

Did paying drugs misuse treatment providers for outcomes lead to unintended consequences for hospital admissions? Difference-in-differences analysis of a pay-for-performance scheme in England

Thomas Mason¹ , William Whittaker¹, Andrew Jones¹  & Matt Sutton^{1,2}

Division of Population Health, Health Services Research and Primary Care, School of Health Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK¹ and Melbourne Institute, Applied Economic and Social Research, Melbourne, Australia²

ABSTRACT

Aims To estimate how a scheme to pay substance misuse treatment service providers according to treatment outcomes affected hospital admissions. **Design** A controlled, quasi-experimental (difference-in-differences) observational study using negative binomial regression. **Setting** Hospitals in all 149 organisational areas in England for the period 2009–2010 to 2015–2016. **Participants** 572 545 patients admitted to hospital with a diagnosis indicating drug misuse, defined based on International Classification of Diseases 10th Revision (ICD-10) diagnosis codes (37 964 patients in 8 intervention areas and 534 581 in 141 comparison areas). **Intervention and comparators** Linkage of provider payments to recovery outcome indicators in 8 intervention organisational areas compared with all 141 comparison organisational areas in England. Outcome indicators included: abstinence from presenting substance, abstinent completion of treatment and non-re-presentation to treatment in the 12 months following completion. **Measurements** Annual counts of hospital admissions, emergency admissions and admissions including a diagnosis indicating drugs misuse. Covariates included age, sex, ethnic origin and deprivation. **Findings** For 37 245 patients in the intervention areas, annual emergency admissions were 1.073 times higher during the operation of the scheme compared with non-intervention areas (95% CI = 1.049; 1.097). There were an estimated additional 3 352 emergency admissions in intervention areas during the scheme. These findings were robust to a range of secondary analyses. **Conclusion** A programme in England from 2012 to 2014 to pay substance misuse treatment service providers according to treatment outcomes appeared to increase emergency hospital admissions.

Keywords Difference-in-differences, drugs misuse, financial incentives, pay-for-performance, substance misuse, unintended consequences.

Correspondence to: Thomas Mason, Division of Population Health, Health Services Research and Primary Care, School of Health Sciences, Faculty of Biology, Medicine and Health, University of Manchester. E-mail: thomas.mason@manchester.ac.uk

Submitted 15 July 2020; initial review completed 24 September 2020; final version accepted 3 March 2021

INTRODUCTION

The evidence base on pay-for-performance (P4P) schemes adopted in healthcare settings internationally has grown over the past 15 years without a consensus as to their effects [1–11]. Schemes have tended to link payment to care processes, because linking to outcomes can be more problematic (particularly for potentially rare health outcomes) [12–15]. Overall, there is no evidence that P4P schemes have a consistent effect on intermediate health outcomes [5].

Most studies have examined P4P schemes within either primary or secondary healthcare—there are relatively few assessments in substance misuse treatment. Evaluations of two United States (US) schemes tentatively suggest that their introduction led to improvements in certain treatment outcomes [16,17], but their interpretation is limited by study design; lack of control group, missing baseline measurements and reliance on self-reported abstinence. A more recent Washington-based study adopted a robust, quasi-experimental design to evaluate a care quality improvement scheme within substance misuse treatment

[18] and found no evidence of improved care quality overall.

New models of contracting in the public sector have become increasingly common internationally [19–22]. In the previous decade the United Kingdom (UK) government increased its focus on paying providers of public services according to outcomes, including a scheme incentivising recovery success within substance misuse treatment. This scheme was adopted amidst development of the UK government's drug strategy, prioritising the goal of 'recovery'—a nebulous term that superseded 'abstinence' [23–25].

The P4P pilot scheme represented a key pillar of the revised drug strategy that would 'incentivise the system to deliver on recovery outcomes' [25] by making treatment providers financially dependent on their achievement. The P4P pilot was also aimed at increasing efficiency and value for money [25]. It was introduced in 2012 and linked payment for treatment to nationally agreed outcomes.

Previous studies have considered the impact of this P4P scheme during its operation, from 2012 to 2014 [23,26,27]. These studies showed that the policy led to: lower treatment initiation; treatment completion (free of dependence); treatment completion without subsequent re-presentation; and higher waiting times. However, these studies did not consider impacts across the wider drugs misusing population, rather being limited to those in treatment. Such an approach fails to identify the impacts that lower initiations and higher waiting times may have on the population at risk of the harms associated with drugs misuse. This wider population is observed across government services such as welfare, criminal justice and hospitals. The latter allows for examination of a wider set of economic consequences and the health outcomes of those experiencing the most severe harms.

Furthermore, previous evaluations only had access to data until 2014. The pilot scheme ended in April 2014 without national adoption of the scheme—although commissioners had autonomy to incorporate elements of P4P in reimbursement models. Analyses of outcomes after the scheme's termination can provide firmer evidence of overall impact. Insight into the wider impact of the scheme can inform local commissioners and decision makers of its potential efficiency; P4P schemes remain popular with decision makers internationally [28] including within substance misuse treatment [18].

This study is therefore focused on whether the P4P scheme, operational between 2012–2013 and 2013–2014, led to unintended consequences in a wider population at risk of drugs misuse using quasi-experimental methods applied to individual-level hospital data for England from 2009–2010 to 2015–2016. This covers the period before, during and after the scheme. Specifically, we examine the following research questions. For

individuals admitted to hospital with a diagnosis indicating substance misuse during the study period, did the introduction of the scheme lead to a differential change in their rate of:

- 1 hospital admissions,
- 2 emergency admissions, and
- 3 hospital admissions including a diagnosis indicating drugs misuse.

METHODS

Intervention

Eight organisational areas (local authorities) participated in the pilot scheme in England from April 2012 to March 2014. Participating areas were selected to be representative from an applicant sample of 29 areas. Non-participating areas were permitted to operate their own commissioning models, which had been traditionally based on block grants [26,27]. The P4P models applied to all services within participating areas; with all service users moved to P4P funding models during the study period.

Participant areas enjoyed a level of autonomy, such as adjusting income weights allocated to particular performance indicators, provided their schemes adhered to a national outcomes framework (Supporting information Tables S1–S4). Payments were conditional on performance indicators in three nationally specified domains: (i) progression toward abstinence from presenting drug(s) of dependence (and crack/heroin); (ii) reduction(s) in offending; and (iii) improvements in health and wellbeing.

Initial assessments and treatment plans were conducted by the newly constituted local area single assessment and referral system (LASARS). Complexity scores for individuals informed the adjustment of provider payments. LASARS could be established independently of treatment service providers, albeit they were often co-located.

Design

To examine the effect of participation in the scheme on hospital admissions, we used difference-in-differences (DiD) estimation to compare changes in participant areas to those not participating.

Participants

There is no authoritative measure of the populations needing drugs misuse treatment in each area in each year. We used International Classification of Diseases 10th Revision (ICD-10) codes related to substance misuse (Table 1) consistent with previous academic studies and government reports [29–33] to identify drug-related hospital admissions. The study population was then defined as individuals

Table 1 ICD-10 codes used to define diagnoses of substance misuse in hospital

| Description | ICD-10 codes |
|---|--------------------------|
| Mental and behavioural disorders due to psychoactive substance use, excluding alcohol and tobacco | F11–16; F18–19 |
| Accidental poisoning by drugs, medicaments and biological substances | X40–X44 |
| Intentional self-poisoning by drugs, medicaments and biological substances | X60–X64 |
| Assault by drugs, medicaments and biological substances | X85 |
| Poisoning by drugs, medicaments and biological substances, undetermined intent | Y10–Y14 |
| Poisoning by drugs, medicaments and biological substances (including overdose and taken in error) | T40.0–40.7; T40.9; T43.6 |

with at least one drug-related hospital admission (a ‘qualifying admission’) during the study period (2009–2010 to 2015–2016). This definition of the ‘at risk’ population allowed us to focus on those who experienced the most severe harms at some point before, during, or after the adoption of the experimental payment system

Intervention and comparison groups were based on area of residence first recorded within the study period. Change of area of residence was possible within the study period and within financial years, possibly representing self-selection into intervention/control groups in response to or anticipation of the policy. We, therefore, fixed individuals to their initial area of residence, but generated a ‘migration’ sub-population for testing the sensitivity of the results to the identifying assumption of fixed residence.

Data

We used activity data from the admitted patient care data files in Hospital Episodes Statistics (HES) for the financial years 2009–2010 to 2015–2016. HES are episode-level data on care received in hospitals across England. This includes; privately paid patients treated in National Health Service (NHS) hospitals, patients resident outside of England and care delivered by non-NHS providers that is funded by the NHS [34]. HES data provide clinical information (such as diagnoses and operations), patient information (such as age and sex), administrative information (such as admission date and discharge method) and information on patient area of residence [34].

Episodes in HES represent a period of care delivered under a specific consultant in a single hospital [34]. These episodes are uniquely identified within spells, defined as, the period from admission to discharge. We aggregated episodes into spells, referred to as admissions herein. From the admission-level data, we constructed a longitudinal data set for each individual in each financial year from 2009–2010 to 2015–2016 (Fig. 1).

We examined the effect of the scheme on the following annual counts of hospital activity per person; the number of hospital admissions, the number of emergency hospital admissions and the number of hospital admissions containing a diagnosis of substance misuse.

Analysis

We used DiD estimation comparing changes in the outcomes before (2009–2010 to 2011–2012), during (2012–2013 to 2013–2014) and after (2014–2015 to 2015–2016) the scheme in the intervention areas (8 local authorities) and comparison areas (141 local authorities). The DiD is the change over time in average outcome in the intervention areas minus that in the comparison areas [35]. DiD assumes that, conditional on the other variables in the model, changes over time would have been the same in the intervention areas as in the comparison areas in the absence of the intervention. We formally tested for parallel pre-trends to assess the plausibility of this assumption.

We also compared intervention areas to a restricted subset of 51 local authorities, matched for similarity on two population domains (% of persons with opiate and crack disorders in drugs misuse treatment populations (selected because of relatively high hospitalisation rates for persons with opiate/crack disorders [36] and 2010 indices of multiple deprivation [IMD] [26,37]).

We estimated all regression models using negative binomial regression appropriate for count data [38–40].

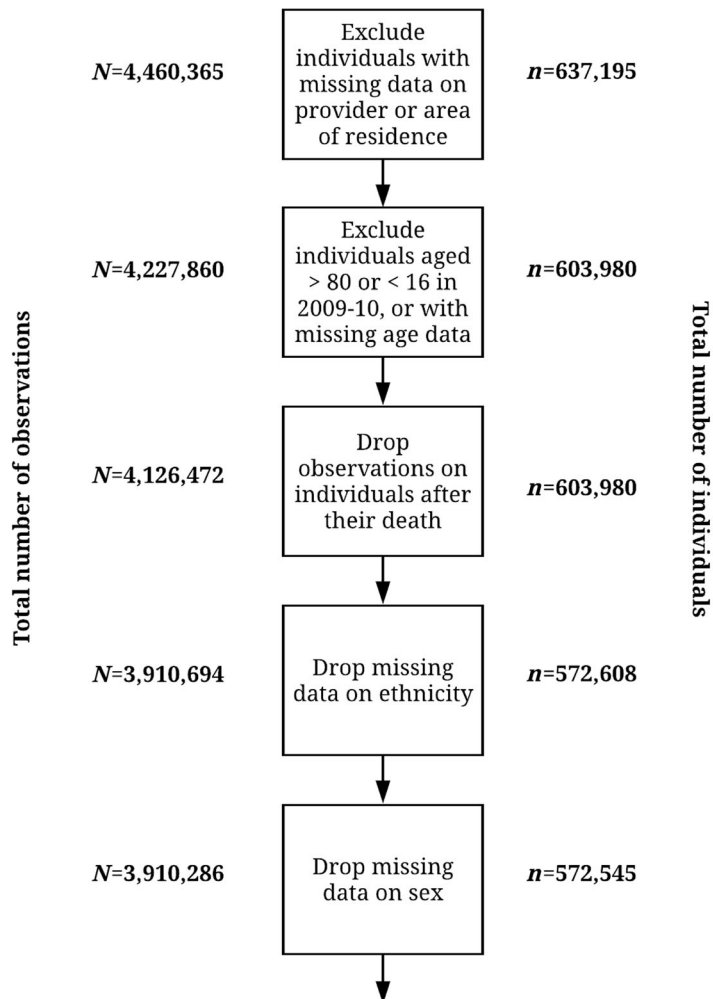
We conducted analyses at the individual-level using annual counts of hospital admissions for each financial year from 2009–2010 to 2015–2016 as the outcomes. We included individual-level controls including age intervals, sex, IMD deprivation deciles and ethnic origin (all of which have been shown to affect health outcomes in substance misusers) [23,26,27,41]; in addition to year indicators, a binary indicator for being initially resident in the intervention areas, an indicator for the intervention period, an interaction between the intervention period and areas (the DiD term) and an interaction between the post-intervention period and intervention areas.

All models were clustered by individual, control for hospital facility fixed effects and were estimated using the *nbreg* command in Stata version 16. Estimating equations for all analyses and statistical tests are presented in the Supporting information.

We present both incident rate ratios (IRRs) and estimated marginal effects (MEs) on the component parts of the DiD term [42–44]. The marginal effects are estimates

Initial inclusion criteria: elective and emergency hospital inpatient admissions between 2009-10 and 2015-16 for individuals who had at least one drugs misuse diagnosed admission over that period; who were resident in an English Local Authority

$N=4,600,127$ annual observations in $T=7$ years on $n=657,161$ individuals



Estimation sample: $N=3,910,286$ observations on $n=572,545$ individuals

Figure 1 Estimation sample construction

of the magnitude of the effects in terms of the number of admissions and were estimated using the margins postestimation command [44].

Secondary analyses

As secondary analyses, we restricted the sample to exclude individuals who changed residence during the study period. We tested the sensitivity to the sample inclusion criteria by restricting the sample to those who had a qualifying admission during the intervention period. We also tested the sensitivity of the results to a lagged definition of the intervention period (2013–2014 and 2014–2015).

Finally, our definition of the ‘at risk’ population could have caused bias if the intervention had prevented

individuals who had not previously had a drugs misuse hospitalisation from having any drugs misuse hospitalisations during and after the intervention—as these individuals would not have been included in the study population, and such an impact would be missed.

We used two approaches to examine this. The first approach examined impacts only on those individuals who had at least one qualifying admission in the pre-intervention period. This measure of the at-risk population is not affected by the introduction of the experimental payment system. The second approach was to estimate a two-part area-level model. The first part of the area-level model considered the impact of the scheme on the number of individuals in each area having at least one qualifying admission in each year, with the at-risk population being

the total resident population. The second part of the area-level model considered the impact of the scheme on the number of hospital admissions amongst those who had at least one qualifying admission within that year. This two-part approach separately identified the effect of the experimental payment system on the size of the population having a qualifying admission during the intervention period.

The full results of secondary analyses are presented in the Supporting information. The analysis in this study was not pre-registered and the results should be considered exploratory.

RESULTS

Descriptive statistics

The intervention areas were less ethnically diverse than the comparison areas and generally had a higher percentage of individuals resident in areas of lower deprivation (Table 2). The populations were generally similar in terms of age and gender in both the intervention and comparison areas. Matched comparison areas were more similar to the intervention areas in terms ethnicity, age and IMD deprivation.

Trends in admissions before, during and after the intervention

For unadjusted mean values of the outcomes, the trends in pre-intervention years (2009–2010 to 2011–2012) were similar in intervention and comparison areas (Table 2; Fig. 2). There was then a differentially higher increase in the intervention areas during the scheme's operation, followed by a reversion to lower levels of admissions after the scheme was discontinued in the intervention areas. For matched comparison areas, pre-intervention trends appeared more parallel, and there was a pronounced differential increase in the pilot areas, with a more noticeable decrease after the cessation of the scheme.

Formal testing of parallel pre-trends could not reject the null hypothesis of parallel pre-trends in all primary and secondary analyses, except one secondary analysis of all admissions (Table 3).

Regression analyses

We estimated regression analyses on $n = 3\,910\,286$ annual observations on $N = 572\,545$ individuals for the period 2009–2010 to 2015–2016. Analyses of matched comparison areas included $n = 1\,458\,038$ observations on $N = 213\,525$ individuals. Data excluded after initial sample construction are outlined in Fig. 1.

Table 3 outlines the estimated effects of the P4P intervention on all outcomes for all primary and secondary analyses in both relative (IRRs) and absolute terms (MEs).

Table 4 presents full regression estimates for primary analyses.

All hospital admissions

For individuals initially resident in P4P pilot areas, annual hospital admissions (of all types) were 1.064 times higher in the intervention period compared with non-intervention areas (95% CI = 1.035, 1.094). This corresponds to 0.062 annual hospital admissions per person in the intervention areas. For analyses restricted to matched comparators, admissions were 1.065 times higher (95% CI = 1.035, 1.096); 0.059 more admissions yearly per person. The largest relative increase in admissions was for the restricted sample who had a qualifying admission in the pre-intervention period (IRR = 1.085; 95% CI = 1.039, 1.133).

Emergency admissions

For individuals originally located in P4P pilot areas, annual emergency admissions were 1.073 times higher in the intervention period compared with non-intervention areas (95% CI = 1.049, 1.097). This corresponds to 0.045 annual emergency admissions per person. For analyses restricted to matched comparators, emergency admissions were also 1.066 times higher (95% CI = 1.041, 1.092); 0.039 more per person yearly.

Drugs misuse diagnosed admissions

For individuals initially resident in P4P pilot areas, annual admissions including a substance misuse diagnosis were 1.055 times higher in the intervention period compared with non-intervention areas (95% CI = 1.026, 1.084). This represents 0.013 annual admissions per person. The largest relative increase in drugs misuse diagnosed admissions was for the restricted sample who had a qualifying admission in the intervention period (IRR = 1.079; 95% CI = 1.023, 1.138). There were no differences in drugs misuse admissions for the restricted sample that had a qualifying admission in the pre-intervention period.

DISCUSSION

Summary of main findings

Hospital admissions were differentially higher during the scheme for individuals initially resident in the intervention areas. For the 37 245 individuals identified in intervention areas, over 2012–2013 and 2013–2014 we estimated; 4 618 additional hospital admissions, 3 352 emergency admissions and 968 with a diagnosis indicating substance misuse.

Secondary analyses showed that the relative increase in qualifying (drugs misuse diagnosed) admissions was largest

Table 2 Outcomes and population characteristics in 2 years before the pilot (descriptive statistics and *t* tests; primary analysis sample)

| | 2011–2012 | | | | | | | | | | | | | | |
|-----------------------------------|--------------------|--------|----------------------|--------|---------------------|---------|--------------------|--------|----------------------|--------|---------------------|--------|--------|---------|--------|
| | 2009–2010 | | | | 2011–2012 | | | | 2011–2012 | | | | | | |
| | Intervention areas | | All comparison areas | | Matched comparators | | Intervention areas | | All comparison areas | | Matched comparators | | | | |
| N | Mean | N | Mean | N | Mean | N | Mean | N | Mean | N | Mean | N | Mean | P value | |
| Outcomes | | | | | | | | | | | | | | | |
| All admissions | 37 964 | 0.77 | 534 581 | 0.81 | 0.014 | 175 561 | 0.789 | 0.22 | 37 578 | 0.905 | 528 938 | 0.92 | 0.425 | 0.425 | 0.796 |
| Emergency admissions | | 0.552 | | 0.548 | 0.583 | | 0.538 | 0.06 | | 0.607 | | 0.611 | 0.631 | 0.631 | 0.374 |
| Drugs misuse diagnosed admissions | | 0.229 | | 0.231 | 0.464 | | 0.224 | 0.178 | | 0.251 | | 0.255 | 0.377 | 0.377 | 0.556 |
| Characteristics | | | | | | | | | | | | | | | |
| Age | 37 964 | 36.320 | 534 581 | 36.308 | 0.882 | 175 561 | 36.832 | <0.001 | 37 578 | 38.150 | 528 938 | 38.126 | 0.769 | 0.769 | <0.001 |
| IMD score | | 26.213 | | 30.335 | <0.001 | | 23.481 | <0.001 | | 26.209 | | 30.331 | <0.001 | <0.001 | <0.001 |
| Male | 19 611 | 0.517 | 280 721 | 0.525 | 0.001 | 89 061 | 0.507 | 0.001 | 19 376 | 0.516 | 277 251 | 0.524 | 0.001 | 0.001 | 0.001 |
| Asian | 554 | 0.015 | 18 351 | 0.034 | <0.001 | 4019 | 0.023 | <0.001 | 548 | 0.015 | 18 244 | 0.034 | <0.001 | <0.001 | <0.001 |
| Black | 521 | 0.014 | 15 784 | 0.030 | <0.001 | 2606 | 0.015 | 0.099 | 518 | 0.014 | 15 681 | 0.030 | <0.001 | <0.001 | 0.106 |
| Mixed | 422 | 0.011 | 6881 | 0.013 | 0.003 | 1709 | 0.010 | 0.014 | 422 | 0.011 | 6844 | 0.013 | 0.004 | 0.004 | 0.011 |
| Other | 836 | 0.022 | 15 680 | 0.029 | <0.001 | 3508 | 0.020 | 0.011 | 834 | 0.022 | 15 604 | 0.030 | <0.001 | <0.001 | 0.01 |
| White | 35 631 | 0.939 | 477 885 | 0.894 | <0.001 | 163 719 | 0.933 | <0.001 | 35 256 | 0.938 | 472 565 | 0.893 | <0.001 | <0.001 | <0.001 |

IMD = index of multiple deprivation; mean values of outcomes are mean counts per person year; mean values of age and IMD are continuous variables; means of male and ethnicity groups are proportions; N = number of individuals in areas in specified year.

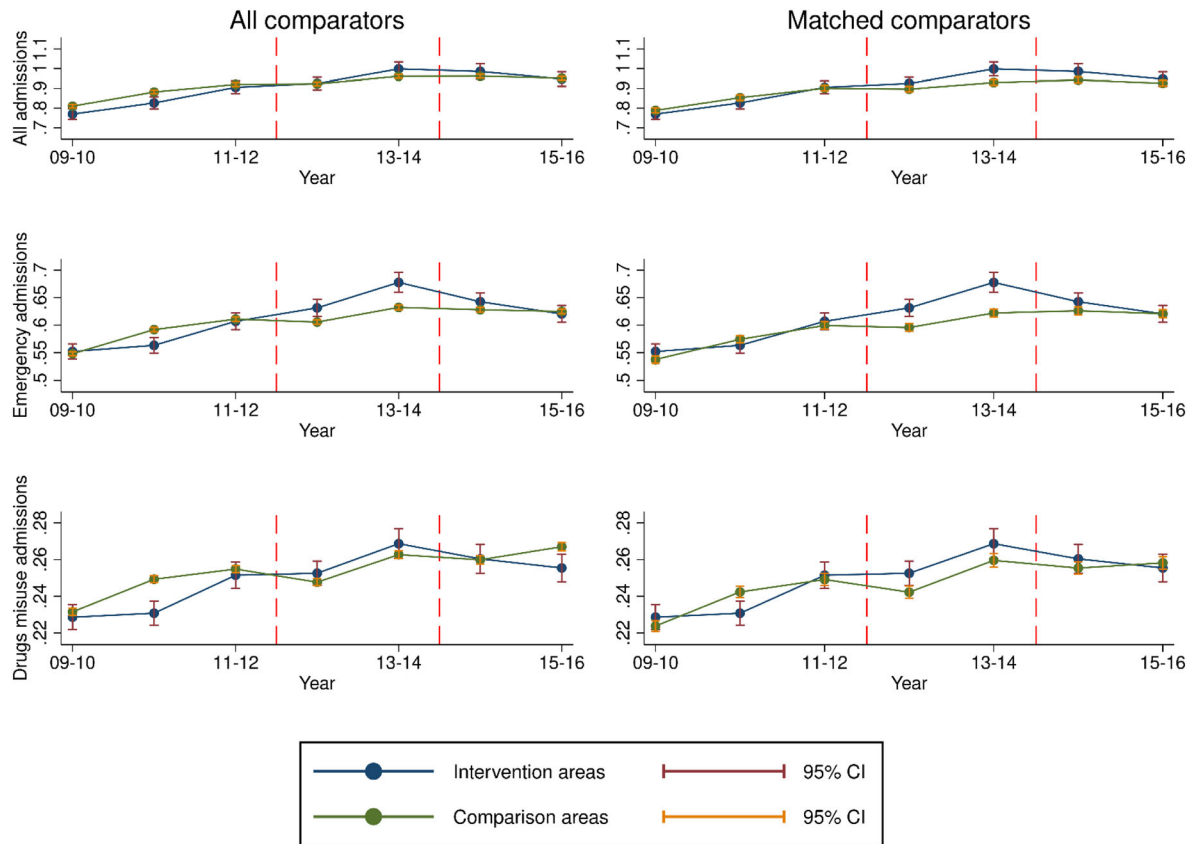


Figure 2 Unadjusted trends in outcomes in intervention and comparison areas. Dotted lines indicate beginning and end of the P4P scheme in the intervention areas [Colour figure can be viewed at wileyonlinelibrary.com]

in those who had at least one qualifying admission in the intervention period; and results from two area-level secondary analyses were consistent with this, showing that the scheme increased the population rate of individuals having at least one qualifying admission in the intervention areas during the scheme. The relative increase in all/emergency admissions was largest in those who had a qualifying admission in the pre-intervention period.

Explanation of notable findings

Differentially increased hospitalisations in the intervention areas may have been partly related to changes in access to treatment (i.e. lower treatment retention and initiation), outlined by previous evaluations of the scheme [23,26,27]. New drug misuse treatments decreased by 6.57% in intervention areas in 2012–2013 and 2013–2014 compared with 2010–2011 and 2011–2012—more than double the relative decrease in comparison areas (Table 5) [23]. Causes of reduced access included; higher waiting times because of implementation difficulties, and changes in the acceptability of treatment for some patients (for example changes in substitute prescribing) [23].

Reduced treatment retention and initiation represent higher levels of unmet need, where need is defined as the

capacity to benefit from treatment [45]. Unmet need in individuals with substance misuse disorders is significant and well established [46], and engagement with treatment services for people with substance misuse disorders has been shown to reduce hospitalisations [47].

The results of primary analyses incorporate the net effects of two impacts of the scheme; differentially increasing the number of individuals having qualifying admissions in the intervention areas during the intervention, and also increasing overall admissions amongst those who had qualifying admissions before the intervention period. The latter effect—especially for emergency admissions—indicates that this latter population's health needs were not being met as well during the scheme.

Evaluations of the scheme did not find evidence of cherry-picking by providers [23]. This may either indicate the effectiveness of (i) the risk adjusted tariff payments, and (ii) the quasi-independence of LASARS, which were introduced as a feature of the intervention with the aim of establishing an independent function responsible for the assessment of all patients in and referred to the treatment system and their subsequent tariffing. Alternatively, cherry-picking may have occurred in ways that were not possible to identify from available data.

Table 3 Difference-in-differences estimates and tests of pre-trends; primary and secondary analyses

| | Secondary analyses | | |
|----------------------------------|--|---|---|
| | Primary analysis | Matched comparators | Qualifying admission in intervention period |
| | 8 intervention vs 141 comparison areas | 8 intervention areas vs 51 comparison areas | 8 intervention vs 141 comparison areas |
| Estimated DID | $n = 3\,910\,286$ $N = 572\,545$ | $n = 1\,458\,038$ $N = 213\,525$ | $n = 1\,385\,442$ $N = 201\,208$ |
| All admissions | | | |
| IRR [95% CI] | 1.064*** [1.035; 1.094] | 1.065*** [1.035; 1.096] | 1.055** [1.013; 1.099] |
| DID of MEs | 0.062 | 0.059 | 0.123 |
| Pre-trends | $P > \chi^2 = 0.564$ | $P > \chi^2 = 0.890$ | $P > \chi^2 = 0.020^*$ |
| Emergency admissions | | | |
| IRR [95% CI] | 1.073*** [1.049; 1.097] | 1.066*** [1.041; 1.092] | 1.070*** [1.033; 1.109] |
| DID of MEs | 0.045 | 0.039 | 0.104 |
| Pre-trends | $P > \chi^2 = 0.142$ | $P > \chi^2 = 0.166$ | $P > \chi^2 = 0.305$ |
| Drug misuse diagnosed admissions | | | |
| IRR [95% CI] | 1.055*** [1.026; 1.084] | 1.042** [1.012; 1.073] | 1.079** [1.023; 1.138] |
| Diff-in-diff of MEs | 0.013 | 0.010 | 0.042 |
| Pre-trends | $P > \chi^2 = 0.649$ | $P > \chi^2 = 0.499$ | $P > \chi^2 = 0.127$ |
| | | | 8 intervention vs 141 comparison areas |
| | | | $n = 1\,904\,020$ $N = 283\,166$ |
| | | | 1.085*** [1.039; 1.133] |
| | | | 0.102 |
| | | | $P > \chi^2 = 0.824$ |
| | | | 1.057** [1.022; 1.094] |
| | | | 0.060 |
| | | | $P > \chi^2 = 0.210$ |
| | | | 1.012 [0.962; 1.064] |
| | | | 0.006 |
| | | | $P > \chi^2 = 0.644$ |

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. n = number of annual observations on, N = number of individuals; DID = difference-in-differences; ME = Marginal effect; models estimated using negative binomial regression; DID of MEs takes the DID between the estimated marginal effects of the components of the DID term (see Supporting information); null hypothesis in test of pre-trends assumes no significant difference in slope comparing intervention and comparators.

Table 4 Regression estimates (primary analyses)

| <i>n</i> = 3 910 286 observations on <i>N</i> = 572 545 individuals | <i>All admissions</i> | | | <i>Emergency admissions</i> | | | <i>Drugs misuse diagnosed admissions</i> | | |
|---|-----------------------|---------------|-------|-----------------------------|---------------|-------|--|---------------|-------|
| | <i>IRR</i> | <i>95% CI</i> | | <i>IRR</i> | <i>95% CI</i> | | <i>IRR</i> | <i>95% CI</i> | |
| Intervention area = 1 | 1.054 | 1 | 1.11 | 1.026 | 0.994 | 1.059 | 1 | 0.974 | 1.027 |
| Intervention period = 1 | 1.005 | 0.995 | 1.015 | 0.994 | 0.986 | 1.002 | 0.935 | 0.925 | 0.945 |
| DiD | 1.064 | 1.035 | 1.094 | 1.073 | 1.049 | 1.097 | 1.055 | 1.026 | 1.084 |
| DiD post intervention | 1.049 | 1.013 | 1.086 | 1.022 | 0.996 | 1.048 | 1.009 | 0.979 | 1.041 |
| Year indicators | | | | | | | | | |
| 2009–2010 | 0.924 | 0.913 | 0.934 | 0.926 | 0.918 | 0.935 | 0.881 | 0.870 | 0.892 |
| 2010–2011 | 0.990 | 0.979 | 1.001 | 0.988 | 0.979 | 0.996 | 0.942 | 0.932 | 0.953 |
| 2011–2012 | 1.020 | 1.010 | 1.031 | 1.011 | 1.003 | 1.020 | 0.964 | 0.954 | 0.975 |
| 2013–2014 | 1.033 | 1.026 | 1.040 | 1.039 | 1.032 | 1.046 | 1.058 | 1.049 | 1.068 |
| 2014–2015 | 1.022 | 1.014 | 1.029 | 1.014 | 1.007 | 1.021 | 0.978 | 0.969 | 0.988 |
| Ethnic origin | | | | | | | | | |
| Asian | 0.981 | 0.936 | 1.027 | 0.850 | 0.831 | 0.869 | 0.799 | 0.785 | 0.813 |
| Black | 1.244 | 1.166 | 1.327 | 0.971 | 0.931 | 1.014 | 0.855 | 0.831 | 0.879 |
| Mixed | 0.947 | 0.889 | 1.008 | 0.884 | 0.853 | 0.915 | 0.915 | 0.887 | 0.943 |
| Other | 0.981 | 0.941 | 1.023 | 0.946 | 0.919 | 0.973 | 0.901 | 0.883 | 0.919 |
| IMD decile | | | | | | | | | |
| 1 | 0.774 | 0.753 | 0.796 | 0.697 | 0.684 | 0.710 | 0.794 | 0.782 | 0.806 |
| 2 | 0.870 | 0.846 | 0.895 | 0.804 | 0.789 | 0.819 | 0.851 | 0.838 | 0.864 |
| 3 | 0.929 | 0.904 | 0.954 | 0.893 | 0.876 | 0.910 | 0.896 | 0.882 | 0.910 |
| 4 | 0.965 | 0.940 | 0.992 | 0.952 | 0.934 | 0.970 | 0.946 | 0.928 | 0.964 |
| 6 | 1.028 | 0.999 | 1.057 | 1.038 | 1.018 | 1.057 | 1.019 | 1.003 | 1.036 |
| 7 | 1.041 | 1.014 | 1.068 | 1.081 | 1.060 | 1.102 | 1.065 | 1.048 | 1.083 |
| 8 | 1.043 | 1.015 | 1.072 | 1.090 | 1.069 | 1.111 | 1.073 | 1.055 | 1.091 |
| 9 | 1.014 | 0.987 | 1.042 | 1.081 | 1.060 | 1.103 | 1.095 | 1.075 | 1.114 |
| 10 | 0.926 | 0.897 | 0.956 | 0.970 | 0.951 | 0.990 | 1.054 | 1.035 | 1.074 |
| Age interval | | | | | | | | | |
| 16–25 | 0.554 | 0.545 | 0.564 | 0.683 | 0.673 | 0.692 | 0.850 | 0.840 | 0.861 |
| 26–35 | 0.661 | 0.649 | 0.674 | 0.760 | 0.751 | 0.770 | 0.975 | 0.963 | 0.987 |
| 36–45 | 0.819 | 0.805 | 0.833 | 0.889 | 0.879 | 0.900 | 1.066 | 1.053 | 1.079 |
| 56–65 | 1.314 | 1.279 | 1.349 | 1.172 | 1.154 | 1.191 | 0.878 | 0.864 | 0.893 |
| 66–75 | 1.644 | 1.588 | 1.703 | 1.374 | 1.349 | 1.399 | 0.755 | 0.744 | 0.767 |
| 76+ | 1.634 | 1.573 | 1.697 | 1.528 | 1.503 | 1.554 | 0.755 | 0.738 | 0.772 |
| Male | 0.902 | 0.891 | 0.913 | 0.911 | 0.904 | 0.919 | 1.010 | 1.003 | 1.018 |
| Constant | 1.301 | 1.142 | 1.481 | 1.002 | 0.879 | 1.141 | 0.354 | 0.304 | 0.413 |
| ln (alpha) | 0.822 | 0.811 | 0.834 | 0.907 | 0.900 | 0.914 | 0.611 | 0.594 | 0.627 |

Models estimated using negative binomial regression; IRR = incidence rate ratio; all models clustered by individual and include hospital facility fixed effects; reference groups: 2011–2012; White; 5th IMD decile; Aged 46–55; Female; 2 year indicators are excluded—one as a reference group, another to allow for estimation of the intervention period indicator.

Table 5 New drugs misuse treatment journeys

| <i>YEAR</i> | <i>Intervention</i> | <i>Comparison</i> |
|------------------|---------------------|-------------------|
| 2010–2011 | 5 295 | 74 377 |
| 2011–2012 | 5 421 | 69 082 |
| Total before P4P | 10 716 | 143 459 |
| 2012–2013 | 5 170 | 69 212 |
| 2013–2014 | 4 842 | 69 717 |
| Total during P4P | 10 012 | 138 929 |
| Change | –704 | –4 530 |
| As % | –6.57% | –3.16% |

Data from Donmall *et al.* [23]. Source used in study is the National Drug Treatment Monitoring System (NDTMS).

Limitations

The generalisability of this study in part depends on the identification of differential changes over time in the appropriate population of interest; individuals with a (potentially unmet) need for substance misuse treatment, identified using hospital diagnosis codes related to substance misuse. Although this allows for the capture of possible wider impacts of the scheme on access to treatment and treatment acceptability in addition to on the hospital sector, it does not precisely capture each individual with a need for substance misuse treatment. Our approach treated admission to hospital including a drugs misuse diagnosis at any point during the study period as a reasonable proxy of a potential

need for drugs misuse treatment. Although this excludes estimating impacts on those not hospitalised at all, it includes those subject to the most severe health harms.

The sample we identified is comparable to the substance misuse treatment population; the relatively young mean age of patients in this study, at 35, is closer to the substance misuse treatment population (age = 33) than the general admitted population in England (age = 53) [34]. However, we sought to also identify potential impacts on unmet need for treatment; and individuals with unmet need are unlikely to be representative of those in treatment [48–50].

Only a relatively small proportion of those in substance misuse treatment in 2012–2013 had planned hospital inpatient detoxification (1.57%) [51], with the vast majority of hospitalisation relating to drugs misuse being independent of planned drugs misuse treatment.

Differential changes over time in factors not captured in the analyses (including factors entirely unrelated to the payment scheme) could in theory provide alternative explanations of some of the estimated effects. For example, there was an increase in opioid/crack using subjects in treatment during the study in the intervention areas. If this increase represented something important about differential comparative change in drug use, it might help explain the increase in hospital admissions. Equally, current evidence on community prevalence of opioid/crack use suggests no significant increase in use in the intervention areas during the pilot (<https://www.ndtms.net/>).

Diagnosis recording in hospital is inexact—in some cases drugs misuse may be suspected but not recorded, and in other cases not suspected despite drugs misuse being a contributing cause of an admission [38].

Implications for policy and research

Although the effect sizes estimated at patient-level in each year may at first appear modest, we estimated that there were 3 352 additional emergency admissions for 37 245 patients in the 8 intervention areas in 2012–2013 and 2013–2014, equivalent to a minimum £2.1 million additional hospital expenditure [52,53]. If the effects on admissions were repeated (even weakly) under national adoption of the scheme across all 149 areas, the economic (and clinical) impact to hospitals would be substantial.

More effective P4P schemes have been targeted on process indicators for health conditions linked to best practice pathways applicable to a majority of the target population (such as Stroke) [54]. Such examples contrast with this scheme, which was focused on outcomes in a setting characterised by a heterogeneous population with complex needs and that involves multiple agencies. Future policies focused on reforming payment systems for substance misuse treatment providers to include P4P may prefer to link

incentive payments to processes evidenced to be universally effective in specific groups of patients (for example, vaccination against Hepatitis B infection [55]).

Considering impacts on hospital admissions and not just those in contact with treatment agencies in research allows for capture of populations at risk of the most severe harms. Administrative data are well suited to this as information is captured on the whole population of interest, not just those in treatment. Furthermore, examination of this wider population is useful as there may be effects on who enters and remains in treatment.

The ‘efficiency savings’ that can be generated by P4P schemes may be offset by costs in other sectors. This scheme is a reminder that research should consider possible effects across various government services—especially for substance misuse policy that affects a range of agencies.

Declaration of interests

None.

Acknowledgements

Department of Health and Social Care, grant number: 011/0055.

Author contributions

Thomas Mason: Conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; project administration; validation. **Andrew Jones:** Conceptualization; data curation; funding acquisition; methodology; project administration; supervision. **Matt Sutton:** Conceptualization; funding acquisition; investigation; project administration; supervision.

References

1. Scott A., Sivey P., Ouakrim D. A., Willenberg L., Naccarella L., Furler J., *et al.* The effect of financial incentives on the quality of health care provided by primary care physicians. *Cochrane Database Syst Rev* [Internet]. 2011 [cited 2020 Jun 9]. Available from: <https://doi.org/10.1002/14651858.CD008451.pub2/abstract>
2. Van Herck P., De Smedt D., Annemans L., Remmen R., Rosenthal M. B., Sermeus W. Systematic review: effects, design choices, and context of pay-for-performance in health care. *BMC Health Serv Res* 2010; **10**: 247.
3. Alshamsan R., Majeed A., Ashworth M., Car J., Millett C. Impact of pay for performance on inequalities in health care: systematic review. *J Health Serv Res Policy* 2010; **15**: 178–84.
4. Eijkenaar F., Emmert M., Scheppach M., Schöffski O. Effects of pay for performance in health care: a systematic review of systematic reviews. *Health Policy* 2013; **110**: 115–30.
5. Mendelson A., Kondo K., Damberg C., Low A., Motúapuaka M., Freeman M., *et al.* The effects of pay-for-performance programs on health, health care use, and processes of care. *Ann Intern Med* 2017; **166**: 341–53.

6. Havers S. M., Kate Martin E., Wilson A., Hall L. A systematic review and meta-synthesis of policy intervention characteristics that influence the implementation of government-directed policy in the hospital setting: implications for infection prevention and control. *J Infect Prev* 2020; 21: 84–96.
7. Zou K., Li H.-Y., Zhou D., Liao Z.-J. The effects of diagnosis-related groups payment on hospital healthcare in China: a systematic review. *BMC Health Serv Res* 2020; 20: 112.
8. Tello J. E., Barbazza E., Waddell K. Review of 128 quality of care mechanisms: a framework and mapping for health system stewards. *Health Policy Amst Neth* 2020; 124: 12–24.
9. Kondo K. K., Wyse J., Mendelson A., Beard G., Freeman M., Low A., *et al.* Pay-for-performance and veteran care in the VHA and the community: a systematic review. *J Gen Intern Med* 2018; 33: 1155–66.
10. Lavergne M. R. Financial incentives for physicians to improve health care. *CMAJ Can Med Assoc J J Assoc Medicale Can* 2017; 189: E1505–E1506.
11. Mandavia R., Mehta N., Schilder A., Mossialos E. Effectiveness of UK provider financial incentives on quality of care: a systematic review. *Br J Gen Pract J R Coll Gen Pract* 2017; 67: e800–e815.
12. Gravelle H., Sutton M., Ma A. Doctor behaviour under a pay for performance contract: treating, cheating and case finding? *Econ J* 2010; 120: F129–F156.
13. Crawley D., Ng A., Mainous A. G., Majeed A., Millett C. Impact of pay for performance on quality of chronic disease management by social class group in England. *J R Soc Med* 2009; 102: 103–7.
14. Karve A. M., Ou F.-S., Lytle B. L., Peterson E. D. Potential unintended financial consequences of pay-for-performance on the quality of care for minority patients. *Am Heart J* 2008; 155: 571–6.
15. Sutton M., Nikolova S., Boaden R., Lester H., McDonald R., Roland M. Reduced mortality with hospital pay for performance in England. *N Engl J Med* 2012; 367: 1821–8.
16. McLellan A. T., Kemp J., Brooks A., Carise D. Improving public addiction treatment through performance contracting: the Delaware experiment. *Health Policy* 2008; 87: 296–308.
17. Shen Y. Selection incentives in a performance-based contracting system. *Health Serv Res* 2003; 38: 535–52.
18. Garnick D. W., Horgan C. M., Acevedo A., Lee M. T., Panas L., Ritter G. A., *et al.* Influencing quality of outpatient SUD care: implementation of alerts and incentives in Washington state. *J Subst Abuse Treat* 2017; 82: 93–101.
19. Sanderson M., Allen P., Gill R., Garnett E. New models of contracting in the public sector: a review of Alliance contracting, prime contracting and outcome-based contracting literature. *Soc Policy Adm* 2018; 52: 1060–83.
20. Lengnick-Hall R., Willging C., Hurlburt M., Fenwick K., Aarons G. A. Contracting as a bridging factor linking outer and inner contexts during EBP implementation and sustainment: a prospective study across multiple US public sector service systems. *Implement Sci* 2020; 15: 43.
21. Maurya D., Ramesh M. Program design, implementation and performance: the case of social health insurance in India. *Health Econ Policy Law* 2019; 14: 487–508.
22. Yerkes M. A., van den Braken L. J. The ambiguity of social return policies in the Netherlands. *Soc Policy Adm* 2019; 53: 113–27.
23. Donmall M., Sutton M., Jones A. *Evaluation of the Drugs and Alcohol Recovery Payment by Results Pilot Programme Final Report* [Internet]. Manchester: University of Manchester; 2017 p. 164. Available from: http://research.bmh.manchester.ac.uk/epidemiology/NDEC/research/publications/PbRDR_Final_Report.pdf
24. NTA. Medications in Recovery: reorientating drug dependence treatment [Internet]. National Treatment Agency; 2012. Available from: <http://www.nta.nhs.uk/uploads/medications-in-recovery-mainreport3.pdf>
25. HMG. Drug Strategy 2010 Reducing Demand, Restricting Supply, Building Recovery: Supporting People to Live a Drug Free Life: (531792013–001) [Internet]. 2010. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/98026/drug-strategy-2010.pdf
26. Mason T., Sutton M., Whittaker W., McSweeney T., Millar T., Donmall M., *et al.* The impact of paying treatment providers for outcomes: difference-in-differences analysis of the 'payment by results for drugs recovery' pilot. *Addiction* 2015; 110: 1120–8.
27. Jones A., Pierce M., Sutton M., Mason T., Millar T. Does paying service providers by results improve recovery outcomes for drug misusers in treatment in England? *Addiction* 2018; 113: 279–86.
28. James J. Pay-for-Performance | Health Affairs [Internet]. 2012 [cited 2020 Jun 9]. Available from: <https://doi.org/10.1377/hpb20121011.90233/full/>
29. ONS. Deaths related to drug poisoning in England and Wales - Office for National Statistics [Internet]. 2018 [cited 2020 Jun 9]. Available from: <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/deathsrelatedtodrugpoisoninginenglandandwales/2017registations>
30. Pierce M., Bird S. M., Hickman M., Marsden J., Dunn G., Jones A., *et al.* Impact of treatment for opioid dependence on fatal drug-related poisoning: a national cohort study in England. *Addiction* 2016; 111: 298–308.
31. Public Health Wales. Drug deaths in Wales 2016 [Internet]. 2017 p. 24. Available from: <http://cardiffandvaleapb.org/download/Welsh%20Government%20Documents/Drug-related-deaths-2016-ONS-Analysis.pdf>
32. Shah A. D., Wood D. M., Dargan P. I. Survey of ICD-10 coding of hospital admissions in the UK due to recreational drug toxicity. *QJM Mon J Assoc Physicians* 2011; 104: 779–84.
33. Mills H., Skodbo S., Blyth P. Understanding organised crime: estimating the scale and the social and economic costs [Internet]. 2013 p. 137. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/246390/horr73.pdf
34. NHS Digital. Hospital Admitted Patient Care Activity 2015–16 [Internet]. 2016. Available from: <https://files.digital.nhs.uk/publicationimport/pub22xxx/pub22378/hosp-epis-stat-admi-summ-rep-2015-16-rep.pdf>
35. Angrist J. D., Pischke J.-S. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press; 2008, p. 393.
36. NHS Scotland & ONS. Drug-Related Hospital Statistics, Scotland 2017/18 [Internet]. 2019 p. 53. Available from: <https://www.isdscotland.org/Health-Topics/Drugs-and-Alcohol-Misuse/Publications/2019-05-28/2019-05-28-DRHS-Report.pdf>
37. Communities and Local Government. The English Indices of Deprivation 2010 [Internet]. 2011. Available from: <https://assets.publishing.service.gov.uk/>

- government/uploads/system/uploads/attachment_data/file/6871/1871208.pdf
38. Basu A., Manning W. G. Issues for the next generation of health care cost analyses. *Med Care* 2009; **47**: S109–S114.
 39. Manning W. G., Mullahy J. Estimating log models: to transform or not to transform? *J Health Econ* 2001; **20**: 461–94.
 40. Cameron A. C., Trivedi P. K. Essentials of Count Data Regression. In: *A Companion to Theoretical Econometrics* [Internet]. John Wiley & Sons, Ltd; 2007. pp. 331–48. Available from: 10.1002/9780470996249.ch16.
 41. Patel R., Wilson R., Jackson R., Ball M., Shetty H., Broadbent M., et al. Association of cannabis use with hospital admission and antipsychotic treatment failure in first episode psychosis: an observational study. *BMJ Open* 2016; **6**: e009888.
 42. Greene W. H. *Econometric Analysis*. Pearson Education India; 2003.
 43. Cornelißen T., Sonderhof K. Partial effects in Probit and logit models with a triple dummy-variable interaction term. *Stata J* 2009; **9**: 571–83.
 44. Williams R. Using the margins command to estimate and interpret adjusted predictions and marginal effects. *Stata J* 2012; **12**: 308–31.
 45. Culyer T. J., Wagstaff A. Need, equity and equality in health and health care [Internet]. Centre for Health Economics, University of York; 1992 Feb [cited 2020 Jun 10]. (Working Papers). Report No.: 095chedp. Available from: <https://ideas.repec.org/p/chy/respap/95chedp.html>
 46. Shepard D. S., Strickler G. K., McAuliffe W. E., Beaton-Blaakman A., Rahman M., Anderson T. E. Unmet need for substance abuse treatment of adults in Massachusetts. *Adm Policy Ment Health Ment Health Serv Res* 2005; **32**: 403–26.
 47. Neighbors C. J., Zywiak W. H., Stout R. L., Hoffmann N. G. Psychobehavioral risk factors, substance treatment engagement and clinical outcomes as predictors of emergency department use and medical hospitalization. *J Stud Alcohol* 2005; **66**: 295–304.
 48. Mulvaney-Day N., DeAngelo D., Chen C., Cook B. L., Alegria M. Unmet need for treatment for substance use disorders across race and ethnicity. *Drug Alcohol Depend* 2012; **125**: S44–S50.
 49. Stringer K. L., Baker E. H. Stigma as a barrier to substance abuse treatment among those with unmet need: an analysis of parenthood and marital status. *J Fam Issues* 2015; **24** [cited 2020 Jun 12]; Available from: <https://doi.org/10.1177/0192513X15581659>
 50. Terplan M., Longinaker N., Appel L. Women-centered drug treatment services and need in the United States, 2002–2009. *Am J Public Health* 2015; **105**: e50–e54.
 51. Public Health England. Drug Statistics from the National Drug Treatment Monitoring System (NDTMS) 1 April 2012 to 31 March 2013 [Internet]. 2013 p. 37. Available from: <http://www.nta.nhs.uk/uploads/annualdrugstatistics2012-13-statisticalreport.pdf>
 52. Curtis L. *Unit costs of health and social care* Place of publication not identified. University Of Kent At Can; 2012.
 53. Curtis L. C. *Unit costs of health & social care*. Kent: Personal Social Services Research Unit, Univ. of Canterbury; 2013, p. 281.
 54. Mason T., Lau Y.-S., Sutton M. Is the distribution of care quality provided under pay-for-performance equitable? Evidence from the advancing quality programme in England. *Int J Equity Health* 2016; **15**: 156.
 55. Budd J., Robertson R., Elton R. Hepatitis B vaccination and injecting drug users. *Br J Gen Pract* 2004; **54**: 444–7.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1 Domain 1 in the national outcomes framework (i).

Table S2 Domain 1 in the national outcomes framework (ii).

Table S3 Domain 2 in the national outcomes framework.

Table S4 Domain 3 in the national outcomes framework.

Table S5 Regression estimates (lagged definition of the intervention period and corresponding DiD estimate).

Table S6 Regression estimates (secondary analysis [matched comparison areas]).

Table S7 Regression estimates (secondary analysis [excluding individuals who moved area residence]).

Table S8 Regression estimates (secondary analysis [individuals with a qualifying admission in the intervention period]).

Table S9 Regression estimates (secondary analysis [individuals with a qualifying admission in the pre-intervention period]).

Table S10 Estimated marginal effects (primary analyses).

Table S11 Estimated marginal effects (lagged definition of the intervention period).

Table S12 Estimated marginal effects (matched comparison areas).

Table S13 Estimated marginal effects (excluding individuals who moved area residence).

Table S14 Estimated marginal effects (individuals with a qualifying admission in the intervention period).

Table S15 Estimated marginal effects (individuals with a qualifying admission in the pre-intervention period).

Table S16 Outcomes and population characteristics pre and post intervention (descriptive statistics and *t* tests; primary analysis sample).

Table S17 Area-level model of the impact of P4P on the number of individuals who had at least one qualifying (drug misuse diagnosed) hospital admission.

Table S18 Area-level model of the impact of P4P on the mean number of admissions for each individual who had at least one qualifying admission.