

# EDITORIALS

## Why pay for performance may be incompatible with quality improvement

Motivation may decrease and gaming of the system is rife

Steffie Woolhandler *professor*<sup>1</sup>, Dan Ariely *James B Duke professor of psychology and behavioral economics*<sup>2</sup>, David U Himmelstein *professor*<sup>1</sup>

<sup>1</sup>City University of New York School of Public Health, New York, NY 10024, USA; <sup>2</sup>Duke University, Durham, NC, USA

In a linked article (doi:10.1136/bmj.e5047), Glasziou and colleagues highlight the tenuous nature of the evidence that financial reward systems work in healthcare settings.<sup>1</sup> They propose that before pay for performance schemes are implemented the potential benefits and harms should be assessed. Such schemes, which aim to improve the quality and efficiency of healthcare by the use of financial incentives to encourage desirable behaviours, have been adopted as a key strategy by the NHS in the United Kingdom, Medicare in the United States, and many private insurers. The schemes are based on a basic tenet of economics and psychology: that people respond to rewards.

Beyond the simple criticism that pay for performance can't operate on an extended time frame and that years may elapse between treatment and outcome, the concept of pay for performance in healthcare rests on flawed assumptions about medicine, measurement, and motivation. Performance based pay may increase output for straightforward manual tasks. However, a growing body of evidence from behavioral economics and social psychology indicates that rewards can undermine motivation and worsen performance on complex cognitive tasks, especially when motivation is high to begin with.

One questionable assumption underlying pay for performance is that measurements of doctors' performance reflect their overall performance and not—for example—their patients' characteristics or their ability to “game” the system. Health outcomes such as death or disability are the most easily measured and unambiguous indicators of overall performance, but they require risk adjustment. Hospital mortality provides the best case scenario for risk adjustment—outcomes are frequent, unambiguous, and likely to reflect performance; time horizons are short; and experts have spent years analysing the rich trove of hospital data. Yet, four widely used algorithms yield divergent rankings of inpatient mortality.<sup>2</sup>

Risk adjustment is devilishly difficult, partly because key inputs—clinical diagnoses—aren't solely patient characteristics

but also reflect the aggressiveness of coding and diagnostic investigations.<sup>3 4</sup> Seeking out and documenting unimportant comorbidities and diagnoses (such as occult prostate cancer in older people) exaggerate the severity of illness and artefactually raise risk adjusted quality scores. Hence, excessive testing inflates quality scores without improving quality.

Similarly, intensive coding—that is, embellishing diagnoses to maximise payment under per case or risk adjusted capitation schemes—also makes patients seem sicker on paper, and hence boosts risk adjusted quality scores. Under US Medicare's DRG (diagnosis related groups) hospital payment system, recoding a diagnosis as “aspiration pneumonia with acute on chronic systolic heart failure” rather than simply “pneumonia with chronic heart failure” triples the payment and increases the risk score.<sup>5</sup> Such “upcoding” is endemic among private health maintenance organisations that contract with Medicare for risk adjusted capitation payments,<sup>6</sup> as well as among hospitals.<sup>7</sup> One Maryland rehabilitation hospital reportedly urged doctors to document “protein malnutrition” in patients' charts, and this enabled the hospital to bill for 287 cases of “kwashiorkor” in 2007 (up from 0 in 2004).<sup>8</sup> Yet pay for performance programmes assume that quality indicators accurately reflect global quality and won't be distorted by payment incentives.

Process based indicators, although easier to calculate than risk adjusted outcomes, are poor proxies for quality of care. Even seemingly clear cut measures have hidden complexity. As examples, total hospital readmission rates correlate poorly with avoidable readmission rates,<sup>9</sup> and starting treatment for patients with pneumonia within four hours of arrival correlates with quality, yet Medicare's incentives for hospitals to do so resulted in the administration of antibiotics to almost any patient in the emergency department with a cough.<sup>10</sup>

Patients' social characteristics also confound process based measures; even excellent doctors who care for disadvantaged or difficult patients may look bad on current pay for performance metrics. Among respected doctors at a flagship Harvard teaching hospital, those who cared for more non-English speaking, poor,

uninsured patients from more minority backgrounds, as well as for patients with infrequent visits, scored low on pay for performance metrics.<sup>11</sup>

Using clinical audit for financial reward and punishment, rather than in a collegial and reflective effort to improve care, amplifies the challenges of performance measurement. Incentives may mutate honesty into legal trickery; gaming can so thoroughly distort reality that rewards become uncoupled from performance.

A second questionable assumption is that traditional payment or funding systems are too simple. Pay for performance replaces simpler more general payment contracts with ones that specify the “deliverables” in greater detail. For instance, contracts for accountable care organisations mandated under the 2010 US health reform incentivise 33 quality standards, whereas the UK’s primary care pay for performance programme initially tabulated 146 parameters,<sup>12</sup> with more to come. However, more may not be better; highly detailed prescriptive contracts may be perceived as controlling and may undermine the intrinsic motivation crucial to maintaining quality when nobody is looking.

Offering financial incentives to doctors, rather than enhancing their intrinsic motivation, may reduce their desire to perform an activity for its inherent rewards (such as pride in excellent work, empathy with patients). Worse still, if poor performance is the result of financial distress that is beyond providers’ control, penalising low scorers can make matters worse and exacerbate disparities, effectively punishing patients who cannot go elsewhere. Hospitals and doctors’ practices that deliver irremediably deficient care should be closed. It makes little sense to put already quality challenged providers on a starvation diet.

Future studies may yet establish the clinical effectiveness of pay for performance, but given its questionable assumptions this may never be the case. Despite a dearth of robust evidence that the system is clinically effective in healthcare, payers charge ahead with implementing everywhere an intervention that has not been proved to work anywhere. We are worried that pay for performance may not work simply because it changes the mindset needed for good doctoring. However, if such schemes must be envisaged, is essential that their likely benefit is

rigorously considered before their implementation. Glasziou and colleagues’ checklist provides a salutary guide to such consideration.

**Competing interests:** All authors have completed the ICMJE uniform disclosure form at [www.icmje.org/coi\\_disclosure.pdf](http://www.icmje.org/coi_disclosure.pdf) (available on request from the corresponding author) and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

**Provenance and peer review:** Not commissioned; externally peer reviewed

- 1 Glasziou PP, Buchan H, Del Mar C, Doust J, Harris M, Knight R, et al. When financial incentives do more good than harm: a checklist. *BMJ* 2012;345: e5047.
- 2 Shahian DM, Wolf RE, Iezzoni LI, Kirle L, Normand S-LT. Variability in the measurement of hospital-wide mortality rates. *N Engl J Med* 2010;363:2530-9.
- 3 Song Y, Skinner J, Bynum J, Sutherland J, Wennberg JE, Fisher ES. Regional variations in diagnostic practices. *N Engl J Med* 2010;363:45-53.
- 4 Welch HG, Sharp SM, Gottlieb DJ, Skinner JS, Wennberg JE. Geographic variation in diagnosis frequency and risk of death among Medicare beneficiaries. *JAMA* 2011;305:1113-8.
- 5 Administrative Consultant Services. Reporting secondary diagnoses that impact MS-DRG payment. *MS-DRG News* 2011. [www.acsteam.net/sites/acs/uploads/documents/newsletters/Secondary\\_2011.pdf](http://www.acsteam.net/sites/acs/uploads/documents/newsletters/Secondary_2011.pdf).
- 6 Brown J, Duggan M, Kuziemko I, Wollston W. How does risk selection respond to risk adjustment? Evidence from the Medicare Advantage program. NBER Working Paper 16977. National Bureau of Economic Research, 2011.
- 7 Silverman E, Skinner J. Medicare upcoding and hospital ownership. *J Health Econ* 2004;23:369-89.
- 8 Anonymous. Maryland hospital accused of inflating treatment claims. *Mod Healthcare* 2011;41:4.
- 9 Van Walraven C, Jennings A, Taljaard M, Dhalla I, English S, Mulpuru S, et al. Incidence of potentially avoidable urgent readmissions and their relation to all-cause urgent readmissions. *CMAJ* 2011;183:E1067-72.
- 10 Kanwar M, Brar N, Khatib R, Fakhri MG. Misdiagnosis of community acquired pneumonia and inappropriate utilization of antibiotics: side effects of the 4-h antibiotic administration rule. *Chest* 2007;131:1865-9.
- 11 Hong CS, Atlas SJ, Chang Y, Subramanian SV, Ashburner JM, Barry MJ, et al. Relationship between patient panel characteristics and primary care physician clinical performance rankings. *JAMA* 2010;304:1107-13.
- 12 Doran T, Fullwood C, Gravel H, Reeves D, Kontopantelis E, Hiroeh U, et al. Pay-for-performance programs in family practices in the United Kingdom. *N Engl J Med* 2006;355:375-84.

Cite this as: *BMJ* 2012;345:e5015

© BMJ Publishing Group Ltd 2012