Technical appendix: The impact of redesigning urgent and emergency care in Northumberland

Stephen O'Neill, Arne Wolters, Adam Steventon

About this technical appendix

This technical appendix provides supplementary information relating to analysis conducted by the Improvement Analytics Unit, a partnership between NHS England and the Health Foundation. It supports the Health Foundation's briefing that considers the findings of the analysis.

The appendix focuses in particular on the following elements of the study:

- the risk-adjustment undertaken
- the synthetic control method
- the sensitivity analyses performed
- the limitations of the study.

The briefing is available from: www.health.org.uk/impact-redesign-care-Northumberland

The Health Foundation

Identifying 'comparable' control areas

To enable the Improvement Analytics Unit to compare Northumberland Clinical Commissioing Group (CCG) with control areas, the unit began by selecting the 20 CCGs in England that were most similar to Northumberland. There are 209 CCGs in England, but the unit initially excluded 30 CCGs in London and 59 CCGs participating in new care models vanguards from the potential list of comparable control areas. This left 120 CCGs, which the unit characterised in terms of variables such as the number of general practitioners per capita and the prevalence of common diseases – these were obtained from a range of sources (see Box 1). The unit then sought to exclude CCGs whose populations were too dissimilar to the population for which Northumberland CCG is responsible.

To determine how similar each CCG was to Northumberland CCG, the unit adapted the method used in NHS England's Commissioning for Value tool.* That method assessed similarity in terms of age structure, ethnic mix and deprivation, to which the unit added variables of relevance that were publicly available at CCG level or that could be derived from Secondary Uses Service (SUS) data, such as pre-intervention hospital utilisation (see Box 1). Where available, annual data for each CCG for the financial years 2011/12, 2012/13, 2013/14 and 2014/15 were included.

Similarity was assessed by calculating the squared Euclidean distance (SED) from each CCG to Northumberland CCG across these variables, with a lower SED indicating greater similarity to Northumberland CCG.† Since the variables were measured on different scales and hence were not directly comparable, the Improvement Analytics Unit standardised the data using inter-decile range standardisation‡ as used by the Office for National Statistics and in the Commissioning for Value tool, prior to calculating the SED.[12,13] The unit also weighted the variables according to how predictive they were of the rate of hospital admissions in 2014/15 (controlling for the other variables) since many variables were included, some of which may be closely related to each other. The weight given to each variable was determined by its squared standardised coefficient in a regression of the rate of hospital admissions in 2015 on the variables for the preceding year.§ Variables that received greater weight included the Quality and Outcomes Framework (QOF) achievement scores and the past rates of elective and emergency admissions.

To incorporate multiple periods within the similarity measure, the Improvement Analytics Unit calculated the geometric mean of the SED for each CCG from 2011/12 to 2014/15. This was then used as the overall measure of similarity over the pre-intervention period. This procedure aimed to ensure that the 'most similar' CCGs were similar to Northumberland CCG across all of the pre-intervention years included in the study. The final 20 CCGs included in the donor pool are illustrated in Figure 2 in the briefing.

---

* The Commissioning for Value tool is used to determine the 10 most similar CCGs in England for a given CCG. See www.england.nhs.uk/rightcare/intel/support

† Here the SED for CCG j across the 56 variables is calculated as: $SED_j = \sqrt{\sum_{k=1}^{56}\left(x_{k,j} - x_{k,Northumberland}\right)^2}$

‡ The standardised value of a variable is calculated as $\frac{x_i - median(x)}{90th\ percentile(x) - 90th\ percentile(x)}$ where i is the subject in question.

§ Alternative methods such as those proposed by Lindeman, Merenda and Gold[14] and the newly proposed method by Feldman[15] are computationally-intensive making them undesirable in contexts with a large number of variables as is the case here.

Box 1: Variables relating to the characteristics of CCGs

- The age profile of the local population – the proportion of the registered patients who fall into the following age groups: <5, 5-14, 15-24 and 75+[1]
- Proportion of men – the proportion of the registered patients who are men[1]
- Proportion of people by ethnicity – proportion of the resident population who fall into the following ethnic groups: white, black, Asian, mixed and other[2]
- Proportion of people with third-level education or higher – proportion of the resident population who have two or more A-levels or equivalent qualifications, or higher[2]
- Population density[2]
- Socioeconomic deprivation – average of Lower Super Output Area (LSOA) level Index of Multiple Deprivation (IMD) scores over all LSOAs in a CCG[3]
- Health deprivation – average of LSOA level IMD scores on health deprivation over all LSOAs in a CCG[4]
- Number of general practitioners – number of general practitioners (full-time equivalent) per 1,000 people in the resident population[5]
- Number of care home beds – number of care home beds (residential and nursing) per 1,000 people in the resident population[6]
- Quality and Outcomes Framework (QOF) achievement scores – total of achievement scores on all QOF indicators across cardiovascular, respiratory, high dependency and other long-term conditions, and musculoskeletal QOF indicator groups[7]
- Disease prevalence – for atrial fibrillation, coronary heart disease, cardiovascular disease, heart failure, hypertension, peripheral arterial disease, stroke and transient ischaemic attack, asthma, COPD, cancer, chronic kidney disease (18+ only), diabetes, palliative care, osteoporosis (50+ only) and rheumatoid arthritis (16+)[6]
- Elixhauser comorbidity score – average number of comorbidities defined by Elixhauser et al,[8] per 1,000 of the resident population[9]
- Rates of admissions by clinical area – number of hospital admissions for each of the following primary diagnoses: cerebrovascular, cardiovascular, respiratory, gastrointestinal, musculoskeletal, renal, per 10,000 of the resident population[10]
- Rates of elective admissions, emergency admissions and outpatient attendances – number of elective hospital admissions, emergency hospital admissions and outpatient attendances, per 1,000 of the resident population.[11]

Comparison of Northumberland CCG, the donor pool and the rest of England

To assess whether the donor pool was more similar to Northumberland CCG than the full set of CCGs, the Improvement Analytics Unit compared the average value of each variable between 2011/12 and 2014/15 for the following:

- Northumberland CCG
- the 20 CCGs in the donor pool
- the remaining 188 CCGs in England.

As shown in Table 1, the 20 CCGs included in the donor pool were more similar to Northumberland CCG than those in the rest of England for 50 of the 56 variables.

Table 1: The characteristics of Northumberland CCG (2011–14) compared with the average for the 20 most similar CCGs and the remaining CCGs in England

Variable	Northumberland	Mean for donor CCGs	Mean for excluded CCGs	Donor more similar than excluded CCGs?
Index of Multiple Deprivation for health domain	0.20	0.16	0.003	Yes
Index of Multiple Deprivation score	20.53	23.76	21.72	No
Percent of population aged under 5	4.97	6.30	6.32	Yes
Percent of population aged 5–14	10.52	11.63	11.50	No
Percent of population 15–24	10.54	12.38	12.49	Yes
Percent of population aged 75 and older	9.88	8.54	7.98	Yes
Percent of population that is male	49.09	49.71	49.77	Yes
Population density	0.60	13.14	21.66	Yes
Black ethnicity (proportion)	0.10	1.82	3.45	Yes
Asian ethnicity (proportion)	0.80	7.24	7.79	Yes
Mixed ethnicity (proportion)	0.50	1.88	2.23	Yes
Other ethnicity (proportion)	0.10	0.63	1.06	Yes
Percent of population with third-level qualification	12.10	12.08	12.28	Yes
Number of full-time equivalent GPs per 1,000 population	0.67	0.57	0.59	No
Number of social beds per 1,000 population	11.92	9.67	8.37	Yes
QOF achievement scores				
Atrial fibrillation	764.24	735.64	611.42	Yes
Blood pressure	665.19	645.65	536.48	Yes
Secondary prevention of coronary heart disease	1,972.04	1,872.70	1,560.42	Yes
Cardiovascular disease – primary prevention	410.00	392.76	320.11	Yes
Heart failure	1,291.92	1,247.66	1,035.99	Yes
Hypertension	1,166.51	1,113.60	928.41	Yes
Peripheral arterial disease	267.68	254.24	212.11	Yes
Stroke/transient ischaemic attack	664.05	636.76	528.80	Yes
Asthma	2,010.86	1,922.88	1,599.04	Yes
COPD	1,548.95	1,478.28	1,225.81	Yes
Cancer	494.13	472.10	393.28	Yes
Chronic kidney disease	1,400.42	1,331.55	1,105.57	Yes

Variable	Northumberland	Mean for donor CCGs	Mean for excluded CCGs	Donor more similar than excluded CCGs?
Diabetes mellitus	3,675.33	3,316.67	2,802.20	Yes
Palliative care	267.00	257.10	213.54	Yes
Osteoporosis	378.00	321.51	266.76	Yes
Rheumatoid arthritis	257.74	249.86	209.40	Yes
Disease prevalence				
Atrial fibrillation	2.10	1.72	1.64	Yes
Coronary heart disease	4.73	3.54	3.27	Yes
Cardiovascular disease	0.59	0.61	0.60	No
Heart failure	0.99	0.78	0.73	Yes
Hypertension	17.21	14.67	13.84	Yes
Peripheral arterial disease	0.93	0.70	0.63	Yes
Stroke and transient ischaemic attack	2.44	1.89	1.73	Yes
Asthma	6.93	6.34	5.97	Yes
COPD	2.55	2.03	1.85	Yes
Cancer	2.92	2.34	2.26	Yes
Chronic kidney disease	4.91	3.70	3.28	Yes
Diabetes	6.02	5.46	5.12	Yes
Palliative care	0.56	0.32	0.31	Yes
Osteoporosis	0.10	0.06	0.06	Yes
Rheumatoid arthritis	0.94	0.65	0.60	Yes
Average number of comorbidities defined by Elixhauser et al (1998), per 1,000 of the resident population	132.51	102.68	96.52	Yes
Rate of cerebrovascular admissions	24.13	20.76	19.60	Yes
Rate of cardiovascular admissions	214.44	179.64	172.32	Yes
Rate of respiratory admissions	178.10	178.70	160.30	Yes
Rate of gastrointestinal admissions	446.03	354.39	338.05	Yes
Rate of muscoskeletal admissions	305.11	242.07	232.08	Yes
Rate of renal admissions	204.57	344.71	302.69	No
Rate of elective admissions	194.63	171.41	168.29	Yes
Rate of emergency admissions	115.69	103.05	96.40	Yes
Rate of outpatient admissions	1,328.07	1,496.89	1,397.20	No

Obtaining person-level data on hospital use

The remaining parts of the study used data from the Secondary Uses Service (SUS) – a national, person-level database that is closely related to the widely used hospital episode statistics (HES). The Improvement Analytics Unit has access to these data for its work, and processes them in a secure environment based at the Health Foundation. All data are pseudonymised, meaning that they have been stripped of fields that can directly identify a patient, such as name, full date of birth and address. The NHS number was replaced with a pseudonym, which the unit used to link records for the same individual over time. The overall approach to information governance has been scrutinised by the programme oversight group and by information governance experts at NHS Digital.

For this study, the Improvement Analytics Unit used data on accident and emergency (A&E) visits and inpatient attendances from May 2011 to July 2016. A&E visits were excluded if they didn't have a pseudonymised NHS number, if the patient left before being seen or refused treatment, or if they were duplicates.

Inpatient data were structured into continuous inpatient spells (CIPS), which may consist of several consultant episodes (since patients may be under the care of multiple consultants during a hospital stay) and stays at several hospitals (if patients are transferred). The unit excluded spells that were missing a pseudonymised NHS number or admission date, or where the discharge date preceded the admission date due to data quality problems.

Finally, the unit excluded a small number of A&E visits and inpatient admissions where gender was given as other than male or female. Although these records were considered valid, they caused technical difficulties for the statistical modelling.

Risk-adjusting the impact metrics

Impact metrics were 'risk-adjusted' to take account of differences over time in patients' demographics, comorbidities and prior hospital use. Such differences, if ignored, would bias estimates of the effect of the intervention, since any change in the impact metrics due to these differences would be (wrongly) attributed to the intervention.

First, the Improvement Analytics Unit predicted the levels of hospital use that it would expect to observe for each hospital visit or spell based on: patient-level characteristics such as age, sex, ethnicity, existing health conditions and past hospital activity; the day and month the visit/spell commenced; and the CCG responsible for the patient's care. These predictions were then aggregated to obtain predicted impact metrics for the CCG each month. Finally, to obtain the risk-adjusted version of the impact metrics, we multiplied the ratio of observed to predicted impact metric by the average impact metric over the pre-intervention period. Risk-adjustment was not applied to the rate of A&E visits and the overall admission rate, since it was considered that the health profile of the population was unlikely to change significantly over the period involved. The following section provides further details of the risk-adjustment approach taken.

Obtaining predicted versions of the impact metrics

For each impact metric listed in column 1 of Table 2 (with the exceptions of the rate of A&E visits and the rate of admissions), we estimated a regression model that explained the corresponding visit- or CIPS-level version of the impact metric (in column 2) based on the observed covariates displayed in Box 2, such as the patients' demographics, comorbidities and prior hospital use. The regression models were estimated after excluding activity in the post-intervention period, or for patients for whom CCGs in London, or those taking part in other new care model vanguards,* were responsible.

For each CCG-level impact metric, the corresponding visit- or spell-level estimand and dependent variable for the regression model are shown in columns 2 and 3 of Table 2. Column 4 shows the sample for which each model was estimated. Logistic regression models were estimated where the dependent variable was binary. A&E visit lengths were modelled using a generalised linear model (GLM). For the other continuous variables, two-part models were used to allow for excess zero counts, consisting of a logistic model for whether the count was zero, and a GLM for the non-zero counts. The Improvement Analytics Unit used the modified Park test[16] to determine the appropriate distribution (Gaussian, Poisson, Gamma or Inverse Gaussian) for the GLMs. In all cases, a log link and Gamma distribution was found to provide the best fit to the data. The unit compared predicted and actual outcomes to assess whether the models provided a good fit to the data and by considering Nagelkerke's R2. The adequacy of the binary models was further assessed using Tjur's coefficient of discrimination, and the concordance statistic ('C-statistic').†

* This included primary and acute care systems, multispecialty community providers, enhanced health in care homes, urgent and emergency care, as well as CCGs linked to two acute care collaborations namely, the Salford and Wigan Foundation Chain and the Healthcare Group in Dartford and Gravesham.

† The C-statistic equals the area under the receiver operating characteristic (ROC) curve.

Box 2: Covariates included in risk-adjustment models

- Age categories (five-year bands, and 90+)
- Gender indicator
- Ethnicity indicators
- CCG indicators
- Month
- Day of the week
- Indicators for each of the Elixhauser comorbidity categories
- Indicator for dementia
- History of elective admissions
- History of emergency admissions in past 24 months
- History of A&E visits in past 24 months
- Indicators for primary diagnosis based on summary hospital-level mortality indictor categories* (excluded from A&E outcome models)
- Number of elective admissions in the preceding 24 months (continuous up to 30 and indicators for 30-60 and more than 60)
- Number of emergency admissions in the preceding 24 months (continuous up to 30 and indicators for 30-60 and more than 60)
- Number of A&E visits in the preceding 24 months (continuous up to 30 and indicators for 30-60 and more than 60)

After obtaining the regression models for the pre-intervention period, we predicted the visit- or spell-level version of the impact metric for each visit or CIPS in the pre-intervention period, and we also predicted the corresponding versions of the impact metrics for the (out-of-sample) post-intervention periods. The predictions were then aggregated to obtain a CCG-level impact metric for each month. The risk-adjusted, CCG-level impact metric for a given period t was calculated as:

$$Y^{RA}(i,t) = \bar{Y}_{All,pre\text{-}intervention} \times \left(\frac{obsY_{i,t}}{E(Y_{i,t})} \right)$$

where $\bar{Y}_{All,pre\text{-}intervention}$ represents the CCG-level impact metric (for example, average length of stay) calculated using data for the pre-intervention months for all CCGs (excluding London and other new care models); $obsY_{i,t}$ is the observed CCG-level impact metric for CCG i calculated using all spells in month t; and $E(Y_{i,t})$ is the corresponding expected value for this impact metric, calculated by aggregating the spell-level GLM model predictions.† The approach taken here is broadly similar to the approaches adopted by Byrne et al,[17] Birkmeyer et al,[18] Tsai et al,[19] Zuckerman et al,[20] and Anselmi et al.[21]

* Indicators for summary hospital-level mortality indictor categories with fewer than three cases per 10,000 population for elective or emergency admissions were excluded to ensure model convergence and prevent overfitting.

† The precise steps taken to calculate a CCG-level impact metric ($Y_{i,t}$) from the visit/CIPS impact metric ($y_{j,t}$) depended on the particular impact metric under consideration. For rates, we summed $y_{j,t}$ (for example, whether the patient was admitted) across all spells occurring in period t among the population of CCG i and divided this by the CCG population/10,000 to express it as rate per 10,000 population. For average length of stay, we summed the length of stay ($y_{j,t}$) for each spell in period t among the population of CCG and divided this by the relevant number of spells to obtain the average length of stay.

Table 2: Model selection and interpretation

Rate of A&E visits	-	-		Not risk-adjusted
Rate of all admissions	-	-		Not risk-adjusted
Rate of elective admissions	Expected probability of the admission being an elective admission	Binary indicator for whether the admission was an elective admission	All admitted patients	Logistic regression
Rate of emergency admissions	Expected probability of the admission being an emergency admission	Binary indicator for whether the admission was an emergency admission	All admitted patients	Logistic regression
Proportion of A&E visits leading to an admission	Expected probability of being admitted for patients that visited A&E	Binary indicator for whether the A&E visit ended with the patient being admitted	All A&E visitors	Logistic regression
Proportion of A&E visits in which patient is seen within four hours	Expected probability of being seen within 4 hours for patients visiting A&E	Binary indicator for whether the A&E visit ended with the patient being admitted, transferred or discharged within 4 hours of attending A&E	All A&E visitors	Logistic regression
Average A&E visit length	Expected A&E visit length for patients visiting A&E	Visit length in hours	All A&E visitors	GLM model with log link and gamma distribution
Average length of stay of all admissions	Expected length of stay for all admitted patients	Length of stay in bed days	All admitted patients	Two-part model: logistic regression + GLM model with log link and gamma distribution
Average length of stay of elective admissions	Expected length of stay for the subset of admitted patients with an elective admission	Length of stay in bed days	All admitted patients with an elective admission	Two-part model: logistic regression + GLM model with log link and gamma distribution
Average length of stay of emergency admissions	Expected length of stay for the subset of admitted patients with an emergency admission	Length of stay in bed days	All admitted patients with an emergency admission	Two-part model: logistic regression + GLM model with log link and gamma distribution

Synthetic controls

After obtaining the risk-adjusted impact metrics for each CCG by month, the Improvement Analytics Unit conducted an analysis to estimate the effects of the intervention using the synthetic control method.[22,23] For a particular impact metric, the synthetic control method finds a weighted average of the relevant metrics for the CCGs in the donor pool that closely tracks the metric for Northumberland CCG in the pre-intervention period. The post-intervention impact metric of this synthetic control area is then taken to represent the level that the unit would have expected to observe for Northumberland CCG in the absence of the intervention. The intervention effect is the difference between the actual impact metric for Northumberland CCG and the impact metric for the relevant synthetic control area.

For each impact metric, using data for k pre-intervention months, the synthetic control area was formed by finding the $(N_{controls} \times 1)$ vector of weights W^* that minimised $(X_1 - X_0 W)' V (X_1 - X_0 W)$ subject to the weights in W being positive and summing to 1, where X_1 and X_0 are $(k \times 1)$ and $(k \times N_{controls})$ matrices that contain the pre-intervention impact metrics for Northumberland CCG and the CCGs in the donor pool respectively, and V is chosen such that the mean squared prediction error of the impact metric variable is minimised for the pre-intervention periods. The optimal set of weights creates a synthetic control area which approximates the (risk-adjusted) versions of the impact metrics for Northumberland CCG in each pre-intervention period ($Y_{Northumberland,t}$):

$$\sum_{j \epsilon Control} w_j Y_{jt} = Y_{Northumberland,t} \quad \forall t \leq T_0 \qquad [Eq\ 1]$$

with $0 \leq w_j \leq 1$ and $\sum_{j \epsilon Donor\ pool} w_j = 1$.

If *Eq 1* holds for a sufficiently long period, it can be assumed that unobserved confounders and their potentially time-varying effects are also balanced between the synthetic control area and Northumberland CCG.[23] The effect of phase 1 of the primary and acute care systems (PACS) implementation can then be estimated by:

$$\hat{\tau} = Y_{Northumberland,t} - \sum_{j \epsilon Control} w_j Y_{jt}$$

The unit conducted a separate synthetic control analysis for each impact metric. For some metrics, it was not possible to find weights that provided a synthetic control area that closely tracked Northumberland CCG.* In these instances, the intervention effect could not be estimated reliably and the corresponding impact metrics were excluded from the remainder of the analysis. Table 3 displays the weights attached to each of the CCGs in the donor pool for the impact metrics for which it was possible to find a synthetic control that fitted adequately.

* These metrics were: the rate of elective admissions; the proportion of A&E visits leading to an admission; the average length of stay of elective admissions; and the average length of stay of emergency admissions.

Table 3: Synthetic control weights

Clinical commissioning group	Rate of A&E visits	Rate of all admissions	Rate of emergency admissions	Proportion of A&E visits in which patient is seen within 4 hours	Average A&E visit length	Average length of stay of all admissions
NHS Bolton	0	0	0.127	0	0	0
NHS Oldham	0	0	0	0.180	0	0
NHS Heywood, Middleton and Rochdale	0.330	0.465	0	0.341	0.054	0
NHS Tameside and Glossop	0	0	0	0.001	0	0
NHS Doncaster	0.390	0	0	0	0.057	0
NHS East Riding of Yorkshire	0	0	0	0	0	0
NHS Rotherham	0	0.380	0.240	0	0.004	0.045
NHS North Derbyshire	0.267	0	0	0.199	0.025	0.666
NHS Birmingham South and Central	0.011	0	0	0	0	0
NHS Shropshire	0	0	0	0	0	0
NHS Stoke-on-Trent	0	0	0	0	0.019	0.093
NHS Wolverhampton	0	0	0	0	0	0.182
NHS Ipswich and East Suffolk	0	0	0	0.001	0	0.012
NHS Mid Essex	0	0	0	0	0.386	0
NHS North East Essex	0	0	0.633	0	0	0
NHS West Essex	0	0	0	0.055	0.002	0
NHS Coastal West Sussex	0	0	0	0	0.389	0
NHS North West Surrey	0	0	0	0	0	0
NHS Bristol	0	0	0	0.139	0	0.002
NHS Basildon and Brentwood	0	0.155	0	0.082	0.061	0

Placebo tests

As explained in the briefing that this appendix accompanies,* the difference between the impact metrics in Northumberland and the relevant synthetic control area provides an estimate of the effect of the changes to urgent and emergency care. However, the precision of this estimate also needed to be assessed. This was important since outcomes vary over time even without changes to care delivery, and it would be misleading to attribute this normal statistical variation to the effect of the changes made. Traditionally, statisticians deal with this issue by reporting the 'p-value', which is the probability that an effect of at least the magnitude observed could have arisen by chance. If this probability is low (for example, less than 5%) then the findings are usually considered to represent a systematic difference between the two groups. However, the synthetic control approach does not lend itself to the calculation of p-values, or related quantities like confidence intervals, and so a different approach was needed.

The Improvement Analytics Unit's method relied on the calculation of placebo effects. These were obtained by repeating the analysis for each of the 20 CCGs in the donor pool – in each case by creating a synthetic area from the combination of the other 19 CCGs. When the CCGs were compared with their synthetic counterparts with respect to the impact metrics, the unit expected to find no systematic differences, since to its knowledge none of the CCGs in the donor pool introduced major changes to health care delivery at the same time as Northumberland. However, in practice the placebo effects will differ from zero due to normal statistical variation, and thus they provide useful information regarding the degree of variation that could arise within the data.

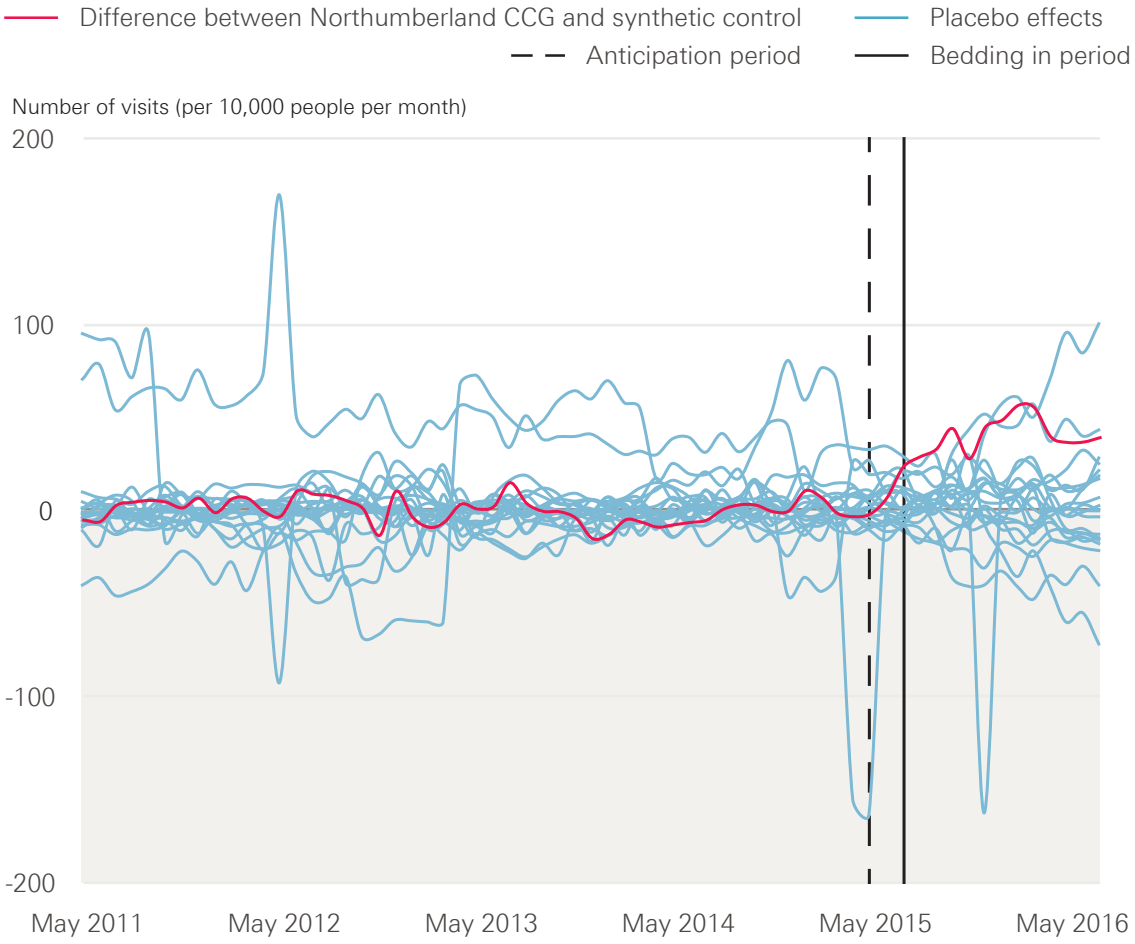
The placebo effects were used to assess a quantity that the unit refers to as the 'significance score'. These scores were calculated separately for each impact metric, and perform a similar role to the p-value. A simple method would be to calculate the proportion of the placebo effects that exceeded the magnitude of the effect for Northumberland over the post-intervention period. However, this would not account for variation in how similar the CCG and its synthetic control was. For example, in some instances there could be differences between a donor CCG and its synthetic control area over the pre-intervention period. Therefore, a slightly more complex method was applied, based on assessing the closeness of the donor CCG and its synthetic control over both periods, and taking the ratio of the Mean Squared Prediction Error (MSPE)† in the post-intervention period to that in the pre-intervention period. The significance score was then calculated as the percentage of placebos with a larger post:pre MSPE ratio than Northumberland. Like the p-value, the lower the significance score, the more confidence the unit can have that the findings reflect a systematic difference in the impact measures between the two areas, rather than chance.

* See www.health.org.uk/impact-redesign-care-Northumberland

† The MSPE refers to the squared deviations between the impact measures for the CCG of interest and the corresponding synthetic control, averaged over all months in the relevant period.

Figure 1 illustrates the approach for one particular impact metric, namely the rate of A&E attendances. This shows the difference between Northumberland CCG and the synthetic control area in a red line. Consistent with the findings shown in the main briefing, this line is close to zero over the first 4 years, but then increases to around 40 A&E visits per 10,000 people per month from August 2015. Also shown in Figure 1 are the 20 placebo effects. Very few of these were as large as the impact observed for Northumberland CCG following August 2015. In fact, the significance score was 0%, meaning that, when judged on the MSPE ratio, the treatment effect for Northumberland CCG was exceeded by none of placebos. Therefore, it is possible to have a high degree of confidence that the findings for A&E visits represented a systematic difference.

Figure 1: Placebo tests for rate of A&E visits and estimated effect for Northumberland CCG



Sensitivity analyses

A number of sensitivity analyses were conducted to assess whether results were sensitive to the assumptions made in the main analysis. The Improvement Analytics Unit assessed the sensitivity of its results to the following changes in approach.

Structuring the hospital data on a quarterly, rather than monthly, basis

The synthetic control method's performance is known to be influenced by the number of pre-intervention periods and the variability of the impact metric. Using quarterly data reduced the variability in impact metrics, at the cost of having fewer pre-intervention periods over which similarity was assessed (16 quarters rather than 48 months). The analysis was repeated using quarterly data and it was found that results are not sensitive to the choice of data frequency (Table 4).

Table 4: Sensitivity to using quarterly instead of monthly data

	Monthly data		Quarterly data	
	% Effect	Significance score	% Effect	Significance score
Rate of A&E visits	13.6%	0%	13.2%	10%
Rate of all admissions	0.8%	95%	0.8%	95%
Rate of emergency admissions	-1.6%	70%	-1.6%	55%
Proportion of A&E visits in which patient is seen within 4 hours	6.9%	0%	7.4%	5%
Average A&E visit length	-10.5%	0%	-9.0%	10%
Average length of stay of all admissions	12%	10%	12.1%	10%

Increasing the duration of the anticipation period when determining the weights used to form the synthetic control areas

The unit also assessed the sensitivity of its estimates to changes in the length of the pre-intervention period by changing the pre-intervention period to end in April 2014 instead of April 2015 (ie using the first 36 months instead of 48 months). As shown in Table 5, results were not sensitive to a shorter pre-intervention period.

Table 5: Sensitivity to using a shorter pre-intervention period

	Using the standard pre-intervention period		Using a shorter pre-intervention period	
	% Effect	Significance score	% Effect	Significance score
Rate of A&E visits (not risk-adjusted)	13.6%	0%	14.6%	0%
Rate of all admissions (not risk-adjusted)	0.8%	95%	0.4%	95%
Rate of emergency admissions	-1.6%	70%	-1.4%	85%
Proportion of A&E visits in which patient is seen within 4 hours	6.9%	0%	6.0%	15%
Average A&E visit length	-10.5%	0%	-14.0%	0%
Average length of stay of all admissions	12%	10%	12.5%	10%

Considering the importance of risk-adjustment by estimating effects without risk-adjusting the impact measures

Although it is more appropriate to risk-adjust the impact metrics as discussed above, the unit assessed whether results would have been very different had it not done so (see Table 6). Where effects were precisely estimated in the baseline analysis, similar effects were found using unadjusted versions of the impact metrics, albeit the effect on the average length of stay of all admissions was about half as large as in the baseline analysis when variation in patients' demographics and past utilisation were not accounted for. Another observation is that the unit's estimate of the impact of the changes on emergency admission rates was a 1.6% reduction with risk-adjustment, but a 10.1% increase without risk-adjustment. However, in both cases the significance score was very high, and the unit's conclusion is the same – namely that there is no evidence of an impact. Where the estimates in the baseline analysis were imprecisely estimated (reflected in high significance scores), the average intervention effects were more sensitive to risk-adjustment.

Table 6: Sensitivity to using unadjusted versions of the impact metrics

	Risk-adjusted impact metrics		Unadjusted impact metrics	
	% Effect	Significance score	% Effect	Significance score
Rate of A&E visits	-	-	13.6%	0%
Rate of all admissions	-	-	0.8%	95%
Rate of emergency admissions	-1.6%	70%	10.1%	85%
Proportion of A&E visits in which patient is seen within 4 hours	6.9%	0%	7.9%	0%
Average A&E visit length	-10.5%	0%	-11.9%	75%
Average length of stay of all admissions	12%	10%	6.4%	5%

Reducing the size of the donor pool to the 10 CCGs most similar to Northumberland

Since the synthetic control area is a weighted average of the CCGs included in the donor pool, the Improvement Analytics Unit assessed the sensitivity of estimates to reducing the number of CCGs in the donor pool, by including only the 10 most similar CCGs instead of the 20 most similar. Estimates are similar to those in the baseline analysis (see Table 7).

Table 7: Sensitivity to using fewer CCGs in the donor pool

	Using 20 CCGs in the donor pool		Using 10 CCGs in the donor pool	
	% Effect	Significance score	% Effect	Significance score
Rate of A&E visits (not risk-adjusted)	13.6%	0%	10.2%	30%
Rate of all admissions (not risk-adjusted)	0.8%	95%	-2.4%	90%
Rate of emergency admissions	-1.6%	70%	-2.5%	20%
Proportion of A&E visits in which patient is seen within 4 hours	6.9%	0%	6.8%	0%
Average A&E visit length	-10.5%	0%	-13.7%	0%
Average length of stay of all admissions	12%	10%	10.2%	20%

Using an alternative donor pool, selected using the Commissioning for Value method

Since results may also be sensitive to the choice of CCGs to include in the donor pool, the analysis was repeated using the 20 most similar CCGs based on the Commissioning for Value tool (after again excluding CCGs in London or in new care models). This donor pool differed substantially from the donor pool in Figure 1, with only five CCGs common to both rankings.

For the risk-adjusted impact metrics, results are robust to the donor pool used (Table 8). However, for the unadjusted impact metrics (the rate of A&E visits and the total rate of admissions), results are quite different, albeit effects are imprecisely estimated using the alterative donor pool. A comparison of the pre-intervention fit, measured by the MSPE, indicates that the study donor pool provides a synthetic control area that is much more similar to Northumberland CCG in the pre-intervention period than that based on the alternative donor pool.

Table 8: Sensitivity to using the 20 most similar CCGs based on the Commissioning for Value method

	Baseline donor pool		Commissioning for Value donor pool		Comparison of model fit
	% Effect	Significance score	% Effect	Significance score	Baseline fits better?
Rate of A&E visits	13.6%	0%	3.2%	60%	Yes
Rate of all admissions	0.8%	95%	-3.4%	85%	Yes
Rate of emergency admissions	-1.6%	70%	-2.3%	25%	Yes
Proportion of A&E visits in which patient is seen within 4 hours	6.9%	0%	4.2%	20%	Yes
Average A&E visit length	-10.5%	0%	-12.5%	5%	Yes
Average length of stay of all admissions	12%	10%	11.6%	5%	No

Limitations

As with all observational studies, this analysis has a number of limitations. While the Improvement Analytics Unit obtained evidence of an intervention effect for some impact metrics, for other measures it cannot be confident that there was an effect, although absence of evidence is not evidence of absence – there may be effects that the unit was unable to pick up. For some impact metrics, it was not possible to find a synthetic control area that was sufficiently similar to Northumberland CCG in the pre-intervention period to provide an adequate counterfactual. It is not possible to say anything about the effects of the intervention for these metrics. Furthermore, for the intervention effects that were observed, it is not possible to separately identify the contribution of each component of the vanguard. The intervention also evolved over time, making it harder still to determine the cause or effects.

A related challenge is that the analysis may capture other changes that occurred in Northumberland CCG in the post-intervention period. This concern was partially addressed by limiting the period under study to that which credibly precedes the introduction of other phases of the PACS implementation.* Nonetheless, it is not possible to rule out confounding by other changes such as the roll out since February 2016 of the 'Medical Interoperability Gateway' (MIG).†

Although a principled approach was used to determine the CCGs to include in the donor pool, the unit cannot rule out the possibility that the CCGs included were not the most appropriate ones. For instance, they may have also implemented changes whose effects could be wrongly attributed to the vanguard. To mitigate this bias, the unit excluded CCGs implementing other new care models. The unit also assessed the sensitivity of estimates to the choice of CCGs included. The unit cannot exclude the possibility that its analysis picked up the effects of other interventions in the trusts or CCGs surrounding Northumberland.

The synthetic control method assumes that similarity of the impact metric over the pre-intervention period is indicative that the impact metrics would have been similar in the absence of the intervention. While this assumption is plausible and the unit used a long pre-intervention period over which to assess similarity, this assumption is fundamentally untestable as it relates to the true counterfactual outcome.

Another limitation is that the true effect of patient characteristics is unknown. While the unit sought to control for variation in patient characteristics prior to estimating effects using a comprehensive set of observed covariates, it is possible that the risk-adjustment equation was mis-specified or that unobserved covariates are not fully accounted for. If this were the case, estimates of effects could be biased, although it should be noted that for impact metrics where the unit found significant effects, results were qualitatively similar even in the absence of risk-adjustment.

---

* The credibility of this assumption was discussed with Northumberland CCG and it felt effects of other phases were unlikely to have occurred in the sample window.

† The MIG allows, with informed patient consent, health care professionals working in a range of urgent and emergency care settings to view essential information on a patient's GP record, helping to inform key clinical decisions about their care and treatment.

Since the analysis relies on data for A&E and inpatient spells, effects of the intervention that may be observed in other domains will not be captured. For instance, impacts on patient satisfaction, staff morale or improved quality of care are not captured by the Improvement Analytics Unit's analysis.

A final point relates to the external validity of the unit's findings. The intervention itself consists of a number of components and evolved over time. Therefore it is unlikely that the intervention could be easily replicated in another location and the specifics of any other implementation would need to be considered before inferring an effect from this study.

References

1. NHS Digital. N*umber of patients registered at a GP practice, Q4 2014/15*.
2. Office for National Statistics. *Census 2011*. 27 March 2011 (census day)
3. Department for Communities and Local Government. *English indices of deprivation*. 2015.
4. Department for Communities and Local Government. *English indices of deprivation*. 2010.
5. NHS Digital. *General and personal medical services*. 2015.
6. Care Quality Commission. *Register of care homes*. 2015.
7. NHS Digital. *Quality Outcome Framework, 2014/15*.
8. Elixhauser A, Steiner C, Harris R, Coffey R. Comorbidity measures for use with administrative data. *Med Care*. 1998;36:8–27
9. Elixhauser score derived by Improvement Analytics Unit using SUS inpatient data for financial years 2011/12–2014/15. Resident population based on NHS Digital patient registration data (2014/15)
10. Number of admissions derived by Improvement Analytics Unit using SUS inpatient data for financial years 2011/12–2014/15. Resident population based on NHS Digital patient registration data (2014/15)
11. Number of admissions and attendances derived by Improvement Analytics Unit, using SUS inpatient data for financial years 2011/12–2014/15. Resident population based on NHS Digital patient registration data (2014/15)
12. Office for National Statistics. *Methods for national statistics 2001 area classification for health areas*. 2001.
13. Office for National Statistics. *Methodology note for the 2011 area classification for local authorities*. 2015. Available from: www.ons.gov.uk/ons/guide-method/geography/products/area-classifications/ns-area-classifications/ns-2011-area-classifications/methodology-and-variables/methodology.pdf
14. Lindeman RH, Merenda PF, Gold RZ. *Introduction to bivariate and multivariate analysis*. 1980. Scott, Foresman, Glenview, IL
15. Feldman B. Relative importance and value. 2005. Manuscript version 1.1, 2005-03-19. Available from: http://dx.doi.org/10.2139/ssrn.2255827
16. Manning WG, Mullahy J. Estimating log models: to transform or not to transform? *Journal of Health Economics*. 2005:20:461–94.
17. Byrne BE, Mamidanna R, Vincent CA, Faiz O. Population-based cohort study comparing 30- and 90-day institutional mortality rates after colorectal surgery. *British Journal of Surgery*. 2013;100(13):1810–1817
18. Birkmeyer JD et al. Surgical skill and complication rates after bariatric surgery. *New England Journal of Medicine*. 2013;369(15):1434–1442
19. Tsai TC et al. Variation in surgical-readmission rates and quality of hospital care. *New England Journal of Medicine*. 2013;369(12):1134–1142
20. Zuckerman RB, Sheingold SH, Orav EJ, Ruhter J, Epstein AM. Readmissions, observation, and the Hospital Readmissions Reduction Program. *New England Journal of Medicine*. 2016;374:1
21. Anselmi L, Meacock R, Kristensen SR et al. Arrival by ambulance explains variation in mortality by time of admission: Retrospective study of admissions to hospital following emergency department attendance in England. *BMJ Quality & Safety*. 2016. Available from: http://dx.doi.org/10.1136/bmjqs-2016-005680
22. Abadie A, Gardeazabal J. The economic costs of conflict: A case study of the Basque Country. *Am Econ Rev*. 2003;93(1):113–32
23. Abadie A, Diamond A, Hainmueller J. Synthetic control methods for comparative case studies: Estimating the effect of California's Tobacco Control Program. *J Am Stat Assoc*. 2010;105(490):493–505. Available from: http://dx.doi.org/10.1198/jasa.2009.ap08746

The Health Foundation
90 Long Acre, London WC2E 9RA
T +44 (0)20 7257 8000
E info@health.org.uk
@HealthFdn
www.health.org.uk

Sign up for our newsletter
www.health.org.uk/enewsletter

The Health Foundation is an independent charity committed to bringing about better health and health care for people in the UK. Our aim is a healthier population, supported by high quality health care.

Registered charity number: 286967
Registered company number: 1714937