

The effects by cohort are reported in the lower panel of Table A3. Only two estimates are significant. Those in cohort four had a longer LoS (1.201 days more, $p<0.05$) and a lower percentage of those in cohort six died in hospital (-0.008, $p<0.05$).

Table A1: Number of Control and Intervention patients (J=36, K+1=maximum)

	2012	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**Total**
Controls (never treated)	91,481	108,395	119,770	128,638	136,944	157,313	179,988	46,371	968,900
Controls (not yet treated)	12,568	33,811	62,168	87,654	72,012	47,573	20,449	0	336,235
Intervention	0	0	0	12,149	46,745	85,928	130,387	32,294	307,503
Cohort 1 - Jan 2015	0	0	0	12,149	14,323	17,075	21,989	5,790	71,326
Cohort 2 - Jan 2016	0	0	0	0	25,582	27,494	28,310	6,645	88,031
Cohort 3 - Sept 2016	0	0	0	0	6,840	21,982	24,756	5,937	59,515
Cohort 4 - May 2017	0	0	0	0	0	12,787	21,745	4,928	39,460
Cohort 5 - Oct 2017	0	0	0	0	0	6,590	26,052	6,399	39,041
Cohort 6 - May 2018	0	0	0	0	0	0	7,535	2,595	10,130

Notes: J indicates pre-invention months, K+1 indicates post-intervention months.

Table A2: Descriptive statistics

	Intervention	Controls (not yet treated and never treated)	t test (p-value)
N observations	307,503	1,305,135	
N sites	66	248	
Outcomes			
LoS (days)	14.63 (17.54)	16.33 (19.85)	0.000
In-hospital mortality (%)	4.04 (19.68)	4.95 (21.70)	0.000
Institutionalisation (%)	1.89 (13.62)	2.73 (16.30)	0.000
30-day readmission (%)	17.87 (38.31)	17.24 (37.77)	0.000
Covariates			
Age	86.27 (5.73)	86.13 (5.69)	0.000
Age 75-80 (%)	13.87 (34.56)	14.20 (34.91)	0.000
Age 80-85 (%)	24.90 (43.25)	25.66 (43.67)	0.000
Age 85-90 (%)	31.17 (46.32)	30.99 (46.24)	0.047
Age 90-95 (%)	22.03 (41.45)	21.88 (41.34)	0.062
Age 95+ (%)	8.03 (27.17)	7.28 (25.97)	0.000
Female (%)	59.81 (49.03)	60.71 (48.84)	0.000
HFRS high risk score	23.28 (6.70)	22.43 (6.15)	0.000
Charlson=0 (%)	9.10 (28.77)	10.67 (30.87)	0.000
Charlson=1 (%)	23.19 (42.21)	25.33 (43.49)	0.000
Charlson=2 (%)	21.91 (41.36)	22.03 (41.44)	0.141
Charlson=3+ (%)	45.80 (49.82)	41.98 (49.35)	0.000
N prev adm=0 (%)	31.06 (46.27)	29.78 (45.73)	0.000
N prev adm=1 (%)	32.71 (46.91)	33.39 (47.16)	0.000
N prev adm=2 (%)	19.22 (39.40)	19.94 (39.96)	0.000
N prev adm=3+ (%)	17.01 (37.57)	16.89 (37.46)	0.105
N diagnoses	13.05 (4.75)	11.92 (4.54)	0.000

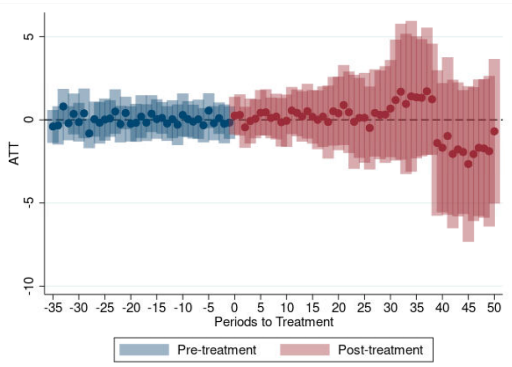
Notes: Mean (std dev). Years: 2012-2019. For intervention, J=36, K+1=maximum where J indicates pre-intervention months, K+1 indicates post-intervention months. N: Number, LoS: Length of stay, HFRS: Hospital Frailty Risk Score; N prev adm: number of previous admissions.

Table A3: Callaway & Sant'Anna results (J=36, K+1=maximum)

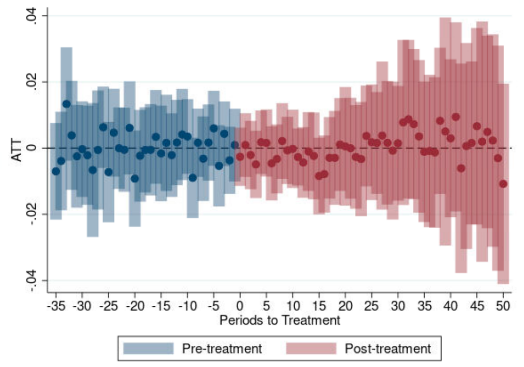
	(1) LoS	(2) In-hospital mortality	(3) Institutionalisation	(4) 30-day readmission
ATT	0.201	-0.001	-0.0003	-0.004
	(0.501)	(0.003)	(0.002)	(0.005)
ATT cohort 1	-1.301	0.002	-0.007	0.007
	(1.027)	(0.007)	(0.006)	(0.008)
ATT cohort 2	1.948	0.004	0.003	-0.015
	(1.008)	(0.006)	(0.004)	(0.008)
ATT cohort 3	-1.179	-0.005	-0.001	-0.002
	(0.929)	(0.006)	(0.004)	(0.011)
ATT cohort 4	1.201*	-0.006	0.003	-0.003
	(0.529)	(0.006)	(0.004)	(0.016)
ATT cohort 5	0.017	-0.008*	0.001	0.002
	(0.588)	(0.003)	(0.004)	(0.008)
ATT cohort 6	-0.039	0.009	0.003	-0.032
	(0.945)	(0.007)	(0.003)	(0.021)
N observations	1,610,173	1,610,176	1,610,176	1,580,553
N sites	249	249	249	249
Covariates	Yes	Yes	Yes	Yes
Cluster sites	Yes	Yes	Yes	Yes
Controls	never + not yet	never + not yet	never + not yet	never + not yet
Months	-36/+all	-36/+all	-36/+all	-36/+all
Years	2012-2019	2012-2019	2012-2019	2012-2019

Notes: ATT: average treatment effect on the treated. Coefficients reported as average marginal effects and standard errors reported in brackets. Significance levels: *** p<0.001, ** p<0.01, * p<0.05, no star = not significant. N: Number. Covariates: age, female, HFRS high risk score, Charlson Comorbodity Index, Number of previous admissions, number of unique diagnoses and died (except for model 2). J indicates pre-intervention months, K+1 indicates post-intervention months.

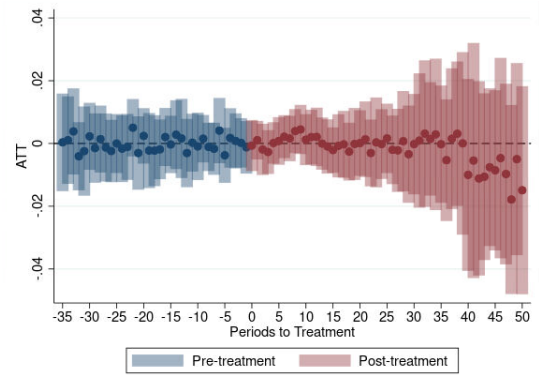
Figure A2: Callaway & Sant'Anna figures (J=36, K+1=maximum)



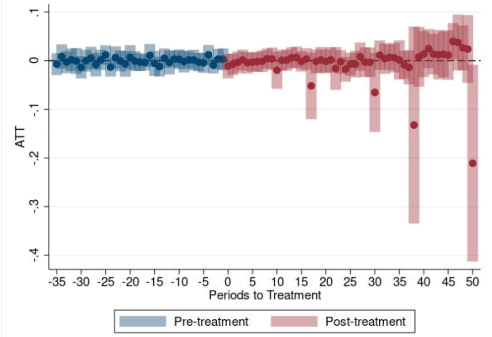
(I) - LoS



(II) - In-hospital mortality



(III) - Institutionalisation



(IV) - 30-day readmission

Notes: The dot represents the mean conditional outcome for those subject to intervention relative to the controls, with the length of the bars indicating the 95% confidence limits. J indicates pre-intervention months, K+1 indicates post-intervention months.

D Robustness check - Twelve month intervention period

We also conducted a robustness check allowing the post-intervention period to cover a full 12 months for all cohorts except cohort six (for which only 11 months of data were available). This resulted in an increased sample of patients considered subject to the intervention, rising from 105,292 in the main analysis to 114,563 here (see Table A6). Descriptive statistics for this new sample of patients are reported in Table A5 compared to the (unchanged) controls.

Table A6 and Figure A3 show that there was no significant overall effect of AFN membership. The cohort effects were also not significant, with one exception. This was for cohort two, in which a lower percentage of patients was readmitted within 30 days (-0.013, $p<0.05$).

Table A4: Number of Control and Intervention patients (J=36, K+1=12)

	2012	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**Total**
Controls (never treated)	91,481	108,395	119,770	128,638	136,944	157,313	179,988	46,371	968,900
Controls (not yet treated)	12,568	33,811	62,168	87,654	72,012	47,573	20,449	0	336,235
Intervention	0	0	0	12,149	32,422	33,653	33,744	2,595	114,563
Cohort 1 - Jan 2015	0	0	0	12,149	0	0	0	0	12,149
Cohort 2 - Jan 2016*	0	0	0	0	25,582	85	0	0	25,667
Cohort 3 - Sept 2016	0	0	0	0	6,840	14,191	0	0	21,031
Cohort 4 - May 2017	0	0	0	0	0	12,787	7,223	0	20,010
Cohort 5 - Oct 2017	0	0	0	0	0	6,590	18,986	0	25,576
Cohort 6 - May 2018	0	0	0	0	0	0	7,535	2,595	10,130

Notes: J indicates pre-intervention months, K+1 indicates post-intervention months. For cohorts one through five, J=36 and K+1=12, except for cohort six where J=36 and K+1=11. * Data start from February 2016 for one site in cohort 2.

Table A5: Descriptive statistics

	Intervention	Controls (not yet treated and never treated)	t test (p-value)
N observations	114,563	1,305,135	
N sites	66	248	
Outcomes			
LoS (days)	15.57 (18.59)	16.33 (19.85)	0.000
In-hospital mortality (%)	4.43 (20.57)	4.95 (21.70)	0.000
Institutionalisation (%)	2.13 (14.45)	2.73 (16.30)	0.000
30-day readmission (%)	17.49 (37.99)	17.24 (37.77)	0.035
Covariates			
Age	86.20 (5.72)	86.13 (5.69)	0.000
Age 75-80 (%)	14.11 (34.81)	14.20 (34.91)	0.368
Age 80-85 (%)	25.15 (43.39)	25.66 (43.67)	0.000
Age 85-90 (%)	31.12 (46.30)	30.99 (46.24)	0.334
Age 90-95 (%)	21.86 (41.33)	21.88 (41.34)	0.905
Age 95+ (%)	7.76 (26.75)	7.28 (25.97)	0.000
Female (%)	60.09 (48.97)	60.71 (48.84)	0.000
HFRS high risk score	23.04 (6.53)	22.43 (6.15)	0.000
Charlson=0 (%)	9.58 (29.43)	10.67 (30.87)	0.000
Charlson=1 (%)	23.81 (42.59))	25.33 (43.49)	0.000
Charlson=2 (%)	22.08 (41.48)	22.03 (41.44)	0.683
Charlson=3+ (%)	44.53 (49.70)	41.98 (49.35)	0.000
N prev adm=0 (%)	31.33 (46.38)	29.78 (45.73)	0.000
N prev adm=1 (%)	32.98 (47.01)	33.39 (47.16)	0.005
N prev adm=2 (%)	19.09 (39.30)	19.94 (39.96)	0.000
N prev adm=3+ (%)	16.61 (37.22)	16.89 (37.46)	0.016
N diagnoses	12.79 (4.73)	11.92 (4.54)	0.000

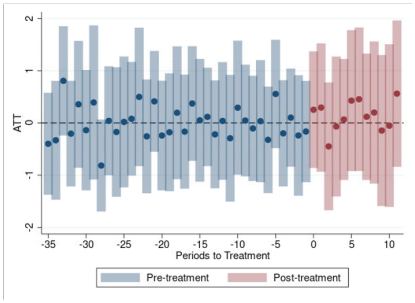
Notes: Mean (std dev). Years: 2012-2019. For intervention, J=36, K+1=12 where J indicates pre-intervention months, K+1 indicates post-intervention months. N: Number, LoS: Length of stay, HFRS: Hospital Frailty Risk Score; N prev adm: number of previous admissions.

Table A6: Callaway & Sant'Anna results (J=36, K+1=12)

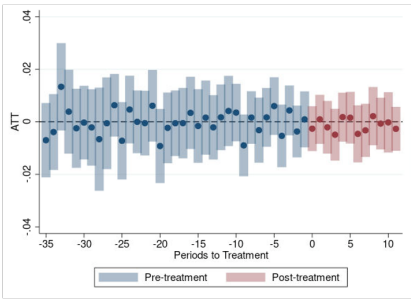
	(1) LoS	(2) In-hospital mortality	(3) Institutionalisation	(4) 30-day readmission
ATT	0.135	-0.001	0.001	-0.003
	(0.340)	(0.002)	(0.002)	(0.005)
ATT cohort 1	-0.311	0.002	-0.002	-0.003
	(0.894)	(0.005)	(0.005)	(0.008)
ATT cohort 2	0.760	0.001	0.001	-0.013*
	(0.881)	(0.005)	(0.003)	(0.006)
ATT cohort 3	-0.814	-0.002	-0.0001	0.001
	(0.933)	(0.005)	(0.004)	(0.012)
ATT cohort 4	0.866	-0.005	0.003	0.0003
	(0.591)	(0.005)	(0.003)	(0.016)
ATT cohort 5	0.030	-0.006	0.001	0.009
	(0.506)	(0.003)	(0.004)	(0.008)
ATT cohort 6	-0.039	0.009	0.003	-0.032
	(0.945)	(0.007)	(0.003)	(0.021)
N observations	1,419,048	1,419,050	1,419,050	1,400,741
N sites	249	249	249	249
Covariates	Yes	Yes	Yes	Yes
Cluster sites	Yes	Yes	Yes	Yes
Controls	never + not yet	never + not yet	never + not yet	never + not yet
Pre/Post Months	-36/+12	-36/+12	-36/+12	-36/+12
Years	2012-2019	2012-2019	2012-2019	2012-2019

Notes: ATT: average treatment effect on the treated. Coefficients reported as average marginal effects and standard errors reported in brackets. Significance levels: *** p<0.001, ** p<0.01, * p<0.05, no star = not significant. N: Number. Covariates: age, female, HFRS high risk score, Charlson Comorbodity Index, Number of previous admissions, number of unique diagnoses and died (except for model 2). J indicates pre-intervention months, K+1 indicates post-intervention months.

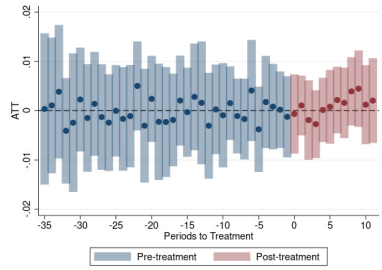
Figure A3: Callaway & Sant'Anna figures (J=36, K+1=12)



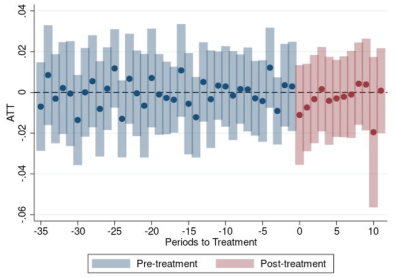
(I) - LoS



(II) - In-hospital mortality



(III) - Institutionalisation



(IV) - 30-day readmission

Notes: The dot represents the mean conditional outcome for those subject to intervention relative to the controls, with the length of the bars indicating the 95% confidence limits. J indicates pre-intervention months, K+1 indicates post-intervention months.

E Robustness check - Restricting controls to those from "Never treated" sites

In this robustness check we confined our selection of controls to "never treated" patients cared for in sites that never became part of the AFN. This reduced the number of control patients from 1,305,135 in the main analysis to 968,900 for this analysis (Table A7), the 336,235 "not-yet-treated" patients having been dropped. Descriptive statistics for this newly-defined control group are presented in Table A8 alongside statistics for patients subject to the intervention, unchanged from the main analysis.

Table A9 and Figure A4 show that the overall effects of AFN were not significant for any of the four outcomes. Nor were there any significant effects for any of the six cohorts, as the lower panel of Table A9 shows.

Table A7: Number of Control and Intervention patients (J=36, K+1=11)

	2012	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**Total**
Controls (never treated)	91,481	108,395	119,770	128,638	136,944	157,313	179,988	46,371	968,900
Intervention	0	0	0	10,966	29,962	31,778	29,991	2,595	105,292
Cohort 1 - Jan 2015	0	0	0	10,966	0	0	0	0	10,966
Cohort 2 - Jan 2016	0	0	0	0	23,122	0	0	0	23,122
Cohort 3 - Sept 2016	0	0	0	0	6,840	12,401	0	0	19,241
Cohort 4 - May 2017	0	0	0	0	0	12,787	5,500	0	18,287
Cohort 5 - Oct 2017	0	0	0	0	0	6,590	16,956	0	23,546
Cohort 6 - May 2018	0	0	0	0	0	0	7,535	2,595	10,130

Notes: J indicates pre-intervention months, K+1 indicates post-intervention months.

Table A8: Descriptive statistics

	Intervention	Controls (never treated)	t test (p-value)
N observations	105,292	968,900	
N sites	66	183	
Outcomes			
LoS (days)	15.57 (18.57)	16.04 (19.74)	0.000
In-hospital mortality (%)	4.44 (20.61)	4.91 (21.62)	0.000
Institutionalisation (%)	2.15 (14.50)	2.81 (16.53)	0.000
30-day readmission (%)	17.46 (37.96)	17.44 (37.9)	0.890
Covariates			
Age	86.20 (5.72)	86.10 (5.69)	0.000
Age 75-80 (%)	14.18 (34.88)	14.25 (34.96)	0.500
Age 80-85 (%)	25.07 (43.34)	25.79 (43.75)	0.000
Age 85-90 (%)	31.17 (46.32)	31.00 (46.25)	0.257
Age 90-95 (%)	21.88 (41.34)	21.69 (41.21)	0.150
Age 95+ (%)	7.71 (26.67)	7.27 (25.97)	0.000
Female (%)	60.09 (48.97)	60.67 (48.85)	0.000
HFRS high risk score	23.03 (6.52)	22.40 (6.13)	0.000
Charlson=0 (%)	9.56 (29.41)	10.75 (30.97)	0.000
Charlson=1 (%)	23.85 (42.61)	25.24 (43.44)	0.000
Charlson=2 (%)	22.08 (41.48)	21.95 (41.39)	0.361
Charlson=3+ (%)	44.51 (49.70)	42.05 (49.36)	0.000
N prev adm=0 (%)	31.36 (46.40)	29.52 (45.61)	0.000
N prev adm=1 (%)	32.96 (47.01)	33.34 (47.14)	0.012
N prev adm=2 (%)	19.07 (39.29)	19.99 (39.99)	0.000
N prev adm=3+ (%)	16.61 (37.22)	17.15 (37.69)	0.000
N diagnoses	12.79 (4.73)	11.92 (4.55)	0.000

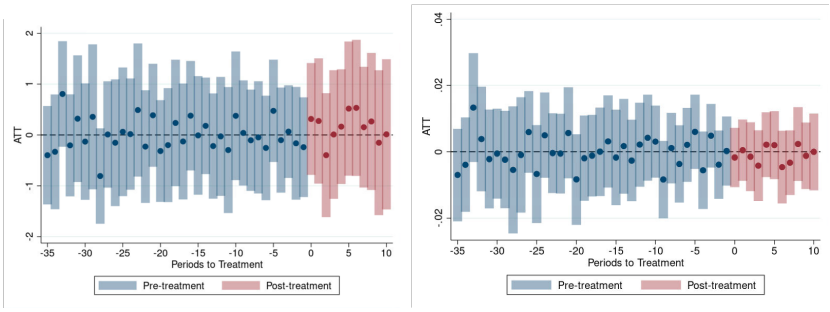
Notes: Mean (std dev). Years: 2012-2019. For intervention, J=36, K+1=11 where J indicates pre-intervention months, K+1 indicates post-intervention months. N: Number, LoS: Length of stay, HFRS: Hospital Frailty Risk Score; N prev adm: number of previous admissions.

Table A9: Callaway & Sant'Anna results (J=36, K+1=11) - never treated

	(1) LoS	(2) In-hospital mortality	(3) Institutionalisation	(4) 30-day readmission
ATT	0.153	-0.001	0.001	-0.003
	(0.343)	(0.002)	(0.002)	(0.005)
ATT cohort 1	-0.442	0.004	-0.002	-0.002
	(0.917)	(0.005)	(0.005)	(0.007)
ATT cohort 2	0.913	0.001	-0.0001	-0.013
	(0.887)	(0.005)	(0.003)	(0.007)
ATT cohort 3	-0.820	-0.002	-0.001	0.003
	(0.938)	(0.005)	(0.004)	(0.012)
ATT cohort 4	0.888	-0.005	0.003	0.001
	(0.620)	(0.005)	(0.003)	(0.016)
ATT cohort 5	0.022	-0.005	0.001	0.008
	(0.485)	(0.004)	(0.004)	(0.008)
ATT cohort 6	-0.039	0.009	0.003	-0.032
	(0.945)	(0.007)	(0.003)	(0.021)
N observations	1,074,184	1,074,186	1,074,186	1,055,897
N sites	249	249	249	249
Covariates	Yes	Yes	Yes	Yes
Cluster sites	Yes	Yes	Yes	Yes
Controls	never treated	never treated	never treated	never treated
Months	-36/+11	-36/+11	-36/+11	-36/+11
Years	2012-2019	2012-2019	2012-2019	2012-2019

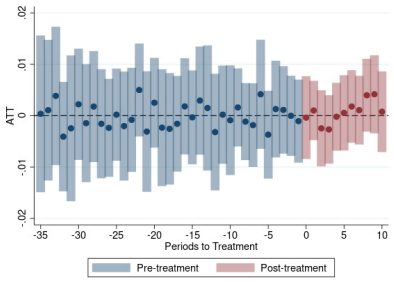
Notes: ATT: average treatment effect on the treated. Coefficients reported as average marginal effects and standard errors reported in brackets. Significance levels: *** p<0.001, ** p<0.01, * p<0.05, no star = not significant. N: Number. Covariates: age, female, HFRS high risk score, Charlson Comorbodity Index, Number of previous admissions, number of unique diagnoses and died (except for model 2). J indicates pre-intervention months, K+1 indicates post-intervention months.

Figure A4: Callaway & Sant'Anna figures (J=36, K+1=11) - never treated

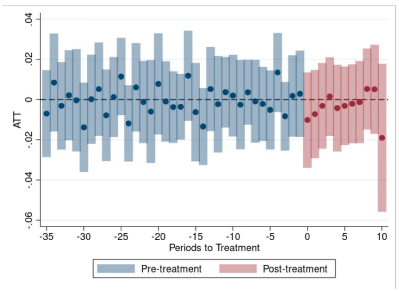


(I) - LoS

(II) - In-hospital mortality



(III) - Institutionalisation



(IV) - 30-day readmission

Notes: The dot represents the mean conditional outcome for those subject to intervention relative to the controls, with the length of the bars indicating the 95% confidence limits. J indicates pre-intervention months, K+1 indicates post-intervention months.

F Robustness check - active engagement

Our final robustness checks repeated the main analysis, but restricted intervention sites to those considered to have actively engaged with the AFN. We used two approaches to identify sites that had most actively engaged with the AFN, both relating to their adoption of the best practice principles set out in Table 1. The first focused on sites that adopted four or more of the principles upon joining. The second definition looked at each principle in turn, with "active" sites adopting that specific principle upon joining.

On average AFN sites adopted four or more of the best practice principles within 12 months of joining, as shown in Table A10, with notable variation across cohorts (information was unavailable for sites in cohort one). By imposing this definition of active engagement, the number of sites considered to have been subject to the AFN intervention was reduced from 66 to 27. To implement this restriction, data relating to the other 39 AFN sites were dropped from the analysis. This reduced (i) the number of patients considered to have been subject to the intervention from 105,292 to 48,372 and (ii) the number of control patients in the "not-yet-treated" group from 336,235 to 115,585 (see Table A11). Descriptive statistics for these smaller intervention and control groups are reported in Table A12.

The overall effects of AFN membership for three of the outcomes remained non-significant, but a higher percentage of patients were discharged to a care or residential home (0.004, $p<0.05$), as reported in the ATT row of Table A13 and Figure A5.

There were, however, some significant cohort effects for those sites that met the inclusion threshold for this sub-group analysis. A significantly lower percentage of patients in the three threshold sites from cohort two were readmitted within 30 days (-0.024, $p<0.001$). Patients in the four threshold sites in cohort three had a significantly longer LoS (1.75 days, $p<0.05$) and a higher percentage were discharged to a care or residential home (0.011, $p<0.001$). Patients in the three threshold sites from cohort six had a significantly shorter LoS (6.579 days, $p<0.001$) and a higher percentage were discharged to a care or residential home (0.014, $p<0.001$).

The second form of this robustness check considered adoption of each best practice principle (BPP) in turn. In this set of ten analyses, only those sites that adopted the specific principle were considered to have been intervention sites. This meant dropping from the analyses those AFN sites that had already adopted the principle prior to joining the AFN and those sites that did not adopt the principle upon joining. Table A14 reports the results: there were no statistically significant effects on the four outcomes associated with adoption of any one of these ten principles.

Table A10: Most engaged sites

	Did the site pass the 4+ threshold? No	Yes
Cohort 1	not known	not known
Cohort 2	9	3
Cohort 3	9	4
Cohort 4	4	7
Cohort 5	2	10
Cohort 6	5	3

Table A11: Number of Control and Intervention patients (J=36, K+1=11) - Most engaged sites

	2012	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**Total**
Controls (never treated)	91,481	108,395	119,770	128,638	136,944	157,313	179,988	46,371	968,900
Controls (not yet treated)	0	7,247	20,051	36,842	31,650	18,979	816	0	115,585
Intervention	0	0	0	0	10,643	17,208	19,866	655	48,372
Cohort 1 - Jan 2015	0	0	0	0	0	0	0	0	0
Cohort 2 - Jan 2016	0	0	0	0	8,597	0	0	0	8,597
Cohort 3 - Sept 2016	0	0	0	0	2,046	3,608	0	0	5,654
Cohort 4 - May 2017	0	0	0	0	0	8,152	3,586	0	11,738
Cohort 5 - Oct 2017	0	0	0	0	0	5,448	14,495	0	19,943
Cohort 6 - May 2018	0	0	0	0	0	0	1,785	655	2,440

Notes: J indicates pre-intervention months, K+1 indicates post-intervention months.

Table A12: Descriptive statistics - Most engaged sites

	Intervention	Controls (not yet treated and never treated)	t test (p-value)
N observations	48,372	1,084,485	
N sites	27	210	
Outcomes			
LoS (days)	15.48 (18.30)	16.17 (19.77)	0.000
In-hospital mortality (%)	4.05 (19.71)	4.88 (21.54)	0.000
Institutionalisation (%)	1.91 (13.69)	2.78 (16.45)	0.000
30-day readmission (%)	17.89 (38.32)	17.36 (37.87)	0.003
Covariates			
Age	86.23 (5.72)	86.11 (5.69)	0.000
Age 75-80 (%)	14.04 (34.74)	14.23 (34.93)	0.243
Age 80-85 (%)	24.99 (43.29)	25.75 (43.73)	0.000
Age 85-90 (%)	31.25 (46.35)	31.01 (46.25)	0.267
Age 90-95 (%)	21.88 (41.34)	21.74 (41.25)	0.462
Age 95+ (%)	7.85 (26.89)	7.27 (25.96)	0.000
Female (%)	60.03 (48.98)	60.64 (48.85)	0.007
HFRS high risk score	23.37 (6.73)	22.42 (6.15)	0.000
Charlson=0 (%)	9.79 (29.72)	10.77 (31.00)	0.000
Charlson=1 (%)	23.92 (42.66)	25.27 (43.46)	0.000
Charlson=2 (%)	22.15 (41.52)	21.96 (43.40)	0.333
Charlson=3+ (%)	44.14 (49.66)	42.00 (49.35)	0.000
N prev adm=0 (%)	31.67 (46.51)	29.68 (45.69)	0.000
N prev adm=1 (%)	32.77 (46.94)	33.36 (47.15)	0.008
N prev adm=2 (%)	18.95 (39.19)	19.96 (39.97)	0.000
N prev adm=3+ (%)	16.60 (37.21)	17.01 (37.57)	0.020
N diagnoses	13.26 (4.82)	11.95 (4.57)	0.000

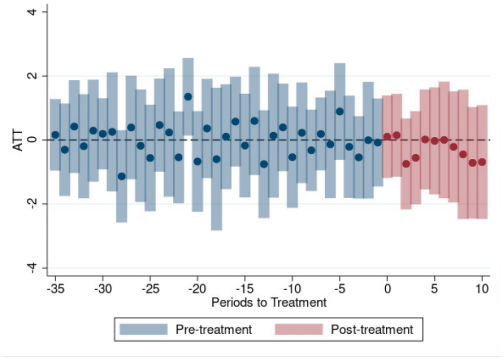
Notes: Mean (std dev). Years: 2012-2019. For intervention, J=36, K+1=11 where J indicates pre-intervention months, K+1 indicates post-intervention months. N: Number, LoS: Length of stay, HFRS: Hospital Frailty Risk Score; N prev adm: number of previous admissions.

Table A13: Callaway & Sant'Anna results (J=36, K+1=11) - Most engaged sites

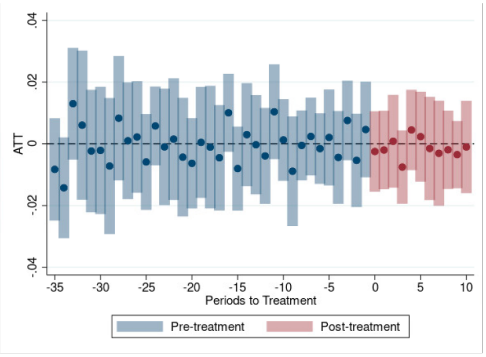
	(1) LoS	(2) In-hospital mortality	(3) Institutionalisation	(4) 30-day readmission
ATT	-0.291	-0.001	0.004*	-0.001
	(0.460)	(0.003)	(0.002)	(0.008)
ATT cohort 2	-0.410	0.008	0.005	-0.024***
	(0.541)	(0.007)	(0.003)	(0.008)
ATT cohort 3	1.750*	0.008	0.011***	-0.025
	(0.831)	(0.004)	(0.002)	(0.021)
ATT cohort 4	0.220	-0.009	0.004	0.010
	(0.691)	(0.006)	(0.003)	(0.016)
ATT cohort 5	-0.409	-0.004	0.001	0.012
	(0.579)	(0.004)	(0.003)	(0.010)
ATT cohort 6	-6.579***	0.0001	0.014***	-0.005
	(1.338)	(0.008)	(0.003)	(0.011)
N observations	1,041,367	1,041,369	1,041,369	1,023,798
N sites	210	210	210	210
Covariates	Yes	Yes	Yes	Yes
Cluster sites	Yes	Yes	Yes	Yes
Controls	never + not yet	never + not yet	never + not yet	never + not yet
Months	-36/+11	-36/+11	-36/+11	-36/+11
Years	2012-2019	2012-2019	2012-2019	2012-2019

Notes: ATT: average treatment effect on the treated. Coefficients reported as average marginal effects and standard errors reported in brackets. Significance levels: *** p<0.001, ** p<0.01, * p<0.05, no star = not significant. N: Number. Covariates: age, female, HFRS high risk score, Charlson Comorbodity Index, Number of previous admissions, number of unique diagnoses and died (except for model 2). J indicates pre-intervention months, K+1 indicates post-intervention months.

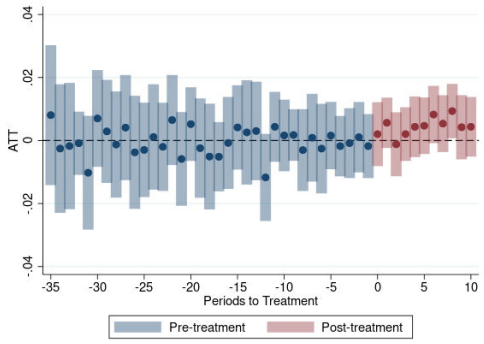
Figure A5: Callaway & Sant'Anna figures (J=36, K+1=11) - Most engaged sites



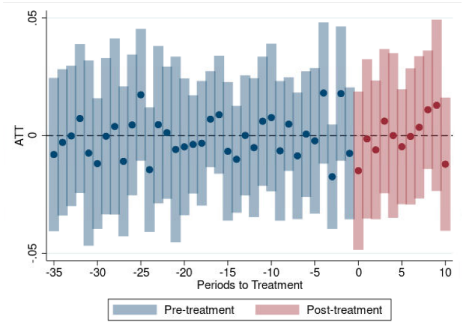
(I) - LoS



(II) - In-hospital mortality



(III) - Institutionalisation



(IV) - 30-day readmission

Notes: The dot represents the mean conditional outcome for those subject to intervention relative to the controls, with the length of the bars indicating the 95% confidence limits. J indicates pre-intervention months, K+1 indicates post-intervention months.

Table A14: Callaway & Sant'Anna results (J=36, K+1=11) - adoption of each BPP

	(1) LoS	(2) In-hospital mortality	(3) Institutionalisation	(4) 30-day readmission
ATT BPP1	0.416	-0.001	0.002	-0.001
	(0.507)	(0.003)	(0.002)	(0.008)
N observations	1,056,105	1,056,107	1,056,107	1,038,599
N sites	214	214	214	214
ATT BPP2	-0.366	-0.001	0.001	-0.007
	(0.495)	(0.003)	(0.003)	(0.008)
N observations	955,875	955,877	955,877	938,348
N sites	199	199	199	199
ATT BPP3	0.079	-0.009	0.003	-0.003
	(0.771)	(0.005)	(0.003)	(0.017)
N observations	817,932	817,934	817,934	817,930
N sites	193	193	193	193
ATT BPP4	-0.241	0.003	0.004	-0.005
	(0.514)	(0.004)	(0.002)	(0.009)
N observations	977,637	977,639	977,639	960,242
N sites	200	200	200	200
ATT BPP5	-0.546	-0.002	0.001	0.003
	(0.481)	(0.003)	(0.003)	(0.008)
N observations	1,030,053	1,030,055	1,030,055	1,012,484
N sites	209	209	209	209
ATT BPP6	0.154	-0.005	0.0003	0.008
	(0.783)	(0.004)	(0.004)	(0.011)
N observations	959,381	959,383	959,383	941,812
N sites	202	202	202	202
ATT BPP7	-0.216	0.002	0.003	-0.006
	(0.395)	(0.003)	(0.002)	(0.009)
N observations	1,031,328	1,031,330	1,031,330	1,013,478
N sites	207	207	207	207
ATT BPP8	-0.369	-0.002	0.001	0.010
	(0.565)	(0.004)	(0.002)	(0.011)
N observations	909,880	909,882	909,882	892,237
N sites	201	201	201	201
ATT BPP9	0.513	-0.004	0.002	0.007
	(0.415)	(0.004)	(0.002)	(0.009)
N observations	993,694	993,696	993,696	976,264
N sites	200	200	200	200
ATT BPP10	-0.369	-0.005	0.002	0.007
	(0.461)	(0.004)	(0.002)	(0.008)
N observations	1,011,229	1,011,231	1,011,231	993,799
N sites	202	202	202	202
Covariates	Yes	Yes	Yes	Yes
Cluster sites	Yes	Yes	Yes	Yes
Controls	never + not yet	never + not yet	never + not yet	never + not yet
Treated leads/lags	-36/+11	-36/+11	-36/+11	-36/+11
Years	2012-2019	2012-2019	2012-2019	2012-2019

Notes: ATT: average treatment effect on the treated. BPP: Best practice principle. Coefficients reported as average marginal effects and standard errors reported in brackets. Significance levels: *** p<0.001, ** p<0.01, * p<0.05, no star = not significant. N: Number. Covariates: age, female, HFRS high risk score, Charlson Comorbodity Index, Number of previous admissions, number of unique diagnoses and died (except for model 2). J indicates pre-intervention months, K+1 indicates post-intervention months.

References

[1] Clarke D, Tapia-Schythe K. Implementing the panel event study. Stata Journal. 2021;21(4):853-84.

[2] Callaway B, Sant'Anna PH. Difference-in-Differences with multiple time periods. Journal of Econometrics. 2021;225(2):200-30.

[3] Baker AC, Larcker DF, Wang CC. How much should we trust staggered difference-in-differences estimates? Journal of Financial Economics. 2022;144(2):370-95.

[4] NHS Digital. Hospital Episode Statistics Data Dictionary; 2022. [accessed: 02.03.2023]. https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics/hospital-episode-statistics-data-dictionary.

[5] Gilbert T, Neuburger J, Kraindler J, Keeble E, Smith P, Ariti C, et al. Development and validation of a Hospital Frailty Risk Score focusing on older people in acute care settings using electronic hospital records: an observational study. Lancet. 2018;391:1775-82.

[6] Street A, Maynou L, Gilbert T, Stone T, Mason S, Conroy S. The use of linked routine data to optimise calculation of the Hospital Frailty Risk Score on the basis of previous hospital admissions: a retrospective observational cohort study. The Lancet Healthy Longevity. 2021;2(3):e154-62.

[7] Ministry of Housing, Communities and Local Government. The English Indices of Deprivation 2019; 2019. [accessed: 02.03.2023]. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/835115/IoD2019_Statistical_Release.pdf.

[8] Charlson M, Pompei P, Ales K, MacKenzie C. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987;40(5):373-83.

[9] Quan H, Sundararajan V, Halfon P, Fong A, Burnand B, Luthi JC, et al. Coding Algorithms for Defining Comorbidities in ICD-9-CM and ICD-10 Administrative Data. Medical Care. 2005;43(11):1130-9.

[10] Stagg V. Charlson: Stata module to calculate Charlson index of comorbidity. Statistical Software Components, Boston College Department of Economics. 2017;S456719.