

Anell, Anders; Dietrichson, Jens; Ellegård, Lina Maria

Working Paper

Can Pay-for-Performance to Primary Care Providers Stimulate Appropriate Use of Antibiotics?

Working Paper, No. 2015:36

Provided in Cooperation with:

Department of Economics, School of Economics and Management, Lund University

Suggested Citation: Anell, Anders; Dietrichson, Jens; Ellegård, Lina Maria (2016) : Can Pay-for-Performance to Primary Care Providers Stimulate Appropriate Use of Antibiotics?, Working Paper, No. 2015:36, Lund University, School of Economics and Management, Department of Economics, Lund

This Version is available at:

<https://hdl.handle.net/10419/260173>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Working Paper 2015:36

Department of Economics
School of Economics and Management

Can Pay-for-Performance to Primary Care Providers Stimulate Appropriate Use of Antibiotics?

Anders Anell
Jens Dietrichson
Lina Maria Ellegård

December 2015
Revised: June 2016



LUND
UNIVERSITY

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

Lina Maria Ellegård[†] Jens Dietrichson[‡] Anders Anell[§]

June 29, 2016

Abstract

Antibiotics resistance is a major threat to public health. We examine if pay-for-performance (P4P) to primary care providers stimulates appropriate prescription of antibiotics; specifically, if P4P induces a substitution of narrow-spectrum antibiotics for broad-spectrum antibiotics (which contribute more to resistance) in the treatment of children with respiratory tract infections (RTI). During 2006-2013, a subset of Swedish healthcare authorities introduced antibiotics-related P4P in their reimbursement schemes for care providers. We employ municipality-level data covering all purchases of RTI antibiotics in a difference-in-differences analysis, and find that P4P significantly increased the share of narrow-spectrum antibiotics. There were no signs that physicians tried to game the system by increasing overall antibiotics use.

Keywords: pay-for-performance, antibiotics resistance, primary care
JEL Classification: D23; I11; I18; J33; J38; H73

1 Introduction

According to a recent report from the World Health Organization (WHO), resistance to commonly used antibiotics is spread all over the world. For instance, resistance to the last-resort antibiotic for life-threatening infections caused by *K. pneumoniae*, a bacteria responsible for major hospital-acquired infections such as pneumonia and infections in frail patients, has been reported in all regions of the world. In some countries, less than half of the population would be helped by the antibiotic used to fight such infections (WHO, 2014). With increasing spread of resistance and few new antibiotics under development, it is clear that antibiotics resistance is a major and rising threat to the effectiveness of modern health care (ECDC/EMA Joint working group, 2009; Carlet et al., 2011).

*We are grateful to Malin Bredenberg for excellent research assistance. The article has benefited from useful comments by 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. The project is funded by the Crafoord Foundation, FORTE, and the Swedish Competition Authority.

[†]Corresponding author. Department of Economics, Lund University. E-mail: lina-maria.ellegard@nek.lu.se

[‡]SFI - The Danish National Centre for Social Research. E-mail: jsd@sfi.dk

[§]Department of Business Administration, Lund University. E-mail: anders.anell@fek.lu.se

Because every antibiotics cure contributes to the development of resistance (Turnidge and Christiansen, 2005), it is essential that health care professionals only prescribe antibiotics when necessary. Overuse of antibiotics is however common: in the US, almost half of all antibiotics prescriptions for acute respiratory tract infections (RTI)¹ – the most common reason for antibiotics use in the US – are judged to be inappropriate (Fleming-Dutra et al., 2016). In addition, the spread of resistance has led physicians to increasingly turn to broad-spectrum antibiotics, which are effective against more bacteria types and thus reinforce the development of resistance (Ranji et al., 2006; Kaier and Moog, 2012). To reduce the inappropriate use of broad-spectrum antibiotics is therefore a central concern.

One reason for the inappropriate antibiotics use is that physicians seek to please their patients (Ashworth et al., 2015), who have weak incentives to internalize the externality of contributing to the spread of resistance (Coast et al., 1996; Elbasha, 2003). To balance this adverse incentive, it is important to find policy tools that encourage antibiotic stewardship, i.e. appropriate use of antibiotics. For economists, it is natural to think of pay-for-performance (P4P) as one such tool. P4P incentivize quality improvements by rewarding health care providers that reach certain pre-defined targets. Such schemes have been extensively applied in health care during the past two decades, particularly in the US and the UK. However, P4P has rarely been tied to physicians’ prescription of antibiotics. For instance, the comprehensive P4P scheme in British primary care, the Quality and Outcomes Framework (QoF), only recently included antibiotics prescription practices as a component (NHS, 2016). Consequently, little is known regarding the role for P4P in promoting antibiotic stewardship. The few studies to date yield mixed results, and suffer from methodological shortcomings impairing the possibility to disentangle the effect of P4P (Mullen et al., 2010; Yip et al., 2014; McDonald et al., 2015; Gong et al., 2016).

This study contributes to the currently small literature by analyzing data from Sweden, where a number of health care authorities – i.e. county councils – introduced P4P in primary care during the past decade. We study the impact of P4P indicators encouraging antibiotic stewardship in the treatment of children with RTI. Specifically, to make physicians select narrow-spectrum instead of broad-spectrum antibiotics when possible, eight county councils tied P4P to the share of narrow-spectrum *penicillin V* (PcV) in childrens’ total consumption of RTI antibiotics. The focus on the PcV share mirrored Swedish RTI treatment guidelines, which advocate PcV as first-line antibiotic in most cases. Importantly, to ensure that broad-spectrum antibiotics would be used when warranted, the P4P targets still allowed for a fairly large share of broad-spectrum drugs.

We analyze municipality-level register data covering all purchases of RTI antibiotics prescribed to Swedish children between 2006 and 2013. The cross-sectional and time variation in P4P policies allows us to use a difference-in-differences (DID) strategy to study the impact of P4P. Despite that the monetary incentives were small, we find that the introduction of P4P was associated with a 1.5-2 percentage point increase in PcV’s share of RTI antibiotics consumption. Our baseline estimate corresponds to about 20% of the pre-P4P standard deviation, and it would be large enough to close almost one third of the gap between the PcV share in the control group and the national target of 80% PcV. Still, the absolute size

¹Examples of RTI’s include sore throat (pharyngitis), ear infection (otitis), certain types of influenza, cough (bronchitis), pneumonia, sinus infection, tonsillitis, laryngitis, and the common cold.

of the P4P effect is small in relation to the initial average PcV share of 61%.

We further find that childrens' overall consumption of RTI antibiotics was unaffected by the incentive, implying that the change in PcV share reflects a substitution of narrow-spectrum for broad-spectrum antibiotics. This pattern is fully consistent with the intention behind the P4P indicator, which was to stimulate such substitution rather than to reduce overall consumption. Importantly, the substitution implies that prescribers did not manipulate the performance measure by issuing more PcV prescriptions – a general concern for performance measures formulated as ratios (e.g. Gravelle et al., 2010).

Our study has important methodological advantages relative to the small set of previous studies on antibiotics-related P4P. Most obviously, only two of the four previous studies include a control group (Yip et al. and Mullen et al.). These studies, in turn, are susceptible to confounding from simultaneous and substantial changes to the reimbursement system, i.e. the removal of a previously very strong incentive for physicians to prescribe antibiotics (Yip et al.), or the introduction of competing incentives that were strong enough to crowd out the antibiotics-related incentives (Mullen et al.). Though our study setting is not immune to confounding threats from other policies, these are less dramatic and evidently not strong enough to crowd out the antibiotic incentives. We also present suggestive empirical evidence that P4P provides the most plausible interpretation of the effect.

Yet another contribution of the study is to bring evidence from a low-prescribing context. The relatively large impacts of P4P found in Chinese studies may be overly optimistic for developed countries with lower prescription rates, which nonetheless have an antibiotic stewardship homework to do. In comparison with other interventions to promote antibiotic stewardship, such as education and feedback to physicians, we conclude that the Swedish P4P policies appear to perform on par with or better than most interventions studied in Scandinavia. At the same time, education and feedback interventions have had larger effects outside Scandinavia (Ranji et al., 2006). If the relatively weak effect of education and feedback in Scandinavia reflects the relatively low antibiotics use in these countries, then our estimate of the P4P effect may be thought of as a lower bound for high-prescribing contexts – though this is clearly a speculative conjecture.

In relation to the general literature on P4P in health care, our study provides corroborating evidence on the effectiveness of P4P tied to process measures. This is a valuable contribution to a literature that is heavily dominated by studies from the US and UK (Eijkenaar et al., 2013). Also, the similarity of findings makes it more plausible to generalize our results regarding antibiotics-related P4P to these countries. This is particularly interesting considering that one of the antibiotics indicators just introduced in the British QoF scheme is very similar to the one studied here.²

To summarize, our analysis suggests that P4P can be a useful tool in the combat against antibiotics resistance, but that one should not be overly optimistic about the size of its impact. Thus, there remains an urgent need for complementary policies to establish antibiotic stewardship and to stimulate the continuous development of new antibiotics.

The next section discusses the medical considerations with regards to antibiotics for RTI patients, with particular emphasis on the treatment guidelines in Sweden. Section 3 reviews

²The QoF 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 2014/15 value.

the literature on P4P in health care, in particular the few previous studies on antibiotics-related P4P, and briefly describes the impact of other interventions to promote antibiotic stewardship. Section 4 gives an institutional background to Swedish primary care and describes the P4P schemes. Section 5 describes the data and sample, Section 6 outlines our empirical strategy and Section 7 lays out the results. Section 8 discusses mechanisms and Section 9 concludes.

2 Antibiotics for respiratory tract infections

Respiratory tract infections is an umbrella term for conditions affecting the respiratory organs; e.g. sore throat (pharyngitis), ear infection (otitis), certain types of 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 percent of children’s total antibiotics consumption (SWEDRES-SVARM, 2013), and it has been estimated that approximately half of all RTI patients in primary care (regardless of age) receive antibiotics (André et al., 2008). Similarly, a recent study of the United States estimated that RTI patients account for almost 45% of all antibiotics prescriptions in US outpatient care, half of all RTI patients are prescribed antibiotics, and, notably, almost half of prescriptions are inappropriate for the diagnosis (Fleming-Dutra et al., 2016) – a figure that has been stable over time (Ranji et al., 2006).

The high rate of inappropriate prescriptions reflects that the clinical basis for prescribing antibiotics for RTI is often weak. 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.³ 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 et al., 2010).

When antibiotics are needed, it is essential to choose the right antibiotic (Ranji et al., 2006). According to the Swedish treatment guidelines, narrow-spectrum antibiotics are often sufficient; in particular, PcV is the recommended first-line antibiotic for most of the common RTI.⁴ The renowned and government-supported network Strama (Strategic Programme against Antibiotic Resistance)⁵ has proposed as a national target that PcV should account for 80% share of RTI antibiotics prescribed to children in the 0-6 year age group.⁶ By allowing for as much as 20% broad-spectrum antibiotics, Strama considers the target to strike a balance between the conflicting goals of impeding resistance and avoiding the rare but severe complications that require broad-spectrum antibiotics.

³See e.g. Spurling et al. (2011); Lemienegre et al. (2012); Spinks et al. (2013); Venekamp et al. (2013); Cronin et al. (2013); Smith et al. (2014) for reviews of antibiotics treatment of different RTI.

⁴Swedish RTI guidelines are available (in Swedish) at <https://www.folkhalsomyndigheten.se/amnesomraden/smittskydd-och-sjukdomar/antibiotika-och-antibiotikaresistens/behandlingsrekommendationer/>, last accessed 2016-05-19.

⁵For more information about Strama, see <http://strama.se/about-strama/?lang=en>, last accessed 2016-05-19.

⁶See Appendix A for a translation of Strama’s three national targets.

Though the 80% PcV target for children has yet not been reached (SWEDRES-SVARM, 2013), PcV has long been the most common antibiotic prescribed for RTI in Sweden (André et al., 2008). Notably, in an international perspective, the strong preference for PcV is a uniquely Scandinavian phenomenon. E.g., narrow-spectrum penicillin accounted for over 50% of total penicillin use in the three Scandinavian countries in 2002-04, to be compared with 7% in the US and 9% in EU-27 (Goossens et al., 2005, 2007).⁷ Though it is clearly outside the scope of this paper to pin down the reasons for these differences,⁸ they suggest that the P4P targets evaluated in this study, aiming at quite high PcV shares, may be too high for other countries.

3 Related literature

In situations characterized by principal-agent problems, i.e. when the interests of a principal and its agents are not perfectly aligned, pay-for-performance may align agents' incentives with the principal's preferences. By making agents' reimbursement depend on their performance with respect to certain performance indicators, the hope is that agents respond by increasing effort on tasks valued by the principal (Prendergast, 1999). In the present study setting, policy-makers seeking to promote antibiotic stewardship can be thought of as the principal, and physicians – the prescribers – as agents.

The numerous reviews of empirical research on P4P in health care point at mixed evidence: there are examples of substantial positive effects, but most studies find small or zero effects (Town et al., 2005; Rosenthal and Frank, 2006; Petersen et al., 2006; Christianson et al., 2008; Van Herck et al., 2010; Scott et al., 2011; Eijkenaar et al., 2013).⁹ To give a sense of magnitude, Van Herck et al. (2010) suggests that P4P generally is associated with improvements of about 5%. P4P has typically had stronger impact when the incentive is tied to process measures, such as recording of diagnoses or vaccination rates, than when tied to broader outcome measures such as mortality (Eijkenaar et al., 2013). One reason may be that physicians exert more control over their performance with respect to process measures, which are less affected by random fluctuations. Random noise devalues physicians' incentives

⁷Antibiotics use measured by Defined Daily Doses (DDD) per inhabitant per day. The cited articles do not distinguish between PcV and other narrow-spectrum penicillins in the J01CE ATC group.

⁸Different treatment guidelines likely play a role. E.g., US guidelines propose PcV as first-line treatment only for sore throat and pneumonia (Bradley et al., 2011, ; see also <http://www.cdc.gov/getsmart/community/for-hcp/outpatient-hcp/pediatric-treatment-rec.html> , last accessed 2016-05-19). RTI guidelines often differ between countries; for instance, rhinosinuitis in children (which is bacterial in less than 15% of cases) motivates antibiotics in the US but generally not in the UK and Sweden (Cronin et al., 2013), and the guidelines for sore throat recommend antibiotics in the US, France and Finland, but not in e.g. Belgium, UK, and Sweden (e.g. Matthys et al., 2007). Resistance to penicillin is further lower in Scandinavia than in some other countries – e.g., less than 8% of *S. Pneumoniae* bacteria are non-susceptible to penicillin in Scandinavia, as opposed to 25% in Spain (ECDC, 2015). Notably though, countries such as the UK, Germany and (more recently) the US report similar resistance rates as the Scandinavian countries (ECDC, 2010, 2015, and <http://www.cdc.gov/abcs/reports-findings/surv-reports.html>). Finally, there are plenty of differences in the health systems and the regulation of and marketing of pharmaceuticals (Goossens et al., 2007), that may explain the differences in PcV use.

⁹These reviews cover mostly studies from the US and UK. The emerging literature on P4P in low- and middle income countries tend to find higher effectiveness of P4P (de Walque et al., 2015)

to make efforts to perform well; partly because adverse random events may over-shade such efforts, partly because the possibility for advantageous random events may lead physicians to relax and hope for high performance. Theoretical work suggests that incentives should be stronger in settings where signals of agents' performance are reliable, especially if agents are risk-averse (Holmström, 1979; Prendergast, 1999).

In this light, the outlooks for an impact of P4P related to physicians' prescription of antibiotics, a process measure which to a large extent is influenced by the physicians themselves, are promising. But only a few previous studies have examined the impact of P4P seeking to promote antibiotic stewardship.¹⁰ Gong et al. (2016) study an antibiotic stewardship program employed by a pediatric hospital in Guangzhou, China. The study compares antibiotics' share of all prescriptions before and after the hospital introduced the program, which coupled audit and feedback with significant fines for physicians who had prescribed unwarranted antibiotics. Antibiotics' share of all prescriptions fell sharply immediately after the policy was introduced – from 10 to 4% in outpatient care and from 6 to 4% in inpatient care – and remained low throughout the policy period (one year). However, the study lacks a control group and the effect of the financial penalties is completely confounded with the effect of audit and feedback, which by itself may have an effect (see below).

Yip et al. (2014) evaluate a randomized experiment in the Chinese province Ningxia. Primary health care centers in the intervention group switched from a fee-for-service (FFS) reimbursement scheme that gave physicians strong monetary incentives to prescribe (via a positive margin on all prescriptions), to a reimbursement regime that combined capitation with a P4P scheme according to which centers whose antibiotics prescriptions were below (above) average were punished (rewarded) financially. Whereas 44% of visits led to an antibiotic prescription in the FFS control group, the rate was 15% lower (i.e. 37%) in the intervention group. Notably though, it is impossible to disentangle the effect of P4P from the effect of removing the previously very strong incentives for physicians to prescribe.

Leaving the Chinese context, Mullen et al. (2010) study the introduction of a set of P4P schemes used by Californian health maintenance organizations contracting with medical groups. One of the P4P schemes included the preferred antibiotics usage rate for bronchitis or pharyngitis (sore throat) as a performance indicator in 2003. Contrary to what was intended, the rate *fell* by 3 percentage points (6%) after P4P was introduced.¹¹ The authors suggest that other components of the P4P scheme were more important than the antibiotics-related indicator, effectively crowding out physicians' attention to their antibiotics prescription behavior.

Finally, the hospital P4P scheme Advancing Quality (AQ), introduced in North West region of England in 2008, included two measures relating to antibiotic stewardship. McDonald et al. (2015), a comprehensive evaluation of AQ, reports that the share of patients receiving appropriate antibiotics increased from 82 to 84% among pneumonia patients and from 85 to 88% among hip and knee replacement patients between the first and fourth quarter in the

¹⁰Beside financial incentives directed to physicians, there is a small literature incentives targeting patients. A study of the RAND insurance experiment found that patients facing a higher co-payment share had a lower antibiotics consumption (Foxman et al., 1987). Similar results were observed in a non-experimental Canadian study (MacCara et al., 2001).

¹¹In addition, the percentage of antibiotics that were prescribed in appropriate cases, a measure that was not rewarded by P4P, fell slightly when P4P was introduced.

first year of AQ. It is however difficult to quantify AQ’s contribution to these improvements, as the performance data were collected only for AQ hospitals, meaning that there is neither a before-AQ measurement nor a control group. Further, there may be seasonal variation between the first and fourth quarters of a year.¹²

To summarize, among the four previous studies, only two (Yip et al. and Mullen et al.) have control groups, and one study (McDonald et al.) even lacks a before-P4P baseline measure. Also, the introduction of antibiotics-related P4P was coupled with other important policy changes – a removal of previous incentives to prescribe antibiotics in Yip et al., the introduction of other P4P indicators, in Mullen et al. and McDonald et al., not to mention the simultaneous introduction of audit and feedback in Gong et al. These constitute serious confounding threats in the estimations of the P4P effect, as illustrated by the counter-intuitive effect reported by Mullen et al. It is also notable that all except one of these studies (McDonald et al.) are from high-prescribing countries, posing questions regarding the external validity to low-prescribing contexts.

Apart from the above studies, Martens et al. (2007) also examines a financial incentive related to antibiotics, though the incentive was not related to physicians’ *performance*. Rather, physicians in the intervention group – a Dutch region – received extra money *ex ante* in return for signing a letter of commitment to a set of prescription guidelines (e.g. substitute narrow- for broad-spectrum antibiotics, choose recommended gastric drugs, avoid costly new drugs). All physicians in the intervention group agreed to sign the letter. The study shows that compared to physicians in a control region, physicians in the intervention group became more adherent to prescription guidelines right after they signed the letter, but the differences were no longer significant after one year.

It may seem unsurprising that a financial incentive paid out *ex ante* failed to have a long-term effect. But, as pointed out by Celhay et al. (2015), temporary monetary incentives can be used as a nudging device to make organizations overcome the inertia otherwise inhibiting the change of inappropriate habits and routines. Put differently, as humans are reluctant to changing routines, a temporary incentive that induces a transition from one routine to another may have long-lasting impact. On the other hand, the removal of financial incentives can have adverse consequences, if the incentives have permanently crowded out previous intrinsic motivation to perform well (Deci et al., 1999). As these two effects are of opposite direction, the existence of long-term effects is an empirical and context-dependent question. Indeed, while there was no long-term effect for antibiotics in Martens et al. (2007), they found a long-term improvement for gastric drugs. Similarly, in a completely different application, Celhay et al. found that a temporary P4P scheme in Argentinian prenatal care had effects lasting at least 24 months.

¹²Appropriate antibiotics for pneumonia was also included in the Hospital Quality Incentive Demonstration (HQID) in the US 2003-2007. We are not aware of any study on the impact of that HQID indicator. Two studies have examined another antibiotics measure in HQID, the share of pneumonia patients getting antibiotics within 4 hours. Notably, the indicator aimed to speed up treatment, rather than to affect resistance. In fact, some have worried that the indicator may inflate antibiotics use (e.g. Drake et al., 2007), though such concerns were not substantiated in a later analysis of pneumonia patients (Friedberg et al., 2009). According to Grossbart (2006), HQID hospitals in Ohio improved less than control hospitals with respect to antibiotic timing during the first year of the program, whereas Lindenauer et al. (2007) showed that after two years, HQID hospitals had improved 4 percentage points more than matched control hospitals.

There is a larger literature on *non-financial* interventions to improve antibiotic stewardship, e.g. education campaigns directed to physicians and/or patients, or audit and feedback to physicians about their prescription patterns. Systematic reviews of this literature suggests that such interventions are moderately effective, with interventions with more than one component having larger effects (Arnold and Straus, 2005; Ranji et al., 2006; van der Velden et al., 2012; Vodicka et al., 2013).¹³ Ranji et al. computed the median effect size among the included studies; the median was -9 percentage points for interventions to reduce overall prescriptions, and 10 percentage points for interventions to increase the rate of appropriate antibiotics (the outcome most closely related to ours). With regards to the latter outcome, it is however notable that studies from Scandinavian contexts, where the initial rate of appropriate prescriptions generally is high, mostly report statistically insignificant effects of less than 5 percentage points. Munck et al. (1999); Ranji et al. (2006); Bjerrum et al. (2011) One exception is a Norwegian study of a recent, large cluster-randomized intervention that led to a 10 percentage point increase in the PcV share (Gjelstad et al., 2013).

4 Institutional background

4.1 Organization of primary care in Sweden

In Sweden, the organization, financing and provision of health care is delegated to 21 *county councils*. Health care is mostly financed by a proportional income tax (the tax rate varies between counties but is typically around 11%), though patients also pay a fee for visits (up to a yearly cap) and part of the price for drugs prescribed in outpatient care. Patients 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.

The counties have considerable discretion regarding how to organize primary care. Primary care is typically provided in group practices, with a handful of general practitioners and additionally a team of specialist or general nurses, social workers and behavioral therapists, physiotherapists and midwives. Though rare, there are also examples of private solo-practices (Anell et al., 2012a). We do not make a distinction between group- and solo-practices in the paper, but refer to all primary care providers as *primary care centers*. In total, there are about 1,500 primary care centers in Sweden. The county councils operate most primary care centers, but they also contract with private providers. Since 2010 (a few years earlier in some county councils), there is free entry for primary care providers that fulfil the minimum requirements defined by the county council in which they want to operate. The minimum requirements, together with the rules for reimbursement, are written down in a directive approved by the county council assembly yearly. The directive pertains to private and public providers alike (Anell, 2011). As of 2013, about 40 percent of primary

¹³More recent studies yield similar conclusions (Meeker et al., 2014; Meeker, D et al., 2016; Hallsworth et al., 2016).

care centers were private, accounting for a quarter of primary care expenditures.¹⁴

Primary care centers are reimbursed by a mix of capitation, fee-for-service (FFS), and P4P. Capitation, i.e. a fixed, though often risk-adjusted, sum per enrolled patient, is the dominant part of reimbursement (70-98% of revenues) in most counties; with a capitation share of only 40%, the county council of Stockholm is the exception to this rule. The remaining part of revenues is covered predominantly FFS, while P4P is more of a complementary reimbursement. It is also common to give extra compensation to care centers located in rural and/or disadvantaged areas (Anell et al., 2012b).

Within the geographical boundaries of each county council, there is a number of *municipalities* (290 in total). The municipalities represent a separate layer of government – they organize public services such as schooling and elderly care¹⁵ – but they are interesting for the present study because our data is aggregated at the municipality level (see Section 5). County council borders never cut across a municipality, i.e. all citizens of a given municipality “belong” to the same county council. Thus, when a county council changes its health care policy, all inhabitants of municipalities within that county council are affected by the policy change.

4.2 P4P in Swedish primary care

During the past ten years, P4P has gained in popularity among the county councils as a complementary reimbursement to primary care centers.¹⁶ In 2006, only five councils applied P4P in primary care, a figure that rose to fifteen in 2009 and twenty in 2012.

Beside performance indicators related to drug prescription patterns (e.g. antibiotics use; see next section), P4P has commonly been tied to e.g. patient satisfaction survey scores, registrations of patients in national quality registers (e.g. for diabetes care), or vaccination rates. The county council assemblies revise the P4P schemes yearly, as the reimbursement scheme is part of the directive for primary care centers. Some indicators remain in the schemes for several years, while others are removed after one or a few years.

P4P only accounts for a minor part of revenues for the primary care centers: in 2012, P4P accounted for 1-5% of total reimbursement (Anell et al., 2012b). Notably, the size of the P4P reimbursement is determined by the performance of the care center as a whole, i.e. the incentive is targeted to the primary care centers. Consequently, the incentives are not by default related to the reimbursement of individual physicians; it is up to each care center to decide whether the reimbursement to individual physicians should depend on their performance. We cannot rule out that some centers may have used P4P as a partial reimbursement of individual physicians, but it is probably rare: in both public and private care centers, physicians are typically reimbursed by a fixed monthly salary (Anell et al., 2012a). At least for the largest private chain, we know that they have not applied

¹⁴This and other publicly released official statistics are available through the database KOLADA, <http://www.kolada.se>.

¹⁵They are responsible only for some health care to a) elderly residing in nursing homes, and b) primary and secondary school children.

¹⁶The prevalence of P4P in 2006, 2009 and 2013 was surveyed by Anell (2006) (unpublished), Anell (2009) and Anell et al. (2012b). A research assistant has complemented the survey material with additional material for other years in the sample period.

antibiotics-related P4P to individual physicians (personal communication with Capiro).

4.2.1 P4P related to antibiotics for children with RTI

Eight county councils decided to tie P4P to the share of PcV during our sample period (Table 1). The incentives aimed to make physicians choose PcV instead of broad-spectrum antibiotics; bringing down the total level of RTI antibiotics consumption was not in focus. Though the eight counties developed their P4P indicators independently, they chose very similar performance measures, i.e. *PcV's share of RTI antibiotics prescriptions to children aged 0-6 years*, with only minor variations with respect to what antibiotics counted as "other" RTI antibiotics and (in two counties) restrictions of included diagnoses. Plausibly, the national PcV target (see Section 2) was a source of inspiration for the P4P indicators.

The link between performance and monetary incentives, i.e. the P4P targets, differed between counties. Two counties promised bonuses to care centers that reached a target of 80% PcV (i.e. the national target); one of these counties also employed another, lower, target level (the national average PcV share). Three county councils used P4P targets of 75% PcV. SLL used a reference target of 70%, with bonus/penalties being linearly related to deviations from this level (bonuses/penalties for higher/lower PcV shares). Bonuses were linearly related to performance also in Kronoberg (range 65-80% PcV). Halland used penalties to incentivize care centers. Care centers with less than 69% PcV were penalized, with larger penalties for PcV shares below 60%.

The PcV-related P4P accounted for a very small share of reimbursement in all eight counties. Unfortunately, we do not have access to information about the actual payment sizes in the counties, 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 for affected care centers. Obviously, the impact on the personal incomes of prescribing physicians – who are typically salaried – was negligible. Thus, failing to reach the P4P target would not be financially detrimental. Though the care centers' were probably aware that P4P was a minor source of revenue, they need not have had a well-grounded belief about exactly how much money to expect. Indeed, in the county of Skåne, the bonus size depended on the number of care centers meeting the target, rendering such calculations almost impossible to make.

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 |

X = County council uses a P4P indicator related to the PcV share

Table 1 shows that four county councils abolished PcV P4P after a few years. In Section 6,

we discuss the implications for how we should specify the empirical estimations.¹⁷

4.2.2 Concurrent antibiotics-related policies

There were some other notable antibiotics-related policies during the study period.

For instance, the national PcV target was only one of three targets proposed by Strama in 2009 (see Section 2 and Appendix A.) The two other targets related to the total number of antibiotics prescriptions and the share of broad-spectrum drugs to women with urinary tract infections (UTI). Probably inspired by the national targets, some counties – among them, some of those using PcV P4P (see Appendix B) – tied P4P to one or two of these other measures. Consequently, we cannot disentangle the independent effect of the PcV P4P by controlling for the presence of another antibiotics-related indicator in the econometric analysis. We do not think that this is an important source of confounding, for two reasons: first, our outcome variable – the PcV share – is considerably more closely related to the PcV P4P than to the other two antibiotics-related measures. Second, when we include an indicator variable for *control group* municipalities that use P4P for either of the other two measures, this has virtually no impact on our estimates of the PcV P4P effect.

Additionally, in late 2010 the national government launched a P4P program (*Patientsäkerhetssatsningen*) directed at the *county councils* (i.e. not directly to primary care centers). The first and most highly profiled target within the program was that each county should reduce the total number of antibiotics prescriptions to a certain level.¹⁸ To receive a bonus, the counties also had to show that they had tried to promote antibiotic stewardship, e.g. by instituting local Strama groups that would work with education and feedback to physicians (Socialdepartementet, 2010). The national program lasted between 2011-2014, with largely the same conditions for each year. Most county councils failed to reach their annual targets for overall antibiotics consumption, while all counties could demonstrate that their local Strama groups had promoted antibiotic stewardship.

Though the national targets and the government’s initiative were directed to all county councils, we cannot rule out that they had differential effects on the counties that implemented PcV-related P4P. As already noted, it is plausible that these national policies inspired politicians and bureaucrats in the eight counties to implement PcV-related P4P. However and importantly, that does not mean that the *prescribers*, i.e. physicians, in these counties reacted differently to national targets and policies than their colleagues in other counties. In Section 7.3, we run estimations suggesting that prescribers in PcV P4P counties did not react differently to the national policies.

¹⁷According to personal communication with staff in Halland (H), Skåne (S) and Dalarna (D), some of the reasons for abolishing the indicator were that the performance measure was viewed as too volatile for small care centers (H, S), that the county’s central administration was content with the PcV share and wanted to prioritize other aspects of antibiotics stewardship (S, D), and that physicians criticized that the measure included all prescriptions without taking diagnosis into account (H). In Halland, the central administration was also worried that care centers would increase PcV prescriptions to improve performance.

¹⁸The long-term goal was that less than 250 prescriptions per 1,000 residents should be prescribed in 2014. For 2011 and 2012, the ceiling was such that each county council’s consumption should decrease by 10% of the difference between its consumption the last year and the long-term goal. For 2013, the ceiling was less ambitious, i.e. the ceiling was set to last year’s level minus 1 prescription.

5 Data

5.1 Data sources

The Swedish Prescribed Drug Register is an official register containing information about all prescribed pharmaceuticals that patients have collected since July 2005. The data is reported by the pharmacies, which are the only retailers of antibiotics in Sweden. As the register records the municipality of residence of the patient collecting antibiotics, we use data aggregated at the municipality level. All municipalities in a county council that implemented P4P at time t , thus become "treated" at that point in time. Our dataset includes yearly municipality-level¹⁹ information for the period 2006 – 2013.

There are three notable limitations of the data. First, the register does not distinguish between prescriptions from primary care and prescriptions from other parts of the healthcare system. It is estimated that primary care accounts for about 50% of outpatient antibiotics (Mölstad et al., 2008), though the dominance of RTI antibiotics in the 0-6 age span suggests that the share is likely higher for children. The implication of noise from outside primary care is that our estimates will be less precise, and the measurement error implies that we underestimate the effect.²⁰

A second data limitation is that patients may not adhere to physicians' prescriptions. Patients who do not bother to collect their prescribed medicine are not covered by the data. Third and finally, the use of aggregated (rather than individual-level) data is a limitation. On the other hand, compared to data aggregated at the health care center level, municipality-level data has the advantage of decreasing the risk of bias stemming from patients sorting across primary care centers. For example, patients may respond to stricter prescription practices by enrolling with another primary care center. As sorting of patients across municipality borders ought to be less common, municipality-level data dampens this channel.

Beside the pharmaceutical data, we also have access to a large set of covariates, including both municipality-level background characteristics (demography, income and education level) and some county council-level policy variables. Because the covariates have little impact on our estimated P4P effect (see Section 7.2), we defer a further description of the covariates to Appendix C.2 and C.4, and only note here that the municipalities in P4P counties were on average more populous and had more highly educated residents.

Our dataset includes information for all 290 municipalities all of the years 2006-2013.

¹⁹Defined by patients' municipality of residence; i.e. if Anna lives in municipality B but is enrolled at a primary care center in municipality C, her medication consumption data will be recorded as part of the data for municipality B. This ought to have minor implications for the analysis: as children in the 0-6 year age span typically spend their time in close proximity to their homes, it is reasonable to assume that they are enrolled at care centers in their municipality of residence. This is most often the case for adults too, for instance, 90 percent of the residents in Skåne are enrolled at a care center in their home municipality (Anell et al., 2016)

²⁰One may be worried that we confound the effect of P4P in primary care with P4P in specialist care. But this is unlikely to be the case, as P4P has been used to much less extent in specialist care: we know of only one instance where P4P currently is related to antibiotics prescriptions (not one of our treated county councils, and not during the study period). Anell (2013); Andersson et al. (2014); Lindgren (2014) offer partial descriptions of P4P in specialist care.

We exclude one municipality with exceptional responsibilities for health care, as we lack information about its reimbursement system.²¹ This leaves us with a sample of 2,312 observations, corresponding to 289 municipalities belonging to 21 county councils over the period 2006-2013.

5.2 Descriptive statistics

Our main dependent variable is the *PcV share*, i.e. the number of PcV prescriptions divided by the total number of RTI antibiotics prescriptions prescribed to children between 0-6 years of age. The exact substances included in the measure are defined in Appendix C.1.²² Table 2 shows the average PcV share in 2006 – the first year in our sample, before any county had introduced PcV P4P, and in 2013 – the last year of our sample. The first two rows of the table shows the averages by treatment group membership, i.e. separate computations for the group that *ever* used PcV P4P and the control group that *never* used P4P. The ever-P4P group had a lower PcV share at both points in time, however only the latter difference is statistically significant (due to the lower cross-sectional variation in 2013).²³ 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 year of our sample (2006), again by treatment group status. The two groups experienced the same average growth between 2006 and 2008, the last year before any county council introduced P4P. Notably, before 2011, only four of the county councils in the P4P group had actually implemented P4P; in fact, up to and including 2011, the trend reflects a mix of the growth in councils with P4P and in councils that not yet had 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).

Table 2: PcV share by treatment group

| | 2006 | | 2013 | | Obs. |
|------------------------|-------|-------|-------|-------|------|
| | Mean | S.d. | Mean | S.d. | |
| 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 |

The first two rows show the mean PcV share by year and treatment group status (ever P4P (treatment) or never P4P (control)), weighted by population size. The third row shows the unconditional mean by year.

²¹Norrtälje municipality.

²²The reason for examining the number of prescriptions, rather than e.g. the number of Defined Daily Doses (DDD), is that it appears to be the policy variable of interest in Sweden. The number of prescriptions are in focus in the P4P schemes, in the national antibiotics targets, in the national initiative *Patientsäkerhetssatsningen*, and in the annual Open Comparisons made by the Swedish Association of Local Authorities and Regions.

²³The significance tests come from regressions on the PcV share on a dummy for belonging to the P4P group. The regression is weighted by population size and the standard errors are clustered at the county council level. $p=0.197$ ($p=0.026$) in the 2006 (2013) regression.

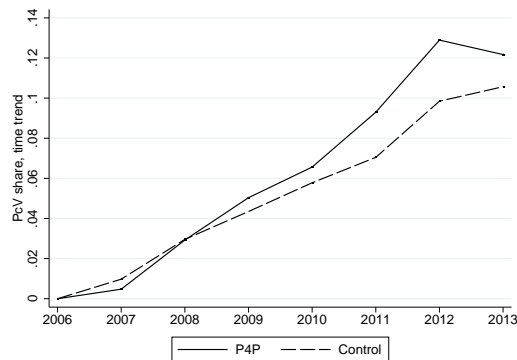


Figure 1: The figure shows the estimated year effects from regressions of the PcV share on a vector of year dummies and municipality fixed effects. Separate estimations for municipalities in county councils that *ever* used PcV P4P (P4P) and municipalities in county councils that *never* used PcV P4P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

To further illustrate the development of the PcV share in the P4P group, Figure 2 plots each municipality's PcV share against the time relative to the year when P4P was implemented ($t=0$). Each circle represents one municipality and the size of the circle is proportional to the population size of the municipality. The plot ranges from 7 years before P4P was implemented, until 4 years after; the number of municipalities at each point in (event) time vary according to the length of time we observe them before/after the P4P implementation year; e.g., all municipalities observed at $t=-7$ are located in Kronoberg, the county that was the last to implement P4P. The largest circle represents Stockholm, the capital, which is observed 6 years before P4P and 2 years after. All P4P municipalities are observed at least 3 years before and 1 year after the year of P4P implementation.

The general increase in the PcV share is clear from the figure, but it is hard to see if the implementation of P4P was associated with an additional jump in the PcV share. To grasp if there was a jump, we also plot two fitted regression lines (weighted by population size): one line for the period before the event, one for the period after. We only include event time periods when all counties are observed, to retain the same sample over (event) time (i.e., $t \in [-3, 2]$). The vertical distance between the endpoint of the first line and the starting point of the second suggests that there was a jump upwards around the time of P4P implementation. The jump is attenuated if we shift back the break between the regression lines one time period, suggesting that something happened at $t=0$ rather than before (see Appendix C.3).²⁴

²⁴The Appendix also shows that we get similar a similar pattern when we exclude the two counties with shortest pre-/post periods (to extend the regression lines), as well as when we remove the largest county council (SLL) from the sample or use unweighted data.

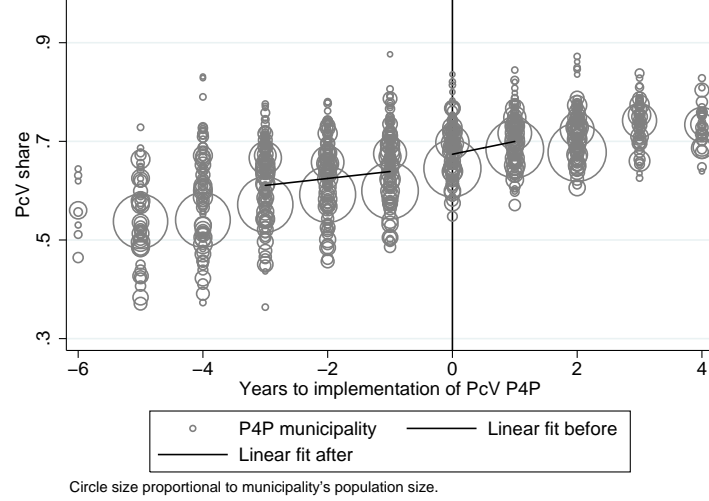


Figure 2: The circles show each P4P municipality’s PcV share, by time to P4P implementation. Linear fit = regression lines calculated for data points when all municipalities are observed (see main text).

6 Empirical strategy

6.1 Identification

We estimate the effect of P4P by comparing the change in PcV share in the treatment group – municipalities in county councils that introduced PcV P4P – to the corresponding change in the control group – municipalities in county councils that did not use PcV P4P. In so doing, we face some empirical challenges in relation to the main identifying assumption in a DID strategy: that the trends in treatment and control groups would have been parallel in the absence of treatment.

First, the adoption of P4P is a choice variable for the county councils. One concern is therefore that the decision to adopt P4P may relate to the ability of increasing the PcV share. For instance, recall that the initial PcV share was relatively low in the P4P counties (though the difference is not significant). Having a low PcV would seem as a good reason to adopt PcV P4P, and, simultaneously, may mean that it is easier for municipalities in these counties to increase the PcV share in response to P4P (say, because there are more patients for which PcV can doubtlessly be substituted for broad-spectrum antibiotics). This is problematic in regards to the external validity of our results; i.e. an estimated positive P4P effect may be smaller or non-existent in other contexts, if our treatment group has unusually high potential to respond to the policy. However, an interpretation of the effects as treatment-on-the-treated would still be valid. In section 7.3, we test if there is a larger treatment effect in municipalities with a low initial share of PcV.

Second, a more general concern is that the municipalities in the treatment group might have increased their PcV share more, even if the counties they belonged to had not implemented P4P. In particular, counties with low initial PcV shares might have been expected to catch up regardless of whether or not they used P4P. It is not possible to directly test for

this threat to the parallel trends assumption, but the timing of the development shown in Figure 1 suggests that catching up is not the whole story. The increase between 2006 and 2008 – before any county had implemented P4P – is almost exactly the same in the two groups, and the deviation is greatest in 2011 – the year when the most county councils had P4P in place. If the divergence mainly reflected catching-up from a initial level, it ought to have been present already from the start. We test the potential pre-P4P divergence more formally in Section 7.3.

Third, we may be confounding the P4P effect with the effects of other factors such as other contemporaneous policies and RTI morbidity patterns. To account for underlying features and trends in individual municipalities, we include municipality fixed effects and municipality-specific linear trends in the estimation equation (cf. Li et al., 2014, for a similar specification in a P4P context).²⁵ To further check whether other county-council level policy changes are driving our results, we look for potential “treatment effects” on outcomes that are unrelated to PcV (Section 7.3). By including year fixed effects, our DID specification addresses the confounding impact of national policies that affected all municipalities similarly. Still, it is possible that different counties were differently affected by e.g. the national incentive program – which did not directly focus on PcV, but might have influenced physicians’ PcV decisions via increased educational activities by the local Strama groups. We examine the basis for such concerns in Section 7.3. Lastly, to check if the P4P effect reflects differential morbidity patterns, we regress the treatment variable on other outcomes that are related to RTI morbidity (Section 7.4)

6.2 Estimation

Our baseline estimation equation can be written as follows:

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

y_{mct} is the value of the dependent variable in year t for municipality m in county council c . μ_{mc} are the municipality fixed effects and λ_t is a vector of year fixed effects, capturing common shocks to all municipalities in year t . θ_{mc} is a vector of municipality-specific coefficients on the linear trend variable t , i.e. the municipality-specific linear trends. ε_{mct} is an idiosyncratic error term. In a robustness check, we include a vector of municipality-and-year covariates and a vector of county council-and-year covariates, described in Appendix C.2.

$PcVP4P_{mc}$ equals one for observations affected by PcV-related P4P and thus α captures the effect of PcV P4P. Because some county councils removed PcV P4P after a few years, we adopt two versions of the treatment variable. Initially, we classify municipalities as affected by PcV P4P ($=1$) in years when the county council they belong to employs P4P, and as unaffected ($=0$) otherwise. This specification is however misleading if P4P had a persistent effect on prescription patterns (c.f. Celhay et al., 2015). To account for this, we then augment the specification with a dummy variable ($postPcVP4P$) for observations in county councils that previously used PcV-related P4P.²⁶ The result of this exercise leads

²⁵Unfortunately, we have too few pre-P4P data points to make meaningful estimations using the synthetic control method, another method that has successfully been used in a P4P application (Kreif et al., 2015).

²⁶I.e., this variable equals 1 in 2012-2013 for municipalities in Blekinge, Dalarna and Halland, and in 2013 for municipalities in Skåne.

us to our preferred specification, in which $PcVP4P_{mc}$ equals one for observations in county councils *that currently use or have previously used* PcV P4P.

With clustered data, the standard errors may be underestimated if we do not adjust for the grouped error structure (Bertrand et al., 2004). In our case, we need to adjust for the correlation within county councils that remains after conditioning on the municipality fixed effects. In the baseline estimations, we cluster the standard errors at the county council level, using Stata’s cluster-robust sandwich estimator (Rogers, 1993). Because the small number of clusters (21) may nonetheless lead to underestimated standard errors (Cameron and Miller, 2015), we also subject our main specification to the wild cluster bootstrap procedure suggested by Cameron et al. (2008).²⁷

We weight the data for each municipality by its population size; by so doing, we assign the same weight to antibiotics consumption stemming from individuals living in larger municipalities as to consumption stemming from inhabitants of smaller municipalities.²⁸ This seems sensible given the large between-municipality variation in population size. We estimate the model using the Stata module *xtivreg2* (Schaffer, 2010), which allows us to easily partial out the municipality linear trends and the year effects from the estimations.

7 Results

Section 7.1 contains our baseline estimates and Section 7.2 presents a set of sensitivity tests. Section 7.3 examines potential confounding factors relating to other policies, whereas Section 7.4 examines how the effect was channelled and if it appears to be driven by differential morbidity. Section 7.5 briefly discusses heterogeneity of the results.

7.1 Main results

Column 1 of Table 3 shows the estimates of our initial model, in which municipalities are classified as treated during the years the county councils apply P4P. This specification yields a positive and statistically significant P4P effect, amounting to a 1.1 percentage points increase in the PcV share ($p=0.008$).

To account for the possibility that the shift in prescription behavior induced by P4P may persist even when the incentive is removed, we next add the *post PcV P4P* dummy variable to the specification. As shown in column 2, the P4P effect then increases to 1.8 percentage points ($p=0.001$). Moreover, the positive and significant estimate on *post PcV P4P* suggests that the impact of P4P persisted even after the incentive has been removed. The persistence explains why the estimate on the main treatment variable increases so much compared to column 1: when municipalities in county councils that later removed P4P are classified as part of the control group, their linear trends are tilted upwards by the post-P4P effect, which makes it harder to distinguish the treatment effect in the years they used P4P.

²⁷We use version 2.0.0 of *cgmwildboot*, a Stata module developed by Judson Caskey. The algorithm yields a p-value of the estimated α parameter and a 95% confidence interval based on the percentiles of the bootstrap distribution of $\hat{\alpha}$.

²⁸We use Stata’s *aweight* command.

Because the two treatment dummies in column 2 are not statistically significantly different from each other ($p=0.48$), we decide to replace them with a single "permanent treatment" dummy. Thus, in column 3 – our preferred specification – and in all estimations below, municipalities are counted as treated if they are situated in county councils that either currently use, or have recently abolished, PcV-related P4P. As seen in column 3, the estimated P4P effect is basically the same – approximately 1.8 percentage points – in this specification.

Table 3: Baseline estimations

| | (1) | (2) | (3) |
|----------------|------------------------|------------------------|------------------------|
| PcV P4P | 0.0106*** (0.00399) | 0.0180*** (0.00562) | 0.0176*** (0.00550) |
| post PcV P4P | | 0.0209** (0.00822) | |
| Observations | 2,312 | 2,312 | 2,312 |
| Municipalities | 289 | 289 | 289 |
| R2 | 0.009 | 0.015 | 0.015 |

Dependent variable: *PcV share*; panel mean=0.69. In column 1 and 2, *PcV P4P*=1 for observations in county councils that currently use PcV P4P. In column 2, *post PcV P4P*=1 for observations in county councils that have previously used PcV P4P. In column 3, *PcV P4P*=1 for observations in county councils that *currently use or have previously used* PcV P4P. Standard errors clustered by county council in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

7.2 Sensitivity checks

We find broadly similar estimates when we modify the baseline specification to check sensitivity. A first issue is the possibility that 21 clusters may not be sufficient to make the cluster-robust estimator of the standard errors consistent. Reassuringly, column 1 of Table 4 shows that the conclusion of a positive treatment effect is retained when we employ the wild cluster bootstrap to draw inference.²⁹

The specification reported in column 2 includes a set of covariates, described in Appendix C.2. The main impact of the covariates appears to be to reduce the degrees of freedom, as shown in the reduced precision of the P4P effect ($p=0.030$), which itself is slightly attenuated (1.6 percentage points instead of 1.8). The covariate estimates are available in Appendix D. Only one covariate (mean personal income in the municipality), is statistically significant. It is further notable that the estimate on a dummy equalling one for municipalities in the control group that use *other* antibiotics-related P4P indicators (see Section 4.2.2) is small in magnitude (0.3 percentage points) and far from significant ($p=0.496$). This result suggests that our estimated P4P effect is not contaminated by effects of other antibiotics-related P4P indicators.

²⁹The reason why the wild bootstrap p-value is so low, $p=0.000$, is that none of the 999 bootstrap samples yielded a smaller t-statistic.

Table 4: Sensitivity

| | (1) | (2) | (3) | (4) | (5) |
|----------------|----------------------|-----------------------|--------------------|---------------------|-----------------------|
| PcV P4P | 0.0176*** (0.000) | 0.0159** (0.00732) | 0.0212 (0.0157) | 0.0192* (0.0112) | 0.0153** (0.00678) |
| CI lower | 0.00698 | | | | |
| CI upper | 0.0282 | | | | |
| Observations | 2,312 | 2,304 | 2,312 | 2,304 | 2,312 |
| Municipalities | 289 | 288 | 289 | 288 | 289 |
| R2 | 0.914 | 0.020 | 0.027 | 0.095 | 0.004 |

Dependent variable: *PcV share*; panel mean=0.69. (1) Wild cluster bootstrap (p-value in parenthesis; CI lower/upper = 95% confidence interval); (2) Covariates; (3) No trends; (4) No trends, covariates; (5) Unweighted data. In columns 2-5, standard errors clustered by county council in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

While the municipality-specific linear trends make us less dependent on the parallel trends assumption, the imposed linearity motivates a sensitivity check. In columns 3-4, we therefore remove the municipality trends. We then find that the estimated P4P increases slightly, but it is only statistically significant when we include the covariates (column 4). In contrast to what we found when including covariates in our trend specification, the covariates thus increase precision when the specification does not include trends. This makes sense: the covariates should increase in importance when we exclude the trends, which otherwise pick up much of the variation.³⁰

Column (5) shows the estimates from a specification in which the data is not weighted by population size. The estimate is slightly smaller compared to the baseline specification, suggesting a stronger P4P effect in larger municipalities.

7.3 Confounding

We next address the identification threats discussed in Section 6.1. To check if the estimate merely picks up already existing differential trends, we augment the main specification with placebo dummies for the two years before the P4P municipalities actually became subject to P4P.³¹ Reassuringly, we see from column 1 of Table 5 that the PcV P4P effect is hardly affected by the inclusion of the two placebo dummies, which themselves have small and insignificant impacts on the PcV share. Notably, this result speaks to the potential worries about a confounding effect of the introduction of the national PcV target in 2009. As the national target was launched during the placebo years for several of the P4P counties, we would expect to see larger placebo effects if confounding with the target was an important issue.

Another concern is that the estimated α may pick up the effect of *other* policy changes

³⁰When we do not partial out the trends from our preferred specification, R^2 is 0.52.

³¹Note that this means that the dummies equal 1 in different calendar years for different P4P counties; see Table 1. We can only include two placebo years, in order to still have a before-P4P period for the first county council that introduced PcV P4P.

that were implemented by the treated county councils contemporaneously with P4P. For instance, many counties increased patients freedom to choose primary care center during the study period, potentially affecting physicians' sensitivity to patients' demands. To check whether we should be concerned about such policy changes, we estimate four models using outcome variables that have nothing to do with RTI (columns 2-5 of Table 5). We consider a set of highly different outcomes: the number of prescriptions of emollients and protectives with no specific therapeutic effect (column 2), antidepressants (3), or sleeping pills (4); and the number of statin users (5).³² The outcome variable in column 2 includes e.g. lotions that do not require a prescription, but patients may still request a prescription to reach the annual cost ceiling in the drug subsidy system. Antidepressants and sleeping pills are widely used in Sweden (used by 10% of the population yearly) and are therefore important to patients. County-wide changes in physicians' response to patients' demand may thus be reflected in these measures.³³ However, we find no statistically significant impact on any of these outcomes. It can also be noted that a dummy for just having introduced patient choice (one of the covariates in Table 4) has no impact on the estimates.

A remaining potential confounder is the national incentive program (see Section 4.2.2), which may have had particularly strong impact on the P4P counties. Though the most salient component of the national program was the incentive to reduce the number of antibiotics prescriptions, one may argue the program might have affected the PcV share via another component, namely the accompanying local educational efforts to stimulate antibiotics stewardship. Indeed, given that the P4P counties had relatively low PcV shares, it is not implausible that educational efforts in these counties were particularly focused on communicating RTI guidelines, potentially resulting in an increased PcV share.

Now, we argue that if this explanation drives our main treatment effect, there should be a gradient in the P4P effect, i.e. larger increases in the PcV shares of municipalities whose initial PcV share was relatively low. The reason is that even within counties with a low average PcV share, there are municipalities whose PcV share was average or high in a national perspective, and there is no reason why the local educational efforts targeting these municipalities would stress (or affect) the PcV share in particular. Consequently, if local educational efforts is the main reason for our main effect, then it ought to be driven by the municipalities with a low initial PcV share. That is not the case however: when we interact our treatment variable with a dummy indicating municipalities whose 2006 PcV share was below the national median (column 6), we find that the interaction is small and insignificant, and it is far from soaking up the baseline effect, which is still significant and almost as large as before.³⁴

As a second approach to examine the potential impact of the national program, we recognize that it placed high emphasis on reducing total antibiotics consumption. Thus, if the PcV P4P counties were relatively sensitive to the national program, we would expect

³²See Appendix C.1 for exact variable definitions.

³³ <http://www.socialstyrelsen.se/publikationer2016/2016-4-25>

³⁴Note that this estimation excludes observations from 2006, i.e. the estimation sample starts in 2007. When estimating our baseline specification (i.e. no interaction with below-median 2006 PcV share) on the 2007-2013 sample, we get a P4P estimate of 0.0177 percentage points. This is more or less the same as when the observations from 2006 are included.

to see a "treatment effect" also with respect to the overall consumption of antibiotics.³⁵ However, the PcV P4P counties did not reduce overall prescriptions more than the control counties. Column 7 shows that there was absolutely no impact on the number of antibiotics prescriptions net of RTI antibiotics for children, and the next section shows that P4P counties did not cut down on RTI antibiotics to children either. In sum, we find little to substantiate the claim that the PcV P4P counties were particularly sensitive to the national incentive program.

A remaining potential confounder is that the P4P policy might have coincided with a period of milder RTI morbidity, warranting less broad-spectrum PcV, in the P4P counties. In the next section, we present evidence suggesting that different morbidity does not explain our results.

7.4 Channels and RTI morbidity

An increase in the PcV share may reflect increases as well as decrease in the consumption *level* of PcV and broad-spectrum antibiotics, and it does not automatically mean that the total consumption of RTI antibiotics has decreased. To see how the P4P effect operated, Table 6 shows the effect on the consumption *levels* of broad-spectrum (*RTI broad*) and narrow-spectrum *PcV* RTI antibiotics, and how these changes affected the overall consumption of RTI antibiotics ($total = RTI\ broad + PcV$). In columns 1 and 2, the outcome variables are the number of prescriptions of broad-spectrum RTI antibiotics and the number of PcV prescriptions respectively.³⁶

Though none of the estimates are statistically significant, they indicate a reduction in the number of broad-spectrum prescriptions and an increase in the number of PcV prescriptions of similar size. Consequently, the overall consumption of antibiotics for children with RTI was hardly affected at all (column 3). Thus, P4P did not lead physicians to turn down parents demanding antibiotics for their children; rather, they prescribed another, less resistance-driving, preparation. This is exactly the behavior that the PcV P4P indicator was designed to stimulate. It is also important *per se* that the improvement was not achieved by physicians gaming the system, i.e. trying to improve the ratio by prescribing more PcV without reducing broad-spectrum prescriptions.

The lack of change in overall RTI antibiotics consumption is also suggestive in relation to concerns that our main P4P effect may reflect lower RTI morbidity in the P4P group. If the main effect merely reflected lower morbidity, we would reasonably have seen a decreasing overall consumption of RTI antibiotics.

³⁵The PcV P4P counties had a relatively high overall consumption too, suggesting that they should be at least as concerned about the overall consumption target as the control group.

³⁶The numbers are divided by the population size per municipality and year (in thousands), and, as before, both variables include only prescriptions prescribes to children between 0-6 years of age. See Appendix C.1.

Table 5: Counfounding 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> |
|----------------------|-------------------------|--------------------|---------------------|--------------------|----------------------|-------------------------|------------------------|
| 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) |
| Placebo t-1 | 0.000607 (0.00790) | | | | | | |
| Placebo t-2 | 0.000426 (0.00444) | | | | | | |
| PcV P4P \times P50 | | | | | | 0.00342 (0.00989) | |
| Mean of <i>y</i> | 0.69 | 97.3 | 56.0 | 496.8 | 97.9 | 0.69 | 350.5 |
| Observations | 2,312 | 2,312 | 2,312 | 2,312 | 2,312 | 2,023 | 2,312 |
| R-squared | 0.016 | 0.008 | 0.001 | 0.001 | 0.000 | 0.018 | 0.000 |
| Municipalities | 289 | 289 | 289 | 289 | 289 | 289 | 289 |

(1) Placebo years; (2)-(5) Placebo substances; (6) Interaction with below-median PcV share in 2006 (P50) – sample starts in 2007; (7) Number of prescriptions of non-RTI antibiotics per 1,000 residents. For definitions for dependent variables in columns 2-5 and 7, see Appendix C.1. Standard errors clustered by county council in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 6: Channels and related substances

| <i>y</i> : | (1) <i>RTI broad</i> | (2) <i>PcV</i> | (3) <i>total</i> | (4) <i>Nasal prep.</i> | (5) <i>Cough & cold</i> | (6) <i>Temp leave</i> |
|------------------|-------------------------|-------------------|---------------------|---------------------------|--------------------------------|--------------------------|
| PcV P4P | -0.509 (0.322) | 0.525 (0.558) | 0.017 (0.792) | 0.417 (0.579) | -3.999 (2.794) | 0.827 (2.166) |
| Mean of <i>y</i> | 9.9 | 20.1 | 30.0 | 116.0 | 159.0 | 184.2 |
| Observations | 2,312 | 2,312 | 2,312 | 2,312 | 2,312 | 2,312 |
| Municipalities | 289 | 289 | 289 | 289 | 289 | 289 |
| R2 | 0.004 | 0.002 | 0.000 | 0.000 | 0.013 | 0.001 |

See Appendix C.1 for variable definitions. Standard errors clustered by county council in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

To further check that differential RTI morbidity is not a confounder, we consider three additional outcome measures that may relate to RTI morbidity. In column 4, the outcome variable is the number of prescriptions of nasal preparations³⁷ per 1000 inhabitants. There is no trace of impact on the prescription of such preparations in county councils that use PcV-related P4P. In column 5, the outcome variable is the number of prescriptions of cough and cold preparations³⁸ per 1000 inhabitants. It should be acknowledged that this measure is not as closely related to children’s RTI morbidity as it seems at first glance, as physicians are recommended not to prescribe such medications to children because of negative side-effects (Läkemedelsverket, 2008). In any case, there is no effect on these medications either.

In column 6, we consider another outcome variable that relates closely to RTI morbidity; namely the number of cases of temporary parental leave (Sw. *Tillfällig föräldrapenning*). In Sweden, where most parents participate on the labor market, children’s RTI’s are inevitably linked to parental work absence. To cover the loss of income, parents can claim reimbursement from the national parental insurance from the first day of absence, importantly without having to first take the child to a physician. It is thus unlikely that physicians’ prescription behavior would affect parents’ decision to initiate a period with temporary parental leave.³⁹ We find no association between our P4P indicator and the number of cases of temporary parental leave, further suggesting that the post-P4P period did not coincide with a period of unusually low RTI morbidity.

7.5 Heterogeneity

Given that the counties’ P4P schemes differed in details such as targets and reimbursement rates, it is tempting to try to identify features that are related to the strength of the P4P effect. Given that there were only eight unique P4P schemes, the feasibility of disentangling the importance of specific features are obviously limited. However, in Appendix E, we show estimations indicating that the effect was larger for municipalities in a few counties whose P4P schemes had theoretically advantageous features, i.e. a continuum of targets or incentives formulated as penalties. According to the so-called *goal-gradient hypothesis*, little

³⁷ATC code R01

³⁸ATC code R05

³⁹Physicians’ prescription behavior can affect the number of *days* of absence, i.e. the intensive margin, but we study the extensive margin here.

behavioral change can be expected if physicians consider the target infeasible, i.e. because their current level of performance is far below or above the target. This suggests that there should be a continuum of targets in order to affect agents at all levels of performance. Further, if loss-aversion is a prominent characteristic of health care providers, one may expect a larger response to financial penalties than to rewards (bonuses) (Kahneman and Tversky, 1979; Eijkenaar, 2013). The counties with theoretically advantageous P4P schemes also include many municipalities whose initial PcV shares were relatively low. Given that the P4P effect was broadly similar for municipalities above and below the median initial PcV share (see Section 7.3), we do not think that this is the main explanation for the heterogeneity though.

We further find that there is a tendency that the effect was stronger in municipalities with a large share of privately owned primary care centers. Private providers may be more sensitive to financial incentives due to profit motives and harder budget constraints. Recognizing that private providers are more common in urban areas, this may be a reason why the P4P estimate is slightly attenuated when we do not apply population weights (see Table 4).

8 Mechanisms

Our findings are consistent with the idea that P4P can affect physicians' prescription of RTI antibiotics. However, recalling that the financial incentives were small, and not directly tied to the prescribing physicians' salaries, it is hard to believe that the outlooks for earning more (or less) money made the whole difference for their prescription behavior (c.f. McDonald et al., 2009). Another mechanism was likely operating.⁴⁰

A plausible mechanism is that P4P made the antibiotics issue more salient for physicians, thus lowering their psychological barriers to changing well-established prescription routines (Celhay et al., 2015). There are at least two reasons why P4P might have increased issue salience. First, the P4P system brought the PcV share to physicians' attention by repeatedly generating feedback to primary care centers about their prescription patterns. Second, care center managers ought to have communicated to their staff that the PcV share might affect revenues, thereby signalling that it was an important outcome. Notably, the second of these reasons suggests that the financial element of P4P was crucial to *make* the issue salient. Although P4P was a minor source of revenue, it at least offered care centers a way to influence their revenues. In a setting otherwise dominated by fixed reimbursement (capitation), care center managers should be expected to respond to this incentive.

It is further conceivable that the increased focus on PcV following the introduction of P4P lead physicians to learn more about how to prescribe in accordance with antibiotic stewardship. In other words, P4P may have worked via an educational mechanism. Though, against the background of the generally small impacts of educational efforts to improve antibiotics stewardship in Scandinavian countries (see Section 3), we believe that the saliency channel was more important than the educational channel.

Regardless of whether the effect was driven by salience or education, P4P implied an impetus to change prescription routines with regards to the selection of antibiotics for chil-

⁴⁰Notably, the limited incentive size was a commonality of the county councils' own P4P schemes and the national incentive programme.

dren with RTI. After having established a new routine, another impetus would be required to reverse the new routine (Celhay et al., 2015). This can explain why the effect persisted in counties that later removed the monetary incentive.

9 Concluding remarks

This study suggests that P4P can stimulate antibiotic stewardship. Our baseline estimate corresponds to an effect size of approximately 20% of a pre-P4P standard deviation. Generalizing to the control group, P4P would close almost one third of the gap between their average PcV share in 2013 (74%) and the national target of 80%. Still, the effect is small in absolute terms, suggesting that larger incentives or complementary policies are required to substantially affect physicians' antibiotic prescription patterns.

As the financial incentives were very small, the effect was achieved at low cost; with so little money at stake, it is rather surprising that we find an effect at all. To make a full economic evaluation of the policy, we would have to translate the increased PcV share into a final outcome, for instance the number of cases of antibiotics-resistant infections avoided due to P4P. Given the stochastic evolution of resistance, this task would require simulation exercises far outside the scope of the present study.

The medical risks associated with the PcV P4P policy appear small, as the P4P targets still allowed for a fairly large share of broad-spectrum drugs. But the exceptionally high reliance on PcV in Scandinavia suggests that the PcV share as such may be less relevant as a P4P indicator outside Scandinavia. Core features of the indicator are however generally applicable. Indeed, one of the antibiotics-related P4P indicators just introduced in the British QoF scheme is very similar to the one study here, differing mainly with regards to what drugs enter the nominator and denominator. Though the external validity of our results are uncertain, it is a promising sign that we find an impact of P4P tied to a process measure, as this aligns with previous research (mainly from the US and UK) on P4P in health care. Moreover, our P4P effect implies a relative improvement close to the 5% suggested as typical by Van Herck et al. (2010) in their review of P4P in health care. Recalling that other interventions to improve antibiotics stewardship (e.g. education and feedback) generally have had larger impact outside Scandinavia, our estimate of the effect of antibiotics-related P4P may even represent a lower bound for non-Scandinavian countries.

Notably, this study suggests that physicians switched antibiotic type, but did not issue fewer prescriptions overall. To impede the spread of resistance, it is important that physicians quit prescribing unwarranted antibiotics. Larger monetary incentives may be needed to influence physicians' decision to prescribe, as patients are probably less bothered by the substitution of one substance for another than by an outright refusal to get a prescription. Indeed, patients tend to be more satisfied with primary care practices that prescribe more antibiotics (Ashworth et al., 2015). In conclusion, research on the importance of the size of incentives, especially in relation to overall prescription, is warranted. Other interesting fields for future research relate to P4P design features, especially the formulation of target and penalties versus rewards, and on differences between private and public providers.

References

- Andersson, F., Janlöv, N., Rehnberg, C., 2014. Konkurrens, kontrakt och kvalitet – hälso- och sjukvård i privat regi.
- André, M., et al., 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., 2006. Tradition och förnyelse i primärvården. Unpublished manuscript.
- Anell, A., 2009. Målrelaterad ersättning i primärvården : kartläggning av mål/indikatorer och former för ersättningen hösten 2009. Sveriges Kommuner och Landsting.
- Anell, A., 2011. Choice and privatization in Swedish primary care. *Health Economics, Policy and Law* 6 (4), 549–569.
- Anell, A., 2013. Vårdval i specialistvården: utveckling och utmaningar. Sveriges Kommuner och Landsting.
- Anell, A., Dackehag, M., Dietrichson, J., 2016. Does risk-adjusted payment influence primary care providers' decision on where to set up practice? Unpublished manuscript.
- Anell, A., Glenngård, A. H., Merkur, S., 2012a. Sweden: Health system review. *Health systems in transition* 14 (5), 1–159.
- Anell, A., Nylinder, P., Glenngård, A., 2012b. Vårdval i primärvården: Jämförelse av uppdrag, ersättningsprinciper och kostnadsansvar. Sveriges Kommuner och Landsting.
- 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., et al., 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*.
- Bertrand, M., Duflo, E., Mullainathan, S., 2004. How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics* 119 (1), 249–275.
- Bjerrum, L., et al., 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.
- Bradley, J. S., et al., 2011. The management of community-acquired pneumonia in infants and children older than 3 months of age: clinical practice guidelines by the Pediatric Infectious Diseases Society and the Infectious Diseases Society of America. *Clinical Infectious Diseases* 53 (7), e25–76.
- 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.

- Cameron, A. C., Miller, D. L., 2015. A practitioner's guide to cluster-robust inference. *Journal of Human Resources* 50 (2), 317–372.
- Carlet, J., et al., 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.
- Coast, J., Smith, R. D., Millar, M. R., 1996. Superbugs: Should antimicrobial resistance be included as a cost in economic evaluation? *Health Economics* 5 (3), 217–226.
- 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 5p.
- de Walque, D., et al., 2015. Using provider performance incentives to increase HIV testing and counseling services in Rwanda. *Journal of Health Economics* 40, 1–9.
- Deci, E. L., Koestner, R., Ryan, R. M., 1999. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin* 125 (6), 627–668.
- Drake, D. E., Cohen, A., Cohn, J., 2007. National hospital antibiotic timing measures for pneumonia and antibiotic overuse. *Quality Management in Health Care* 16 (2), 113–122.
- ECDC, 2010. Resistance surveillance in Europe 2009. Annual Report of the European Antimicrobial Resistance Surveillance Network (EARS-Net).
- ECDC, 2015. Resistance surveillance in Europe 2014. Annual Report of the European Antimicrobial Resistance Surveillance Network (EARS-Net).
- ECDC/EMA Joint working group, 2009. ECDC/EMA Joint Technical Report: The bacterial challenge: time to react.
- Eijkenaar, F., 2013. Key issues in the design of pay for performance programs. *The European Journal of Health Economics* 14 (1), 117–131.
- 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.
- Elbasha, E. H., 2003. Deadweight loss of bacterial resistance due to overtreatment. *Health Economics* 12 (2), 125–138.
- Fleming-Dutra, K., Hersh, A., Shapiro, D., et al, 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.
- Foxman, B., et al., 1987. The effect of cost sharing on the use of antibiotics in ambulatory care: results from a population-based randomized controlled trial. *Journal of Chronic Diseases* 40 (5), 429–437.
- Friedberg, M. W., Mehrotra, A., Linder, J. A., 2009. Reporting hospitals’ antibiotic timing in pneumonia: adverse consequences for patients? *The American journal of managed care* 15 (2), 137–144.
- Gjelstad, S., et al., 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., Li, M., Luo, R., He, L., Wei, Q., Shen, S., Liu, Y., Zhang, L., Zhou, W., Huang, P., Mai, J., Liu, L., Xu, Y., Liang, H., 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. *Public Health Policy*, 99.
- 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.
- Gravelle, H., Sutton, M., Ma, A., 2010. Doctor Behaviour under a Pay for Performance Contract: Treating, Cheating and Case Finding?*. *The Economic Journal* 120 (542), F129–F156.
- Grossbart, S. R., 2006. What’s the Return? Assessing the Effect of “Pay-for-Performance” Initiatives on the Quality of Care Delivery. *Medical Care Research and Review* 63 (1 suppl), 29S–48S.
- Hallsworth, M., et al., 2016. Provision of social norm feedback to high prescribers of antibiotics in general practice: a pragmatic national randomised controlled trial. *The Lancet* 387 (10029).
- Holmström, B., 1979. Moral Hazard and Observability. *The Bell Journal of Economics* 10 (1), 74–91.
- Kahneman, D., Tversky, A., 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47 (2), 263–292.

- 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*, n/a–n/a.
URL <http://onlinelibrary.wiley.com/doi/10.1002/hec.3258/abstract>
- Lemienegre, M. B., et al., 2012. Antibiotics for clinically diagnosed acute rhinosinusitis in adults. In: *The Cochrane Collaboration* (Ed.), *Cochrane Database of Systematic Reviews*. Chichester, UK.
- 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.
- Lindenauer, P. K., Remus, D., Roman, S., Rothberg, M. B., Benjamin, E. M., Ma, A., Bratzler, D. W., 2007. Public Reporting and Pay for Performance in Hospital Quality Improvement. *New England Journal of Medicine* 356 (5), 486–496.
- Lindgren, P., 2014. Ersättning i sjukvården - modeller, effekter, rekommendationer. SNS Förlag.
- Läkemedelsverket, 2008. Farmakologisk behandling av nedre luftvägsinfektioner i öppen vård. Information från Läkemedelsverket 3:2008.
- MacCara, M. E., Sketris, I. S., Comeau, D. G., Weerasinghe, S. D., 2001. Impact of a limited fluoroquinolone reimbursement policy on antimicrobial prescription claims. *The Annals of Pharmacotherapy* 35 (7-8), 852–858.
- 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.
- Matthys, J., De Meyere, M., van Driel, M. L., De Sutter, A., 2007. Differences among international pharyngitis guidelines: not just academic. *Annals of Family Medicine* 5 (5), 436–443.
- McDonald, R., Boaden, R., Roland, M., Kristensen, S. R., Meacock, R., Lau, Y.-S., Mason, T., Turner, A. J., 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*. NIHR Journals Library, Southampton (UK).

- 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., et al., 2014. Nudging guideline-concordant antibiotic prescribing: A randomized clinical trial. *JAMA Internal Medicine* 174 (3), 425–431.
- Meeker, D, Linder, JA, Fox, CR, et al, 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., et al., 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., Søgaaard, P., Søgaaard, 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>.
- 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.
- Prendergast, C., 1999. The Provision of Incentives in Firms. *Journal of Economic Literature* 37 (1), 7–63.
- Ranji, S. R., Steinman, M. A., Shojania, K. G., et al., 2006. Closing the Quality Gap: A Critical Analysis of Quality Improvement Strategies (Vol. 4: Antibiotic Prescribing Behavior). AHRQ Technical Reviews. Agency for Healthcare Research and Quality (US), Rockville (MD).
- 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.
- Schaffer, M., 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>.

- Scott, A., Sivey, P., Ait Ouakrim, D., et al., 2011. The effect of financial incentives on the quality of health care provided by primary care physicians. In: Cochrane Database of Systematic Reviews.
- Smith, S. M., Fahey, T., Smucny, J., Becker, L. A., 2014. Antibiotics for acute bronchitis. In: Cochrane Database of Systematic Reviews.
- Socialdepartementet, 2010. Bemyndigande att underteckna en överenskommelse om patientsäkerhetssatsning 2011.
- 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.
- 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.
- Turnidge, J., Christiansen, K., 2005. Antibiotic use and resistance—proving the obvious. *The Lancet* 365 (9459), 548–549.
- van der Velden, A. W., Pijpers, E. J., Kuyvenhoven, M. M., et al., 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., et al., 2013. Antibiotics for acute otitis media in children. The Cochrane Database of Systematic Reviews 1.
- Vodicka, T. A., Thompson, M., Lucas, P., et al., 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.
- WHO, 2014. Antimicrobial resistance: global report on surveillance. World Health Organization.
- Yip, W., Powell-Jackson, T., Chen, W., et al., 2014. Capitation combined with pay-for-performance improves antibiotic prescribing practices in rural China. *Health Affairs (Project Hope)* 33 (3), 502–510.

A Translation of Strama’s national targets

Strama’s national targets are available in Swedish at their webpage and from the corresponding author upon request.⁴¹ The following is a direct translation of the document containing the targets:

”Strama’s targets for outpatient antibiotic use

1. Five years from now, the total prescription of antibiotics in Sweden should not be higher than 250 prescriptions per 1,000 inhabitants on a yearly basis.

The target comprises the whole ATC group J01 excluding methenamine and refers to the national level. The target is not applicable to individual health care units, but may be an indicator at the county council level.

2. Penicillin V should account for 80 percent of antibiotics for respiratory tract infections to children aged 0-6 years.

The target is not directly connected to the reason for the prescription, but is indirectly related to diagnosis as it is based on the pharmacies’ sales data on typical “respiratory tract substances” (see below). Sales are measured by the number of prescriptions per 1,000 inhabitants and year.

Nominator: Prescriptions of penicillin V (J01CE02) expedited by pharmacies. All package sizes. Children 0-6 years.

Denominator: Prescriptions of amoxicillin (J01CA04), penicillin V (J01CE02), amoxicillin with clavulanic acid (J01CR02), cephalosporines (J01DB-DE) and macrolids (J01FA) expedited by pharmacies. All package sizes. Children 0-6 years.

3. Fluoroquinolones should account for no more than 10 percent of prescribed antibiotics for women (ages 18-79) with urinary tract infections”

The motivation for the PcV target is found later in the document, under the heading ”Luftvägsinfektioner”:

”Respiratory tract antibiotics

Antibiotics only contribute slightly to the recovery from most of the common RTIs in children. Pneumonia is the exception. The common cold and acute bronchitis in children should not be treated with antibiotics. Otitis in children above 2 years of age often does not require treatment. It is extremely rare that children require treatment for acute rhinosinusitis.

Penicillin V is the first-line drug when otitis and tonsillitis require treatment. Only a small share of the children get relapsed or complicated otitis or relapsed tonsillitis, in which case other drugs may be needed. Also for pneumonia, PcV is the first-line drug.

For these reasons, we estimate that PcV could account for more than 80 percent of all RTI antibiotics for children.”

⁴¹<http://strama.se/wp-content/uploads/2016/04/Stramas-mal-for-antibiotikaanvandningen-beskrivning.pdf> (last accessed 2016-05-19)

B Antibiotics-related P4P indicators in PcV P4P group, by year

Figure B.1: P4P indicators in the treatment group.

| PcV-related P4P | | | | | | | | |
|-----------------|----------|---------|-------|----------------|---------|-----------|-----------------|----------|
| Year | Blekinge | Dalarna | Skåne | Västernorrland | Halland | Kronoberg | Stockholm (SLL) | 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 |

X = County council uses a P4P indicator related to the PcV share

| Total antibiotics consumption P4P | | | | | | | | |
|-----------------------------------|----------|---------|-------|----------------|---------|-----------|-----------------|----------|
| Year | Blekinge | Dalarna | Skåne | Västernorrland | Halland | Kronoberg | Stockholm (SLL) | Sörmland |
| 2006 | | | | | | | | |
| 2007 | | | | | | | | |
| 2008 | | | | | | | | |
| 2009 | | | | | | | | |
| 2010 | | | | | | | | |
| 2011 | X | | X | | | | | X |
| 2012 | X | | X | | | X | | X |
| 2013 | X | | X | | | X | | X |

X = County council uses a P4P indicator related to total antibiotics consumption

| UTI-related P4P | | | | | | | | |
|-----------------|----------|---------|-------|----------------|---------|-----------|-----------------|----------|
| Year | Blekinge | Dalarna | Skåne | Västernorrland | Halland | Kronoberg | Stockholm (SLL) | Sörmland |
| 2006 | | | | X | | | | |
| 2007 | | | | X | | | | |
| 2008 | | | | X | | | | |
| 2009 | X | | X | | X | | | |
| 2010 | X | X | X | X | X | | | |
| 2011 | X | X | X | X | | | X | |
| 2012 | | X | X | X | | X | X | |
| 2013 | | | | X | | X | X | |

X = County council uses a P4P indicator related to UTI antibiotics consumption

C Variables: definitions and summary statistics

C.1 Dependent variable definitions

Table C.1: Definitions of antibiotics-related dependent variables

| Variable | Description | Age group |
|------------------|--|-----------|
| <i>PcV share</i> | Nominator: <i>PcV</i> (see below) Denominator: <i>total</i> (see below) | 0-6 years |
| <i>PcV</i> | # per 1,000 of J01CE02 | 0-6 years |
| <i>RTI total</i> | # per 1,000 of J01CE02/J01AA02/J01CA04/J01CR02/J01DB-DE/J01FA | 0-6 years |
| <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 years |

The table shows definitions of the antibiotics-related dependent variables. # per 1,000 = number of 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.

Table C.2: 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 cap | 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 |

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

C.2 Covariate definitions

Table C.3: 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 group council uses <i>total</i> or <i>UTI</i> P4P (Section 4.2.2)) | 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 |

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

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 4.2.2 (i.e., P4P related to either total antibiotics prescriptions or to UTI quinolones). *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). A recent study has shown that the introduction of such reforms led to a temporary increase in overall antibiotics use in some municipalities (Fogelberg, 2014). Finally, the covariate set also includes the dummy variable *cost responsibility*, which equals one if the county council has delegated the budget responsibility for drugs to primary care centers.

C.3 Summary statistics: dependent variable

This section presents additional descriptives on the main dependent variable. Figure C.1 shows that the jump between the regression lines in Figure 2 or Section 5, which is reproduced in panel (a), is attenuated when we shift back the cut-off for the regressions one time period. Figure C.2 shows that we get similar a similar pattern when we exclude the two counties with shortest pre-/post periods (Skåne and Kronoberg) to extend the regression lines. Figure C.3 shows that the pattern and jump are similar with unweighted data, and when we remove municipalities located in the largest county council (SLL) from the sample.

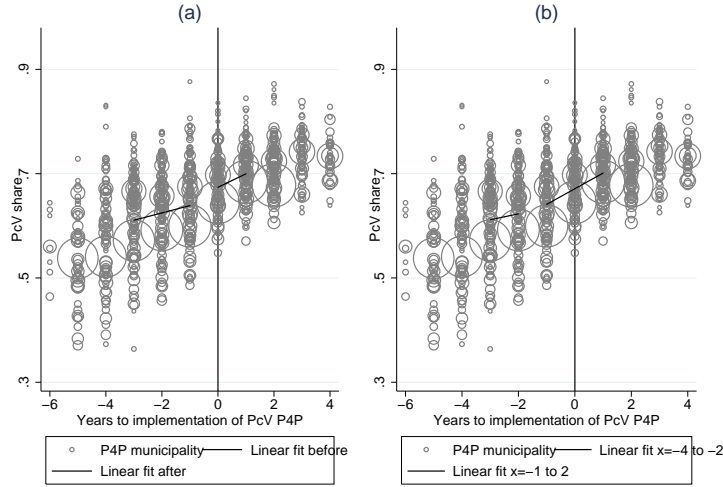


Figure C.1: (a) Reproduction of Figure XX in Section 5; (b) Same, but shift back regression line cut-off one year.

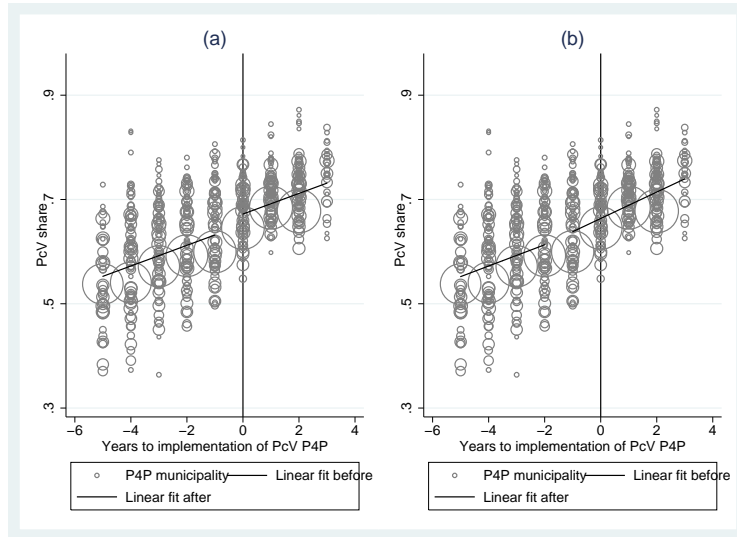


Figure C.2: (a) Longer regression period; (b) Same, but shift back regression line cut-off one year.

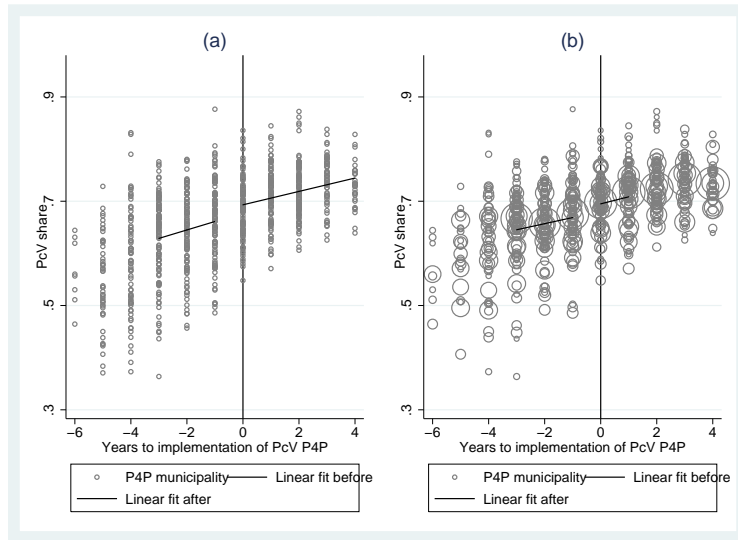


Figure C.3: (a) Unweighted data; (b) Excluding SLL

C.4 Summary statistics: covariates

Table C.4: Summary statistics for 2008, by later P4P status

| Variable | Ever PcV P4P | | Never PcV P4P | | <i>p-value of difference</i> |
|--------------------------------|--------------|---------|---------------|--------|------------------------------|
| | Mean | S.d. | Mean | S.d. | |
| <i>share children (%)</i> | 11.6 | (1.5) | 10.6 | (1.1) | 0.072 |
| <i>share elderly (%)</i> | 17.0 | (3.3) | 18.5 | (3.2) | 0.178 |
| <i>log(population)</i> | 11.3 | (1.3) | 10.8 | (1.3) | 0.043 |
| <i>mean income</i> | 250.9 | (40.0) | 230.0 | (15.8) | 0.191 |
| <i>share secondary edu (%)</i> | 42.8 | (6.4) | 46.3 | (5.4) | 0.056 |
| <i>share tertiary edu (%)</i> | 33.3 | (10.6) | 28.5 | (9.1) | 0.074 |
| <i>other P4P</i> | 0.05 | (0.226) | 0 | (0) | 0.001* |
| <i>choicereform</i> | 0.42 | (0.50) | 0.05 | (0.23) | 0.330* |
| <i>cost responsibility</i> | 0.15 | (0.35) | 0.80 | (0.40) | 0.000* |
| Observations | 108 | | 181 | | |

See Appendix C.2 for variable definitions. Column *p-value of difference* shows p-values from regressions of each variable on a dummy for PcV P4P county council, i.e. 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. * County-council level variable. The difference between P4P and control counties is not statistically significant when contrasted in t-tests at the county level (n=21).

D Estimates on covariates

Table D.1: 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 the covariates. Column 1 (2) refers to estimations of P4P effect in column 2 (4) of Table 4. Standard errors clustered by county council in parentheses. *** p<0.01, ** p<0.05, * p<0.1

E Heterogeneity estimations

This Appendix explores heterogeneous responses to P4P. First, we examine whether the P4P effect was stronger in municipalities with a large share of private providers. Second, we try to test the goal-gradient hypothesis, which states that the P4P effect should be stronger for observations close to their county's P4P target. We then identify a few counties that are important for our main result.

To examine the first hypothesis, we interact the treatment variable with a variable measuring the share of *private* primary care centers.⁴² In this specification, there is thus within-county council variation in the intensity of treatment. The base effect of P4P is smaller than in the baseline specification and statistically insignificant (column 1 of Table E.1). Thus, P4P makes a smaller difference in municipalities where there are only public providers (about one third of the municipalities). The interaction term is statistically insignificant; however,

⁴²This information comes from a register over all primary care centers operating in Sweden between 2005-2013, which we have constructed by augmenting information from two previous registers with information obtained through direct communication with providers and authorities.

the total marginal effect is positive and significant for realistic values of the share of private providers; for instance, the total marginal effect is 1.3 percentage points ($p=0.0235$) for municipalities with the average share of private providers (25% over the whole period).⁴³ In the extreme case of only private providers (21 municipalities), the P4P effect amounts to 3 percentage points (see row *total ME* in the table).

We next try to examine the goal-gradient hypothesis. We estimate a model where the treatment dummy is interacted with the dummy variable $affected_{mc}$, which equals one for municipalities that were at most one standard deviation away from the P4P target of their council *in the year before P4P was introduced*.⁴⁴ We choose a time-invariant specification of $affected$ to avoid post-treatment bias. The estimates in column 2 suggest a positive but insignificant ($p=0.498$) incremental effect of 0.5 percentage points for municipalities that were classified as affected. Thus, this estimation yields no strong evidence in favour of the goal-gradient hypothesis.⁴⁵

Table E.1: Extensions

| | (1) | (2) | (3) | (4) |
|------------------|----------------------|-----------------------|----------------------|----------------------|
| PcV P4P | 0.00783 (0.00909) | 0.0146** (0.00723) | 0.0110 (0.00887) | 0.00901 (0.00816) |
| PcV P4P×Private | 0.0224 (0.0164) | | | |
| PcV P4P×affected | | 0.00520 (0.00850) | -0.0103 (0.00801) | |
| Total ME | 0.0303 | 0.0198 | 0.000708 | |
| p | 0.0018 | 0.0014 | 0.9413 | |
| Observations | 2,312 | 2,312 | 2,000 | 2,000 |
| Municipalities | 289 | 289 | 250 | 250 |
| R2 | 0.017 | 0.016 | 0.003 | 0.002 |

(1) Interaction with *private*. (2)-(3) Interaction with *affected*. (3)-(4) Excluding municipalities located in Halland, Kronoberg and SLL counties. Standard errors clustered by county council in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Total ME = total marginal effect (see text).

Theoretically, one advantage of the goal-gradient specification is that it introduces within-county council variation in the intensity of treatment. Table E.2 however shows that the amount of such variation should not be overstated: in three of the county councils in our sample (SLL, Kronoberg and Halland), all municipalities are classified as close to their P4P target according to our definition. If we estimate the model using only observations from county councils that contribute to the within-council variation in treatment intensity (column 3), the interaction effect goes in the opposite direction from what we expected,

⁴³The total marginal effect is calculated as the base effect + interaction effect×0.25

⁴⁴The standard deviation is calculated separately for each county council, pooled over the years 2006-2008.

⁴⁵Though the total marginal effect is significant, $p=0.001$. Note that it is the combination of a significant base effect and an insignificant interaction effect that make us draw the conclusion that there is no gradient. In the previous estimation of the importance of private providers, in which we also found an insignificant interaction term, the base effect was insignificant too but the total marginal effect was significant; hence, there was a gradient in that case.

though it is still insignificant. Moreover, the base effect is also attenuated and statistically insignificant, suggesting that the municipalities located in the three excluded county councils are important for our main result.

Table E.2: Closeness to target

| | Blekinge | Dalarna | Skåne | V-norrland | Halland | SLL | Södermanland | Kronoberg |
|-----------------|----------|---------|-------|------------|---------|-----|--------------|-----------|
| <i>affected</i> | 0.2 | 0.87 | 0.33 | 0.86 | 1 | 1 | 0.44 | 1 |

Notes: The table shows the share of municipalities where the PcV share was less than 1 standard deviation away from the P4P target (share of *affected* municipalities).

The latter suspicion is confirmed in column 4. When we estimate our preferred model (Eq. (1)) on the same restricted sample, the P4P effect is just below 1 percentage point (column 3 of Table E.1). The estimate is statistically insignificant, which is not surprising as we lose more than one third of the treatment group in this specification (39 of 109 municipalities). In relation to our baseline model, it should be noted that the excluded municipalities belong to the control group before they implement P4P and thus the estimates are not directly comparable. Nonetheless, there are some intuitive reasons why these counties would be important for the main effect. First, SLL and Kronoberg had a continuum of targets, and as shown by Table E.2, all observations in Halland were affected by the incentive. This suggests that the goal-gradient hypothesis may have some relevance, after all. Second, SLL and Halland were the only councils that operationalized their incentives (at least partly) as penalties, suggesting that loss-aversion may be crucial for the success of P4P. Third, there were on average more private providers in these three councils (mean 53%) than in the other five P4P councils (mean 33 %). Fourth, the three councils had relatively low PcV shares before they introduced P4P: 53% on average, to be compared with 65% in the other P4P councils.