

Can pay-for-performance to primary care providers stimulate appropriate use of antibiotics?

Lina Maria Ellegård¹  | Jens Dietrichson² | Anders Anell³

¹Department of Economics, Lund University, Lund, Sweden

²SFI, The Danish National Centre for Social Research, Copenhagen, Denmark

³Department of Business Administration, Lund University, Lund, Sweden

Correspondence

Lina Maria Ellegård, Department of Economics, Lund University, P.O. Box 7082, S-220 07 Lund, Sweden.

Email: linamaria.ellegard@nek.lu.se

JEL Classification: D23; I11; I18; J33; J38; H73

Summary

Antibiotic resistance is a major threat to public health worldwide. As the health-care sector's use of antibiotics is an important contributor to the development of resistance, it is crucial that physicians only prescribe antibiotics when needed and that they choose narrow-spectrum antibiotics, which act on fewer bacteria types, when possible. Inappropriate use of antibiotics is nonetheless widespread, not least for respiratory tract infections (RTI), a common reason for antibiotics prescriptions. We examine if pay-for-performance (P4P) presents a way to influence primary care physicians' choice of antibiotics. During 2006–2013, 8 Swedish healthcare authorities adopted P4P to make physicians select narrow-spectrum antibiotics more often in the treatment of children with RTI. Exploiting register data on all purchases of RTI antibiotics in a difference-in-differences analysis, we find that P4P significantly increased the share of narrow-spectrum antibiotics. There are no signs that physicians gamed the system by issuing more prescriptions overall.

KEYWORDS

antibiotic resistance, pay-for-performance, primary care

1 | INTRODUCTION

Pay-for-performance (P4P), monetary incentives to reach predefined targets, is ubiquitous in health care. Literature reviews conclude that incentives tied to process measures, such as screening rates and guideline adherence, often have small to moderate effects, whereas P4P has no impact on broad outcome measures such as mortality (Christianson, Leatherman, & Sutherland, 2008; Eijkenaar, Emmert, Scheppach, & Schöski, 2013; Ogundeji, Bland, & Sheldon, 2016; Petersen, Woodard, & Urech, 2006; Rosenthal & Frank, 2006; Scott, Sivey & Ait Ouakrim, 2011; Town, Kane, Johnson & Butler, 2005; Van Herck, De Smedt, Annemans, & Remmen, 2010). The reviews also point at gaps in the literature: Many studies lack adequate control groups (Ogundeji et al., 2016), and there are few studies from outside the UK or the United States (Eijkenaar et al., 2013). This study provides evidence from Sweden, where P4P became widely adopted in primary care during the past decade. With one exception, Ödesjö, Anell, Gudbjörnsdóttir, Thorn, and Björck (2015) who examine incentives tied to diabetes registry entries in one county, Swedish applications of P4P have not been studied before. We take advantage of the unique institutional setting, in which the responsibility for health care is decentralized to 21 county councils. Due to the decentralization, monetary incentives have been introduced in some counties but not in others, making it possible to evaluate P4P by comparing counties that use specific incentives with counties that do not.

Another contribution of the study is to focus on a scarcely studied P4P application of great policy relevance, namely, incentives related to antibiotic prescriptions. Antimicrobial resistance is one of the greatest threats to the effectiveness of health care

.....
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2017 The Authors Health Economics Published by John Wiley & Sons Ltd.

(ECDC/EMEA Joint working group, 2009; Carlet et al., 2011). To impede the spread of resistance, physicians ought to engage in antibiotic stewardship: that is, prescribe antibiotics only when necessary and select antibiotics that act against as narrow a spectrum of bacteria as possible (Kaier & Moog, 2012; Ranji, Steinman, & Shojanian, 2006). However, inappropriate use of antibiotics is common.¹ Results for other process measures cannot readily be extrapolated to antibiotics-related P4P, which imply more than “box-ticking”: Physicians may bear a real cost from denying patients desired antibiotics (Ashworth et al., 2015). There are only four previous studies of P4P tied to antibiotic stewardship, of which three find positive associations (Gong, Qiu, Song, & Sun, 2016; McDonald, Boaden, Roland, & Kristensen, 2015; Yip, Powell-Jackson, & Chen, 2014) and one finds a negative relationship (Mullen, Frank, & Rosenthal, 2010).

We study the effect of P4P for antibiotic stewardship in the treatment of children with respiratory tract infections (RTI), introduced by eight counties in a staggered manner. Using municipality-level register data on all purchases of RTI antibiotics prescribed to Swedish children aged 0–6 years in 2006–2013, we estimate a difference-in-differences (DID) model of the impact of P4P. The incentives we study encouraged physicians to select narrow-spectrum antibiotics in the treatment of children with RTI. Specifically, the incentives were tied to the share of narrow-spectrum penicillin V (PcV) prescriptions in childrens' total consumption of RTI antibiotics. The focus on the PcV share mirrored Swedish RTI treatment guidelines, which advocate PcV as the first-line antibiotic in most cases because of its smaller impact on resistance.

The next section gives a background to RTI and Swedish treatment guidelines. Section 3 reviews related literature. Section 4 gives an institutional background and describes the P4P schemes. Section 5 describes the data, Section 6 outlines our empirical strategy, and Section 7 presents the results. Section 8 concludes.

2 | RTI AND ANTIBIOTICS

Respiratory tract infections is an umbrella term for conditions affecting the respiratory organs, for example, sore throat (pharyngitis), ear infection (otitis), influenza, cough (bronchitis), pneumonia, sinus infection, tonsillitis, laryngitis, and the common cold. RTI is highly associated with antibiotics prescriptions. In Sweden, antibiotics that are typically prescribed for RTI account for 90% of children's total antibiotics consumption (SWEDRES-SVARM, 2013) and approximately half of all RTI patients in primary care receive antibiotics (André et al., 2008). However, there is sometimes a weak clinical basis for prescribing antibiotics. Many patients seek care for viral RTI, such as the common cold, which are not cured by antibiotics. For bacterial RTI, antibiotics may reduce the symptom spell by a few days but in most cases do not yield large health gains.² Nevertheless, some bacterial infections carry a small risk for detrimental complications that require antibiotic treatment. Fear of such consequences, together with diagnostic uncertainty, may explain the widespread use of antibiotics for RTI (e.g., Keith, Saxena, Murray, & Sharland, 2010).

The use of antibiotics gives an evolutionary advantage to resistant bacteria strains (Kaier & Moog, 2012). To impede the spread of resistance, it is advisable to choose an antibiotic that acts against a narrow spectrum of bacteria whenever possible (Hersh, Shapiro, Pavia, & Shah, 2011). According to the Swedish treatment guidelines, narrow-spectrum antibiotics are often sufficient treatment for patients with bacterial RTI; specifically, PcV is the recommended first-line antibiotic for most common RTIs.³ In 2009, the renowned and government-supported network Strategic Programme against Antibiotic Resistance (Strama)⁴ proposed as a national target that PcV should account for 80% of RTI antibiotics prescriptions to children aged 0–6 years. The reason for the focus on children was that most RTI patients in this age group benefit little from antibiotics and that PcV is often sufficient when antibiotics are needed.⁵ Notably, the target still allowed for broad-spectrum antibiotics when they were needed.

3 | RELATED LITERATURE

There is one previous study of P4P in Swedish primary care. Ödesjö et al. (2015) study financial incentives for register entry of diabetes patients introduced in the county Västra Götaland in 2011. They compare the 1-year development with one other

¹In the United States, almost half of all prescriptions for acute respiratory tract infections—the most common reason for antibiotics use in the United States—are inappropriate (Fleming-Dutra, Hersh & Shapiro, 2016).

²See for example, Spurling, Doust, Del Mar and Eriksson (2011); Lemiengre et al. (2012); Spinks, Glasziou and Del Mar, (2013); Venekamp, Sanders and Glasziou, (2013); Cronin, Khan and Saeed, (2013); Smith, Fahey, Smucny, and Becker, (2014).

³Swedish RTI guidelines are available at <https://www.folkhalsomyndigheten.se/smittskydd-beredskap/antibiotika-och-antibiotikaresistens/behandling-srekommendationer/>, last accessed 2016-12-06. PcV has long been the most common RTI antibiotic in Sweden (André et al., 2008) but is much less used outside Scandinavia (Goossens et al., 2005, 2007).

⁴<http://strama.se/about-strama/?lang=en>, last accessed 2016-05-19.

⁵See Online Appendix for English translation of targets.

county (Skåne) that did not use this type of P4P and find increased register entry, increased completeness of data, and altered blood pressure entry behavior.

Of the four previous studies addressing incentives for antibiotic stewardship, two Chinese studies find reduced prescription rates (2–7 percentage points, or 15–60%) (Gong et al., 2016; Yip et al., 2014), a British study reports a 2–3 percentage point (3%) improvement of the rate of appropriate antibiotics (McDonald et al., 2015), and a U.S. study finds deteriorations of a similar outcome (Mullen et al., 2010).⁶ These studies have notable methodological weaknesses. Only two (Yip et al., 2014; Mullen et al., 2010) have control groups, and one study (McDonald et al., 2015) even lacks a before-P4P baseline measure. In all four studies, P4P was coupled with other important policy changes—complete restructuring of the payment scheme from fee-for-service (FFS) to capitation (Yip et al., 2014), the introduction of other P4P indicators (McDonald et al., 2015; Mullen et al., 2010), and the simultaneous introduction of audit and feedback (Gong et al., 2016).

The literature on *non-financial* interventions for antibiotic stewardship suggests that education and audit and feedback have moderate effects (Arnold and Straus, 2005; Ranji et al., 2006; van der Velden, Pijpers, & Kuyvenhoven, 2012; Vodicka, Thompson & Lucas, 2013).⁷ Notably though, studies in Scandinavian settings typically report insignificant effects (Bjerrum et al., 2011; Munck, Gahrn-Hansen, Sögaard, & Sögaard, 1999; Ranji et al., 2006). One exception is an educational intervention in Norway, which increased the PcV share from 45% to 54% (20%) (Gjelstad et al., 2013), though part of this may be a Hawthorne effect.

4 | INSTITUTIONAL BACKGROUND

4.1 | Organization of primary care

In Sweden, health care is the responsibility of 21 independent county councils. Primary care is typically provided in group practices, with a handful of general practitioners and a team of nurses, social workers and behavioral therapists, physiotherapists, and midwives. Solo practices are very rare (Anell, Glengård, & Merkur, 2012a). Primary care providers are referred to as *primary care centers* in this paper. The counties operate most of the about 1,500 primary care centers, though an increasing share (about 40% as of 2013) is private.

In all counties, patients pay visit fees and part of the price for outpatient drugs,⁸ but income taxes are the dominant source of funding. Each county sets its own tax rate and designs the reimbursement scheme for primary care centers. The reimbursement is a mix of capitation, FFS, and P4P. Capitation is the dominant payment type, accounting for 70–98% of reimbursement in most counties during the sample period (one exception is Stockholm county, where the capitation share was 40%). The remaining part of revenues is predominantly FFS (Anell et al., 2012a). P4P is a complementary reimbursement, which gained in popularity during the past decade: The number of counties using P4P increased from 5 to 20 between 2006 and 2012. Incentives are commonly tied to, for example, prescription targets, patient ratings, registrations in quality registers, and vaccination rates (Anell, Nylander, & Glengård, 2012b). In 2012, P4P accounted for 1–5% of total reimbursement. Notably, the P4P reimbursement is determined by the performance of the care center, that is, the payment goes into the care center's budget. Physicians typically earn a fixed salary, and their reimbursement is not directly affected by the P4P incentive.

4.1.1 | PcV-related P4P

To increase PcV's share of RTI antibiotics to children aged 0–6 years, eight counties introduced P4P during our sample period (Table 1). The counties developed their P4P indicators independently but chose similar performance measures, with only minor variations regarding the definition of “other” RTI antibiotics and (in two counties) restrictions of included diagnoses. The indicators were likely inspired by Strama's national PcV target (see Section 2).

Because each county developed its own scheme, P4P targets differed between counties. Most used a target close to 80%. Stockholm county used a reference target of 70%, with bonus/penalties linearly related to deviations from this level (bonuses/penalties for higher/lower performance). Bonuses were linearly related to performance also in Kronoberg (in the range of 65–80% PcV). Another notable exception is Halland, using penalties rather than bonuses, with 69% PcV as the threshold.

⁶Martens, Werkhoven, Severens, and Winkens (2007) examine an antibiotics-related financial incentive that was paid ex ante (thus unrelated to performance). They found an advantageous but short-lasting effect.

⁷For recent studies, see Meeker et al. (2014); Meeker, Linder, and Fox (2016); Hallsworth et al. (2016).

⁸Patients pay the full price for prescription drugs up to a yearly cap, above which the county subsidizes a progressively larger part. Expenditure above an upper cap (currently about 240 euros per year) are fully subsidized. The caps and subsidy rates for prescription drugs are nationally regulated and thus the same in all county councils, as are pharmacies' prices of subsidized drugs.

TABLE 1 Prevalence of PcV P4P by year and county council

Year	Blekinge	Dalarna	Skåne	Västernorrland	Halland	Kronoberg	Stockholm	Sörmland
2006								
2007								
2008								
2009			X					
2010	X	X	X	X				
2011	X	X	X	X	X		X	X
2012			X	X		X	X	X
2013				X		X	X	X

Notes. X = county council uses a P4P indicator related to the PcV share. PcV share = Penicillin V prescriptions' share of total consumption of antibiotics prescribed for respiratory tract infections (ATC codes: J01CE02/J01AA02/J01CA04/J01CR02/J01DB-DE/J01FA), children 0–6 years. PcV = penicillin V; P4P = pay-for-performance.

We unfortunately lack information about actual payments, but back-of-the-envelope calculations suggest that for a care center of average size, the PcV-related P4P would account for between 0.05% and 1.2% of total reimbursement. Obviously, the impact on the personal incomes of prescribing physicians—who are typically salaried—was negligible. Though the care centers were probably aware that P4P was a minor source of revenue, it was likely difficult to estimate its size. Indeed, in the county of Skåne, the bonus size depended on the number of care centers meeting the target, rendering such calculations impossible to make.⁹

4.1.2 | Potentially confounding policies

Simultaneously with the national PcV target in 2009, the Strama network launched targets related to the total antibiotic use and prescriptions for urinary tract infections. Some counties—e.g., some using PcV P4P—also tied P4P to these other measures.¹⁰ Further, in 2011–2014, the national government ran a P4P program (*Patientsäkerhetsnsatsningen*) targeting the counties (i.e., not the primary care centers). The overarching goal of that program was to reduce the overall number of antibiotic prescriptions. But to receive a bonus, counties also had to institute local Strama groups that should promote antibiotic stewardship (Swedish Government, 2010).

5 | DATA

5.1 | Data sources

The Swedish Prescribed Drug Register comprises all purchases of prescribed pharmaceuticals since July 2005. The data are reported by the pharmacies, the only retailers of antibiotics in Sweden. Our main analysis uses yearly data on RTI antibiotics consumption of all children aged 0–6 years in 2006–2013. Further analyses use data on the consumption of all antibiotics, and some other substances, for residents of any age (Section 7.3 and Appendix A.1). The register records the patients' municipality of residence; thus, we use data aggregated at municipality level. County borders never cut across a municipality, that is, all citizens of a given municipality "belong" to the same county. We exclude one municipality (Norrtälje), where the responsibility for health care is shared by the county and municipality and for which we lack information about the reimbursement system. Our sample contains 2,312 observations: 289 municipalities in 21 county councils observed yearly from 2006 to 2013.

There are some limitations of the register: It only includes redeemed prescriptions and holds no diagnosis information. Further, the data include prescriptions from all provider types (primary care, outpatient specialist care, and dentistry).¹¹ The register holders advised us against using the variable that indicates provider type, as they do not check its quality. However, assuming that the provider type variable gives a rough indication (fewer than 2% of prescriptions lack information on provider type),

⁹Some counties abolished the indicator after a few years. According to personal communication with county administrators, some of the reasons were dissatisfaction with the measure and/or a wish to prioritize other aspects of antibiotic stewardship. In Halland, the administration worried that care centers would game the system by issuing more PcV prescriptions.

¹⁰See online Appendix. The reason we did not evaluate the P4Ps related to urinary tract infections is that we lack outcome data for the relevant patient group (women). The main reason we do not evaluate P4P indicators for total consumption is that they were introduced simultaneously with a national incentive with the same purpose (see main text), meaning that we could only study the marginal effect of adding one incentive to another. Such an analysis would be difficult to interpret, and the policy implications are limited.

¹¹In specialist care, antibiotics-related P4P is rare: We know of only one example, introduced in 2014.

TABLE 2 PcV share by treatment group

	2006		2013		Obs.
	Mean	SD	Mean	SD	
Ever P4P	0.585	0.086	0.706	0.043	109
Never P4P (Control)	0.638	0.078	0.743	0.055	181
Unconditional	0.612	0.086	0.724	0.053	290

Note. The first two rows show the means and standard deviations (SD) of the PcV share^a by year, conditional on treatment group status (*ever* P4P (treatment) or *never* P4P (control)). The third row shows the unconditional mean and SD by year. Municipality-level data weighted by municipality population size. ^aPcV share = penicillin V prescriptions' share of total consumption of antibiotics prescribed for respiratory tract infections (ATC codes: J01CE02/J01AA02/J01CA04/J01CR02/J01DB-DE/J01FA), children 0–6 years. P4P = pay-for-performance.

around 80% of children's RTI antibiotics are prescribed in primary care (similar in P4P and non-P4P counties) and dentistry accounts for a negligible share.¹²

Our data also include official statistics on municipality characteristics in terms of demography, income and educational level, and county council level policy variables collected by ourselves (Appendix A.2). We also use data from the Social Insurance Agency on the number of cases of temporary parental leave due to care-taking of sick children.

5.2 | Descriptive statistics

Our main dependent variable is the *PcV share*, that is, the number of PcV prescriptions divided by the total number of RTI antibiotics prescriptions to children between 0 and 6 years of age. Appendix A.1 lists included substances. Table 2 shows the average PcV share in 2006, before any county had introduced PcV P4P, and in 2013. The group that *ever* used P4P had a lower PcV share at both points in time, though only the latter difference is significant.¹³ The table further indicates that the PcV share increased over time in both groups. This is confirmed by Figure 1, which shows the yearly growth of the PcV share in relation to the initial sample year, again by treatment group status. The two groups experienced the same average growth between 2006 and 2008, the last year before any county introduced P4P ($\chi^2 = 2.62$, $p = 0.269$ in joint test). Notably, before 2011, only four of the counties in the P4P group had actually implemented P4P; in fact, up to and including 2011, the trend reflects a mix of the growth in counties using P4P and counties not yet having introduced P4P (cf. Table 1). The P4P group diverged substantially from the control group in 2011, the year when the number of councils using P4P rose substantially (from 4 to 7).

To see more clearly how the PcV share relates to the implementation of P4P, Figure 2 plots each treated municipality's PcV share in event time (i.e., $t = 0$ is the P4P implementation year). Each circle represents a municipality year, and the circle size is proportional to the municipality's population size. The number of municipalities at each point in (event) time vary according to the length of time we observe them before/after the implementation of P4P; for example, Stockholm, the largest circle, is observed 6 years before and 2 years after. The figure also shows two regression lines: one for the period up to $t = 0$, one for $t = 0$, and onwards. The vertical distance between the endpoint of the first line and the starting point of the second indicates that the PcV share jumps when P4P is introduced.¹⁴

¹² The information disaggregated by provider type mentioned here was provided to us by the Swedish eHealth Agency, one of the register holders, for 2011–2013. For the population as a whole, irrespective of age, estimates from other sources suggest that primary care accounts for about 50% of outpatient antibiotics (Mölstad, Erntell, & Hanberger, 2008). The difference between the estimates is reasonable, as RTI antibiotics heavily dominate children's consumption and RTI typically is treated by primary care (Section 2).

¹³ The additional data disaggregated by provider type (unfortunately not available before 2011, see Section 5.1 and footnote¹²) suggest that there is practically no difference between the PcV shares of primary care providers in P4P and non-P4P counties. Instead, the difference is due to specialists in P4P counties prescribing more broad-spectrum drugs than specialists in non-P4P counties.

¹⁴ The online appendix shows that the jump is attenuated if we shift back the break 1 year, suggesting that something happened at $t = 0$ rather than before.

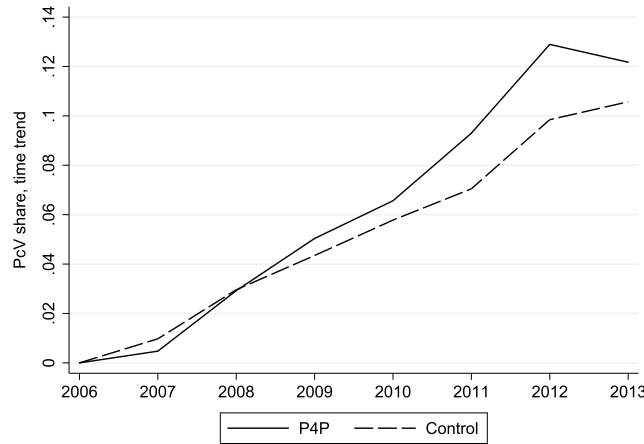


FIGURE 1 The figure shows the estimated year effects from regressions of the penicillin V (PcV) share on a vector of year dummies and municipality fixed effects. Separate estimations for municipalities in county councils that *ever* used pay-for-performance (P4P) and municipalities in county councils that *never* used P4P (control). The estimates are weighted by population size

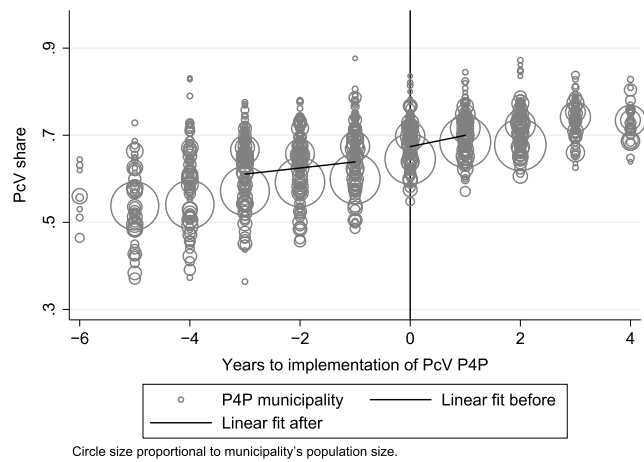


FIGURE 2 The circles show each pay-for-performance (P4P) municipality's penicillin V (PcV) share, plotted against the time (in years) to P4P implementation ($t = 0$). Circles are proportional to municipality population size

6 | EMPIRICAL STRATEGY

6.1 | Identification issues

We use a DID strategy to estimate the P4P effect (Angrist & Pischke, 2008). The implementation of P4P was exogenous for primary care centers, which have no influence over their county's reimbursement system. But as the adoption of P4P was a choice variable for the counties, a discussion of identification issues is still warranted.

First, the decision to adopt P4P may relate to the ability to improve. Recall that the initial PcV share was relatively low in the P4P counties. This may mean that care centers in these counties could respond to the policy with relative ease, for example, if they had a large stock of patients that could be given PcV instead of broad-spectrum antibiotics. In this case, we overestimate the effect in other contexts, though a treatment-effect-on-the-treated interpretation is still valid.

Second, the municipalities in the treatment group might have increased their PcV share more, even if the counties they belonged to had not implemented P4P. For instance, a positive effect may reflect catching-up from a low initial level. If there are catching-up effects, then we would expect a divergence between P4P and non-P4P counties already before P4P was introduced. We test this formally in Section 7.3.

Third, we may be confounding the P4P effect with the effects of other policies or differential RTI morbidity. To account for underlying features and trends in individual municipalities, our specification includes municipality fixed effects and linear

trends (cf. Li, Hurley, DeCicca, & Buckley, 2014).¹⁵ We also include year dummies to capture national trends. In Section 7.2, we further account for background characteristics, first by adding time-varying covariates to the baseline specification, and second by assigning greater weight to control group observations that are similar to the P4P municipalities in terms of (a) the covariates and (b) the PcV share in 2006 (this estimation uses data for 2007–2013). We use the entropy balancing algorithm developed by Hainmueller (2012) to obtain weights. The entropy weights change yearly, as the DID strategy implies that P4P municipalities belong to the control group before they introduce P4P. In Section 7.3, we run placebo estimations to assess the probability of confounding with other policies and differential morbidity.

6.2 | Estimation

We estimate the following equation:

$$y_{mct} = \alpha \times (PcV P4P_{mct}) + \theta_{mc} \times t + \lambda_t + \mu_{mc} + \varepsilon_{mct}, \quad (1)$$

where y_{mct} is the value of the dependent variable in year t for municipality m in county council c . In our main estimations, y_{mct} equals the PcV share of children aged 0–6 years, but we also use the same model to consider other outcomes, for example, the number of PcV and other RTI antibiotics prescriptions for children (Sections 7 and 7.3 and Appendix A.1). μ_{mc} are municipality fixed effects and λ_t year fixed effects. θ_{mc} is a vector of municipality-specific coefficients on the linear trend variable t . ε_{mct} is an idiosyncratic error term.

The dummy $PcVP4P_{mct}$ equals one for observations affected by P4P, and thus, α captures the P4P effect. Because some counties removed P4P after a few years, we adopt two treatment definitions. First, we classify municipalities as affected by P4P in years when their county employed P4P and as unaffected otherwise. To account for the possibility that P4P had a persistent effect on prescription routines (cf. Celhay, Gertler, Giovagnoli, & Vermeersch, 2015), we then augment the specification with a dummy (*post PcV P4P*) for observations previously affected by P4P. This leads to our second treatment definition, in which $PcVP4P_{mct}$ capture observations *currently or previously* affected by P4P.

The model is estimated with Stata module *xivreg2* (Schaffer, 2010). Standard errors are clustered at the county level using Stata's cluster option (Rogers, 1993). Because the small number of clusters (21) may lead to underestimated standard errors (Cameron & Miller, 2015), we also subject our preferred specification to the wild cluster bootstrap with 999 replications (Cameron, Gelbach, & Miller, 2008).¹⁶ Regressions are weighted by the population size.

Acknowledging that our main dependent variable is a fraction bounded between 0 and 1, we estimate a fractional response model as a sensitivity check. We estimate the model proposed by Papke and Wooldridge (2008), which accounts for unobserved time-invariant features by conditioning on time-averages of covariates instead of municipality fixed effects. We calculate the average partial effect according to the formula for nonlinear DID models in Puhani (2012) and bootstrap the average partial effect with 999 replications.

7 | RESULTS

7.1 | Main results

In column 1 of Table 3, municipalities are classified as treated during the years their county employs P4P. The P4P effect of 1.1 percentage points is positive and statistically significant ($p = .008$). In column 2, we add the *post PcV P4P* dummy variable. The P4P effect increases to 1.8 percentage points ($p = .001$), and the estimate on *post PcV P4P* indicates persistence. Because the two dummies are not significantly different from each other ($p = .48$), we replace them with a joint dummy for municipalities in counties that currently or previously have used P4P. Column 3 thus shows our preferred specification, in which the P4P effect is estimated at approximately 1.8 percentage points.

It is possible to increase the PcV share simply by prescribing more PcV. If that is the mechanism behind the increased PcV share, it is hardly consistent with antibiotic stewardship. In the next columns of the table, we change the dependent variable to the number of redeemed prescriptions of PcV (column 4) and other RTI antibiotics (column 5), respectively. Though insignificant, the estimates are of opposing sign and approximately same size. The zero net effect is confirmed in column 6, in which the

¹⁵Unfortunately, we have too few pre-P4P data points to apply the synthetic control method (Kreif, Grieve, Hangartner, & Turner, 2015).

¹⁶We use v.2.0.0 of *cgmmwildboot*, a Stata module developed by Judson Caskey.

TABLE 3 Baseline estimations

<i>y</i>	(1) <i>PcV share</i>	(2) <i>PcV share</i>	(3) <i>PcV share</i>	(4) <i>PcV</i>	(5) <i>RTI broad</i>	(6) <i>total</i>
PcV P4P	0.0106*** (0.00399)	0.0180*** (0.00562)	0.0176*** (0.00550)	0.525 (0.558)	−0.509 (0.322)	0.017 (0.792)
post PcV P4P		0.0209** (0.00822)				
Mean (<i>y</i>)	0.69	0.69	0.69	20.1	9.9	30.0
Observations	2,312	2,312	2,312	2,312	2,312	2,312
Municipalities	289	289	289	289	289	289
<i>R</i> ²	0.009	0.015	0.015	0.002	0.004	0.000

Note. The table shows baseline estimations of the effect of P4P related to the PcV share (= penicillin V prescriptions' share of total consumption of antibiotics prescribed for respiratory tract infections (ATC codes: J01CE02/J01AA02/J01CA04/J01CR02/J01DB-DE/J01FA), children 0–6 years). In columns 1–3, the dependent variable is the PcV share. In columns 4–6, the dependent variables are the number of PcV prescriptions (col. 4; ATC code J01CE02), other RTI antibiotics prescriptions (col. 5; ATC codes: J01AA02/J01CA04/J01CR02/J01DB-DE/J01FA), and the total of PcV and other RTI antibiotics prescriptions (col. 6) in ages 0–6 years. In columns 1 and 2, *PcV P4P* = 1 for observations in county councils that currently use P4P. In column 2, *post PcV P4P* = 1 for observations in county councils that have previously used P4P. In column 3–6, *PcV P4P* = 1 for observations in county councils that *currently use or have previously used* P4P. All specifications include municipality and year fixed effects and municipality-specific linear trends and are weighted by the municipality population size. Standard errors clustered by county council in parentheses. PcV = penicillin V; P4P = pay-for-performance; RTI = respiratory tract infections. *** $p < .01$, ** $p < .05$, * $p < .1$.

dependent variable is the sum of PcV and other RTI antibiotic prescriptions. Thus, the main result reflects a substitution of PcV instead of broad-spectrum antibiotics, exactly the behavior that the incentive was designed to stimulate.

An objection against interpreting the increasing PcV share of the treatment group as due to P4P is that it could reflect a contemporaneous drop in RTI morbidity. That is, with fewer RTI patients, there would be fewer complicated cases and thus less need for broad-spectrum drugs. But if morbidity had decreased, we would also expect fewer prescriptions in total, contrary to what column 6 shows.

7.2 | Sensitivity checks

Column 1 of Table 4 shows that the significant P4P effect remains when we employ the wild cluster bootstrap.¹⁷ The specification reported in column 2 includes covariates, leading to a slightly attenuated and less precise ($p = .030$) estimate. Of the covariate estimates, reported in the online Appendix, only one is significant (the mean personal income). Notably, the estimate on a dummy for control group municipalities using other antibiotics-related P4P indicators (Section 4.1.2) is small and insignificant.

In columns 3–4, we remove the municipality linear trends. The P4P effect increases slightly but is only significant when covariates are included (column 4). In contrast to the preferred model, the covariates thus improve precision when trends are excluded. This is reasonable: The covariates should be more important in the absence of trends, which pick up much of the variation.¹⁸ Column 5 shows unweighted regression results. The estimate is slightly smaller than the baseline, suggesting a stronger effect for larger municipalities. In column 6, we apply the entropy balancing weights; the P4P effect is even larger than in the baseline.¹⁹ This finding aligns with the smaller estimate in column 5: The balancing algorithm gives higher weight to large control municipalities, as the average population size is smaller in the control group than in the treatment group.²⁰ Indeed, only 5% of the control observations (about 100) get weights that are at least as high as the unit weight of treated observations. Finally, in column 7, we show the results from the Papke–Wooldridge fractional response model. The point estimate is slightly larger than the baseline, though it is just above the 10% significance level ($p = .118$). A disadvantage of the fractional response model is that it does not account for municipality fixed effects or linear trends.

¹⁷The reason for the low bootstrap p -value is that none of the 999 bootstrap samples yielded a smaller t -statistic.

¹⁸We partial out the trends in the reported estimations; otherwise R^2 is 0.52 in our preferred specification.

¹⁹Note that this estimation sample starts in 2007. Excluding 2006 from our baseline specification does not affect the P4P estimate. But for this reason, and because a few control municipalities receive relatively large weights, we prefer our baseline specification.

²⁰The online Appendix shows the balance of the treated and entropy-weighted control group in terms of mean, variance, and skewness.

TABLE 4 Sensitivity

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
PcV P4P	0.0176*** (0.000)	0.0159** (0.0073)	0.0212 (0.0157)	0.0192* (0.0112)	0.0153** (0.0068)	0.0252*** (0.0063)	0.0189 (0.0120)
CI lower	0.00698						
CI upper	0.0282						
Municipality FE	X	X	X	X	X	X	
Year FE	X	X	X	X	X	X	X
Municipality trends	X	X			X	X	
Covariates		X		X			X
Population weights	X	X	X	X		X	X
Observations	2,312	2,304	2,312	2,304	2,312	2,016	2,312
Municipalities	289	288	289	288	289	289	289
R ²	0.914	0.020	0.027	0.095	0.004	0.059	

Note. Sensitivity tests for the main dependent variable, that is, the PcV share (= penicillin V prescriptions' share of total consumption of antibiotics prescribed for respiratory tract infections (ATC codes: J01CE02/J01AA02/J01CA04/J01CR02/J01DB-DE/J01FA), children 0–6 years; mean = 0.69). (1) Wild cluster bootstrap on the preferred specification (999 replications, *p*-value in parenthesis; CI lower/upper = 95% confidence interval); (2) preferred specification + time-varying covariates (share of children and elderly, population share with secondary/tertiary education, log population and mean income in municipality, I (control county with other antibiotic-related P4P), I (implementation year of entry/choice reform), I (drug cost responsibility of care centers)); (3) preferred specification but excluding municipality linear trends; (4) preferred specification, excluding municipality linear trends, and including covariates; (5) preferred specification not using population size weights; (6) preferred specification using entropy balancing weights (2007–2013 sample; see the online Appendix for entropy balancing results); (7) fractional response model. Columns 2–7: standard errors clustered by county council in parentheses (cluster-bootstrapped with 999 replications in col. 7). P4P = pay-for-performance; FE = fixed effects. ****p*<.01, ***p*<.05, **p*<.1.

7.3 | Confounding

We next address the identification threats discussed in Section 6.1. To check if the estimate picks up already existing differential trends, we include placebo dummies for the 2 years before P4P was introduced. Column 1 of Table 5 shows that the P4P effect is hardly affected and the placebo estimates are small and insignificant. Importantly, this speaks against worries that Strama's national PcV target drives the effect: The target was launched during the placebo years for seven of the eight P4P counties; thus, we would expect large placebo effects if the national target drove the effect.

One concern is that the estimated effect picks up *other* county-specific policy changes. For instance, recent patient choice reforms might have affected physicians' sensitivity to patients' demands. To check this, we estimate models of placebo outcomes unrelated to RTI (columns 2–5): prescriptions of emollients and protectives with no specific therapeutic effect, antidepressants, sleeping pills, and the number of statin users (see Appendix A.1 for exact definitions). The outcome variable in column 2 includes substances that can be bought over the counter, but patients may still request a prescription to reach the annual cost ceiling in the drug subsidy system. Antidepressants, sleeping pills, and statins are used by many Swedes (around 10% of the population, Socialstyrelsen 2016). There is no significant "treatment effect" on any of these outcomes.

Another concern is that the result may be driven by county-level incentives tied to other antibiotics indicators (Section 4.1.2). We do not think that this is an important issue, as the PcV share is clearly most related to the PcV-related P4P. Also, a dummy variable for control group municipalities using P4P for other antibiotics indicators has no impact on our estimate (results not shown).

A related issue is that Strama's national targets and the national government's incentive program might have had particularly strong impact on the P4P counties. Likely, the national policies inspired the eight counties to implement antibiotics-related P4P. Although that does not mean that the prescribers in these counties reacted differently to national policies than prescribers in other counties, we cannot fully rule it out. However, two pieces of evidence suggest that the national incentive program was of limited importance. First, one potential confounding mechanism is that the national program encouraged counties to promote antibiotic stewardship via local Strama groups, which may have focused on disseminating RTI guidelines in places where the PcV share was low. If this mechanism is the real cause of the treatment effect, it should be driven by the P4P municipalities with low initial PcV shares. That is not the case, though: When we interact the treatment variable with a dummy for municipalities whose PcV share was below the national median in 2006, the interaction is small, insignificant, and far from soaking up the baseline effect (column 6).²¹ However, this still does not rule out that the Strama groups might have been more effective in P4P counties than in non-P4P counties.²²

²¹Note that the estimation sample starts in 2007.

²²Although when asked directly, experts at the Public Health Agency of Sweden, who monitored the Strama groups' work during the national initiative, did not point out any of the P4P counties as exceptionally successful (personal communication, Nov. 21 2016).

TABLE 5 Counting checks

<i>y</i>	(1) <i>PcV share</i>	(2) <i>D02A</i>	(3) <i>N06AA</i>	(4) <i>N05C</i>	(5) <i>statin</i>	(6) <i>PcV share</i>	(7) <i>NonRTIAb</i>	(8) <i>Nasal prep.</i>	(9) <i>Cough & cold</i>	(10) <i>Temp leave</i>
PcV P4P	0.0179* (0.00913)	1.808 (1.141)	-0.628 (0.446)	-2.415 (2.288)	0.193 (0.543)	0.0155** (0.00781)	-0.109 (1.822)	0.417 (0.579)	-3.999 (2.794)	0.827 (2.166)
Placebo t-1	0.000607 (0.00790)									
Placebo t-2	0.000426 (0.00444)									
PcV P4P × P50						0.00342 (0.00989)				
Mean of <i>y</i>	0.69	97.3	56.0	496.8	97.9	0.69	350.5	116.0	159.0	184.2
Observations	2,312	2,312	2,312	2,312	2,312	2,023	2,312	2,312	2,312	2,312
<i>R</i> ²	0.016	0.008	0.001	0.001	0.000	0.018	0.000	0.000	0.013	0.001
Municipalities	289	289	289	289	289	289	289	289	289	289

Note. (1) Preferred specification including two placebo year effects, that is, dummy variables for the 2 years before implementation of P4P. (2)–(5) Preferred specification with various placebo substances as dependent variables (per 1,000 residents): (2) D02A = prescriptions of emollients and protectives with no specific therapeutic effect. (3) N06AA = prescriptions of antidepressants. (4) N05C = prescriptions of sleeping pills; (4) statin = number of statin users (ATC codes: C10AA, C10BA). (6) Preferred specification including an interaction between P4P and a dummy for municipalities whose PcV share⁺ in 2006 was below the national median (2007–2013 sample). (7) Preferred specification with dependent variable = all antibiotic prescriptions except RTI⁺⁺ antibiotic prescriptions for children aged 0–6 years (per 1,000 residents). (8)–(9) Preferred specification using other dependent variables related to RTI⁺⁺ morbidity (per 1,000 residents): (8) prescriptions of nasal preparations (ATC R01); (9) Prescriptions of cough and cold preparations (ATC R05). (10) Cases of temporary parental leave due to care-taking of sick children. Standard errors clustered by county council in parentheses. + PcV share = penicillin V prescriptions' share of total RTI⁺⁺ antibiotic consumption of children 0–6 years. ++ RTI = respiratory tract infections. ****p* < .01, ***p* < .05, **p* < .1.

A second indication that the P4P counties did not react strongly to the national incentive program is that their overall antibiotic consumption did not decrease more than in other counties. Reducing overall consumption was the most salient target of the national program. If P4P counties were relatively responsive to the national program, we would expect larger reductions there.²³ But we saw in column 6 of Table 3 that the total RTI antibiotic consumption of children developed similarly in P4P and non-P4P counties, and column 7 of Table 5 shows that the same is true for the remaining part of total antibiotic consumption.

Lastly, to further check that differential RTI morbidity is not a confounder, we consider additional morbidity correlates: the number of prescriptions of nasal preparations (column 8) or cough and cold preparations (9), and the number of cases of parental work absence due to sick children (10). All estimates are insignificant.

8 | CONCLUDING REMARKS

We find that P4P encourages primary care practices to increase the share of narrow-spectrum antibiotics prescribed to children with RTI, and there are no signs that physicians gamed the system by issuing more prescriptions overall. As the financial incentives were small and not tied to physicians' salaries, the P4P effect was likely driven by something else than just a desire to boost income (cf. McDonald, White, & Marmor, 2009). A plausible mechanism is that P4P made the antibiotics issue more salient, thus lowering physicians' psychological barriers to changing prescription routines (Celhay et al., 2015). After having established a routine, another impetus is required to reverse it. This may explain why the effect persisted in counties that later removed the monetary incentive.

Among the limitations of the study is that it only addresses the selection of antibiotic type. Another important part of antibiotic stewardship is to limit unnecessary prescriptions. A further limitation is that the studied antibiotics may also be prescribed for other reasons than RTI. Notably, the limitation was built into the P4P systems of most of the P4P counties, which relied on the same data to follow up care center performance (one of the reasons why the targets were set far below 100%). To our knowledge, there were no incentives for other diagnoses that may affect the PcV share. A related limitation is that our data include prescriptions from both primary and secondary care, though primary care is the most important source of RTI prescriptions to children, and the share of RTI specialists was very similar in P4P and non-P4P counties throughout the sample period (Socialstyrelsen, 2016). Finally, our empirical strategy does not allow us to rule out that the differences between P4P and comparison counties reflect differential reactions to national targets and incentives that were introduced during the study period. However, we show that the P4P counties did not react differently to a more salient aspect of the national incentive program to promote antibiotic stewardship (a target for the total number of antibiotics prescriptions).

Our baseline effect of 1.8 percentage points amounts to 3% of the pre-P4P PcV share, or 20% of a standard deviation. It is similar to the improvement found in a previous study of P4P in the UK hospital sector (McDonald et al., 2015) but smaller than the effects in recent studies from China, where the overall prescription rate is considerably higher (Yip et al., 2014; Gong et al., 2016). According to the European Centre for Disease Prevention and Control, Sweden is among the European countries with the lowest overall antibiotics consumption and highest reliance on narrow-spectrum drugs (see e.g., ECDC, 2016). Possibly, P4P yields substantial gains when applied in high-prescribing contexts but more limited effects elsewhere. This would align with the observation that non-financial interventions to improve antibiotic stewardship have typically had larger impact outside Scandinavia (e.g., Ranji et al., 2006). In the global fight against resistance, it is of foremost importance to find policies for high-prescribing contexts, where resistance is higher and more likely to develop further (Goossens et al., 2005); we provide a conservative estimate for that policy discussion.

In relation to the general P4P literature, the magnitude of our estimate is similar to the typical impact of P4P for process measures in other contexts (Van Herck et al., 2010; Ogundeji et al., 2016). The result also aligns with the only previous Swedish study on P4P, which concerns diabetes-related process measures (Ödesjö et al., 2015). Our main finding may be particularly interesting for English readers, as a similar indicator was recently introduced in the English Quality and Outcomes Framework scheme (NHS, 2016).²⁴ The Swedish and English primary care systems share many features, for example, publicly financed general practitioners operating in group practices with average patient stocks of similar size (Santos, Gravelle, & Propper, 2017; Anell, 2015). However, there are also notable differences, such as the higher reliance on narrow-spectrum drugs and larger share of publicly owned practices in Sweden. P4P also accounts for a much larger share of total reimbursement in the UK, although the importance of each P4P indicator is small in both countries.

²³Not least as they prescribe above the national average: in 2010, 383 prescriptions per 1,000 residents in P4P counties, 357 in control counties ($p = .000$).

²⁴The Quality and Outcomes Framework target is that the share of certain broad-spectrum antibiotics in total prescriptions should either be less than 10%, or reduced by 20% from the 2014/2015 value.

ACKNOWLEDGEMENTS

We are grateful to Malin Bredenberg for excellent research assistance. The article has benefited from useful suggestions from two anonymous referees and by comments from Marika Cabral, Jon Sussex, Carl Hampus Lyttkens, participants at the 2016 IFN Vaxholm Conference, HESG in Manchester, and seminar participants at the Lund University Division of Health Economics and at SFI.

CONFLICT OF INTEREST

There were no financial and personal relationships between the authors and others that might have biased the work.

ETHICAL STATEMENT

The study deals with aggregated data, and it was impossible to identify any particular individual. Therefore, an approval from an ethical committee was not required.

REFERENCES

- André, M., Vernby, A., Odenholt, I., Lundborg, C. S., Axelsson, I., Eriksson, M., ... Mölstad, S. (2008). Diagnosis-prescribing surveys in 2000, 2002 and 2005 in Swedish general practice: Consultations, diagnosis, diagnostics and treatment choices. *Scandinavian Journal of Infectious Diseases*, 40(8), 648–654.
- Anell, A., Glengård, A. H., & Merkur, S. (2012a). Sweden: Health system review. *Health Systems in Transition*, 14(5), 1–159.
- Anell, A. (2015). The public–private pendulum – Patient choice and equity in Sweden. *New England Journal of Medicine*, 372(1), 1–4.
- Anell, A., Nylinder, P., & Glengård, A. (2012b). *Vårdval i primärvården: Jämförelse av uppdrag, ersättningsprinciper och kostnadsansvar*. Stockholm: Sveriges Kommuner och Landsting.
- Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics* (1st ed). Princeton: Princeton University Press.
- Arnold, S. R., & Straus, S. E. (2005). Interventions to improve antibiotic prescribing practices in ambulatory care. *The Cochrane Database of Systematic Reviews* (4), CD003539.
- Ashworth, M., White, P., Jongasma, H., Schofield, P., & Armstrong, D. (2015). Antibiotic prescribing and patient satisfaction in primary care in England: Cross-sectional analysis of national patient survey data and prescribing data. *British Journal of General Practice*, 66(642), e40–e46. <https://doi.org/10.3399/bjgp15X688105>
- Bjerrum, L., Munck, A., Gahrn-Hansen, B., Hansen, M. P., Jarbol, D. E., Cordoba, G., ... Rønning, M. (2011). Health Alliance for prudent antibiotic prescribing in patients with respiratory tract infections (HAPPY AUDIT) - Impact of a non-randomised multifaceted intervention programme. *BMC Family Practice*, 12(1), 52.
- Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, 50(2), 317–372.
- Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics*, 90(3), 414–427.
- Carlet, J., Collignon, P., Goldmann, D., Goossens, H., Gyssens, I. C., Harbarth, S., ... Voss, A. (2011). Society's failure to protect a precious resource: Antibiotics. *The Lancet*, 378(9788), 369–371.
- Celhay, P., Gertler, P., Giovagnoli, P., & Vermeersch, C. (2015). Long run effects of temporary incentives on medical care productivity. (NBER Working Paper 21361).
- Christianson, J. B., Leatherman, S., & Sutherland, K. (2008). Lessons from evaluations of purchaser pay-for-performance programs: A review of the evidence. *Medical Care Research and Review: MCRR*, 65(6 Suppl), 5S–35S.
- Cronin, M. J., Khan, S., & Saeed, S. (2013). The role of antibiotics in the treatment of acute rhinosinusitis in children: A systematic review. *Archives of Disease in Childhood*, 98(4), 299–303.
- Ödesjö, H., Anell, A., Gudbjörnsdóttir, S., Thorn, J., & Björck, S. (2015). Short-term effects of a pay-for-performance programme for diabetes in a primary care setting: An observational study. *Scandinavian Journal of Primary Health Care*, 33(4), 291–297.
- ECDC (2016). <http://ecdc.europa.eu/en/healthtopics/antimicrobial-resistance-and-consumption/antimicrobial-consumption/esac-net-database/Pages/quality-indicators-primary-care.aspx>, Accessed: November 14, 2016.
- ECDC/EMA Joint working group (2009). ECDC/EMA Joint Technical Report: The bacterial challenge: Time to react.
- Eijkenaar, F., Emmert, M., Scheppach, M., & Schöffski, O. (2013). Effects of pay for performance in health care: A systematic review of systematic reviews. *Health Policy*, 110(2-3), 115–130.
- Fleming-Dutra, K. E., Hersh, A. L., Shapiro, D. J., Bartoces, M., Enns, E. A., File, T. M. Jr., ... Hicks, L. A. (2016). Prevalence of inappropriate antibiotic prescriptions among US ambulatory care visits, 2010–2011. *JAMA*, 315(17), 1864–1873.
- Fogelberg, S. (2014). Effects of competition between healthcare providers on prescription of antibiotics. (IFN Working Paper 949). Research Institute of Industrial Economics.
- Gjelstad, S., Høye, S., Straand, J., Brekke, M., Dalen, I., & Lindbæk, M. (2013). Improving antibiotic prescribing in acute respiratory tract infections: Cluster randomised trial from Norwegian general practice (prescription peer academic detailing (Rx-PAD) study). *BMJ (Clinical research ed.)*, 347, f4403.
- Gong, S., Qiu, X., Song, Y., Sun, X., He, Y., Chen, Y., ... Xia, H. (2016). Effect of financially punished audit and feedback in a pediatric setting in China, within an antimicrobial stewardship program, and as part of an international accreditation process. *Frontiers in Public Health*, 4(99), eCollection 2016. <https://doi.org/10.3389/fpubh.2016.00099>
- Goossens, H., Ferech, M., Coenen, S., Stephens, P., & European Surveillance of Antimicrobial Consumption Project Group (2007). Comparison of outpatient systemic antibacterial use in 2004 in the United States and 27 European countries. *Clinical Infectious Diseases*, 44(8), 1091–1095.

- Goossens, H., Ferech, M., Vander Stichele, R., & Elseviers, M. (2005). Outpatient antibiotic use in Europe and association with resistance: A cross-national database study. *The Lancet*, *365*(9459), 579–587.
- Hainmueller, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political Analysis*, *20*(1), 25–46.
- Hallsworth, M., Chadborn, T., Sallis, A., Sanders, M., Berry, D., Greaves, F., ... Davies, S. C. (2016). Provision of social norm feedback to high prescribers of antibiotics in general practice: A pragmatic national randomised controlled trial. *The Lancet*, *387*(10029).
- Hersh, A. L., Shapiro, D. J., Pavia, A. T., & Shah, S. S. (2011). Antibiotic prescribing in ambulatory pediatrics in the United States. *Pediatrics*, *128*(6), 1053–1061.
- Kaier, K., & Moog, S. (2012). Economic consequences of the demography of MRSA patients and the impact of broad-spectrum antimicrobials. *Applied Health Economics and Health Policy*, *10*(4), 227–234.
- Keith, T., Saxena, S., Murray, J., & Sharland, M. (2010). Risk-benefit analysis of restricting antimicrobial prescribing in children: What do we really know?. *Current Opinion in Infectious Diseases*, *23*(3), 242–248.
- Kreif, N., Grieve, R., Hangartner, D., Turner, A. J., Nikolova, S., & Sutton, M. (2015). Examination of the synthetic control method for evaluating health policies with multiple treated units. *Health Economics*, *25*, 1514–1528.
- Lemienge, M. B., van Driel, M. L., Merenstein, D., Young, J., & De Sutter, A. I. M. (2012). Antibiotics for clinically diagnosed acute rhinosinusitis in adults. *Cochrane Database of Systematic Reviews*, *10*, CD006089. <https://doi.org/10.1002/14651858.CD006089.pub4>
- Li, J., Hurley, J., DeCicca, P., & Buckley, G. (2014). Physician response to pay-for-performance: Evidence from a natural experiment. *Health Economics*, *23*(8), 962–978.
- Martens, J. D., Werkhoven, M. J., Severens, J. L., & Winkens, R. A. G. (2007). Effects of a behaviour independent financial incentive on prescribing behaviour of general practitioners. *Journal of Evaluation in Clinical Practice*, *13*(3), 369–373.
- McDonald, R., Boaden, R., Roland, M., Kristensen, S. R., Meacock, R., Lau, Y.-S., ... Sutton, M. (2015). *A qualitative and quantitative evaluation of the Advancing Quality pay-for-performance programme in the NHS North West*, Health Services and Delivery Research. Southampton (UK): NIHR Journals Library.
- McDonald, R., White, J., & Marmor, T. R. (2009). Paying for performance in primary medical care: Learning about and learning from "success" and "failure" in England and California. *Journal of Health Politics, Policy and Law*, *34*(5), 747–776.
- Meeker, D., Knight, T. K., Friedberg, M. W., Linder, J. A., Goldstein, N. J., Fox, C. R., ... Doctor, J. N. (2014). Nudging guideline-concordant antibiotic prescribing: A randomized clinical trial. *JAMA Internal Medicine*, *174*(3), 425–431.
- Meeker, D., Linder, J. A., Fox, C. R., Friedberg, M. W., Persell, S. D., Goldstein, N. J., ... Doctor, J. N. (2016). Effect of behavioral interventions on inappropriate antibiotic prescribing among primary care practices: A randomized clinical trial. *JAMA*, *315*(6), 562–570.
- Mölstad, S., Erntell, M., Hanberger, H., Melander, E., Norman, C., Skoog, G., ... Cars, O. (2008). Sustained reduction of antibiotic use and low bacterial resistance: 10-year follow-up of the Swedish Strama programme. *The Lancet. Infectious Diseases*, *8*(2), 125–132.
- Mullen, K. J., Frank, R. G., & Rosenthal, M. B. (2010). Can you get what you pay for? Pay-for-performance and the quality of healthcare providers. *The Rand Journal of Economics*, *41*(1), 64–91.
- Munck, A. P., Gahrn-Hansen, B., Sjøgaard, P., & Sjøgaard, J. (1999). Long-lasting improvement in general practitioners' prescribing of antibiotics by means of medical audit. *Scandinavian Journal of Primary Health Care*, *17*(3), 185–190.
- NHS. (2016). Quality premium guidance for 2016/17. <https://www.england.nhs.uk/wp-content/uploads/2016/03/quality-prem-guid-2016-17.pdf> Accessed May 24, 2016.
- Ogundeji, Y. K., Bland, J. M., & Sheldon, T. A. (2016). The effectiveness of payment for performance in health care: A meta-analysis and exploration of variation in outcomes. *Health Policy*, *120*(10), 1141–1150.
- Papke, L. E., & Wooldridge, J. M. (2008). Panel data methods for fractional response variables with an application to test pass rates. *Journal of Econometrics*, *145*(1–2), 121–133.
- Petersen, L. A., Woodard, L. D., Urech, T., Daw, C., & Sookanan, S. (2006). Does pay-for-performance improve the quality of health care? *Annals of Internal Medicine*, *145*(4), 265–272.
- Puhani, P. (2012). The treatment effect, the cross difference, and the interaction term in nonlinear “difference-in-differences” models. *Economics Letters*, *115*(1), 85–87.
- Ranji, S. R., Steinman, M. A., Shojania, K. G., Sundaram, V., Lewis, R., Arnold, S., & Gonzales, R. (2006). *Closing the quality gap: A critical analysis of quality improvement strategies (vol. 4: Antibiotic prescribing behavior)*, AHRQ Technical Reviews No. 9.4. Rockville (MD): Agency for Healthcare Research and Quality (US).
- Rogers, W. H. (1993). Regression standard errors in clustered samples. *Stata Technical Bulletin*, *13*, 19–23.
- Rosenthal, M. B., & Frank, R. G. (2006). What is the empirical basis for paying for quality in health care? *Medical Care Research and Review: MCRR*, *63*(2), 135–157.
- Santos, R., Gravelle, H., & Propper, C. (2017). Does quality affect patients' choice of doctor? Evidence from England. *Economic Journal*, *127*, 445–494. <https://doi.org/10.1111/eoj.12282>
- Schaffer, M. E. (2010). xtivreg2: Stata module to perform extended IV/2SLS, GMM and AC/HAC, LIML and k-class regression for panel data models. <http://ideas.repec.org/c/boc/bocode/s456501.html>. (Accessed June 12 2017)
- Scott, A., Sivey, P., Ait Ouakrim, D., Willenberg, L., Naccarella, L., Furler, J., & Young, D. (2011). The effect of financial incentives on the quality of health care provided by primary care physicians. *Cochrane Database of Systematic Reviews*, *9*(9), CD008451. <https://doi.org/10.1002/14651858.CD008451.pub2>
- Smith, S. M., Fahey, T., Smucny, J., & Becker, L. A. (2014). Antibiotics for acute bronchitis. *Cochrane Database of Systematic Reviews*, *3*(3), CD000245. <https://doi.org/10.1002/14651858.CD000245.pub3>

- Socialstyrelsen (2016). <http://www.socialstyrelsen.se/statistik/statistikdatabas/halsoochsjukvardspersonal>. (Accessed November 14, 2016).
- Socialstyrelsen (2016). Statistics on pharmaceuticals 2015. <http://www.socialstyrelsen.se/publikationer2016/2016-4-25>. (Accessed August 18, 2016).
- Spinks, A., Glasziou, P. P., & Del Mar, C. B. (2013). Antibiotics for sore throat. *The Cochrane Database of Systematic Reviews*, 11, CD000023.
- Spurling, G. K., Doust, J., Del Mar, C. B., & Eriksson, L. (2011). Antibiotics for bronchiolitis in children. *The Cochrane Database of Systematic Reviews*, 6, CD005189.
- Swedish Government. (2010). Bemyndigande att underteckna en överenskommelse om patientsäkerhetsåtgärder 2011. <http://www.regeringskansliet.se/49b695/contentassets/7b321cef1a814194babb11a788f385ef/bemyndigande-att-underteckna-en-overenskommelse-om-patientsakerhetsatsning-2011>. (Accessed June 12, 2017)
- SWEDRES-SVARM. (2013). *Use of antimicrobials and occurrence of antimicrobial resistance in Sweden*: Public Health Agency of Sweden and National Veterinary Institute. Solna/ Uppsala.
- Town, R., Kane, R., Johnson, P., & Butler, M. (2005). Economic incentives and physicians' delivery of preventive care: A systematic review. *American Journal of Preventive Medicine*, 28(2), 234–240.
- van der Velden, A. W., Pijpers, E. J., Kuyvenhoven, M. M., Tonkin-Crine, S. K., Little, P., & Verheij, T. J. (2012). Effectiveness of physician-targeted interventions to improve antibiotic use for respiratory tract infections. *The British Journal of General Practice*, 62(605), e801–807.
- Van Herck, P., De Smedt, D., Annemans, L., Remmen, R., Rosenthal, M. B., & Sermeus, W. (2010). Systematic review: Effects, design choices, and context of pay-for-performance in health care. *BMC Health Services Research*, 10(1), 247.
- Venekamp, R. P., Sanders, S., Glasziou, P. P., Del Mar, C. B., & Rovers, M. M. (2013). Antibiotics for acute otitis media in children. *The Cochrane Database of Systematic Reviews*, 1, CD000219. <https://doi.org/10.1002/14651858.CD000219.pub3>
- Vodicka, T. A., Thompson, M., Lucas, P., Heneghan, C., Blair, P. S., Buckley, D. I., ... Hay, A. D. (2013). Reducing antibiotic prescribing for children with respiratory tract infections in primary care: A systematic review. *The British Journal of General Practice*, 63(612), e445–454.
- Yip, W., Powell-Jackson, T., Chen, W., Hu, M., Fe, E., Hu, M., ... Hsiao, W. C. (2014). Capitation combined with pay-for-performance improves antibiotic prescribing practices in rural China. *Health Affairs*, 33(3), 502–510.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Ellegård LM, Dietrichson J, Anell A. Can pay-for-performance to primary care providers stimulate appropriate use of antibiotics? *Health Economics*. 2018;27:e39–e54. <https://doi.org/10.1002/hec.3535>

APPENDIX A: VARIABLES: DEFINITIONS AND SUMMARY STATISTICS

A.1 DEPENDENT VARIABLE DEFINITIONS

A.2 COVARIATE DEFINITIONS

We include three covariates adjusting for confounding policies at the county council level. *otherP4P_{mct}* indicates control group observations in county councils that use at least one of the *non-PcV* P4P indicators described in Section 4.1.2 (i.e., P4P related to either total antibiotics prescriptions or to urinary tract infection patients). *choicereform* equals one during the year when a county council implemented a reform that expanded patient choice and instituted free entry in primary care (all county councils implemented such reforms during the sample period), which in some municipalities was associated with a temporary increase in overall antibiotics use (Fogelberg, 2014). Finally, the covariate set also includes the dummy variable *cost responsibility*, which equals one if the county council has delegated the drug budget responsibility to primary care centers.

TABLE A1 Definitions of antibiotics-related dependent variables

Variable	Description	Age group (years)
<i>PcV share</i>	Nominator: <i>PcV</i> (see below) Denominator: <i>total</i> (see below)	0–6
<i>PcV</i>	# per 1,000 of J01CE02	0–6
<i>total</i>	# per 1,000 of J01CE02/J01AA02/J01CA04/J01CR02/J01DB-DE/J01FA	0–6
<i>NonRTIab</i>	# per 1,000 of J01 minus prescriptions in variable <i>total</i> ,	All
<i>RTI broad</i>	# per 1,000 of J01AA02/J01CA04/J01CR02/J01DB-DE/J01FA	0–6

Note. The table shows definitions of the antibiotics-related dependent variables. # per 1,000 = number of redeemed prescriptions per 1,000 residents. *RTI* = respiratory tract infections. J01 is the category for antibiotics in the ATC (Anatomic Therapeutic Chemical) classification system. J01CE02 = phenoxymethylpenicillin (PcV), J01AA02 = doxycycline, J01CA04 = amoxicillin, J01CR02 = amoxicillin and enzyme inhibitor, J01DB-DE cephalosporins, J01FA = macrolides. Data source: The Swedish Prescribed Drug Register.

A.3 SUMMARY STATISTICS: COVARIATES

TABLE A2 Definitions of other dependent variables

Variable	Description	Age group
<i>D02A</i>	# per 1,000 of emollients and protectives with no specific therapeutic effect	All
<i>N06AA</i>	# per 1,000 of antidepressants	All
<i>N05C</i>	# per 1,000 of sleeping pills	All
<i>Statin</i>	# individuals with statins (ATC C10AA, C10BA), number per capita	All
<i>Nasal prep.</i>	# per 1,000 of nasal preparations (ATC R01)	All
<i>Cough & cold</i>	# per 1,000 of cough medicine (R05)	All
<i>Temp leave</i>	# per 1,000 cases of temporary parental leave	All

Note. The table shows definitions (ATC codes/abbreviations) of the dependent variables in Tables 4 and 5. # per 1,000 = number of redeemed prescriptions/cases per 1,000 residents. Data source: The Swedish Prescribed Drug Register, The Social Insurance Agency (*Temp leave*).

TABLE A3 Covariate definitions

Variable	Description	Aggregation
<i>share children</i>	Share of residents 0–9 years (%)	m
<i>share elderly</i>	Share of residents >65 years (%)	m
<i>log(population)</i>	Log of population size in thousands	m
<i>mean income</i>	Average personal taxable income (thousands of SEK)	m
<i>share secondary edu</i>	Share of population (ages 16–74) with secondary education (%)	m
<i>share tertiary edu</i>	Share of population (ages 16–74) with tertiary education (%)	m
<i>other P4P</i>	I(Control council using other antibiotics-related P4P (Section 4.1.2))	c
<i>choicereform</i>	I(Implementation year of entry/choice reform)	c
<i>cost responsibility</i>	I(primary care centers have budget responsibility for drug costs)	c

Note. The table shows definitions and aggregation level (m = municipality; c = county council) of covariates. Data source: own data collection (*other P4P*, *choicereform*, *cost responsibility*), Statistics Sweden, and Swedish Association of Local Authorities and Regions (SALAR).

TABLE A4 Summary statistics for 2008, by later P4P status

Variable	Ever PcV P4P		Never PcV P4P		<i>p</i> -value of difference
	Mean	SD	Mean	SD	
<i>share children (%)</i>	11.6	(1.5)	10.6	(1.1)	.072
<i>share elderly (%)</i>	17.0	(3.3)	18.5	(3.2)	.178
<i>log(population)</i>	11.3	(1.3)	10.8	(1.3)	.043
<i>mean income</i>	250.9	(40.0)	230.0	(15.8)	.191
<i>share secondary edu (%)</i>	42.8	(6.4)	46.3	(5.4)	.056
<i>share tertiary edu (%)</i>	33.3	(10.6)	28.5	(9.1)	.074
<i>other P4P^a</i>	0.05	(0.226)	0	(0)	.001
<i>choicereform^a</i>	0.42	(0.50)	0.05	(0.23)	.330
<i>cost responsibility^a</i>	0.15	(0.35)	0.80	(0.40)	.000
Observations	108		181		

^aCounty-council level variable. Summary statistics for 2008, separate for the municipalities in counties that did ("ever") or did not ("never") introduce P4P tied to the PcV share during the period 2009–2013. PcV share = penicillin V prescriptions' share of total consumption of antibiotics prescribed for respiratory tract infections (J01CE02/J01AA02/J01CA04/J01CR02/J01DB-DE/J01FA), children 0–6 years. Column *p*-value of difference shows *p*-values from regressions of each variable on a dummy for PcV P4P county council, that is, the *p*-values of the difference between the treated group (Ever PcV P4P) and the control group (Never PcV P4P). The regressions are weighted by the square root of the population size and standard errors are clustered at the county council level. The difference between P4P and control counties is not statistically significant when contrasted in *t*-tests at the county level ($n = 21$). PcV = penicillin V; P4P = pay-for-performance.

APPENDIX B: ESTIMATES ON COVARIATES AND ENTROPY BALANCING RESULTS

TABLE B1 Estimates on covariates

	(1)	(2)
<i>other P4P</i>	0.00278 (0.00386)	0.00655 (0.00768)
<i>choicereform</i>	-0.000568 (0.00252)	-0.00235 (0.00368)
<i>cost responsibility</i>	0.00559 (0.00800)	-0.0122 (0.00973)
<i>log(population)</i>	-0.0359 (0.382)	-0.0195 (0.202)
<i>share children</i>	0.00429 (0.00899)	-0.0222*** (0.00701)
<i>share elderly</i>	-0.00804 (0.00618)	-0.00679 (0.00584)
<i>share secondary edu</i>	-0.00105 (0.00563)	-0.0139** (0.00543)
<i>share tertiary edu</i>	0.00288 (0.00541)	-0.00550 (0.00799)
<i>mean income</i>	-0.000920** (0.000453)	0.000708 (0.000595)
Observations	2,304	2,304
Municipalities	288	288

Notes. The table shows the parameter estimates on covariates from two specifications: column 1 (2) is the specification referred to in column 2 (4) of Table 4. Standard errors clustered by county council in parentheses. P4P = pay-for-performance. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

TABLE B2 Covariate balance after entropy balancing

	P4P			Control		
	Mean	Variance	Skewness	Mean	Variance	Skewness
<i>PcVshare</i> ₂₀₀₆	0.6238	0.009355	-0.6173	0.6239	0.008396	-0.9057
<i>other P4P</i>	0.1044	0.0937	2.588	0.1045	0.09363	2.586
<i>choicereform</i>	0.1456	0.1247	2.009	0.1456	0.1245	2.01
<i>cost responsibility</i>	0.7015	0.2099	-0.8805	0.7015	0.2095	-0.8806
<i>log(population)</i>	10.16	0.722	0.9165	10.16	1.087	0.3527
<i>share children</i>	11.56	3.052	0.3717	11.56	4.544	0.7567
<i>share elderly</i>	20.77	14.27	0.04682	20.77	19.27	0.1577
<i>share secondary edu</i>	47.35	33.54	-1.506	47.36	41.27	-1.03
<i>share tertiary edu</i>	28.12	90.09	1.334	28.12	99.88	1.159
<i>mean income</i>	245.5	1518	2.071	245.5	2127	2.251

Note. The table shows the covariate balance in terms of mean, variance, and skewness when control municipalities are weighted using weights from the entropy balancing algorithm (Hainmueller, 2012). All treated observations have a weight of 1. The balancing algorithm is run on the 2007–2013 sample, in order to be able to include *PcVshare*₂₀₀₆, the municipality's PcV share in 2006, among the covariates. PcV = penicillin V; P4P = pay-for-performance.