Health Economics, Policy and Law (2007), **2:** 429–433
© Cambridge University Press 2007 doi:10.1017/S1744133107004288

Response

Hospital cost differences and payment by results

JAMES M. MALCOMSON*
University of Oxford, Oxford, UK

Street and Maynard (2007) argue the case for refining the system of payment by results currently being introduced for English hospitals. On the basis of international experience, they recommend, among other things, moving away from setting prices equal to average cost towards setting them equal to best practice costs, adjusting them for quality and using them to signal what activities are desirable. They also discuss how to control total expenditure for primary care trusts (PCTs) under payment by results and question the merits of extending payment by results to mental health services, ambulances, community services, and long-term conditions, 'where it is difficult to describe patient care requirements and cost variations may be high' (p. 9).

What Street and Maynard (2007) do not consider, however, is why costs vary between providers in the first place. This is an important issue because the underlying reasons for cost differences are crucial to knowing how best to respond to them. There are at least four reasons why costs may vary:

1. inherent cost characteristics of providers;
2. economies or diseconomies of scale;
3. case-mix;
4. management practices (including decisions about the quality of services).

Street and Maynard (2007: 9) note the intention for inherent cost differences between English hospitals to be handled by 'compensatory payments for unavoidable costs that some hospitals have to pay for staff, land and buildings, which are partly driven by differences across the country in local market conditions'. They write little about economies of scale. Case-mix, particularly the difference in cost between day-case and in-patient treatment, they view as being handled by a price that is a weighted average of the costs. A weighted average is seen as a way of encouraging hospitals to move towards more use of cheaper day-case treatment. Most of their focus is on management practices. If all the other reasons for cost variation are handled in other ways, a natural response is to make use of best-practice costs to drive suppliers towards more efficient

*Correspondence to: Professor James M. Malcomson, Department of Economics, University of Oxford, Manor Road Building, Manor Road, Oxford OX1 3UQ, UK. Email: james.malcomson@economics.ox.ac.uk
I would like to thank Martin Karlsson and Barry McCormick for very helpful comments.

https://doi.org/10.1017/S1744133107004288 Published online by Cambridge University Press

management practices by yardstick competition along the lines in Shleifer (1985).

Using compensatory payments to take account of differences in inherent cost characteristics is not always straightforward because observed costs are the result of decisions taken as well as inherent cost characteristics, and inherent cost characteristics may depend on things not known to purchasers. I take up these issues further below. Economies of scale certainly deserve consideration because hospitals in England vary substantially in scale, both overall and for individual types of treatment. Standard textbook theory indicates the problems that arise from setting price equal to average cost when marginal cost is increasing unless there is sufficient competition to drive all providers to the point at which marginal cost equals average cost. Perhaps more importantly in the context of hospitals, if marginal cost is below average cost, setting price equal to average cost results in the marginal revenue from carrying out procedures exceeding the marginal cost. That gives hospitals an incentive to provide more services than are strictly necessary and thus excessive expenditure for PCTs. If marginal cost is below average cost, setting prices at what look like 'best practice' levels makes them closer to marginal cost and hence reduces the incentive for oversupply. But this has to be associated with additional payments to lower volume providers to cover the difference between average and marginal cost, for example, a two-part tariff consisting of a block payment plus a payment per patient (or other unit of activity) or a lower differential payment for activity above a specified threshold of the type to be used in 2007/8 for emergency admissions. Without that, lower volume providers will run up deficits that will need to be met in some way (by, for example, closing wards for part of the financial year, reducing services, and increasing waiting times).

Two-part tariffs, as Chalkley and Malcomson (1998b) show, are also a powerful tool for inducing appropriate quality decisions, even when patients are not themselves paying for treatment, and quality cannot be monitored effectively by purchasers – *if* where patients (referred by their GPs) go for specialist treatment responds to quality to some degree. Any demand response to quality, however small, can be made to provide suppliers with an appropriate incentive for quality by careful choice of the payment-per-patient part of a two-part tariff, with the total payment corrected by the amount of the block payment. The 'if' may, though, as Newhouse (2002) has argued, be a big one. When the number of patient referrals cannot or does not respond to quality differences between hospitals, the two-part tariff may usefully be supplemented by some cost sharing, as Chalkley and Malcomson (1998a) show, provided those hospitals care to some extent about the quality of patient care (or physician behaviour ensures they act as if they care). The literature on this has been surveyed in detail in Chalkley and Malcomson (2000) and Newhouse (2002), so I do not dwell further on it here.

https://doi.org/10.1017/S1744133107004288 Published online by Cambridge University Press

Case-mix may be more problematic than Street and Maynard (2007) suggest. Of course, if providers do not have any idea which patients are likely to be more expensive to treat before deciding on treatment, those differences cannot influence their decisions and setting price equal to average cost makes sense. But they surely have some idea; indeed, that they do is central to the argument in Street and Maynard (2007) that lowering prices will induce greater use of lower cost day-case treatment where appropriate. That brings the analysis squarely into the Laffont and Tirole (1993) framework of procurement with asymmetric information about costs for which there is an elegant theory of how payments should be determined. Payment solely on the basis of output provides strong incentives for suppliers to keep costs down but, if set high enough to cover the cost of inherently high-cost suppliers, results in a rent to inherently low-cost suppliers. An optimal contract trades off that rent against the incentive to keep costs down by basing payment partly on the observed cost of provision (supply-side cost sharing in the terminology of Ellis and McGuire, 1993). Chalkley and Malcomson (2000) survey the applications of this approach to purchase of health services. The trade-off is straightforward: if payments are set too low, providers may not be able to cover costs (resulting in deficits, reduced services and longer waiting times); but the higher they are set, the greater the rent received by providers. Chalkley and Malcomson (2002) adapt this framework to the case-mix issue that arises when patients have different costs of treatment for the same basic diagnosis – low treatment cost plays the role of inherently low supply cost in the Laffont and Tirole (1993) model. It is apparent from US Medicare data that costs differ markedly even within diagnosis-related groups (DRGs). By calibrating the model to costs for Medicare DRGs, Chalkley and Malcomson (2002) show very substantial potential savings from optimal cost sharing for DRGs with high cost variation if all that variation were observable to providers before treatment.

Important as it is for deciding on payment arrangements, there is currently little information about the extent to which providers are aware of cost variations before deciding treatments. To the extent that they are not, potential gains from cost sharing are less because there is a smaller difference between high and low cost providers in the rents they anticipate at the time they make decisions. That reduces the gains from using cost sharing to reduce overall rents. Given the problems of collecting cost data on each patient that are sufficiently robust to be used as the basis for payment, it may thus not be worth doing. But that does not mean it is safe to ignore the incentive issues arising from differences between patients. Since clinicians are likely to be better informed than anybody else about which treatment is medically appropriate, the decision about precisely what care should be provided is typically left to them. Street and Maynard (2007) discuss policy towards treatment choice primarily as a matter of persuading suppliers to move towards lower cost day-case treatments, away from higher cost in-patient treatments, by lowering the payment per treatment.

https://doi.org/10.1017/S1744133107004288 Published online by Cambridge University Press

Implicit in this is that suppliers are sensitive to payment-cost differences in their choices between treatments, but then, unless payments are set carefully, the result may be that they provide day-case treatment even when in-patient treatment would be more appropriate. So it is important to analyse how suppliers can be induced to provide patients with the appropriate treatment (and certainly not just the cheapest one that is considered acceptable given the diagnosis). Malcomson (2005) and Siciliani (2006) do just that. It is critical to make payment depend on the treatment actually given, not just on the initial diagnosis: Malcomson (2005) gives an example of precisely how to do this in the case of test procedures for coronary artery disease. Medicare discovered this conclusion from experience. Despite the original idea that DRGs should be based solely on initial diagnosis in order to provide incentives to keep costs down, DRGs soon became broken down into finer categories that depended on the treatment given. As McClellan (1997: 93) comments, 'over 40 percent of DRGs are related not to diagnoses . . . but to the performance of specific intensive procedures' (over 220 out of 480 DRGs in 1993). From this perspective, fixing a price for each English healthcare resource group (HRG) that is just a weighted average of day-case and in-patient costs is not a sensible way forward.

One of the accounting nightmares Street and Maynard (2007) highlight of moving from block contracts to payment by results is that it makes it harder for PCTs to control their expenditures. Critical to this in any system of payment to hospitals based on activity is ensuring that GPs act as effective gatekeepers, referring patients only when appropriate. But, as Street and Maynard (2007) write, 'PCTs have few levers to influence GP referrals, the main hope being Practice Based Commissioning (PBC) under which GPs manage budgets for their patients.' If those holding the budget are those making decisions about specialist referrals, their incentives are to make fewer. PBC harks back to the GP fundholder arrangement used briefly in the 1990s, which, as Street and Maynard (2007) explain, does not appear to have provided very strong incentives to reduce referrals. Interestingly in this context, popular concern with HMOs in the US seems to be in precisely the opposite direction: that too few patients receive appropriate, but expensive, treatments. How to get the balance right is the concern in Malcomson (2004). What is optimal naturally depends on precise details of things that may not be known in practice, so one wants conclusions that are robust to a variety of circumstances. Two stand out from the analysis. First, the arrangement by which the gatekeeper receives a fixed fee and must pay for the cost of treatment (as with HMOs that act as both full insurer and gatekeeper) results in too strong incentives not to refer patients. Second, it is not sensible to introduce incentive contracts for just some GPs when patients can choose between them. Dusheiko *et al.* (2007: 742) report that 'the probability that patients were overall at least very satisfied with their GP practice was ... smaller in fundholding practices'. More specifically, 'patients of fundholders were less satisfied with . . . their GP's ability to arrange

https://doi.org/10.1017/S1744133107004288 Published online by Cambridge University Press

tests and to refer to a specialist'. Who wouldn't be if the purpose of fundholding is to reduce the specialist referrals that patients want?

The model that seems implicit in Street and Maynard (2007) is one in which recalcitrant clinicians have to be induced by financially incentivized hospital managers to take greater notice of costs than they traditionally have (perhaps as a result of the 'technological imperative' discussed by Fuchs (1986)). Their prescriptions make a lot of sense if that is indeed the right model. But they don't cite evidence that it is the right model and their prescriptions could have unintended consequences if it isn't. In the latter case, it might be more appropriate to refine payment by results in the other ways discussed here.

References

Chalkley, M. and Malcomson, J. M. (1998a), 'Contracting for health services when patient demand does not reflect quality', *Journal of Health Economics*, **17**(1): 1–19.

Chalkley, M. and Malcomson, J. M. (1998b), 'Contracting for health services with unmonitored quality', *Economic Journal*, **108**(449): 1093–1110.

Chalkley, M. and Malcomson, J. M. (2000), 'Government purchasing of health services', in Culyer A. J. and Newhouse J. P. (eds), *Handbook of Health Economics*, Vol. 1A, Amsterdam: Elsevier Science, chapter 15, pp. 847–890.

Chalkley, M. and Malcomson, J. M. (2002), 'Cost sharing in health service provision: an empirical assessment of cost savings', *Journal of Public Economics*, **84**(2): 219–249.

Dusheiko, M., Gravelle, H., Yu, N., and Campbell, S. (2007), 'The impact of budgets for gatekeeping physicians on patient satisfaction: evidence from fundholding', *Journal of Health Economics*, **26**(4): 742–762.

Ellis, R. P. and McGuire, T. G. (1993), 'Supply-side and demand-side cost sharing in health care', *Journal of Economic Perspectives*, **7**(4): 135–151.

Fuchs, V. R. (1986), *The Health Economy*, Cambridge, MA: Harvard University Press.

Laffont, J.-J. and Tirole, J. (1993), *A Theory of Incentives in Procurement and Regulation*, Cambridge, MA: MIT Press.

Malcomson, J. M. (2004), 'Health service gatekeepers', *RAND Journal of Economics*, **35**(2): 401–421.

Malcomson, J. M. (2005), 'Supplier discretion over provision: theory and an application to medical care', *RAND Journal of Economics*, **36**(2): 412–432.

McClellan, M. (1997), 'Hospital reimbursement incentives: an empirical analysis', *Journal of Economics and Management Strategy*, **6**(1): 91–128.

Newhouse, J.P. (2002), *Pricing the Priceless: A Health Care Conundrum*, Cambridge, MA: MIT Press.

Shleifer, A. (1985), 'A theory of yardstick competition', *RAND Journal of Economics*, **16**(3): 319–327.

Siciliani, L. (2006), 'Selection of treatment under prospective payment systems in the hospital sector', *Journal of Health Economics*, **25**(3): 479–499.

Street, A. and Maynard, A. (2007), 'Activity based financing in England: the need for continual refinement of payment by results', *Health Economics, Policy and Law*, this issue.

https://doi.org/10.1017/S1744133107004288 Published online by Cambridge University Press