

© Health Research and Educational Trust
DOI: 10.1111/1475-6773.12033
RESEARCH ARTICLE

An Examination of Pay-for-Performance in General Practice in Australia

Jessica Greene

Objective. This study examines the impact of Australia's pay-for-performance (P4P) program for general practitioners (GPs). The voluntary program pays GPs A\$40 and A\$100 in addition to fee-for-service payment for providing patients recommended diabetes and asthma treatment over a year, and A\$35 for screening women for cervical cancer who have not been screened in 4 years.

Design. Three approaches were used to triangulate the program's impact: (1) analysis of trends in national claims for incentivized services pre- and postprogram implementation; (2) fixed effects panel regression models examining the impact of GPs' P4P program participation on provision of incentivized services; and (3) in-depth interviews to explore GPs' perceptions of their own response to the program.

Results. There was a short-term increase in diabetes testing and cervical cancer screens after program implementation. The increase, however, was for all GPs. Neither signing onto the program nor claiming incentive payments was associated with increased diabetes testing or cervical cancer screening. GPs reported that the incentive did not influence their behavior, largely due to the modest payment and the complexity of tracking patients and claiming payment.

Implications. Monitoring and evaluating P4P programs is essential, as programs may not spark the envisioned impact on quality improvement.

Key Words. Pay-for-performance, quality of care, quality improvement, chronic disease care

Research has documented deficiencies in health care quality in many industrialized countries (Seddon et al. 2001; McGlynn et al. 2003; The Study Group of Diagnosis of the Working Group on Heart Failure of the European Society of Cardiology et al. 2003; Hussey et al. 2004; Tomio et al. 2010). Health policy experts often attribute the suboptimal quality to fee-for-service compensation, which incentivizes visit quantity rather than quality (Harris and Zwar 2007; Collier 2009; Stremikis, Davis, and Guterman 2010). Pay-for-performance (P4P) programs, which pay clinicians based upon achieving or improving specific quality metrics, are increasingly being used to improve quality of care. P4P programs range from offering small bonuses (approximately \$400)

to clinicians who reach threshold performance on a few indicators, such as breast and cervical cancer screening rates, to the Quality and Outcomes Framework in the United Kingdom that ties approximately one-fourth of clinicians' incomes to reaching over 100 quality indicators (Pearson et al. 2008; Campbell et al. 2009).

Recent reviews of the impact of P4P in primary care suggest that the programs generally have had limited, positive impacts (Petersen et al. 2006; Van Herck et al. 2010; de Bruin, Baan, and Struijs 2011; Scott et al. 2011). Van Herck and colleagues estimated that P4P programs have resulted in a 5 percent improvement in quality indicators on average, although they document substantial variation—from no impact to large improvements (Van Herck et al. 2010). Many questions remain unanswered about P4P programs (Petersen et al. 2006; Christianson, Leatherman, and Sutherland 2008). Little is known, for example, about the characteristics that distinguish highly successful programs, the factors that limit the impact of programs, and which clinicians are more responsive to P4P. In addition, while there is evidence from the United Kingdom that P4P programs can narrow socio-economic disparities in health care, there is broad concern that P4P programs may widen disparities (Casalino et al. 2007; Doran et al. 2008; Australian Department of Health and Ageing 2009).

This study uses mixed methods to evaluate the impact of a voluntary P4P program that incentivizes Australian general practitioners (GPs) to provide recommended diabetes and asthma care, as well as cervical cancer screening. The goals of this study were to (1) examine the impact of the P4P program on incentivized quality measures; (2) investigate whether there is a differential program impact based upon GP characteristics; and (3) explore GPs' perception of the program's impact on their practice.

AUSTRALIA'S PRACTICE INCENTIVES PROGRAM

In 2001, the Australian government initiated a financial incentive program for "improved management of diseases such as asthma and diabetes and increased screening for cervical cancer" for GPs, who are paid for each patient visit on a fee-for-service basis (Medicare Australia 2001). The voluntary program, called the Practice Incentives Program or PIP,¹ is open to GPs in

Address correspondence to Jessica Greene, Ph.D., Professor, George Washington University School of Nursing, 2030 M Street NW, Suite 300, Washington, DC 20036; e-mail:jessgreene@gwu.edu. At the time the research was conducted she was an Associate Professor at the University of Oregon.

practices that are accredited or undergoing accreditation and it continues today (Medicare Australia). Participation is incentivized with one-time, sign-on payments to medical practices, which are approximately \$250 per full-time GP in the practice for the asthma and cervical cancer programs and \$1,000 per GP for the diabetes program.

The diabetes and asthma incentives are based upon GPs providing patients with a cycle of care over a 12-month period. For diabetes, the cycle includes providing patients recommended tests (HbA1c, microalbuminuria, total cholesterol, HDL cholesterol, triglycerides, and blood pressure); relevant examinations (feet and eyes); reviewing medications, diet, physical activity, and smoking status; measuring body mass index; and providing diabetes self-care and related education. A GP earns A\$40 for each patient who completes the cycle of care, in addition to the regular consultation fee. To earn P4P incentive payments, the GP has to bill specific PIP-related codes for the visit. There is also a practice-level outcome incentive for practices that complete cycles of care for 20 percent or more of their patients with diabetes in a year (A\$20 per diabetes patient).

For asthma, the incentive is A\$100 for each patient who completes the cycle of care. The asthma cycle includes having two consultations with a patient during the year, one of which must be a planned check-up (prior to November, 2006, the requirement was three visits and two check-ups); reviewing the patient's use of medication and devices; assessing the patient's level of asthma control and severity; providing and reviewing an asthma action plan; and providing the patient with self-management education. Unlike the diabetes incentive, the asthma incentive does not include a practice-level payment.

The cervical cancer screening incentive is a reward of A\$35 when a GP screens a woman who has not been screened in the prior 4 years. The practice can also earn A\$3 per female patient in the practice if 65 percent (50 percent before August, 2011) of the women aged 20–69 in a practice have been screened.²

For all the PIP incentive payments there is a rural practice loading, which increases the payments to GPs practicing in rural areas. This ranges from a 15 percent increase for practicing in large rural centers like Cairns, to a 50 percent increase for practicing in very remote areas.

PRIOR RESEARCH ON AUSTRALIA'S P4P PIPS

There have been several prior studies on the Australian P4P program. Scott and colleagues published an evaluation of the diabetes incentive that found

“the average GP...that joined the PIP program is more than 20 percentage points more likely to order an HbA1c test than a comparable GP in a practice that has not joined” (Scott et al. 2009). The study, however, did not have data from prior to the implementation of the P4P program and used a proxy measure for program participation.

Other studies have been more process oriented. One documented that the diabetes incentive was claimed far more often than the asthma PIP (Zwar et al. 2005). This study’s survey of GPs found that “the workload and paperwork, and the perceived administrative burden of the incentive are seen as major barriers to [asthma PIP] implementation.” A qualitative study examining the diabetes incentive identified several characteristics GPs identified as important for practices claiming the PIP, including having a champion within the practice, having a computerized electronic medical record, and having a practice nurse handle some details of the cycle of care (Saunders, Schattner, and Mathews 2008). Two other studies relied upon regional primary care organization-level data to examine characteristics related to higher P4P PIP claims. These studies suggested that there were more PIP diabetes claims by GPs practicing in lower socio-economic status regions and fewer claims in remote areas (Georgiou, Burns, and Harris 2004; Scott and Coote 2010).

METHODS

Three methodological approaches were used to triangulate the impact of the Australian P4P program on quality improvement. First, we used publicly available Medicare claims data to track the number of incentivized services claimed by all doctors in Australia before PIP implementation and afterward (1995–2010). Second, using a panel dataset of GPs followed from 2000 to 2009, we analyzed whether GPs’ P4P program participation was related to increases in diabetes testing and cervical cancer screening. Third, we conducted in-depth interviews with GPs and practice managers in 2011, to explore their perceptions of program impact on their current practice.

Component 1: Tracking Incentivized Services

The Australian Government publicly reports the number of claims for specific procedures, which enabled us to track trends in the provision of four incentivized services before P4P program implementation (1995–2000) and afterward (2001–2010). Two of the incentivized services were diabetes tests (HbA1c and

microalbumin) and the other two services were cervical cancer screens (diagnostic and treatment).³ We examined the two cancer screens separately to be able to detect whether an increase in diagnostic screens led to an increase in treatment screens. The asthma cycle of care does not include any procedures that are billed separately, so we could not track changes in asthma quality of care.

We used descriptive statistics to examine whether there was an increase in the annual number of incentivized diabetes-related tests and cervical cancer screens after program implementation. As each of the P4P PIPs has unique billing codes, we also were able to track the number of claims made for each of the three incentive programs, including asthma, from the inception of the program through 2010.

Component 2: The Relationship between GPs' P4P Program Participation and Provision of Incentivized Services

The second component used a unique panel dataset of 1,131 GPs followed from 2000 to 2009, provided by the Australian Department of Health and Ageing. These data, which were obtained at the GP-site-year level, enabled us to track the extent to which changes in a GP's P4P program participation were related to increases in diabetes testing (HbA1c and microalbumin) or cervical cancer screening (diagnostic and treatment).

For inclusion in the randomly selected study sample, GPs had to have billed a minimum of 375 Medicare claims of any type in 2000 at a given practice site.⁴ We oversampled GPs who practiced in areas of low socio-economic status (using the SEIFA index of relative socio-economic advantage and disadvantage of 5 or lower [Pink 2008]) to examine whether there was a differential program impact for GPs practicing in areas of lower and higher socio-economic status. All analyses were weighted to account for this oversampling.

We used two measures of P4P program participation in our analyses. The first was whether the GP's practice was signed on to the specific incentive (diabetes or cervical cancer screening) in the year. The second was the number of diabetes or cervical cancer PIP claims the GP made in the year. The dependent measures for diabetes were the number of HbA1c and microalbumin tests that a GP's patients had in the year, and for cervical cancer, the measure was the number of patients who had diagnostic and treatment screens. Again, unfortunately, we could not examine asthma quality of care because none of the incentivized behaviors were billed separately. We did not examine the diabetes or cervical cancer practice-level outcome payments.

We conducted descriptive analyses and found that the GPs' demographics were very similar to the published data on all GPs in Australia (Britt et al. 2000). The GPs were mostly male (66 percent), just over half (51 percent) were 45 years of age or older, and the vast majority (79 percent) practiced in New South Wales, Queensland, or Victoria.

We then examined trends over time in P4P participation status. In each year, we computed the percentage of GPs who fell into each of the following four levels of program participation: not signed on to the incentive, signed on but made no incentive claims in the year, made a small number of incentive claims (1–10 in the year), and made a larger number of incentive claims (11 or more).

Next we examined annual testing/screening claims over time, based upon participation status in 2005. For example, we examined the pre- to post-intervention trends in HbA1c testing for GPs who did not participate in the diabetes incentive in 2005, and compared them to trends for those who participated but had no claims, those who had a small number of claims, and to those with a larger number of claims. If the program participants' trend lines were steeper than the trend lines for nonparticipants, that would suggest the program sparked greater increases for participants.

This analysis was conducted using the 541 GPs who continually practiced (billing at least 375 Medicare claims in the year) between 2000 and 2005. In supplementary analyses, we repeated the analysis categorizing GPs by their participation status in 2003 and 2009, and we found substantively similar results. We also found similar results when examining the dependent variable microalbumin.

We then developed multivariate fixed effects panel regression models, which assess the average within-GP impact of P4P program participation. Poisson models were used because the dependent variables were counts of tests or screens. The models controlled for the total number of Medicare claims the GP had in the year as well as for annual time trends. The analyses presented here are for the full sample over the entire 10-year study period, although supplementary analyses using a balanced panel and examining a shorter time frame yielded substantively similar results.

We further developed regression models to test interactions between GP characteristics and P4P program participation. We tested whether GPs practicing in high or low socio-economic status areas, measured using the SEIFA index of socio-economic advantage and disadvantage, had similar response to the P4P program. We also tested program interactions with the ruralness of the GP's practice, to see whether the enhanced incentive level in rural areas

resulted in greater program impact. In addition, we tested program interactions with GP gender, age, and state/territory.

Component 3: General Practitioners' Perceptions of Program Impact

For our third approach, we conducted in-depth interviews with a convenience sample of 13 GPs and 2 practice managers in five cities (Sydney, Brisbane, Perth, Alice Springs, and Darwin) in 2011.

To recruit participants, we compiled a list of GPs in each city (in larger cities, we compiled lists of GPs in high- and low-income neighborhoods) using online searches. Forty-seven GPs were faxed letters requesting their participation in a short in-person interview, and follow-up phone calls were made to arrange interviews. In two cases, the practice manager rather than the GP participated, and four interviews were conducted by telephone rather than in person. Interviewees ranged from nonparticipants in the P4P programs to GPs who routinely claimed incentive payments.

The semi-structured interview guide included questions about why the practice did or did not sign-on to the P4P programs; what the GP's (or practice manager's) experience had been with each of the three incentive programs, including the extent to which the incentive programs influenced their individual behavior and practice norms; and what were challenges to participating in the P4P programs. Interviews typically lasted 15–20 minutes, although some lasted as long as 45 minutes.

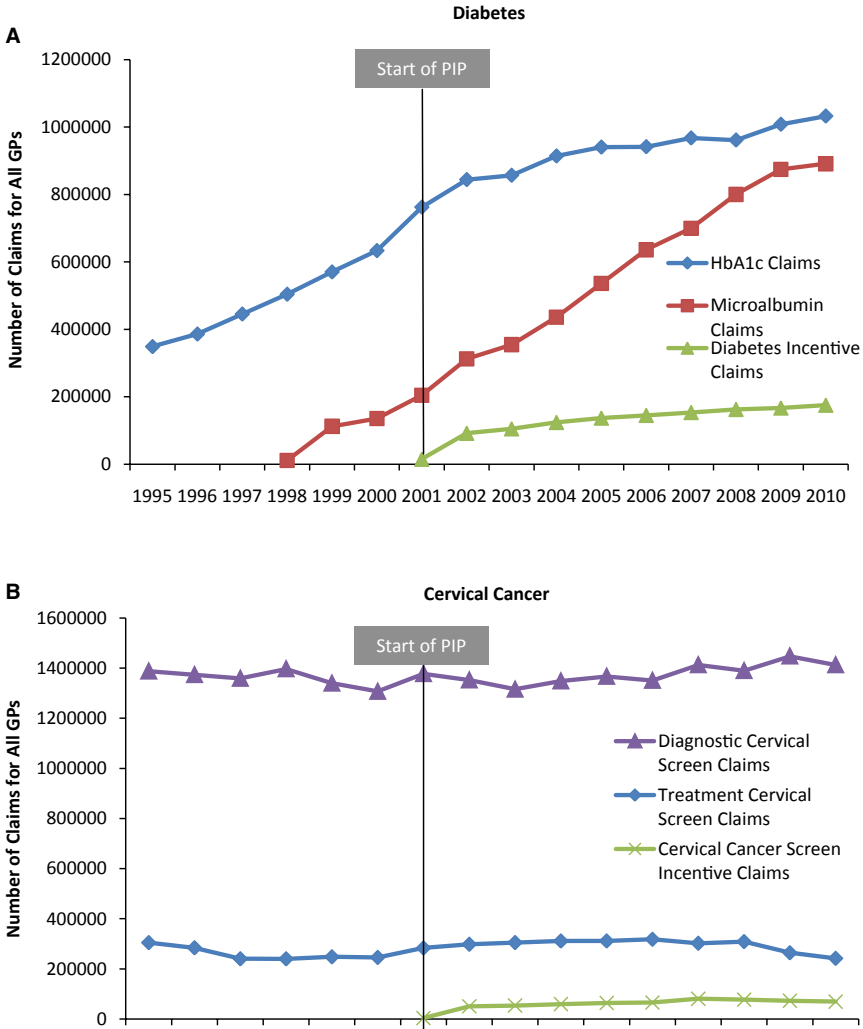
All interviews were audio recorded and transcribed. Data were iteratively collected and analyzed. Transcripts were reviewed to identify key recurring themes, and then the emerging themes were further explored in subsequent interviews. Relevant verbatim text blocks were identified for each theme, and the text blocks were reviewed to test the robustness of themes and to identify subthemes.

RESULTS

Component 1: Tracking Incentivized Services

Figure 1, Panel A shows the number of Medicare claims for HbA1c and microalbumin tests from all doctors in Australia from 1995 and 2010. In the period prior to the P4P program implementation, there were steady increases in the number of both tests: 11–15 percent annual increases for HbA1c and 20 percent increase between 1999 and 2000 for microalbumin. The year the

Figure 1: (A) Number of Diabetes-Related Claims and Diabetes Incentive Claims 1995–2010, and (B) Number of Cervical Cancer Screen Claims and Incentive Claims 1995–2010



incentive began, 2001, there were substantial increases above the existing trend lines: a 20 percent increase for HbA1c and 51 percent increase for microalbumin tests. These increases soon slowed, however, to pre-implementation levels or lower.

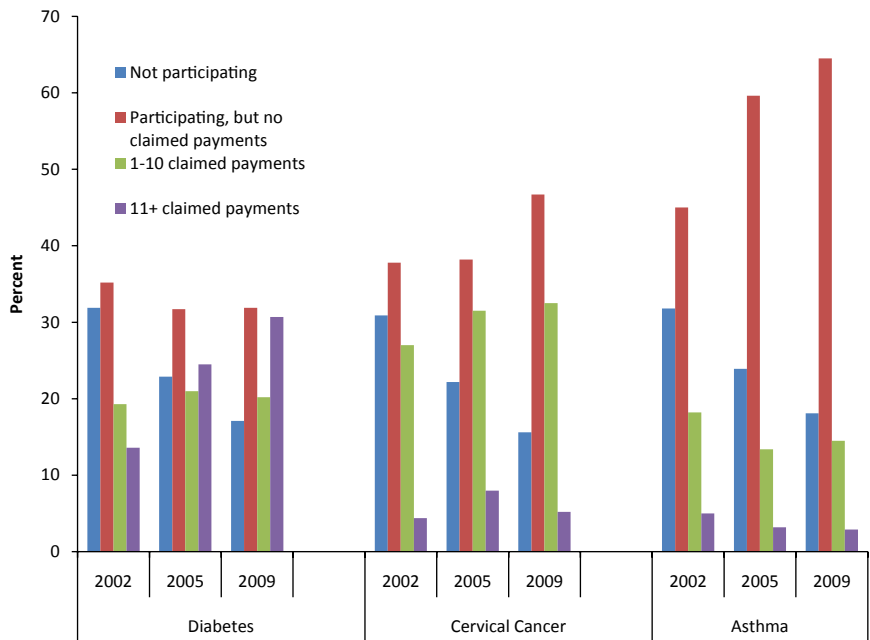
Diabetes P4P incentive claims were infrequent relative to the number of HbA1c tests (approximately 11–17 percent as often) and microalbumin tests billed (20–30 percent). While the number of diabetes incentive claims was low, it was as much as six times higher than the number of asthma incentive claims.

For cervical cancer, there were small decreases in diagnostic and treatment cervical cancer screenings in four of the five pre-implementation years (Figure 1, Panel B). In 2001, when the incentive began, there was a 5 percent increase in the number of both types of cervical cancer screens. Increases, however, did not continue consistently in future years.

Component 2: The Relationship between GPs’ P4P Program Participation and Provision of Incentivized Services

GPs in the longitudinal sample signed on to the P4P program quickly (Figure 2). In 2002, the first full year of the program, over two-thirds of GPs had signed on to each of the three incentives. The majority of those signed on,

Figure 2: Participation Status in Diabetes, Cervical Cancer, and Asthma Incentive Programs Participation Status

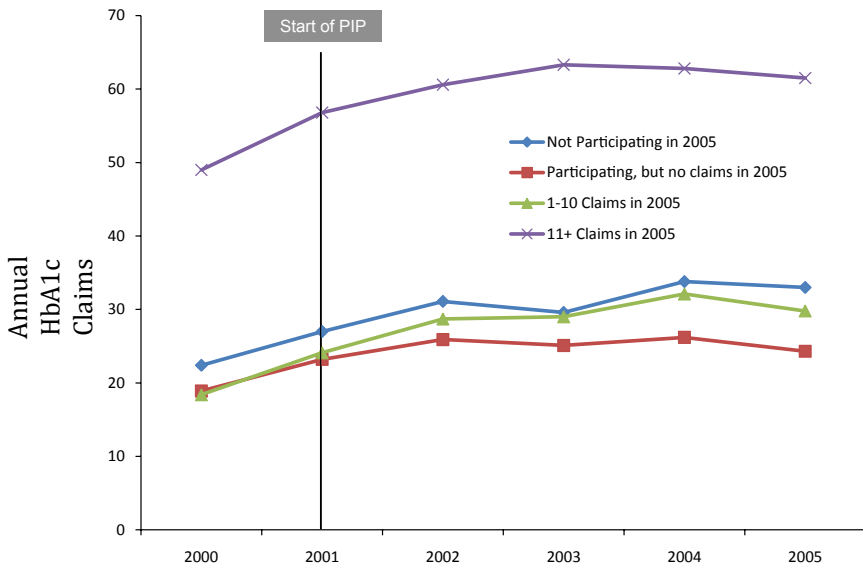


however, made no incentive claims in the year. It was a small minority of GPs who actively participated in the incentive program, claiming more than 10 incentive claims (14, 5, and 4 percent for diabetes, asthma, and cervical cancer, respectively).

Over time, GPs increasingly claimed the diabetes incentive. By 2009, 31 percent of GPs claimed more than 10 incentive payments, up from 14 percent in 2002. The increase in GPs claiming cervical cancer incentive payments was more modest (from 31 to 38 percent), and there was a decrease in the percent claiming the asthma incentive.

Figure 3 shows the change in provision of HbA1c tests from 2000 to 2005, by how actively GPs participated in the diabetes P4P incentive in 2005. Participants claiming 10 or more diabetes incentive payments in 2005 had baseline HbA1c testing levels twice that of all other groups. The increase in HbA1c tests for this group was 26 percent over the 6 years, which was a smaller increase than for all other groups, including those who did not sign on to the incentive program. In other words, those who were actively participating in the incentive program in 2005 had much higher baseline levels of HbA1c testing, and they did not increase testing more than other GPs.

Figure 3: Average Number of HbA1c Test Claims in Each Year, Based upon 2005 Diabetes Incentive Participation Status in 2005



Fixed effects regression models tested the within-GP impact of changing program participation on the number of the tests/screenings provided (Table 1). Neither signing on to the incentive program nor the number of incentive payments claimed by the GP in a year was significantly associated with the number of diabetes tests or cervical cancer screens provided. For diabetes, there were significant increases in the number of HbA1c and microalbumin tests for all GPs between 2000 (pre-implementation) and 2002 (first full year of implementation). While the incident rate ratios continued to rise after 2002, the increase observed in the first years of PIP implementation was much greater than during the rest of the study period. There were more modest increases in cervical cancer screening numbers between 2000 and 2002, which were not statistically significant.

In models testing whether there were differential impacts for subgroups of GPs, none of the interaction terms between program participation and GP characteristics were significant.

Table 1: Incidence Rate Ratios from Poisson Fixed Effects Models That Examine the Relationship between Incentive Participation and Incentivized Tests and Screens

	<i>Incentivized Diabetes Tests</i>				<i>Cervical Cancer Screening</i>			
	<i>HbA1c Test</i>		<i>Microalbumin Test</i>		<i>Diagnostic Screening</i>		<i>Treatment Screening</i>	
Signed on to P4P program	1.00		1.07		1.01		1.04	
Number of P4P incentive claims	1.00		1.00		1.00		1.00	
MBS claims/100 Year	1.01***	1.01***	1.01***	1.01***	1.01***	1.01***	1.01***	1.01***
2000 (Ref)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)	(1.00)
2001	1.08	1.08*	1.19*	1.24***	1.02	1.03	1.04	1.06
2002	1.14**	1.14**	1.44**	1.49***	1.01	1.01	1.06	1.09
2003	1.15**	1.15***	1.58***	1.64***	1.01	1.01	1.11	1.14*
2004	1.18**	1.18***	1.72***	1.79***	1.01	1.02	1.12	1.15*
2005	1.18**	1.18***	1.86***	1.92***	1.00	1.02	1.10	1.12
2006	1.17**	1.17***	1.93***	2.01***	1.00	1.01	1.10	1.13*
2007	1.16**	1.16**	1.98***	2.05***	0.99	1.01	1.04	1.06
2008	1.14*	1.14**	2.06***	2.13***	0.97	0.99	0.99	1.01
2009	1.13*	1.13**	2.07***	2.15***	0.96	0.98	0.93	0.96

* $p < .05$; ** $p < .01$; *** $p < .001$.

Component 3: General Practitioners' Perceptions of Program Impact

GPs who claimed the P4P incentive payments reported that the incentives did not fundamentally influence the way that they treat patients. One GP explained, "It's money for doing what you were already doing." Others comments included: "I don't think it changes my practice," and "I'm not driven by the PIP, I do it anyway." Several GPs did report, although, that the diabetes cycle of care list of behaviors had helped them "audit" themselves, regardless of whether they claimed the incentive payment. One explained, "It might have [changed my behavior] in small ways. Even though I don't actually often use that [PIP billing code]...I went through it all [the cycle of care] with my patient this morning." When asked whether the P4P PIPs were discussed with colleagues as part of practice management, the response was consistently negative. "The doctors, we never talk about it," said one GP. Another said, "I don't know how we fare in it; it's not discussed."

GPs, including program participants and nonparticipants, spoke favorably about the components of the diabetes cycle of care. "From a clinical point of view, yes it's a priority and we do it," explained one GP. In contrast, GPs did not agree with some components of the asthma cycle of care. The concern, voiced by many, was the requirement for a well visit. One GP explained, "They [patients] feel you are wasting their time if you ask them to come in for an asthma check when they feel perfectly well. That's how they feel. They think we are trying to drum up business, and you are busy enough already, you don't need to drum up business. It's not easy to do. It feels wrong. Not so wrong. Anyhow, I don't want to do it. They don't want to have it done to them." Several GPs also questioned the value of the asthma action plans for adults.

The burden of tracking patients and the complexity of billing, relative to the modest incentive payments, were reasons GPs did not more actively participate in the P4P programs. "I'm not improving their care, I'm just doing paperwork. And for that minimal return, it's just not worth it," one GP explained. Another explained, "In the time I could muck around working out whether I've done it this year, have I done all the bits and pieces, ticked all the boxes, rather than do that, I'll just see someone else." One practice manager said his practice had tried in the past to actively follow up with patients through mailings to encourage them to come in for well visits, but had stopped because of the low yield and the cost.

GPs described tracking the cycle of care as time consuming, and few had nurses available to assist them. Needing to bill separate codes for the incentive

payments was a source of frustration. One GP said, "It's not in the forefront of my brain to remember to claim that other set of numbers.... There are too many numbers and too many rules and not enough time." One said it would be easier if it were an added-on claim to a normal office visit, rather than separate codes. Another GP said developing automatic systems that work with electronic health records or the billing system would be a substantial improvement.

GPs mentioned several other challenges preventing them from more frequently claiming the P4P PIPs. One was that patients often seek care at multiple practices. Another challenge was that patients make decisions that impact whether the cycles of care are completed: "I can recommend that they get an eye check, and tell them why. But at the end of the day, they are the ones that have to organize it and get it done." An opportunity identified by two GPs was to leverage GPs' inherent competitive nature. One explained, "I think if you got feedback against your peers...then merely the pride of achieving better outcomes for your population of patients would probably be enough to ensure that you achieved."

DISCUSSION

This study used three approaches to discern the impact on quality of Australia's P4P program for GPs. The quantitative and qualitative results tell different aspects of a consistent story: that the program has not catalyzed long-term improvement in quality of care.

Australian Medicare data documents that the initial P4P program implementation was associated with short-term increases in diabetes testing and cervical cancer screens nationally. Data on a panel of GPs tracked over time suggest that the increases in diabetes-related tests reflected increases among all GPs, not just those participating in the P4P program. It may have been that the program's publicity raised GPs' awareness of recommended diabetes care, for program participants as well as nonparticipants. The increases in cervical cancer screening at the time of PIP implementation were smaller in magnitude than for diabetes and not statistically significant. The asthma incentive, which was much less frequently claimed than the other two incentives, is unlikely to have substantially impacted asthma quality of care.

We found no significant differences in program impact for GPs working in lower and higher socio-economic status areas. Nor were there differences in

impact detected for GPs working in rural areas, where the PIP incentive was higher, compared with those in urban areas. The concern that P4P programs would widen socio-economic disparities seems to be unfounded in this Australian case.

This study contributes to a growing literature documenting that P4P programs have only limited impact on primary care quality improvement (Petersen et al. 2006; Van Herck et al. 2010; de Bruin, Baan, and Struijs 2011; Scott et al. 2011). The program's limited impact underscores the importance of evaluation and monitoring of P4P programs to ensure that they, in fact, change clinician behavior as anticipated. This study also highlights the importance of considering what criteria need to be met for P4P programs to continue, be altered, and be ended. The PIP P4P program has operated over a decade, and while there were modest, short-term impacts, the program does not appear to have improved quality over the long term despite large public investments.

For those designing and managing P4P programs, this study's findings translate into several programmatic recommendations. First, P4P programs should be as automated as possible, so that claiming the payment is not burdensome and a deterrent to participation. The findings also highlight that the incentive's absolute size, as well as the relative size, need to be sufficiently large to spark GPs to change their behavior. In the case of the diabetes incentive, the incentive was only slightly more than the government reimbursement for a typical patient visit, and GPs reported opting to see another patient rather than track and bill for the incentive. The lack of buy-in to the incentivized asthma behaviors highlights the importance of involving clinicians in the development of P4P programs.

Given the limited impact of P4P programs in a number of primary care studies, alternative strategies to improve quality of care should be developed and tested. One approach suggested by GPs in this study was leveraging GPs' competitive nature by providing them with reports on their performance, their clinic's performance, and possibly their colleagues' performance. This type of reporting has the potential to spark discussion within the practice on how GPs are faring toward quality goals, and it can catalyze friendly peer pressure (Greene, Hibbard, and Overton 2012). Other approaches to quality improvement will likely include more comprehensive overhauls of clinician compensation, which are beginning in a number of innovative delivery systems (PRNewswire 2010; Lerner 2011; Oregon House Bill 2012).

The study's findings should be interpreted in light of its limitations. Because the P4P program was voluntary, our findings may be subject to

selection bias. We did observe that those who were active P4P program participants in 2005 provided more diabetes tests to their patients prior to program implementation. If program participants had a slower underlying trend of improving quality than nonparticipants, then our conclusion that the program had no impact would be erroneous. We think this scenario is unlikely as we found only a small, short-term postimplementation increase in total claims for the incentivized services in component 1. In addition, nonparticipants and low-level participants also had very similar baseline profiles of testing/screening (Figure 3), and their trajectories postimplementation were very similar. Another key limitation was that the qualitative component was relatively small in size and relied upon a convenience sample. Despite this, the responses from GPs were very consistent. While the methodological components were each imperfect, a key strength of the study is that the three components had consistent and complementary findings.

CONCLUSION

P4P programs have been developed to create economic incentives for physicians to provide higher quality care. This study provides more empirical evidence that despite the intuitive appeal of P4P programs, they do not necessarily translate to substantial improvements in quality. Future studies are needed to identify the components of successful P4P programs and to explore alternative approaches to improving quality of care.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: I would like to acknowledge the Australian Department of Health and Ageing (DoHA) and The Commonwealth Fund for their support of this work through the Australian-American Health Policy Fellowship. The views presented here are my own and not those of the Fellowship supporters. I would like to thank the Centre for Health Economic Research and Evaluation at The University of Technology Sydney for collegiality and support during my visiting appointment in 2010–2011. I would also like to acknowledge the GPs who very generously made time to be interviewed for this study.

Disclosures: None.

Disclaimers: None.

NOTES

1. The PIP also has nine other incentive payments that are not pay-for-performance oriented. These include incentives for providing after-hours care, employing practice nurses, and teaching medical students.
2. The \$3 is for each female standardized whole patient equivalent, which is a measure of practice size that is weighted by age.
3. Because 45 percent of cervical cancer screens were not billed through Medicare in the state of Victoria in the early years of the program, we have run all analyses both with and without including services or GPs in Victoria. As there were no substantive differences in the results, we have presented data for the entire country.
4. Having at least 375 claims is a benchmark that has been previously used to identify active general practitioners (Britt et al. 2000).

REFERENCES

- Australian Department of Health and Ageing. 2009. "Primary Health Care Reform in Australia. Report to Support Australia's First National Primary Health Care Strategy." Commonwealth of Australia.
- Britt, H., G. C. Miller, J. Charles, S. Knox, G. P. Sayer, L. Valenti, J. Henderson, and Z. Kelly. 2000. "General Practice Activity in Australia 1999–2000." AIHW and Sydney University.
- de Bruin, S., C. Baan, and J. Struijs. 2011. "Pay-for-Performance in Disease Management: A Systematic Review of the Literature." *BMC Health Services Research* 11 (1): 272.
- Campbell, S. M., D. Reeves, E. Kontopantelis, B. Sibbald, and M. Roland. 2009. "Effects of Pay for Performance on the Quality of Primary Care in England." *New England Journal of Medicine* 361: 368–78.
- Casalino, L. P., A. Elster, A. Eisenberg, E. Lewis, J. Montgomery, and D. Ramos. 2007. "Will Pay-For-Performance and Quality Reporting Affect Health Care Disparities?" *Health Affairs* 26 (3): w405–14.
- Christianson, J. B., S. Leatherman, and K. Sutherland. 2008. "Lessons from Evaluations of Purchaser Pay-for-Performance Programs: A Review of the Evidence." pp. 5S–35S.
- Collier, R. 2009. "Shift toward Capitation in Ontario." *Canadian Medical Association Journal* 181 (10): 668–9.
- Doran, T., C. Fullwood, E. Kontopantelis, and D. Reeves. 2008. "Effect of Financial Incentives on Inequalities in the Delivery of Primary Clinical Care in England: Analysis of Clinical Activity Indicators for the Quality and Outcomes Framework." *Lancet* 372: 728–36.
- Georgiou, A., J. Burns, and M. Harris. 2004. "GP Claims for Completing Diabetes' Cycle of Care'." *Australian Family Physician* 33 (9).
- Greene, J., J. Hibbard, and V. Overton. 2012. "Early Experiences with an Innovative Primary Care Compensation Model." Presented at the AcademyHealth Annual Research Meeting.

- Harris, M., and N. Zwar. 2007. "Care of Patients with Chronic Disease: The Challenge for General Practice." *Medical Journal of Australia* 187 (2): 104–7.
- Hussey, P. S., G. F. Anderson, R. Osborn, C. Feek, V. McLaughlin, J. Millar, and A. Epstein. 2004. "How Does the Quality of Care Compare in Five Countries?" *Health Affairs* 23 (3): 89–99.
- Lerner, M. 2011. "Doctors' Pay Plan Is Cutting Edge." *Star Tribune* September 14. [accessed on December 22, 2012]. Available at: <http://www.startribune.com/business/129592298.html?refer=y>
- McGlynn, E. A., S. M. Asch, J. Adams, J. Keesey, J. Hicks, A. DeCristofaro, and E. A. Kerr. 2003. "The Quality of Health Care Delivered to Adults in the United States." *The New England Journal of Medicine* 348 (26): 2635–45.
- Medicare Australia. 2010. "Practice Incentives Program" [accessed on December 22, 2012]. Available at <http://www.medicareaustralia.gov.au/provider/incentives/pip/index.jsp>
- Medicare Australia. 2001. "Practice Incentives Program (PIP) November 2001 Quarterly News Update" [accessed on November 30, 2001]. Available at http://www.medicareaustralia.gov.au/provider/incentives/pip/files/pip_news_nov_2001.pdf
- Oregon House Bill 3650. 2012. "Health System Transformation."
- Pearson, S. D., E. C. Schneider, K. P. Kleinman, K. L. Coltin, and J. A. Singer. 2008. "The Impact of Pay-for-Performance on Health Care Quality in Massachusetts, 2001–2003." *Health Affairs* 27 (4): 1167–76.
- Petersen, L. A., L. D. Woodard, T. Urech, C. Daw, and S. Sookanan. 2006. "Does Pay-for-Performance Improve the Quality of Health Care?" *Annals of Internal Medicine* 145 (4): 265–72.
- Pink, B.. 2008. "Socio-Economic Indexes for Areas (SEIFA) – Technical Paper." Australian Bureau of Statistics: Commonwealth of Australia.
- PRNewswire. 2010. "Independence Blue Cross Invests \$47 Million in New Physician Payment Model to Help Strengthen Primary Care." *Fierce Health Care* [accessed on December 22, 2012]. Available at <http://www.prnewswire.com/news-releases/independence-blue-cross-invests-47-million-in-new-physician-payment-model-to-help-strengthen-primary-care-92496489.html>
- Saunders, M., P. Schattner, and M. Mathews. 2008. "Diabetes 'Cycles of Care' in General Practice – Do Government Incentives Help?" *Australian Family Physician* 37 (9): 781–4.
- Scott, A., and W. Coote. 2010. "Do Regional Primary-Care Organisations Influence Primary-Care Performance? A Dynamic Panel Estimation." *Health Economics* 19 (6): 716–29.
- Scott, A., S. Schurer, P. H. Jensen, and P. Sivey. 2009. "The Effects of an Incentive Program on Quality of Care in Diabetes Management." *Health Economics* 18 (9): 1091–108.
- Scott, A., P. Sivey, D. Ait Ouakrim, L. Willenberg, L. Naccarella, J. Furler, and D. Young. 2011. "The Effect of Financial Incentives on the Quality of Health Care Provided by Primary Care Physicians." *Cochrane Database of Systematic Reviews* 2011(9) : 1–59.

- Seddon, M., M. Marshall, S. Campbell, and M. Roland. 2001. "Systematic Review of Studies of Quality of Clinical Care in General Practice in the UK, Australia and New Zealand." *Quality in Health Care* 10 (3): 152–8.
- Stremikis, K., K. Davis, and S. Guterman. 2010. *Health Care Opinion Leaders' Views on Transparency and Pricing*. New York: Commonwealth Fund.
- The Study Group of Diagnosis of the Working Group on Heart Failure of the European Society of Cardiology, K. Swedberg, J. Cleland, J. C. Aguilar, A. Cohen-Solal, R. Dietz, A. Gavazzi, W. H. Van Gilst, R. Hobbs, J. Korewicki, H. C. Madeira, V. S. Moiseyev, I. Preda, J. Widimsky, N. Freemantle, J. Eastaugh, and J. Mason. 2003. "The EuroHeart Failure Survey Programme's Survey on the Quality of Care among Patients with Heart Failure in Europe." *European Heart Journal* 24(5): 464–74.
- Tomio, J., S. Toyokawa, S. Tanihara, K. Inoue, and Y. Kobayashi. 2010. "Quality of Care for Diabetes Patients Using National Health Insurance Claims Data in Japan." *Journal of Evaluation in Clinical Practice* 16 (6): 1164–9.
- Van Herck, P., D. De Smedt, L. Annemans, R. Remmen, M. Rosenthal, and W. Sermeus. 2010. "Systematic Review: Effects, Design Choices, and Context of Pay-for-Performance in Health Care." *BMC Health Services Research* 10 (1): 247.
- Zwar, N. A., E. J. Comino, I. Hasan, and M. F. Harris. 2005. "General Practitioner Views on Barriers and Facilitators to Implementation of the Asthma 3 + Visit Plan." *Medical Journal of Australia* 183 (2): 4.