

Health Outcomes in Chronic Care Management: a British perspective

Alan Maynard

University of York

Introduction

Whilst much is talked about chronic disease management, service delivery models still fail to provide integrated and demonstrably cost effective care that would improve the health of patients efficiently. This paper focuses in particular on three issues identified in the remit.

Firstly what is implied by measurement and management of outcomes? Secondly how have recent UK reforms, in particular the 2004 contract for general practitioners, affected the delivery of chronic care to patients? The final section draws out some implications for the delivery of chronic care. Throughout the emphasis will be on the ubiquitous nature of social experimentation in health and social care reform and the inadequacy in investing in the evaluation of success in terms of efficiency and equity.

Measuring outcomes

What does the patient value when seeking care from health professionals? As Donabedian emphasised it is important to distinguish between process and outcome. (1). The way in which patients are greeted and treated during an episode of illness is something which they usually value. Poor reception at the office or the hospital, rude practitioners and deficiencies in informing patients about diagnosis, treatment options and outcomes may not only affect patient welfare directly but also affect their compliance with treatment regimens.

Thus for instance, reception and staff attitudes and behaviour may be exemplary but a failure to inform patients about the side effects of pharmaceutical hypertension treatment may contribute to low levels of compliance and avoidable morbidity and mortality. Part of the treatment process is good, and performance in terms of identifying and offering treating to the hypertensive patient may be excellent. However even though there is this process success, patient utilisation may be poor and programme outcomes not achieved.

The nice issue is what are the programme outcomes? Samuel Smiles argued in the nineteenth century that economics was the dismal science. The scientific basis of medicine is less than complete, with one study guesstimating that over 45 per cent of interventions were of unknown clinical effectiveness :see figure 1 (2). However even if this figure is sidelined by argument that colleagues in the Cochrane Collaboration are too focused on “gold standard” evidence , or randomised controlled trials (RCTs) and that really medicine is scientific, there can be no doubt that there **real** dismal science is not economics but medicine!

Outcome measures in medicine traditionally focus on indicators of **failure** such as mortality, complications and readmissions. On sunnier days, physicians may focus on lives saved and life years gained but their attention to the quality of enhanced life, which is of obvious importance to patients, tends to be downplayed, particularly in the delivery of routine health care.

This is surprising as measures of **success** in health care have existed for decades. An example of this is the generic patient reported outcome instrument, short form 36 (www.sf36.org). This was devised in the US Rand Insurance experiment over 30

years ago (3) and subsequently it has been translated into dozens of languages (including English!) and used in thousands of clinical trials across the world.

SF36, also available in shorter forms (e.g. with 12 questions: SF12), can be used to generate profiles of mental and physical well being. Used before and after interventions its scores can be used to indicate improvements or not, in the functioning of patients in the discrete physical and mental domains

EQ5D is the rival generic patient reported outcome measure (PROM) (www.euroqol.org). Like SF36 it is easy for patients to complete this instrument and it has mental and physical domains. As a consequence of research into population values of physical and mental attribute tradeoffs, this PROM can be used to generate a single measure of quality of life.

These instruments have been used translated into dozens of languages and used in thousands of clinical trials (4, 5) There have been validated in this work, although there continues to be debate about sensitivity in all therapeutic areas. However remarkably they have not been adopted in routine clinical practice. This is what makes medicine the truly dismal science. Practitioners are focused on failure and even though they use PROM measures of success in trials, they do not use such measures to determine success in treating patients routinely. Perhaps this is due to their fears that comparative data will not serve their interests even though in terms of patient safety and clinical governance, let alone value for money, measurement and management of success is essential.

The inference is that insurers and governments are not interested in improving the quality of care delivered to patients. However this measurement is essential in all care sectors and especially in chronic care, if we are to ensure efficiency. Merely focusing on process and what health professionals do or do not do, is incomplete if it is not complemented with management of outcomes for patients. The curious issue is why public and private health care systems do not measure PROM. In both types of system there are calls for transparency and value for money, both of which are unattainable without PROM.

In England the private insurer, the British United Provident Association (BUPA) is now using a generic PROM (SF12) and for cataracts a specific PROM (VH14) and scores for a range of procedures by each BUPA hospital are displayed on their website (www.bupahospitals.co.uk/asp/hospital_performance). Also the commercial information firm CHKS is currently running a three year pilot in 4 NHS acute hospitals where SF12 and EQ5D are being used before and after elective surgery to elicit PROM measures of change in patient's quality of life (www.chks.co.uk).

An essential ingredient into the efficient management of chronic disease is the complementing of process measurement with patient reported outcome measurement. This is necessary not only to measure and manage success in the treatment of patients but also the manage safety and errors and provide consumer protection through clinical governance. Without PROM the tradition in medicine of "flying blind" will continue to undermine the quality of patient care.

Incentivising improved treatment of chronic care.

There is a worldwide lament about the failure of health care systems, public and private, to translate evidence into practice in the treatment of tens of millions of patients with chronic disease. Whether in China, Kyrgyzstan, Egypt or the countries

such as Australia, Britain, Canada, New Zealand and the United States the point emphasised is that the gaps between those who could benefit and those who receive treatment is great. Obviously in some low and middle income countries, there may be price barriers to consumption. However these are minimal in “civilised” countries such as Australia, Britain, Canada and New Zealand and even in “less civilised” countries such as the United States, even though chronic disease care may be in the insurance benefit package, it is routinely not delivered to patients.

There are many policy instruments that can be used to improve the efficiency of health care delivery to the chronically ill. The focus here will be on the use of targeted payments to primary care provider practices in Britain. This partial focus should not obviate attention being paid to the evidence base around the use of community nurses and protocol led packages of care such as Evercare. Some comments are made about these instruments and their evidence base below (conclusion 5 at page 6)

In Britain, there have been radical and expensive policy changes in the way primary care providers are paid and many lessons can be learnt internationally from this social experiment. The contract for general practitioners (GPs) was reformed in 2004 with the implementation of the Quality Outcomes Framework (QOF). The GP-QOF is a target payment system that rewards performance of group practices (i.e. multi-disciplinary teams, “led” by GPs and including practice nurses and administrative staff) in relation to their achievement in relation initially to 10 clinical targets and some service targets. These are listed, with the number of points achievable for each item in table 1.

The previous contracts had failed to produce levels of chronic care that were consistent with good medical practice. British GPs can be paid a salary (PMS) or on the basis of the 1948 contract with amendments (GMS), capitation with fee for service and lump sum supplements. The former grew rapidly after its introduction in 1997 and covers over 35 per cent of GPs. The rest of the GP stock operate under the GMS contract and most work in group practices.

The GMS contract was explicit about the delivery of some services which were remunerated with fee for service (ffs) and target payment systems e.g. cervical cytology, vaccination and immunisation services. However it was for the most part implicit about what other services should be provided: this has been called the “John Wayne” contract: a GP has to do what a GP has to do, but there was no specification of what they have to do! The Government’s expectation was that they would do everything necessary, and practiced the principle of leaving service delivery choices to the practitioners. This led to significant variation in service delivery and rising Government concern about the delivery of “essential “ services, in particular care of the chronically ill.

The PMS contract was more specific about service delivery in exchange for payment by salary, but as with the GMS contract, benefit packages varied as did the delivery of appropriate care.

Instead of specifying and enforcing these contracts, the Government chose to adopt a radical target payment innovation and in doing so risked paying practitioners for what they were already doing or more critically paid them for what they should have already been providing if they were delivering care efficiently

Any policy change has to be informed by accurate and timely baseline data. This is necessary not only to implement a system of ffs but also to evaluate its success in achieving policy targets.

Furthermore such baseline data was essential for determining how to determine the targets of individual practices. The contract for the initial 10 clinical areas used a system of points, the allocation of which is dependent on the level of achievement of each individual practice. Each clinical area has a number of specific indicators of performance, a total of 146 spread across the ten clinical areas. Each point was initially valued at £75, subsequently raised to £125. Determining the baseline and targets for practices relied on often incomplete data and trust. Practices are rewarded on a sliding scale for their performance on the clinical indicators within a range of performance. For instance in the 2004 QOF, practices were rewarded for screening asthma patients over the period of the last 15 months. If a 70 per cent target was hit for this service, the practice was allocated the maximum 20 points. ($20 \times £75 = £1500$). If its performance was above 25 and less than 70 per cent the point allocation was proportionate to the level of performance.

Two nice issues arise from the detail on table 1. Firstly there is the issue of the robustness of the evidence base for the items selected to be incentivised. Some of these such as screening and control in hypertension and diabetes are quite well evidence based but their relative contribution to population health gain differs. Items such as asthma cover a range of presentations, ranging from perhaps two annual attacks to repeated and regular attacks. The former may be more easily managed than the latter, particularly if it is associated with high risk behaviours such as smoking. However the health gain from successful intervention with the latter may be greater.

The issue of the evidence base on which this contract was based has been subject to debate. Fleetcroft and Cookson (6) raised major questions about the selection and definition of the 10 clinical targets. For instance they argued that

“Maximum payments for eight interventions examined make up 57% of the total maximum payments for all clinical interventions in the (QOF). There appears to be no relationship between pay and health gain in these eight interventions” (6)

They concluded that payments in the GP-QOF were not well aligned to improvements in population health. This is a product not just of the epidemiology and clinical evidence, but also of the weighting of the items in table 1. These weights reflect not potential population health gain but the estimates of a group of “experts” as to workload consequences of each of the items. There is a nice balance to be struck between rewarding the pursuit of health gain and workload balance.

Currently the National Institute of Clinical Excellence (NICE) is addressing the issue of congruence between the QOF and the evidence base as set out in NICE technology appraisals and clinical guidelines (7). This focus on evidence will no doubt inform more efficiently future changes in the QOF as the evidence sources cited by Lester for instance on the Manchester evidence base are quite uneven ([www.npcrdc.ac.uk/QualityandOutcomesFramework\(QOF\).htm](http://www.npcrdc.ac.uk/QualityandOutcomesFramework(QOF).htm))

In the first two years of the new contract GP income grew very rapidly and in each year achievement levels exceeded 90 per cent. This meant that initial estimates of the cost of the contract were exceeded by £250 million. This money was paid to GP practices and in the GMS these are partnerships and the GPs decide how to allocate their income. Undoubtedly there were increases in nurse recruitment and activity as

essentially the QOF requires limited medical input and can be delivered by nurses and other non physician professionals.

However to the current distress of Government practices allocated much of this increased practice revenue to increasing GP incomes. The income of GPs rose by over 20 per cent in a year with the current average being reported as being in excess of £100,000 per annum (circa US\$190,000). Currently the Government, having recognised the fact that large sums of taxpayers money has gone into practitioners' pockets rather than enhancements in patient services, is squeezing incomes, with for instance in 2007-8 a zero increase in GP funding.

The full opportunity cost of the QOF is quite difficult to determine given the lax way in which the contract was designed and implemented. For instance as GPs were induced to pursue the QOF target, it is unclear about what if any other services were withdrawn, let alone whether these withdrawn services were cost effective or not.

This confusion is a product of the loose way in which primary care in the UK has been funded and delivered. Until quite recently it has been a demand led service but its essence is that it is largely a "black box" in terms of national activity, let alone outcome data. This product of the "John Wayne contract" and this tradition of trust that practitioners provide what is needed by patients continues to permeate the current contractual agreements.

This is epitomised by the regulation of the QOF. Practitioners are being trusted to self report honestly and accurately, and probably the majority of practitioners do behave in this way. However the contracting authority, in England the Primary Care Trusts and in Scotland the Health Boards, have limited capacity to audit reporting practice. As the Canadian economist Barer and Evans have emphasised "the only way to pay doctors is to change the system every three years" as by then these able professionals will have learnt to game it to their advantage! "Light touch" regulation of the 2004 contract may yet be regretted by the Department of Health and the medical practitioners as minority, rogue elements denigrate the profession with deviant behaviour.

An emerging example of the potential for poor behaviour is the problem of exclusions. This form of gaming the contract appears to affect a minority of practices and it is a matter for further research as to whether this is due to income generation or clinical reasons.(8) It may be appropriate clinically to exclude patients with multiple morbidities from part of the QOF. Such patient characteristics will produce legitimate exclusions. However policing to identify illegitimate exclusions is now a nice policy challenge.

In conclusion in regard to the UK GP or "family medicine" contract:

- 1) Pay for performance (**P4P**) is a **powerful incentive**. The new UK-GP contract has produced high levels of performance change in relation to 10 clinical conditions and the 146 clinical indicators within these conditions. Performance in relation to achievable points has generally exceeded ninety per cent with signs of improvement between 2005 and 2006 (see www.qof.ic.nhs.uk which reports outcomes from the QOF contract)
- 2) P4P in the UK NHS was **expensive**. For a variety of reasons, including some naiveté in the Department of Health about likely responses from practitioners

geared up for change and the income associated with it, the contract exceeded estimated costs. The cost of services provided by 35,000 GPs rose by one third between 2003-4 and 2006-7 and the overspend following the introduction of the new contract was £250 million.

- 3) P4P has to be focused on **evidence of health gain**. The continuing debate about the imprecision of the QOF in targeting health gain raises the twin issues of service opportunity cost (what was given up by practices in the pursuit of QOF income) and how to relate such activity orientated indicators to measures of patient outcome. An obvious reform would be to incentivise practices to focus on outcomes by incorporating PROM into the QOF i.e. paying practices to monitor regularly the mental and physical functioning of their patients
- 4) There are **alternative ways of improving chronic care management**. The results of any evaluation of the use of P4P should be compared with the cost effectiveness of other interventions so as to ensure the identification of the “best buy”. For instance community nurses may be able to implement protocols and reduce hospital referrals amongst the elderly. Programmes imported in to the UK-NHS from the USA such as United Health’s Evercare system has proved ineffective in early trials. (9) Structured educational outreach may be cost effective (10) although financial incentives to reduce hospital referrals appears to reduce both necessary and unnecessary referrals (11). As ever the evidence base is incomplete as providers innovate freely but fail to evaluate with robust methods. Myopic evaluation of policy options in chronic care has to be avoided.
- 5) The **regulation of P4P** is likely to involve considerable cost. The current British QOF will change over time; hopefully incorporating more evidence based clinical areas and indicators, as well as better consideration of evidence and outcomes. Negotiating such “changes in the goal posts” will be complex and costly but essential as the evidence base improves. The second element of regulatory cost is associated with the audit and policing of the QOF. If current public perceptions of the high levels of GP pay become associated at the margin with illegitimate gaming by professionals, it is likely that the cost of regulation and policing will rise
- 6) It seems fair to conclude that the clinical effectiveness of the provision of chronic care may have improved but that **better planning and implementation** could have induced change more cost effectively.

Implications for the delivery of chronic care

The implications for policy to be derived from the UK reforms are partly obvious and often contentious.

- 1) The planning and execution of the GP reform was flawed. The new contract not only introduced the QOF but also introduced other changes in practice. One of the most significant was the opt-out for out of hours cover for patients. Under the GMS GPs had an obligation to provide directly or through locums, access to care for patients 24/7 for 365 days a year. The 2004 reform offered an opt out from this obligation if practitioners gave up £6000 per annum. This was very attractive and many GPs opted out. The cost of substituting for this consequent lack of provision at night and over weekends was very high. Provider prices rose and some opted out GPs then returned at prices much higher than those paid previously. This combined with relative ignorance of what GPs do in terms of service delivery outside the QOF; means that the opportunity cost of the new contract was high. GPs seem to have had a large

pay increase, have met QOF targets largely but activity outside the QOF may have declined. It remains unclear what cost effective services have declined because of QOF substitution. The QOF has achieved change but have the distortions created by the GP contract produced overall efficiency?

- 2) As Campbell remarked decades ago “specific reforms are advocated as though they were certain to be successful” and “there is safety under the cloak of ignorance” (12). This British P4P reform was bold but it is an experiment whose cost and ambition requires considerable investment in evaluation. Some of this will be done by the Manchester-York National Primary Care Research and Development Centre. The scope and ambition of this work should be wide if British policy makers and those overseas are to learn more about the cost effectiveness of this large investment of resources. With the US experiencing an “epidemic” of P4P enthusiasm as the latest “magic bullet” to solve cost inflation inefficiency and inequity (e.g. 13), it is essential that policy is informed by fact rather than fiction!
- 3) The evaluation task is demanding. The QOF incentivises activity e.g. the periodic monitoring of hypertension of those with identified blood pressure on the practice patient list and intervention to control it at 140/90. Identification of hypertension should lead to the prescribing of changed behaviour (e.g. diet, weight and smoking) and pharmaceutical interventions. The effect of both behavioural interventions and prescribing may be limited if the patient does not comply e.g. anti-hypertension drugs often have side effects that reduce compliance and estimates of compliance usually identify less than fifty per cent of patients take their hypertension drugs. If such effects are not detected by regular BP monitoring requirements, there is a risk that QOF boxes can be ticked but without the full benefits of chronic disease management being achieved. Ensuring an effective link between activity (e.g. monitoring BP), outputs (e.g. controlled BP) and outcomes (e.g. reduced incidence of stroke and heart attack) is makes design and implementation complex.
- 4) Who should be incentivised? Much of what is often not delivered to patients in terms of identifying and controlling chronic disease is well established in terms of an evidence base. Are physicians irrelevant to the delivery of these interventions, provided nurse practitioners and other health professions work in well triaged systems where complicated case are diverted to medical practitioners? Another element of the NHS reforms has been the creation of nurse prescribers and pharmacists also being given access to the full formulary. The evaluation of these initiatives is less than complete and are characterised, following Campbell (12) by being assumed to be cost effective! If the evidence base improves, the policy issue will become even clearer, should we have primary care practices that are nurse led to protocol (nurses tend to adhere to protocols more assiduously of course!) with minimal use of GPs, who following the 2004 contract, are now very expensive to employ?. The nice challenge is to ensure that if substitution takes place it is really substitution and not development of service packages in a complementary and more expensive manner. In England, Whitehall policy is hinting at toying with the use of private providers in primary care, partly in the hope that they will radically alter skill mix in the delivery of chronic and other primary care. Whether such catalysts can be introduced into a public system cost effectively is unclear.
- 5) Economists tend to favour the use of incentives and might be expected to be advocates of P4P. As ever they are reluctant to offer unambiguous advice always emphasising that most policies have costs and benefits and the challenge is to evaluate and weigh up carefully the balance of advantages and disadvantages. Robinson has noted (14) that “there are many mechanisms for paying physicians: some are good and some are bad. The

worst three are fee for service, capitation and salary"! The problem is epitomised in table 2 which chronicles the advantages and disadvantages of these pay methods. The choice of "best" method depends on social objectives e.g. cost containment and activity control or activity enhancement and cost inflation. Alternative payment systems are double edged: they bring cost and benefits which have to be mapped and quantified with care. The likely effects of the GP contract were anticipated (15)

- 6) A nice concluding issue is the role of positive and negative incentives. Prospect theory in economics analyses individual's attitudes to risk and postulates that their valuation of gains may not be equal to the value of losses (16). For instance a physician's valuation of a gain in salary of £1000 may be the same as her valuation of a loss of income of £100. Individuals may value losses higher than gains, particularly perhaps if such losses are indicators of poor clinical practice. Putting some small part of a physician's salary at risk may be a more cost effective way of inducing behaviour change than simple P4P bonuses to induce compliance with obvious and necessary practice standards. The nice test to be carried out is whether reducing a practice's income by say 1 per cent if it fails any target, would be more or less effective than rewarding the practice with large pay increase like in Britain to achieve compliance. This is an empirical issue and we look forward to experimentation and evidence of the relative merits of these alternatives. Small "demerit" P4P with penalties may be more cost effective than "merit" P4P with expensive bonuses?

Conclusions

For decades physicians internationally have failed to translate evidence into practice about what is clinically and cost effective in the care of the chronically ill. Such behaviour is both unethical and inefficient. Where protocols, education and training and the prayers of policy makers have failed, P4P may induce significant change but not costlessly.

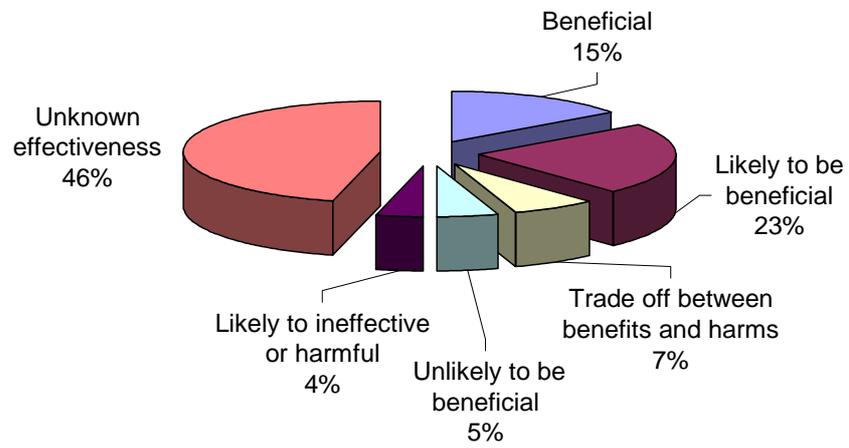
Such incentives have to be devised and implemented with great care. The Minister of Health Aneurin Bevan is alleged to have remarked whilst implementing the NHS in 1948, "the only message understood by a doctor has to be written on a cheque". This may be true but the problem is that such bribes, to which we are all susceptible, have to be targeted with precision and achieve not only changes in the processes of patient care but also maximise patient reported outcomes (PROM) at least cost.

References

- 1) Donabedian, A, Evaluating the quality of medical care, Milbank Memorial Fund Quarterly, 44, 166-206, 1966
- 2) BMJ, Clinical Evidence Concise, 14, December 2005, BMJ Publishing (www.clinicalevidence.com)
- 3) Hays, RD, Sherbourne, CD and Mazel, RM, The Rand 36-item health survey, Health Economics, 2, 3, 217-228, October 1993
- 4) Kind, P and Williams, A, Measuring success in health care: the time has come to do it properly, Health Policy Matters, 9, University of York, March 2004 (accessible at www.york.ac.uk/healthsciences/pubs/hpminindex.htm)
- 5) Appleby, J and Devlin, N, Measuring success in the NHS, Kings Fund and Dr Foster, London, 2005

- 6) Fleetcroft, R and Cookson, R, Do incentives payments in the new NHS contract for primary care reflect likely population health gains? *Journal of Health Service, Research and Policy*, 11,1,27-31, January 2006
- 7) Crawford, R, The NICE way to let people know what is good for them, *Parliamentary Brief*, 11,2, 15-17, March 2007.
- 8) Doran,T, Fullwood,C, Gravelle,H, Reeves, D, Kontopantleis,E Hiroeh, U and Roland, M, Pay-for performance programs in family practices in the United Kingdom, *New England Journal of Medicine*, 355,375-384, July 27th, 2006
- 9) Gravelle, H, Dusheiko, M Sheaff, R, Sargent, P, Boaden, R Pickard, S, Parker, S and Roland, M, Impact of case management (Evercare) on frail elderly patients: controlled before and after analysis of quantitative outcome data, *British Medical Journal*,334,31-34, January 2007.
- 10) Mason, J, Freemantle, N, Nazareth, I, Eccles, M, Haines, A and Drummond, M, When is it cost effective to change the behaviour of health professionals? *Journal of The American Medical Association*, 286, 1988-2992, December 2001
- 11) NHS Service Delivery and Organisation Programme, Can primary care reform reduce demand for hospital outpatient departments? Research summary, March 2007. (www.sdo.lshtm.ac.uk/researchsummaries.html)
- 12) Campbell, DT, Reforms as experiments, *American Psychologist*, 1969,24,4,409-429
- 13) Epstein, AM, Paying for performance in the United States and abroad, *New England Journal of Medicine*, 355,4,406-8, 2006.
- 14) Robinson,JC, Theory and practice in the design of physician payment incentives, *Milbank Quarterly*, 79,149-177,2001
- 15) Maynard, A and Bloor, K, Do those who pay the piper call the tune? *Health Policy Matters* 9,October 2003, (accessible at www.york.ac.uk/healthsciences/pubs/hpindex.htm)
- 16) Kahnman, D and Tversky, A, Prospect theory: analysis of decision making under risk, *Econometrica*, 47, 2, 263-92, March, 1979.

Figure 1: Uncertainty about clinical effectiveness



Source: BMJ publishing group 2005

Table 1: The New GP Contract's Quality and Outcomes Framework

A.	Clinical indicators, categories and points	
1)	Coronary heart disease including left ventricular dysfunction	121
2)	Stroke or transient ischaemic attack	31
3)	Cancer	12
4)	Hypothyroidism	8
5)	Diabetes	99
6)	Hypertension	105
7)	Mental health	41
8)	Asthma	72
9)	COPD	45
10)	Epilepsy	16
		550
B.	Organisational indicators	
	▪ records	85
	▪ patient communication	8
	▪ education and training	29
	▪ practice management	20
	▪ medicines managed	42
		184
C.	Additional services	
	▪ cervical screening	22
	▪ child health surveillance	6
	▪ maternity services	6
	▪ contraceptive services	2
		36
D.	Patient experience	
	▪ patient survey	70
	▪ consultation length	30
		100
E.	Holistic payments	100
F.	Quality practice payments	30
	Sub total	1000
G.	Access bonus	50
	Overall total	1050

Table 2 Effects of differing payment systems

payment type	definition	incentive effects				
		incentive to increase activity	incentive to decrease activity	incentive to shift patients' costs to others	incentive to target the poor	controls cost of doctor employment
fee-for-service	payment for each medical act	yes	no	no	maybe*	No
salary	payment per unit of time input (e.g. per month)	no	yes	yes	no	Yes
capitation	payment per patient for care within a given time period (e.g. a year)	no	yes	yes	no	yes

* If fee-for-service payments for treating poor patients exceed those for treating the middle classes