Financial Penalties for Performance in Health Care

Søren Rud Kristensen
Manchester Centre for Health Economics, University of Manchester

Over the past 15 years, pay for performance (P4P) schemes have been widely employed by health care payers attempting to improve the quality of care. These schemes link a proportion of providers’ income to performance on indicators that are thought to reflect quality of care. The popularity and use of these schemes flourished after the U.S. Institute of Medicine (2001) published its report Crossing the Quality Chasm which suggested a misalignment of incentives in the payment schemes of the time.

Health care payers in the US were quick to respond to the challenge, and numerous quality payment schemes emerged. The largest scheme – the Medicare Hospital Quality Incentive Demonstration framework famously produced only modest improvements in measured process indicators of quality in the short run (Lindenauer et al., 2007) and none in outcomes (Jha et al., 2012).

Although an adaptation of the scheme in one English region was initially deemed a success (Sutton et al., 2012) the findings were not robust to alternative evaluation methods (Kreif et al., 2015) and in the long run the initially identified improvements were not sustained (Kristensen et al., 2014). Experiences from across the OECD countries are still inconclusive (Milstein and Schreyögg, 2015). Like any incentive

---

1 Funding received from the Danish Council from Independent Research, Social Sciences. Part of this work was carried out while I was a visiting scientist at the Harvard T. H. Chan School of Public Health. Helpful comments from Meredith Rosenthal, Matt Sutton, and two anonymous referees are gratefully acknowledged.
scheme, P4P has numerous design levers that might affect the performance of the schemes, and current consensus seems to be that 1) P4P is not a magic bullet (Roland, 2012) and 2) The devil is in the detail (Epstein, 2012).

One of these details is the very mechanism that gave P4P its’ name: the payment. It is an emerging trend to incentivise health care providers not via bonuses for desired performance but using penalties for poor performance. Such schemes, are increasingly being used by health care payers on both sides of the Atlantic and have been highlighted as a potentially cost-effective alternative to using bonuses to improve performance (Maynard, 2012).

Rosenthal (2007) referred to such penalties as nonpayment for performance (NP4P) and the terms nonpayment and penalty are often used interchangeably. One might distinguish between penalties as incentives that are applied ex-post, and non-payments as ex-ante payment changes that can be viewed as penalties only with the baseline payment system as a reference point.

**Examples of penalties for performance in Health Care**
Penalties for readmissions have attracted particular attention in recent years (Kristensen et al., 2015). The English National Health Service (NHS) introduced penalties for readmission from financial year 2011/12 (Department of Health, 2011). The payment change represents a break from the usual activity based reimbursement scheme employed in the NHS where hospitals receive a fixed tariff per admission based
on the patient healthcare resource group (HRG)—the English equivalent of diagnosis related groups.

After the payment reform, hospitals are no longer reimbursed for emergency readmissions occurring within 30 days of discharge from the initial admission when the readmission rate is above a locally set threshold. With a few exceptions (cancer, maternity and child care) the policy targets all-cause readmissions after both elective and emergency index admissions irrespective of whether a clear clinical link has been established between the index and re-admission.

A penalty approach was also taken to reduce readmission in the US Medicare in 2012 but the implementation was different. Rather than targeting all readmissions, the policy focused initially on emergency readmissions within 30 days of discharge after an index admission for heart failure, acute myocardial infarction and pneumonia. Rather than locally agreed thresholds, the Medicare policy relies on an algorithm based on the previous 3 years’ data to set an expected readmission rate for each hospital taking casemix but not socioeconomic status into account (PPACA, 2010).

The evidence base for these reforms is still scarce, but suggests that while readmissions decreased in Medicare for both targeted and non-targeted conditions after the penalties were announced (Zuckerman et al., 2016) the English reform was less successful in decreasing hospital readmission rates (Kristensen and Sutton, n.d.)

In another set of penalty based reforms both Medicare and the English NHS have stopped paying for some hospital acquired conditions (HAC) or never events, that are seen as preventable results of faulty care or medical error. For example, Medicare no

---

2 Although, the incentives embedded in payment schemes such as fee-for-service and capitation can also be seen as implicitly penalising certain types of behaviour, here, I shall limit my focus to payment mechanisms that are explicitly framed as penalties.
longer pays for the cost of treating pressure ulcers acquired during an inpatient stay (Federal Register, 2007) and the NHS no longer reimburses providers for patients incurring death or severe harm due to maladministration of insulin by a health professional (Department of Health, 2011). From 2015, Medicare has reinforced the penalty approach to hospital acquired conditions by also applying penalties for higher than expected HAC rates in an algorithm based approach similar to the one used for readmissions. Evaluations of version one of the Medicare HAC reforms suggests none to moderate effects of the reforms (Lee et al., 2012; Meddings et al., 2012; Schuller et al., 2014; Waters TM et al., 2015) but the evidence to date focuses on a subset of conditions and mostly lacks controls.

Other schemes such as the Hospital Value Based Purchasing (HVBP) program (Federal Register, 2011) contains a combination of bonuses for top performers and penalties for low performers. Individual physicians and group practices are already subject to penalties for (lack of) reporting under the Physician Quality Reporting System (Centers for Medicare & Medicaid Services, 2016a) and to a mix of bonuses and penalties for performance through the the Value-based Payment Modifier (Centers for Medicare & Medicaid Services, 2016b) which from 2019 will continue in the Merit-based Incentive Payment System (MIPS) (Federal Register, 2016) introduced as part of the Medicare Access and CHIP Reauthorization Act of 2015 (MACRA) (U. S. Congress, 2015).

What to expect from penalty based incentive schemes? The standard economic expected utility framework (Von Neumann and Morgenstern, 1947) does not predict different behavioural responses to incentives framed as penalties or bonuses as any penalty scheme can be re-written into a bonus scheme with
the same expected utility (Aron and Olivella, 1994; Becker, 1968; Camerer, 2003; Lazear, 1991).

In Camerer’s (2003) presentation, the expected utility of a risky outcome is evaluated as

\[ E(U) = \sum p_i u_i(x_i) \]

where \( p_i \) is the probability of outcome \( x_i \) from which the agent derives utility \( u_i(x_i) \). Accordingly, this framework does not suggest that framing of the contract as a bonus or penalty affects agents’ evaluation of a contract. However, as suggested by Kahneman and Tversky’s (1979) prospect theory, reference points and loss aversion may affect how agents respond to an incentive scheme.

While a few authors (De Geest et al., 2009; De Geest and Dari-Mattiacci, 2009) have attempted to model differences between bonuses and penalties within an EUT framework, most of this work has taken point of departure in Kahneman and Tversky’s (1979) work. Again relying on Camerer’s representation, in prospect theory, the expected utility is assumed to take the form \( E(U) = \sum \pi(p_i) v(x_i - r) \) with \( r \) being a reference point and \( \pi \) being a nonlinear weight assigned to probabilities so that low probabilities are over-weighted and high probabilities are underweighted. The value function \( v(x_i - r) \) exhibits a reflection effect where \( v''(x - r) > 0 \) for \( x < r \) and \( v''(x - r) < 0 \) for \( x > r \) and loss aversion, so that \( -v(-x) > v(x) \) for \( x > 0 \). In this perspective, the evaluation of an uncertain outcome is thus dependent on the reference point and the associated risk of outcome, and differently framed incentive schemes are predicted to have different behavioural implications.

At the individual level, this notion is supported by some evidence from behavioural economics. Lab experiments (Dickinson, 2001; Hannan et al., 2005; Luft, 1994) and field experiments (Hossain and List, 2009) suggest that employees prefer ‘bonus’ contracts over ‘penalty’ contracts, even when they are monetary equivalents, but exert more
effort under penalty contracts (Dickinson, 2001; Hannan et al., 2005). Other experimental studies have analysed the combination of penalty and bonus schemes (Andreoni et al., 2003; Fehr and Schmidt, 2007), bonuses and penalties in an incomplete contracting framework (Christ et al., 2012), under imperfect monitoring (Van de Weghe and Bruggeman, 2006) and have compared potential explanations for the framing effect (Brooks et al. 2012).

**Questions for future research**

While individual level responses to framing has been given considerable attention in the literature, the behavioural expectations when the unit of analysis changes from individual to the organisation level remain largely unexplored, despite the frequency of regulators’ use of reimbursement incentives to achieve specific policy goals.

Indeed, it is unclear how and whether the essentially psychological underpinnings of prospect theory are applicable to organisational responses to reimbursement contracts. Jensen and Meckling (1976) warned against the ‘personalisation’ of the firm and stressed that ”The firm is not an individual” but a “nexus of contracts” to be analysed as a set of “contracting relationships among individuals” (p. 311). And even if we were to take a personalised view of the firm, it is unclear how loss aversion would generate the same behavioural responses in organisations which are typically thought to be risk neutral.

A key question for future research is thus to address how concepts such as reference dependence and risk should be thought of when the unit of analysis is the organisation. Camerer and Malmendier (2007) point out that this is an aggregation problem, and suggest that organisations can exploit the heterogeneity in the cognitive biases of the individuals that make up the organisation to match individuals to tasks where their
individual biases are less costly, hereby potentially safeguarding the organisation against the irrational behaviour individuals might exhibit on their own. The forthcoming Medicare MIPS scheme which targets both individual physicians and group practices of varying sizes generate exciting possibilities for more empirical research on this topic.

With the range of penalty based performance schemes that are emerging, more empirical estimates of the effect of those schemes on provider performance are likely to appear in the near future. An important question that this research will not necessarily provide an answer to, is the relative effect of penalties versus bonuses. If loss aversion is a driver in the incentive effect of performance penalties, it may be that putting $x$ percent of providers’ income at risk has the same effect as paying out a bonus of $y>x$ percent.

The incentive schemes that rely on a combination of bonuses and penalties may deliver a useful vehicle for empirical researchers to be able to answer this question although the endogeneity of exposure to bonuses or penalties is a concern in this setup. Randomised staged introduction of new schemes or regional pilot schemes that address selection concerns would provide better possibilities for researchers to examine this and other questions of incentive design.

It is a key challenge for researchers to isolate the effect of framing payment incentives as penalties from the effect of the many other elements of incentive scheme design. For example, although both the HRRP and HVBP programmes rely on penalties, they differ in other important aspects, such as complexity and penalty size (Norton et al., 2016). In understanding the relative contribution of penalties compared to other design elements, additional insights from behavioural economics such as mental accounting and choice
overload (Emanuel et al., 2016) and fairness and reciprocity (Fehr and Falk, 2002) might provide useful starting points.

Finally, it is of no less significance to understand how penalty based schemes compare to schemes relying on bonuses in their likelihood of inducing unintended behaviour. There is plenty of evidence to suggest that providers respond to incentive schemes in unanticipated and undesirable ways. If loss aversion is the mechanism relied on to generate behaviour change, might it be that the unintended consequences are even stronger under penalty schemes compared to bonus schemes?

Appropriate risk adjustment may play a crucial role in avoiding such consequences. A review of the impact of P4P on inequalities, found weak evidence that P4P had the potential to decrease inequalities in health (Alshamsan et al., 2010). However, descriptive work by Figueroa et al. (2016) suggests that hospitals serving more vulnerable populations are more likely to receive penalties under the HRRP, HVBP and HAC programmes. The extent to which this finding reflects differences in quality or differences in case mix is still poorly understood, but work by Barnett et al. (2015) on the HRRP suggests that differences in patient characteristics between hospitals currently unaccounted for in the risk adjustment algorithm used by CMS may explain much of the variation in the penalties to hospitals with high readmission rates.

Penalty based incentive schemes may appear to be a low-cost alternative to bonus schemes, but further research is needed to fully appreciate the cost effectiveness of this type of reform.

References


Van de Weghe, P., Bruggeman, W., 2006. The impact of uncertainty on incentive framing effectiveness in a multidimensional task environment. Working Papers of Faculty of Economics and Business Administration, Ghent University, Belgium.

