UNIVERSITY of York

Discussion Papers in Economics

No. 18/03

Hospital Competition under Pay-for-Performance: Quality, Mortality and Readmissions

Domenico Lisi Luigi Siciliani
Odd Rune Straume

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD

Hospital Competition under Pay-for-Performance: Quality, Mortality and Readmissions*

Domenico Lisi† Luigi Siciliani‡ Odd Rune Straume§

April 11, 2018

Abstract

Health outcomes, such as mortality and readmission rates, are commonly used as indicators of hospital quality and as a basis to design pay-for-performance (P4P) incentive schemes. We propose a model of hospital behaviour under P4P where patients differ in severity and can choose hospital based on quality. We assume that risk-adjustment is not fully accounted for and that unobserved dimensions of severity remain. We show that the introduction of P4P which rewards lower mortality and/or readmission rates can weaken or strengthen hospitals' incentive to provide quality. Since patients with higher severity have a different probability of exercising patient choice when quality varies, this introduces a selection bias (patient composition effect) which in turn alters quality incentives. We also show that this composition effect increases with the degree of competition. Critically, readmission rates suffer from one additional source of selection bias through mortality rates since quality affects the distribution of survived patients. This implies that the scope for counterproductive effects of P4P is larger when financial rewards are linked to readmission rates rather than mortality rates. We also show that our results are robust in the presence of public reporting, and discuss welfare implications.

Keywords: quality; pay-for-performance; health outcomes; performance indicators; heterogeneous severity; selection bias.

JEL Classification: I12; I18.

*Odd Rune Straume acknowledges funding from COMPETE (ref. no. POCI-01-0145-FEDER-006683), with the FCT/MEC's (Fundação para a Ciência e a Tecnologia, I.P.) financial support through national funding and by the ERDF through the Operational Programme on Competitiveness and Internationalization – COMPETE 2020 under the PT2020 Partnership Agreement.

†Department of Economics and Business, University of Catania, 95129, Italy. E-mail: domenico.lisi@unict.it.

‡Department of Economics and Related Studies, University of York, Heslington, York YO10 5DD, UK. E-mail: luigi.siciliani@york.ac.uk.

§Department of Economics/NIPE, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal; and Department of Economics, University of Bergen. E-mail: o.r.straume@eeg.uminho.pt.

1 Introduction

The ageing population and the rising prevalence of chronic conditions are putting healthcare systems under pressure. In light of the recent financial downturn, cost containment and value for money in health spending are common objectives across a range of health systems. Since the early eighties, many OECD countries have introduced prospective payment systems to reimburse providers, and stimulate cost efficiency in the provision of healthcare services. However, there are concerns that such payment systems could come at the expense of quality reductions. To address this concern, governments increasingly combine prospective systems with further regulatory mechanisms, including a variety of pay-for-performance (P4P) programmes that explicitly align financial incentives with quality objectives (Busse et al., 2011; OECD, 2010).

The main difficulty in designing P4P programmes is the identification of reliable performance indicators which reflect providers' quality. One appealing option is to measure performance based on health outcomes. Two common indicators are mortality rates and readmission rates which are measured through routinely collected administrative databases (Cashin et al., 2014; Milstein and Schreyoegg, 2016).[1] One advantage of using mortality rates is that they are unequivocal. Higher mortality implies, everything else constant, poorer quality of care. But mortality rates are only relevant for a subset of patient care. Patients receive many treatments for which the mortality rate is zero or negligible (e.g., a cataract surgery, a hip replacement). For such treatments readmission rates is a valid option, and these have been used both for treatments with negligible mortality risk and for treatments with a significant mortality risk (e.g., heart attack, hip fracture).

The main limitation of performance measures based on health outcomes is that they may reflect patient case-mix in addition to hospital quality. They are reliable measures of quality only if appropriate risk-adjustment is made to account for patients' heterogeneity in the risk of mortality and readmission (McClellan and Staiger, 1999; Shahian et al., 2010; Laudicella et al., 2013; Papanicolas and McGuire, 2017). It is recognised that unobserved dimensions of severity remain even after risk adjustment (Mohammed et al., 2009; Berenson et al., 2013; Wennberg et al., 2013) and this reduces the reliability of these performance measures. Moreover, readmission rates suffer from one additional limitation. If some patient characteristics are unobserved (to

[1] Health outcome measures are also largely used in the health literature to evaluate and study hospitals' performance. In particular, low mortality and readmission rates for selected diagnosis (such as, acute myocardial infarction, coronary artery bypass grafting, strokes, hip fracture) are often used as a proxy measure for good inpatient care quality within the hospital competition literature (Kessler and McClellan, 2000; Propper et al., 2004, 2008; Cooper et al., 2011; Gravelle et al., 2014; Gaynor et al., 2016; Lisi et al., 2017).

the regulator or the researcher) and these are correlated with the mortality risk, variations in readmission rates across hospitals may be confounded by variations in mortality rates. For example, a hospital with a lower mortality rate may face a larger share of unobservable sicker patients, resulting in higher readmission rates (McClellan and Staiger, 1999; Laudicella et al., 2013).

Although the empirical literature has highlighted the limitations of developing performance measures based on health outcomes, such as mortality and readmission rates, little is known about how these limitations affect hospital incentives to provide quality. This study fills this gap in knowledge. We propose a model of hospital behaviour under P4P. Patients differ in their severity and, thus, in their probability of negative outcomes, i.e., in their mortality and readmission risk. Patients maximise their expected utility, which depends on the mortality and readmission risk, and choose hospitals based on quality and distance to the hospital. Hospitals operate under a P4P programme, which provides a bonus for a reduction in mortality and/or readmission rates (or, equivalently, a penalty for an increase in mortality and/or readmission rates), and choose how much to invest in the quality of the treatment offered.

Critically, we assume that risk-adjustment is not fully accounted for and that unobserved dimensions of severity remain, so that outcome measures suffer from selection biases. More precisely, unobserved severity affects patient choices and implies different demand responsiveness to quality across different patient severity types, which in turn generates patient composition (selection) effects through *demand* in hospitals' mortality and readmission rates. Consequently, a change in treatment quality of a hospital may not always translate into the expected change in its mortality and readmission rates. Our model also allows for readmission rates to suffer from patient composition (selection) effect through *mortality*, since quality affects the distribution of survived patients.

Our main research question investigates how hospitals' incentives for quality provision are affected by P4P in this rich environment. We obtain several policy-relevant results. First, we show that whether demand responsiveness to quality is stronger for high-severity or for low-severity patients is *a priori* ambiguous. This implies that the direct effect of P4P on hospitals' incentives for quality provision can be either counteracted or reinforced by patient composition (selection) effects through *demand*. Second, if the demand responsiveness to quality is stronger for high-severity than for low-severity patients, we show that a financial bonus related to reductions in *mortality rates* will be counterproductive, and instead lead to lower quality provision (and thus higher mortality

rates), if the difference in mortality risks across severity types is sufficiently large relative to the effect of treatment quality on individual mortality risk.

Third, we show that the relationship between quality and *readmission rates* is affected by selection bias through mortality, and this bias increases the scope for counterproductive effects of P4P. Even if the demand responsiveness to quality is stronger for low-severity than for high-severity patients, P4P linked to readmission rates might nevertheless lead to lower quality provision if the difference in mortality risks across severity types is sufficiently large. Fourth, we show that the presence of this additional selection bias can make readmission rates an unreliable measure to design P4P programmes. In contrast to P4P linked to mortality, where an observed reduction in mortality rates implies a higher treatment quality, a reduction in readmission rates cannot, by itself, be taken as evidence of higher treatment quality. In order to make reliable inferences about treatment quality in this case, the effects of P4P on readmission *and* mortality rates must be seen in conjunction.

We also investigate how the degree of competition in the market interacts with P4P incentive schemes. Stronger competition is usually deemed to have beneficial effects insofar as it stimulates quality provision (Gaynor, 2006; Gravelle and Sivey, 2010; Brekke et al., 2010, 2011). However, we show that the presence of P4P can either dampen or reinforce the positive effect of competition on quality provision, depending on whether the demand responsiveness to quality is respectively stronger for high-severity or low-severity patients. This holds regardless of whether P4P bonuses are linked to mortality or readmission rates. Similarly, the effect of P4P on hospitals' incentives for quality provision is also affected by the degree of competition. If demand responsiveness to quality is stronger for high-severity than for low-severity patients, a higher degree of competition will increase the patient composition effect through demand, thereby increasing the scope for counterproductive effects of P4P. In this case, P4P schemes are more likely to succeed in markets with less competition.

In our main analysis, we assume that hospital quality is observable to patients. In the final part of the paper we relax the assumption of observability (e.g., Gravelle and Sivey, 2010) and assume that patients only have access to a publicly reported measure of hospital quality, either mortality rate or readmission rate. Overall, we show that the effects of a unilateral quality change on health outcomes, as well as the implications for the equilibrium effects of P4P programmes, are similar to the ones derived in the main model. Finally, we investigate the welfare implications of P4P schemes under different perspectives and scenarios. We discuss these in the concluding section.

1.1 Institutional background

OECD countries have recently introduced P4P programmes in the hospital sector with the aim of improving quality of health care. Several of these schemes reward health outcomes (Cashin et al., 2014; Milstein and Schreyoegg, 2016). Common indicators include 30-day mortality and readmission rates for specific planned procedures (e.g., coronary artery bypass graft, hip replacement), emergency ones (e.g., acute myocardial infarction, stroke) or all procedures (planned and emergency ones). Our model applies to both emergency and planned procedures, though it is more relevant for planned ones where selection effects due to patient choice are likely to be more important and for emergency ones where the mortality rate is not negligible so that selection effects through mortality are likely (e.g., readmission rates for acute myocardial infarction and stroke).

The US Premier Hospital Quality Incentive Demonstration (HQID) was launched in 2003 and rewarded reductions in mortality rates for acute myocardial infarction (AMI), coronary artery bypass graft (CABG), hearth failure (HF) and pneumonia (PN), as well as reductions in 30-day readmission rates for hip replacement (Lindenauer et al., 2007). There is evidence suggesting that hospital performance improved (Werner et al., 2011).

The Centers for Medicare & Medicaid Services (CMS) introduced in 2012 the Hospital Value-Based Purchasing (VBP) scheme which rewards lower 30-day mortality rates for AMI, HF and PN, and the Hospital Readmission Reduction Program (HRRP) which penalises hospitals with higher 30-day readmission rates for elective and emergency procedures.[2] The HRRP initially included readmission rates for AMI, HF and PN, and was then extended to CABG, chronic obstructive pulmonary disease (COPD), and total hip and knee replacements. Recent empirical evidence suggests that HRRP has been effective in reducing readmission rates for AMI, but not for HF and PN (Mellor et al., 2017).

The English NHS has also introduced a number of P4P schemes for both primary care (Gravelle et al., 2010) and secondary care (Sutton et al., 2012). In 2008 the NHS introduced the hospital Advancing Quality (AQ) programme, which includes the same quality metrics as the HQID.[3] Sutton et al. (2012) find that the AQ programme was associated with a reduction in mortality rates of 1.3 percentage points for AMI, HF and PN. The English NHS has also introduced in 2011 financial measures linked to readmissions, establishing that hospitals do not receive anymore payment for

[2] More information on Value-Based Programs and CMS quality strategy is available here: https://www.cms.gov.

[3] More information on the AQ programme can be found here: https://www.aquanw.nhs.uk/membership/advancing-quality.htm

emergency readmissions within 30 days of discharge, following either an elective admission or a non-elective admission, when the readmission rate is above a given threshold (Department of Health, 2011). A similar policy was introduced in Germany (Geissler et al., 2011).

Other examples of hospital P4P programmes that reward health outcomes include the Swedish *Målrelaterat Ersättning* program introduced in 2005, which measures revision rates within 2 years following hip replacement, and the Korean Value Incentive Program (VIP) established in 2007, which rewards reductions in 30-day mortality rates for AMI. In the Netherlands (Custers et al., 2007) and Brazil (La Forgia and Couttolenc, 2008), the contracts between purchasers and providers can explicitly include P4P schemes with mortality and readmission rates as quality indicators. Finally, some countries such as Germany (Busse et al., 2009), Italy (Cavalieri et al., 2013) and the Netherlands (Varkevisser et al., 2012) monitor and release in the public domain indicators related to mortality and readmission rates, but these are so far not an integral part of the hospitals payment system.

Despite the increasing uptake of hospital P4P, its success remains to be established. The empirical evidence remains limited and shows at best modest and short-lived effects on quality of care (Mullen et al., 2010; Flodgren et al., 2011; Werner et al., 2011; Eijkenaar et al., 2013; Mellor et al., 2017).[4]

1.2 Related literature

Our study contributes and integrates two strands of the literature. The first relates to the design of optimal pay-for-performance schemes, and more broadly payment schemes, to incentivise quality in the health sector. The second relates to hospital competition and its effect on quality. We discuss these in turn. Recent contributions adapt the seminal study by Holmstrom and Milgrom (1991) on multitasking to the health sector. Eggleston (2005) shows that incentivising some dimensions of quality might come at the cost of reducing unincentivised dimensions of quality when qualities are substitutes, though cost sharing might mitigate this. In turn, Kaarboe and Siciliani (2011) show that this implies that the power of the incentive scheme is generally low (and in some cases it might completely break down). Similarly, Sherry (2016) finds that when P4P programmes reward multiple health care services in a setting characterized by joint production, the impact on both rewarded and unrewarded services remains ambiguous. Kuhn and Siciliani (2009, 2013) show that gaming of the

[4] See Cox et al. (2016) for recent experimental evidence which mimics the HRRP and incentivises cost-effective reductions in hospital readmission rates.

indicators also generally leads to lower powered incentive schemes. Mak (2018) studies a managed healthcare market with two differentiated hospitals, and analyses the interaction of P4P schemes on contractible quality with other features of the market, such as the presence of copayment and consumers' misperception of quality.

These studies assume that quality is directly (though imperfectly) contractible through an incentive scheme. An older literature has shown that hospitals have incentives to provide quality if hospital demand responds to quality (Ma, 1994) but only in health systems with excess capacity (Chalkley and Malcomson, 1998a). However, providers may still provide quality, to some extent, driven by altruistic motives (Ellis and McGuire, 1986; Chalkley and Malcomson, 1998b). Empirical demand elasticities suggest that these are generally low (see Brekke et al., 2014, for an overview) and this may be the reason for governments to wanting to supplement schemes which reward activity with those that reward quality directly. Guccio et al. (2016) show that reducing the tariff for readmitted patients is welfare improving if demand responsiveness to quality is low. Chalkley and Khalil (2005) show that if patients are responsive to variations in treatment, it can be worthwhile to base payment on the health outcome achieved rather than upon the treatment delivered. Differently from the above literature, we focus on the limitations of P4P schemes which directly reward improvements in health outcomes, as opposed to process measures of quality, and investigate the implications for the appropriateness and effectiveness of the incentive scheme.[5]

Our study also relates to an extensive literature on hospital competition. This literature shows that when hospitals are profit maximisers an increase in competition generally increases quality (Ma and Burgess, 1993; Wolinsky, 1997; Gravelle, 1999; Beitia, 2003; Nuscheler, 2003; Brekke et al., 2006, 2007; Gaynor, 2006; Karlsson, 2007; Bardey et al., 2012), though this may not be the case if providers face capacity constraints and altruistic motives, and may also depend on the competition measures or concept (Brekke et al., 2010, 2011; Siciliani et al., 2013). None of these studies focuses on P4P schemes which directly reward quality.

Finally, our study provides a theoretical framework to interpret the mixed empirical evidence on the impact of P4P programmes which reward health outcomes (Mullen et al., 2010; Van Herck et al., 2010; Flodgren et al., 2011; Werner et al., 2011; Eijkenaar et al., 2013; Mellor et al., 2017).

[5] In the context of the HRRP in the US, Zhang et al. (2016) provide a model of hospital behaviour under a P4P scheme which rewards reductions in readmission rates. However, in their model there is neither a microfounded generating process of readmission rates, nor a role for patient choice. This implies that there is no place for patient composition (selection) effects through neither demand nor mortality. Indeed, we show that these effects are the core of the analysis of P4P rewarding directly improvements in health outcomes.

It has been argued that accurate performance metrics are crucial to capture variations in provider effort (Eijkenaar, 2011; Epstein, 2012; Cashin et al., 2014; Conrad, 2015). Our results highlight that the extent to which a change in provider effort translates into a change in health outcomes depends also on the demand responsiveness to quality across severity types, and the degree of competition in the healthcare market.

The rest of the paper is organised as follows. In Section 2, we describe the key assumptions of the model on hospitals' behaviour under P4P. In Section 3, we define performance indicators based on mortality and readmission rates and their relationship with the quality provided by hospitals. Section 4 describes the hospital's maximisation problem, derives equilibrium quality and discusses how P4P bonuses affect equilibrium quality. Sections 5 is devoted to welfare analysis. Section 6 extends the model to the case where health outcomes are also used for public reporting. Section 7 draws conclusions and discusses policy implications.

2 Model

Consider a market for healthcare treatment where two hospitals, indexed by i and j, are located at the endpoints of the unit line $[0, 1]$, and provide treatment qualities q_i and q_j, respectively. Patients differ in severity, s, which can take two values, low and high: $s \in \{L, H\}$. Both patient types are uniformly distributed over the unit line, and the densities of low-severity and high-severity patients are γ and $(1-\gamma)$, respectively, implying that the total mass of patients is normalised to 1. Each patient demands one unit of treatment from the most preferred hospital.

Each patient of severity s undergoing treatment in Hospital i has a probability of dying (e.g., mortality risk during surgery) equal to $M(s, q_i)$. Conditional on surviving with probability $S(s, q_i) = 1 - M(s, q_i)$, the patient has a probability of being readmitted to a hospital (e.g., following complications) equal to $R(s, q_i)$. Each patient has therefore three possible health states after treatment:[6]

1. State M, in which the patient dies. This occurs with probability $M(s, q_i)$ and yields patient utility U^M (which can be thought of as a large negative number).

2. State R, in which the patient survives but is readmitted. This occurs with probability $S(s, q_i) R(s, q_i)$ and yields patient utility $U^R(s, q_i)$.

[6] We rule out the possibility that the patient is readmitted and then dies.

3. State N, in which the patient survives and is not readmitted. This occurs with probability $S(s,q_i)[1-R(s,q_i)]$ and yields patient utility $U^N(s,q_i)$.

We make the following assumptions regarding the mortality and readmission *probabilities*:[7]

A1 $M(H,q_i) > M(L,q_i)$ and $R(H,q_i) > R(L,q_i)$.

A2 $M_{q_i}(s,q_i) < 0$ and $R_{q_i}(s,q_i) < 0$.

A3 $|M_{q_i}(H,q_i)| > |M_{q_i}(L,q_i)|$ and $|R_{q_i}(H,q_i)| > |R_{q_i}(L,q_i)|$.

These are intuitive assumptions. For a given level of treatment quality, more severely ill patients face a higher risk of mortality and readmission (A1). A higher quality level reduces both risks (A2), but more so for high-severity patients (A3).

We make further assumptions regarding the state-contingent *utilities*:

A4 $U^N(s,q_i) > U^R(s,q_i) > U^M$.

A5 $U^R_{q_i}(s,q_i) > 0$ and $U^N_{q_i}(s,q_i) > 0$.

A6 $U^R_{q_i}(H,q_i) > U^R_{q_i}(L,q_i)$ and $U^N_{q_i}(H,q_i) > U^N_{q_i}(L,q_i)$.

Again, these are intuitive assumptions. For surviving patients, the utility of not being readmitted is higher than the utility of being readmitted, and both states yield higher utility than not surviving (A4). Furthermore, higher treatment quality increases the utility in both surviving states (A5), but more so for high-severity patients (A6). Notice that A6 implies $U^k(L,q_i) \neq U^k(H,q_i)$, $k = N, R$. For completeness, we therefore introduce the following assumption:

A7 $U^k(L,q_i) > U^k(H,q_i)$, $k = N, R$.

Thus, conditional on survival, post-treatment utility is higher for a low-severity patient (A7), but higher treatment quality reduces the difference in post-treatment utility for high- and low-severity patients (A6).

For the subsequent analysis, the following definitions will be useful: (i) the expected utility gain if the patient survives the treatment is given by

$$\Delta^{SM}_U(s,q_i) := (1-R(s,q_i))\,U^N(s,q_i) + R(s,q_i)\,U^R(s,q_i) - U^M > 0; \tag{1}$$

[7] We use the notational shorthand $f_x(\cdot)$ to denote the partial derivative of f with respect to x.

(ii) the ex post utility gain of not being readmitted for a surviving patient is given by

$$\Delta_U^{NR}(s, q_i) := U^N(s, q_i) - U^R(s, q_i) > 0; \tag{2}$$

and (iii) the expected marginal utility of quality for a surviving patient is given by

$$\overline{U}_{q_i}(s, q_i) := (1 - R(s, q_i)) U_{q_i}^N(s, q_i) + R(s, q_i) U_{q_i}^R(s, q_i) > 0. \tag{3}$$

2.1 Demand for hospital treatment

Consider a patient of type s who is treated at Hospital i. Net of travelling costs, the expected (gross) utility of this patient is given by

$$V(s, q_i) = M(s, q_i) U^M + S(s, q_i) R(s, q_i) U^R(s, q_i) + S(s, q_i) [1 - R(s, q_i)] U^N(s, q_i), \tag{4}$$

and depends on treatment quality via three different channels, represented by each of the three terms in the following expression:

$$\frac{\partial V(s, q_i)}{\partial q_i} = -M_{q_i}(s, q_i) \Delta_U^{SM}(s, q_i) - R_{q_i}(s, q_i) S(s, q_i) \Delta_U^{NR}(s, q_i) + S(s, q_i) \overline{U}_{q_i}(s, q_i). \tag{5}$$

The first and second terms in (5) are the expected utility gains from a reduction in mortality risk and readmission risk, respectively, whereas the third term is the expected utility gain from higher treatment quality for given mortality and readmission probabilities. All three effects are positive, implying that higher treatment quality by Hospital i increases the expected utility of being treated at that hospital for all patients of both severity types, which in turn increases the demand for Hospital i.

In line with the existing theoretical literature on hospital competition, we assume that quality is observable to patients. Patients may learn about hospital quality through word of mouth (from family and friends), and based on the advice of their family doctors who have long-term and repeated interactions with each hospital through past referrals.[8] In Section 6, we extend the model where quality is observable to patients only through public reporting, e.g., by publishing mortality and readmission rates in the public domain, and we show that the key results are robust to this

[8] For example, Berta et al. (2016) show that patient choice of hospital responds to quality even in health systems with no public reporting of performance indicators in Italy.

alternative set-up.

We assume that patients have travelling costs equal to t per unit of distance travelled when attending a hospital. Since Hospitals i and j are located at the left and right endpoints of the unit line, a patient located at $x \in [0,1]$ derives utility $V(s,q_i) - tx$ and $V(s,q_j) - t(1-x)$ from going to Hospital i and Hospital j, respectively. For a given severity s we can then identify the location of the patient who is indifferent between the two hospitals, defined as $x(s)$, which is implicitly given by

$$V(s,q_i) - tx(s) = V(s,q_j) - t(1-x(s)),$$

and explicitly given by

$$x(s) = \frac{1}{2} + \frac{V(s,q_i) - V(s,q_j)}{2t}. \tag{6}$$

The demand for Hospital i is then given by

$$D_i(q_i,q_j) = \frac{1}{2} + \frac{\gamma \Delta V(L) + (1-\gamma)\Delta V(H)}{2t}, \tag{7}$$

where $\Delta V(s) := V(s,q_i) - V(s,q_j) > (<)\, 0$ if $q_i > (<)\, q_j$. Since total demand is, by assumption, fixed, the demand for Hospital j is $D_j = (1 - D_i)$. It follows directly from (7) that the hospital with the higher quality has the higher demand.

2.2 Hospital objectives and payment schemes

We assume that hospitals are financed by a third-party payer through a prospectively regulated price p per treatment and potentially a lump-sum transfer T.[9] Moreover, hospitals are involved in a pay-for-performance (P4P) programme where they are incentivised through bonuses for improved health outcomes (i.e., mortality and readmission) for their patients. We therefore assume that, although quality is observable to patients, it is not directly contractible. The payer can however contract with the provider based on health outcomes measures, though these are only imperfectly related to treatment quality (as illustrated below, health outcomes also depend on the distribution of severity types).[10] We define p_m and p_r as the P4P bonuses associated with a marginal reduction in the hospitals' mortality and readmission rates, respectively.

[9] The lump-sum transfer T is to ensure that the non-negative profit constraint is satisfied (e.g., Brekke et al., 2008, 2011; Ma and Mak, 2015).

[10] We therefore focus on P4P schemes which reward health outcomes, and not process measures of quality. For a model and discussion of the latter see Kristensen et al. (2016).

The objective function of Hospital i is assumed to be given by

$$\Omega_i\left(q_i,q_j\right)=T+pD_i\left(q_i,q_j\right)+\alpha B(q_i)-C\left(D_i\left(q_i,q_j\right),q_i\right)+p_m(\overline{m}-m_i(q_i,q_j))+p_r(\overline{r}-r_i(q_i,q_j)), \quad (8)$$

where m_i and r_i are, respectively, the mortality and readmission rates of Hospital i, which we will define more explicitly in the next section, and where $\overline{m}$ and $\overline{r}$ are highest acceptable mortality and readmission rates. We can interpret these benchmark levels as the mortality/readmission rates associated with minimum (enforceable) quality $\underline{q}$. The cost of supplying hospital treatments is given by the cost function $C\left(\cdot\right)$, which is assumed to be increasing and (weakly) convex in output and convex in quality.[11] We also allow quality and output to be substitutes in costs (e.g., if the unit cost of quality is increasing in quality, i.e., $C\left(D_i,q_i\right)=c(q_i)D_i$ with $c_{q_i}>0$ and $C_{D_iq_i}>0$) or complements (e.g., due to learning by doing effects, so that $C_{D_iq_i}<0$). The parameter $\alpha\in[0,1]$ measures the degree of provider altruism or intrinsic motivation to provide quality, with $B_{q_i}>0$ and $B_{q_iq_i}<0$, in line with previous literature (Ellis and McGuire, 1986; Chalkley and Malcomson, 1998; Brekke et al., 2011).

3 Quality, patient choices and health outcomes

Before analysing optimal hospital behaviour, we will in this section investigate how a unilateral change in treatment quality affects readmission and mortality rates. Since this depends partly on how the patient composition is affected, we start out by investigating how a quality change affects patient choices.

3.1 Demand responses to quality

A unilateral increase in treatment quality by Hospital i implies that more patients will choose this hospital. The corresponding effect on patient composition depends, *inter alia*, on the relative strength of the demand response to quality between high- and low-severity patients. The relationship is clear-cut if $q_i=q_j$ initially, which implies that average patient severity is equal for both hospitals. In this case, a unilateral quality increase by Hospital i will increase average patient

[11] The cost function is likely to be strictly convex in output in health systems where hospitals face capacity constraints, such as the English National Health Service, and, thus, economies of scale are quickly exhausted (e.g., Brekke et al., 2008). The cost function is more likely to be linear in output in health systems with higher capacity such as Germany and the US.

severity at this hospital (and reduce average patient severity at the other hospital) if high-severity patients respond stronger to quality, i.e., if $V_{q_i}(H, q_i) > V_{q_i}(L, q_i)$, and *vice versa*.[12]

Using (5), we can express the difference in responsiveness to quality between the two severity types as follows:[13]

$$\begin{aligned}\frac{\partial (V(H) - V(L))}{\partial q_i} &= -\left[M_{q_i}(H)\,\Delta_U^{SM}(H) - M_{q_i}(L)\,\Delta_U^{SM}(L)\right] \\ &\quad -\left[R_{q_i}(H)\,S(H)\,\Delta_U^{NR}(H) - R_{q_i}(L)\,S(L)\,\Delta_U^{NR}(L)\right] \\ &\quad +\left[S(H)\,\overline{U}_{q_i}(H) - S(L)\,\overline{U}_{q_i}(L)\right].\end{aligned} \tag{9}$$

From (5) we know that higher quality affects expected patient utility through three different channels: (i) reduced mortality risk, (ii) reduced readmission risk, and (iii) higher expected utility for given mortality and readmission probabilities. The difference between severity types in the effect of quality on expected patient utility through each of these three channels is represented by the first, second and third line in (9), respectively.

Regarding the first channel, higher quality will lead to a larger reduction in mortality risk for a high-severity than for a low-severity patient (by A3). On the other hand, the expected utility gain in case of survival is higher for low-severity patients (i.e., $\Delta_U^{SM}(H) < \Delta_U^{SM}(L)$). There are two reasons for this. First, low-severity patients have a lower probability of readmission (by A1). Second, the utility gain of survival, with or without readmission, is higher for low-severity patients (by A7). Thus, whether the effect of quality on expected utility through this first channel is higher or lower for high-severity types is *a priori* ambiguous.

Regarding the second channel, higher quality leads to a larger reduction in (ex post) readmission risk for high-severity patients (by A3). However, whether high-severity patients also experience a larger reduction in the *ex ante* probability of being readmitted is not clear, since high-severity patients have a lower survival probability (by A1). This makes the overall effect through this second channel also ambiguous, regardless of how the ex post utility gain of not being readmitted (Δ_U^{NR}) depends on severity, which is not intuitively obvious.

Finally, regarding the third channel, higher quality leads to a larger utility gain for high-severity than for low-severity patients in both surviving states (by A6) and, although the readmission risk

[12] If $q_i > q_j$, then $V_{q_i}(H, q_i) > V_{q_i}(L, q_i)$ implies $\Delta V(H) > \Delta V(L)$ and therefore $x(H) > x(L)$. *Vice versa*, if $q_i < q_j$, then $V_{q_i}(H, q_i) > V_{q_i}(L, q_i)$ implies $x(H) < x(L)$.

[13] To save notation, we suppress the indication of q_i as an argument in the various functions. This practice is adopted whenever deemed necessary throughout the paper.

is different for the two patient types, it is probably reasonable to assume that $\overline{U}_{q_i}(H) > \overline{U}_{q_i}(L)$. On the other hand, though, the probability of survival, and thereby the probability of experiencing this increase in expected (ex post) utility, is lower for high-severity patients. This implies that, for given survival and readmission probabilities, whether the effect of higher quality on the expected (ex ante) utility gain is higher or lower for high-severity patients, is also *a priori* ambiguous. Thus, in all three channels through which quality affects expected patient utility, the relative magnitude of the effect for high- and low-severity patients is ambiguous.

3.2 Mortality and readmission rates

The mortality rate in Hospital i is given by

$$m_i(q_i, q_j) = \frac{\gamma x(L)}{D_i} M(L, q_i) + \frac{(1-\gamma) x(H)}{D_i} M(H, q_i), \tag{10}$$

and is the sum of the mortality rates for the two patient types weighted by the proportion of high- and low-severity patients in the hospital. Similarly, the readmission rate in Hospital i is given by[14]

$$r_i(q_i, q_j) = \frac{\gamma x(L) S(L, q_i)}{D_i^S} R(L, q_i) + \frac{(1-\gamma) x(H) S(H, q_i)}{D_i^S} R(H, q_i), \tag{11}$$

where $D_i^S := \gamma x(L) S(L, q_i) + (1-\gamma) x(H) S(H, q_i)$ is the survived population of patients in Hospital i.

3.2.1 The relationship between quality and mortality

From (10), the effect of treatment quality on the mortality rate of Hospital i is given by

$$\begin{aligned} \frac{\partial m_i(q_i, q_j)}{\partial q_i} &= \frac{\gamma x(L)}{D_i} M_{q_i}(L) + \frac{(1-\gamma) x(H)}{D_i} M_{q_i}(H) \\ &\quad + \frac{\gamma(1-\gamma)\Delta M(q_i)}{D_i^2} \left[x_{q_i}(H) x(L) - x_{q_i}(L) x(H) \right], \end{aligned} \tag{12}$$

where $\Delta M(q_i) := M(H, q_i) - M(L, q_i) > 0$. The first line in (12) is the reduction in mortality due to a higher quality for a given patient composition. The second line in (12), instead, captures the change in mortality risk due to changes in the patient composition of Hospital i. The sign of this

[14] Notice that it is irrelevant whether or not the patient is readmitted to the same hospital where the initial treatment took place, since readmissions are attributed to the treating hospital regardless of where the patient was readmitted.

patient composition effect depends on the sign of $[x_{q_i}(H)\,x(L) - x_{q_i}(L)\,x(H)]$. More specifically, a higher treatment quality, which leads to higher demand from both patient types, changes the patient composition in the direction of higher (lower) average severity if

$$x_{q_i}(H)\,x(L) > (<)\, x_{q_i}(L)\,x(H) \Leftrightarrow \frac{x_{q_i}(H)}{x(H)} > (<)\, \frac{x_{q_i}(L)}{x(L)}, \tag{13}$$

i.e., if the percentage increase in the number of high-severity patients is higher (lower) than the percentage increase in the number of low-severity patients. This depends, in turn, on the responsiveness to quality for the two patient types and on the initial patient composition at the hospital.

We can distinguish between three different cases:

(i) Suppose that $q_i < q_j$. If high-severity patients are more quality-sensitive, which holds when $\frac{\partial(V(H)-V(L))}{\partial q_i}$ in (9) is positive, so that $x_{q_i}(H) > x_{q_i}(L)$, then $q_i < q_j$ must necessarily imply $x(H) < x(L)$, which in turn implies $x_{q_i}(H)\,x(L) > x_{q_i}(L)\,x(H)$. Conversely, if $x_{q_i}(H) < x_{q_i}(L)$, then $q_i < q_j$ implies $x(H) > x(L)$ and therefore $x_{q_i}(H)\,x(L) < x_{q_i}(L)\,x(H)$. Thus, a marginal quality increase in the *low-quality* hospital implies that the average severity increases (decreases) if high-severity patients are more (less) quality-sensitive than low-severity patients.

(ii) Suppose instead that $q_i > q_j$. If high-severity patients are more (less) quality-sensitive, this implies $x(H) > (<)\, x(L)$, which in either case makes the sign of $x_q(H)\,x(L) - x_q(L)\,x(H)$ indeterminate. Thus, a marginal quality increase in the *high-quality* hospital implies that whichever severity type responds more strongly to quality, this type is already over-represented in the patient population at this hospital, which implies that the effect on patient composition, and thus average severity, is *a priori* ambiguous.

(iii) Finally, if $q_i = q_j$, then $x(L) = x(H)$, which implies that the sign of the second line in (12) is unambiguously determined by the difference between $x_{q_i}(H)$ and $x_{q_i}(L)$. If the demand from high-severity patients responds stronger to quality then the demand from low-severity patients, i.e., if $x_{q_i}(H) > x_{q_i}(L)$, then a unilateral quality improvement by one of the hospitals will change the patient composition at this hospital in the direction of higher average severity.

We summarise the relationship between treatment quality and mortality rate as follows:

$$\frac{\partial m_i(q_i, q_j)}{\partial q_i} = \underbrace{\textit{Direct Quality Effect}}_{<0} + \underbrace{\textit{Patient Composition Effect through Demand}}_{\lessgtr 0}.$$

The overall effect depends on what we dub the *Direct Quality Effect*, which is negative, and the *Patient Composition Effect through Demand*, which is ambiguous as explained above. The ambiguous nature of the latter effect implies that higher quality can, in principle, lead to a higher mortality rate through changes in patient composition. Two necessary (but not sufficient) conditions for the *Patient Composition Effect* to dominate the *Direct Quality Effect* are that (i) demand is sufficiently quality elastic and that (ii) demand responsiveness to quality differs across severity types to a sufficient degree. If either of these conditions is not met, higher quality will always lead to a lower mortality rate.

3.2.2 The relationship between quality and readmission

From (11), the effect of treatment quality on the readmission rate of Hospital i is given by

$$\begin{aligned}\frac{\partial r_i(q_i,q_j)}{\partial q_i} &= \frac{\gamma x(L) S(L)}{D_i^S} R_{q_i}(L) + \frac{(1-\gamma) x(H) S(H)}{D_i^S} R_{q_i}(H) \\ &+ \frac{\gamma(1-\gamma) S(L) S(H) \Delta R(q_i)}{\left(D_i^S\right)^2} \left[x_{q_i}(H) x(L) - x_{q_i}(L) x(H)\right] \\ &+ \frac{\gamma(1-\gamma) x(L) x(H) \Delta R(q_i)}{\left(D_i^S\right)^2} \left[S_{q_i}(H) S(L) - S_{q_i}(L) S(H)\right],\end{aligned} \tag{14}$$

where $\Delta R(q_i) := R(H,q_i) - R(L,q_i) > 0$. The first line in (14) is the reduction in readmissions due to higher quality for a given patient composition and a given mortality risk. The second and third line in (14), on the other hand, are the effects on the readmission rate due to changes in patient composition in Hospital i. The second line captures the *Patient Composition Effect through Demand,* analogous to the one already discussed for mortality, and the sign of this effect is determined by the condition in (13).

The third line in (14) captures a different kind of composition effect, which we dub the *Patient Composition Effect through Mortality*. This is the worsening of the risk distribution of survived patients induced by a lower mortality rate. More specifically, since a quality increase leads to a larger improvement in survival probability for high-severity than for low-severity patients, and since high severity patients have a higher risk of readmission, this particular change in the patient composition unambiguously leads to an increase in the readmission rate.[15]

[15] Since $S_{q_i}(H) > S_{q_i}(L)$ and $S(L) > S(H)$, it follows that $S_{q_i}(H) S(L) > S_{q_i}(L) S(H)$.

We summarise the relationship between treatment quality and readmission rate as follows:

$$\frac{\partial r_i(q_i, q_j)}{\partial q_i} = \underset{\substack{\textit{Quality Effect} \\ < 0}}{\textit{Direct}} + \underset{\substack{\textit{through Demand} \\ \lessgtr 0}}{\textit{Patient Composition Effect}} + \underset{\substack{\textit{through Mortality} \\ > 0}}{\textit{Patient Composition Effect}} .$$

As for the case of mortality, the overall effect of higher quality on the readmission rate is *a priori* indeterminate and depends on whether indirect effects through changes in patient composition outweigh the direct effect of higher quality. Compared to the mortality rate, however, there is an additional source of bias in the readmission rate that goes in the opposite direction of the direct quality effect, namely that higher quality implies higher readmission rates because of the *Patient Composition Effect through Mortality*. Thus, even if the demand responsiveness to quality is very low, which would make the *Patient Composition Effect through Demand* negligible, a positive relationship between treatment quality and readmission rates might nevertheless appear because of the *Patient Composition Effect through Mortality*, which dominates the *Direct Quality Effect* if quality has a sufficiently large effect on the mortality risk relative to the direct effect on the probability of readmission.

4 Hospital competition and quality under P4P

In this section we analyse how P4P bonuses related to mortality and readmission affect equilibrium quality choices, and how equilibrium mortality and readmission rates, in turn, are affected. We consider a symmetric simultaneous-move game in which each hospital independently chooses its treatment quality to maximise the objective function given by (8).

4.1 Nash equilibrium

The first order condition for Hospital i's maximisation problem is given by[16]

$$\begin{aligned}\frac{\partial \Omega_i\left(q_i,q_j\right)}{\partial q_i} &= \left(p-\frac{\partial C\left(D_i\left(q_i,q_j\right),q_i\right)}{\partial D_i}\right)\frac{\partial D_i\left(q_i,q_j\right)}{\partial q_i}+\alpha\frac{\partial B\left(q_i\right)}{\partial q_i}-\frac{\partial C\left(D_i\left(q_i,q_j\right),q_i\right)}{\partial q_i} \\ &\quad -p_m\frac{\partial m_i\left(q_i,q_j\right)}{\partial q_i}-p_r\frac{\partial r_i\left(q_i,q_j\right)}{\partial q_i}=0\end{aligned} \tag{15}$$

In the absence of P4P, the optimal quality choice is determined by the first line of (15), where the marginal monetary and non-monetary benefits are optimally weighed against the marginal cost of quality provision. However, the P4P programme introduces two additional terms which can be either positive or negative, depending on the relationship between quality and health outcomes (mortality and readmission) as analysed in the previous section. If higher quality leads to a reduction (increase) in mortality and readmission rates, the P4P programme increases the marginal benefit (cost) of quality provision.

Applying symmetry, i.e., $q_i=q_j=q^*$, which implies $D_i=D_j=\frac{1}{2}$ and $x\left(L\right)=x\left(H\right)=\frac{1}{2}$, the Nash equilibrium quality, q^*, is implicitly given by[17]

$$\widehat{\Omega}:=\frac{\partial \Omega_i\left(q^*\right)}{\partial q_i}=\left(p-\frac{\partial C\left(1/2,q^*\right)}{\partial D_i}\right)\frac{\partial D_i\left(q^*\right)}{\partial q_i}+\alpha\frac{\partial B\left(q^*\right)}{\partial q_i}-\frac{\partial C\left(q^*\right)}{\partial q_i}-p_m\frac{\partial m_i\left(q^*\right)}{\partial q_i}-p_r\frac{\partial r_i\left(q^*\right)}{\partial q_i}=0 \tag{16}$$

where

$$\frac{\partial m_i\left(q^*\right)}{\partial q_i}=\gamma M_{q_i}\left(L\right)+\left(1-\gamma\right)M_{q_i}\left(H\right)+2\gamma\left(1-\gamma\right)\Delta M(q^*)\left[x_{q_i}\left(H\right)-x_{q_i}\left(L\right)\right] \tag{17}$$

[16] The second-order condition,

$$\frac{\partial^2\Omega}{\partial q_i^2}=\left(p-\frac{\partial C}{\partial D_i}\right)\frac{\partial^2 D_i}{\partial q_i^2}+\alpha\frac{\partial^2 B}{\partial q_i^2}-2\frac{\partial^2 C}{\partial D_i\partial q_i}\frac{\partial D_i}{\partial q_i}-\frac{\partial^2 C}{\partial D_i^2}\left(\frac{\partial D_i}{\partial q_i}\right)^2-\frac{\partial^2 C}{\partial q_i^2}-p_m\frac{\partial^2 m_i}{\partial q_i^2}-p_r\frac{\partial^2 r_i}{\partial q_i^2}<0,$$

is satisfied if the cost function is sufficiently convex in quality.

[17] For simplicity, we use the notational shorthand $f\left(x^*\right)$ to denote $f\left(x_i,x_j\right)|_{x_i=x_j=x^*}$.

and

$$\begin{aligned}\frac{\partial r_i(q^*)}{\partial q_i} &= \frac{\gamma S(L)}{2D_i^S} R_{q_i}(L) + \frac{(1-\gamma) S(H)}{2D_i^S} R_{q_i}(H) \\ &+ \frac{\gamma(1-\gamma) S(L) S(H) \Delta R(q^*)}{2\left(D_i^S\right)^2} \left[x_{q_i}(H) - x_{q_i}(L)\right] \\ &+ \frac{\gamma(1-\gamma) \Delta R(q^*)}{\left(2D_i^S\right)^2} \left[S_{q_i}(H) S(L) - S_{q_i}(L) S(H)\right].\end{aligned} \tag{18}$$

The signs of (17) and (18) are generally ambiguous, as previously explained. However, at the symmetric Nash equilibrium, the sign of the *Patient Composition Effect through Demand* depends unambiguously on which patient type is more responsive to changes in quality. More specifically, if demand from high-severity patients is more (less) quality-responsive than demand from low-severity patients, i.e., if $x_q(H) > (<) x_q(L)$, then the *Patient Composition Effect through Demand* is unambiguously positive (negative).

4.2 The effects of P4P on equilibrium quality

We can assess the effects of the P4P bonuses on the hospitals' quality provision by considering marginal changes in p_m and p_r at the Nash equilibrium. By totally differentiating (16) with respect to p_m, p_r, q_i and q_j, and evaluating the subsequent expressions at $q_i = q_j = q^*$, we derive

$$\frac{\partial q^*}{\partial p_m} = -\frac{\partial \widehat{\Omega}/\partial p_m}{\partial \widehat{\Omega}/\partial q_i} = \frac{\partial m_i(q^*)/\partial q_i}{\partial \widehat{\Omega}/\partial q_i} > (<) 0 \quad if \quad \frac{\partial m_i(q^*)}{\partial q_i} < (>) 0, \tag{19}$$

and

$$\frac{\partial q^*}{\partial p_r} = -\frac{\partial \widehat{\Omega}/\partial p_r}{\partial \widehat{\Omega}/\partial q_i} = \frac{\partial r_i(q^*)/\partial q_i}{\partial \widehat{\Omega}/\partial q_i} > (<) 0 \quad if \quad \frac{\partial r_i(q^*)}{\partial q_i} < (>) 0, \tag{20}$$

where[18]

$$\frac{\partial \widehat{\Omega}}{\partial q_i} = \frac{\partial^2 \Omega_i(q^*)}{\partial q_i^2} + \frac{\partial^2 \Omega_i(q^*)}{\partial q_i \partial q_j} < 0.$$

[18] The negative sign of $\frac{\partial \widehat{\Omega}}{\partial q_i}$ is ensured by the second-order condition of the hospitals' optimisation problem, $\frac{\partial^2 \Omega_i(q^*)}{\partial q_i^2} < 0$, and the equilibrium stability condition

$$\frac{\partial^2 \Omega_i(q^*)}{\partial q_i^2} \frac{\partial^2 \Omega_j(q^*)}{\partial q_j^2} - \frac{\partial^2 \Omega_i(q^*)}{\partial q_j \partial q_i} \frac{\partial^2 \Omega_j(q^*)}{\partial q_i \partial q_j} > 0.$$

Since, at the symmetric equilibrium, $\frac{\partial^2 \Omega_i(q^*)}{\partial q_i^2} = \frac{\partial^2 \Omega_j(q^*)}{\partial q_j^2}$ and $\frac{\partial^2 \Omega_i(q^*)}{\partial q_j \partial q_i} = \frac{\partial^2 \Omega_j(q^*)}{\partial q_i \partial q_j}$, the stability condition implies $\left|\frac{\partial^2 \Omega_i(q^*)}{\partial q_i^2}\right| > \left|\frac{\partial^2 \Omega_i(q^*)}{\partial q_i \partial q_j}\right|$.

Thus, the effect on equilibrium quality of an increase in the bonus related to mortality (or readmission) is determined by the relationship between quality and the mortality (or readmission) rate. If a unilateral quality increase by Hospital i leads to higher mortality (or readmission) rate at that hospital, a higher bonus related to mortality (or readmission) will have a counterproductive effect and lead to lower quality provision in equilibrium.

The relationship between quality and mortality/readmission rates depends, in turn, crucially on the demand-responsiveness to quality of high-severity versus low-severity patients. Let us consider each of the two possible scenarios.

Proposition 1 *Suppose that the demand-responsiveness to quality is stronger for low-severity than for high-severity patients, $x_q(L) > x_q(H)$:*

(i) A larger bonus related to mortality will unambiguously lead to higher quality provision.

(ii) A larger bonus related to readmission has an a priori ambiguous effect on quality provision:

(ii.a) Quality increases if one of the following two conditions are met: (1) the effect of quality on survival probabilities, $S_q(s)$, is sufficiently low; (2) the difference in readmission probability between high- and low-severity patients, ΔR, is sufficiently low.

(ii.b) Quality decreases if both of the following conditions are met: (1) the direct effect of quality on readmission risks, $R_q(s)$, is sufficiently low; (2) the difference in demand responsiveness to quality between low- and high-severity patients, $x_q(L) - x_q(H)$, is sufficiently low.

If demand from low-severity patients is more quality-responsive than the demand from high-severity patients, the *Patient Composition Effect through Demand* is negative, which yields an unambiguous relationship between quality and mortality rates. A quality improvement at Hospital i will directly reduce the mortality risk for each patient at this hospital. In addition, the hospital will attract new patients in a way that changes the patient composition at the hospital in the direction of lower average patient severity. Both effects contribute to a lower mortality rate at the hospital. In this case, both hospitals will respond to a larger bonus related to mortality by improving their treatment quality.

The relationship is less clear-cut for readmissions. Although the *Patient Composition Effect through Demand* is negative and therefore contributes towards a negative relationship between quality and readmission rates, the *Patient Composition Effect through Mortality* is positive and therefore pulls in the opposite direction. The latter effect is sufficiently small to be dominated by

the two other effects if $S_q(s)$ is sufficiently low, which implies that a quality improvement has a sufficiently small impact on the risk distribution of the survived patients, or if ΔR is sufficiently small, which implies that the worsened risk distribution of survived patients has a sufficiently small impact on the readmission rate. If this is the case, a larger bonus related to readmission will also have the desired effect of stimulating quality provision.

The scope for counterproductive effects of P4P bonuses is larger in the other scenario, in which a unilateral quality improvement by one of the hospitals attracts a disproportionately larger share high-severity patients:

Proposition 2 *Suppose that the demand-responsiveness to quality is stronger for high-severity than for low-severity patients, $x_q(H) > x_q(L)$:*

(i) A larger bonus related to mortality leads to lower (higher) quality provision if the difference in mortality risk between high- and low-severity patients, ΔM, is sufficiently large (small) relative to the direct effect of quality on the mortality risk of each patient, $M_q(s)$.

(ii) A larger bonus related to readmission leads to lower (higher) quality provision if the difference in readmission risk between high- and low-severity patients, ΔR, is sufficiently large (small) relative to the direct effect of quality on the readmission risk of each patient, $R_q(s)$.

In the scenario characterised by $x_q(H) > x_q(L)$, the (indirect) patient composition effects of a quality increase (through demand or mortality) go in the opposite direction of the direct effects of quality on mortality and readmissions, respectively. The overall effect is therefore determined by the relative size of the direct and indirect effects. As is evident from (19) and (20), the latter effects depend on the difference between high-severity and low-severity patients with respect to mortality and readmission probabilities, respectively. If these differences are sufficiently large, the overall effects are driven by the patient composition effects, implying that a unilateral quality increase by one of the hospitals will increase mortality and readmission rates at this hospital. In this case, a larger bonus related to mortality or readmission will lead to lower quality provision by both hospitals in equilibrium.

The previous analysis suggests that the scope for a counterproductive P4P policy is larger when bonuses are related to readmission rather than mortality. The reason is the *Patient Composition Effect through Mortality*, which comes into effect only when bonuses are related to readmissions and which contributes in the direction of a positive relationship between unilateral quality improvements

and readmission rates. The next proposition identifies a case in which the two different legs of a P4P programme have opposite effects on the hospitals' incentives for quality provision.

Proposition 3 *Suppose that the difference in demand-responsiveness to quality between high- and low-severity patients is negligible, $x_q(H) \approx x_q(L)$. In this case, a larger bonus related to mortality will always increase quality, whereas a larger bonus related to readmissions will reduce quality if the effects of quality on survival probabilities, $S_q(s)$, are sufficiently large relative to the effects of quality on readmission probabilities, $R_q(s)$.*

4.3 The effects of P4P on equilibrium mortality and readmission rates

In the previous subsection we have shown that P4P bonuses related to mortality and readmission can have counterproductive effects in the sense that equilibrium treatment quality goes down. But what are the corresponding effects on equilibrium mortality and readmission rates? In the symmetric Nash equilibrium, mortality and readmission rates are given by, respectively

$$m(q^*) = \gamma M(L, q^*) + (1-\gamma) M(H, q^*) \tag{21}$$

and

$$r(q^*) = \frac{\gamma S(L, q^*)}{D^S(q^*)} R(L, q^*) + \frac{(1-\gamma) S(H, q^*)}{D^S(q^*)} R(H, q^*), \tag{22}$$

where $D^S(q^*) = \gamma S(L, q^*) + (1-\gamma) S(H, q^*)$. From (21)-(22) we derive:

$$\frac{\partial m(q^*)}{\partial q^*} = \gamma M_q(L) + (1-\gamma) M_q(H) < 0 \tag{23}$$

and

$$\begin{aligned} \frac{\partial r(q^*)}{\partial q^*} &= \frac{\gamma S(L) R_q(L) + (1-\gamma) S(H) R_q(H)}{D^S} \\ &\quad + \frac{\gamma(1-\gamma)\Delta R}{(D^S)^2} \left[S_q(H) S(L) - S_q(L) S(H)\right] \gtrless 0. \end{aligned} \tag{24}$$

Higher equilibrium treatment quality always reduces the equilibrium mortality rates. The reason is that, in a symmetric equilibrium, where each hospital has an equal amount of high- and low-severity patients, and therefore equal average patient severity, the *Patient Composition Effect through Demand* is neutralised. This leaves only the direct effect of quality on the mortality

risk of each individual patient, which establishes an unambiguously negative relationship between treatment quality and mortality rates in equilibrium.

The relationship between equilibrium quality and readmission rates is ambiguous, though. Although the *Patient Composition Effect through Demand* is neutralised in equilibrium, the *Patient Composition Effect through Mortality* is still present. Higher equilibrium treatment quality reduces the readmission probability for a given patient, but, because of an increase in the survival probability, the average severity level in the survived patient population is worse, which – all else equal – leads to a higher frequency of readmissions.

The implications for the effects of P4P on equilibrium mortality and readmission rates follow straightforwardly:

Proposition 4 *(i) If a larger bonus related to mortality leads to higher (lower) equilibrium treatment quality, the equilibrium mortality rates go down (up). (ii) Regardless of the effect of a larger bonus related to readmission on equilibrium quality, the resulting effect on equilibrium readmission rates is generally ambiguous.*

These results have implications for empirical evaluations of P4P programmes. The unambiguously negative relationship between equilibrium quality and equilibrium mortality rates implies that, keeping other factors constant, if the introduction of a P4P scheme related to mortality leads to lower mortality rates, this must be caused by an improvement in treatment quality. In other words, the effects of a mortality-based P4P scheme on quality provision can be reliably assessed by measuring the effects on observed mortality rates.

However, this is not the case for P4P schemes related to readmission rates. Even if the introduction of such a scheme leads to a reduction in readmission rates, this does not necessarily mean that treatment quality has improved. On the contrary, a reduction in readmission rates might be explained by higher mortality rates (via the *Patient Composition Effect through Mortality*) caused by *lower* treatment quality. Thus, the effects of a readmission-based P4P scheme on quality provision *cannot* be reliably assessed by measuring the effects only on readmission rates. This information needs to be combined with information on mortality rate effects in order make reliable inferences about whether or not the scheme has been successful in terms of improving treatment quality.

The first part of Proposition 4 also illustrates that the hospitals are potentially caught in a Prisoners' Dilemma if they are subject to P4P programmes with bonuses related to mortality.

Suppose that the *Patient Composition Effect through Demand* dominates the *Direct Quality Effect*, such that higher bonuses related to mortality lead to less quality provision in equilibrium. In this case, each hospital has an incentive to reduce its quality provision in order to improve its patient composition and thereby receive a higher bonus. However, since both hospitals have the same incentives, which exactly cancel each other in the symmetric Nash equilibrium, the resulting outcome is a higher mortality rate with a correspondingly lower bonus received by each hospital.

4.4 Intensity of competition

Under prospectively regulated prices per treatment, hospitals' incentives to provide treatment quality depends on the intensity of competition, which, in the current modelling framework, is usually measured by the transportation cost parameter t. A reduction in t, which makes provider-specific demand more elastic and therefore intensifies competition, will induce both hospitals to increase quality provision if the treatment price, p, is above marginal costs. Thus, if this condition is satisfied, policy makers can stimulate quality provision by introducing measures that increase the degree of patient choice and therefore intensify the degree of competition in the market. However, the effects of such measures can be influenced by the presence of P4P schemes.

To see this, consider the effects of a marginal reduction in t on the equilibrium quality given by (16). By totally differentiating (16) with respect to t, q_i and q_j, and evaluating the subsequent expressions at $q_i = q_j = q^*$, we derive

$$\frac{\partial q^*}{\partial t} = -\frac{\partial \widehat{\Omega}/\partial t}{\partial \widehat{\Omega}/\partial q_i}, \tag{25}$$

where

$$\begin{aligned} \frac{\partial \widehat{\Omega}}{\partial t} &= -\frac{1}{t}\left(p - \frac{\partial C\left(1/2, q^*\right)}{\partial D_i}\right)\frac{\partial D_i\left(q^*\right)}{\partial q_i} \\ &\quad + \frac{\gamma\left(1-\gamma\right)}{t}\left[2p_m \Delta M\left(q^*\right) + \frac{S\left(L\right)S\left(H\right)}{2\left(D_i^S\right)^2} p_r \Delta R\left(q^*\right)\right]\left(x_{q_i}\left(H\right) - x_{q_i}\left(L\right)\right). \end{aligned} \tag{26}$$

In the absence of P4P, the second term in (26) vanishes and only the standard competition effect remains. However, the presence of P4P introduces an indirect effect through patient composition. More specifically, a reduction in t magnifies the *Patient Composition Effect through Demand*, which in turn affects the hospitals financially through P4P. As is evident from (26), the sign of this indirect

effect depends on whether demand from high-severity patients is more or less responsive to quality changes than demand from low-severity patients. If $x_{q_i}(H) > x_{q_i}(L)$, each hospital's incentive to attract patients by increasing quality will be dampened by the financial costs (through P4P) of increased mortality and readmission rates, because each hospital, by unilaterally increasing quality, will attract a disproportionately large share of high-severity patients. However, this incentive is *strengthened* if $x_{q_i}(H) < x_{q_i}(L)$. We summarise as follows:

Proposition 5 *If demand-responsiveness to quality is stronger (weaker) for high-severity than for low-severity patients, then the presence of P4P will dampen (reinforce) the positive effect of competition on quality provision.*

The above analysis also suggests that the effects of P4P on quality provision might depend on the degree of competition in the market. As explained above, a reduction of t increases the magnitude of the *Patient Composition Effect through Demand*, which – all else equal – increases the scope for counterproductive effects of P4P. This suggests that, in terms of stimulating quality provision, P4P is more likely to succeed (or is likely to succeed to a larger extent) in markets where the intensity of competition is more limited.

5 Welfare analysis

In this section we provide a welfare analysis. Our analysis is positive, rather than normative, and investigates how P4P affects welfare. We define welfare in two different ways. We start out by defining it as the difference between patient benefits and provider costs. We then compare it to the perspective of the funder, whose objective function is defined as the difference between patient benefits and the transfers to the providers. We also compare P4P schemes with bonuses versus penalties. Finally, we discuss P4P schemes which are budget neutral, so that (an increase in) a bonus is accompanied by a reduction in the basic DRG tariff, which is in line with some schemes and countries which have a fixed budget for healthcare.

Throughout this section, we consider the welfare effects of a marginal increase in P4P bonuses (or penalties). These effects can either, if evaluated at $p_m = p_r = 0$, be interpreted as the effects of *introducing* a (small) P4P bonus related to mortality or readmission rates, or, alternatively, they can be interpreted as the effects of making an already existing P4P scheme more high-powered along either of the two dimensions (mortality and readmissions).

Suppose that welfare W is defined for each of the two identical providers as the difference between patient benefits and the cost to the provider. More formally,

$$W(q^*(p_m, p_r)) = \overline{V}\left(q^*(p_m, p_r)\right) - C\left(1/2, q^*(p_m, p_r)\right), \tag{27}$$

where $\overline{V}(q^*) := \gamma V\left(L, q^*\right) + (1-\gamma)V\left(H, q^*\right)$ is the total patient benefit across severity types. The marginal effects of higher P4P bonuses are then given by

$$\frac{\partial W}{\partial p_m} = \left[\overline{V}_q\left(q^*\right) - C_q\left(1/2, q^*\right)\right]\frac{\partial q^*}{\partial p_m}, \tag{28}$$

$$\frac{\partial W}{\partial p_r} = \left[\overline{V}_q\left(q^*\right) - C_q\left(1/2, q^*\right)\right]\frac{\partial q^*}{\partial p_r}. \tag{29}$$

In both cases, welfare increases if (i) the marginal patient benefit of quality is higher than the marginal cost of quality provision, and (ii) a (larger) P4P bonus increases quality. Welfare instead reduces if (i) holds but the P4P bonus reduces quality, or if (ii) holds but the marginal patient benefit is below the marginal cost of quality provision. Notice that under (i) and (ii) a marginal increase in P4P bonuses is always welfare improving regardless of the effectiveness of P4P in increasing quality, though the welfare gain is amplified when the P4P scheme is more effective in increasing quality.

5.1 The purchaser perspective

We now take the purchaser (funder) perspective and assume that the utility of the purchaser, $\widetilde{W}$, is defined as the difference between patient benefit and the *transfer*, rather than the cost, to the provider:

$$\widetilde{W}(q^*(p_m, p_r)) = \overline{V}\left(q^*(p_m, p_r)\right) - p_m\left[\overline{m} - m^*(q^*(p_m, p_r))\right] - p_r(\overline{r} - r^*(q^*(p_m, p_r))) - \frac{p}{2}. \tag{30}$$

The effects of a marginal increase in P4P bonuses on purchaser utility are given by

$$\frac{\partial \widetilde{W}}{\partial p_m} = \left[\overline{V}_q\left(q^*\right) + p_m\frac{\partial m^*}{\partial q^*} + p_r\frac{\partial r^*}{\partial q^*}\right]\frac{\partial q^*}{\partial p_m} - \left[\overline{m} - m^*(q^*)\right], \tag{31}$$

$$\frac{\partial \widetilde{W}}{\partial p_r} = \left[\overline{V}_q\left(q^*\right) + p_m\frac{\partial m^*}{\partial q^*} + p_r\frac{\partial r^*}{\partial q^*}\right]\frac{\partial q^*}{\partial p_r} - \left[\overline{r} - r^*(q^*)\right]. \tag{32}$$

For a given quality level (and thus for given mortality and readmission rates), a higher bonus is obviously costly for the purchaser, as indicated by the last term in (31) and (32). Whether these

costs are justified from the purchaser's perspective depends on both the direction and the strength of the effect of P4P bonuses on quality provision, as indicated by the first term in (31) and (32).

Consider first the effects of introducing a small P4P bonus (i.e., evaluating (31)-(32) at $p_m = p_r = 0$). A *necessary* condition for this to increase purchaser utility is that the introduction of such a bonus leads to higher quality provision. A *sufficient* condition for purchaser utility to increase is that the increase in patient benefits (from higher quality) outweighs the cost of introducing the P4P scheme. These effects are also present if we consider making an existing P4P scheme more high-powered. In addition, though, notice that a positive quality response to a more high-powered P4P scheme ($\partial q^*/\partial p_m > 0$ or $\partial q^*/\partial p_r > 0$) will amplify the cost of P4P related to mortality (since $\partial m^*/\partial q^* < 0$) while the indirect effect on readmission bonuses is a priori ambiguous (since the sign of $\partial r^*/\partial q^*$ is indeterminate).

Finally, it is worth noticing that the providers are always better off with higher bonuses, since (by the Envelope Theorem) $\partial\Omega\left(q^*\right)/\partial p_m = [\overline{m} - m^*(q^*)] > 0$ and $\partial\Omega\left(q^*\right)/\partial p_r = [\overline{r} - r^*(q^*)] > 0$.

5.1.1 Penalties versus bonuses

Some P4P schemes introduce *penalties* as opposed to *bonuses*. This case can be obtained analytically by setting $\overline{m} = \overline{r} = 0$. The results in Section 4 remain unchanged, and the introduction of a penalty has the same effect on quality as the introduction of a bonus. The only difference is that the prices p_m and p_r are now interpreted as penalties. This gives the following purchaser utility function:

$$\widehat{W}(q^*(p_m, p_r)) = \overline{V}\left(q^*(p_m, p_r)\right) + p_m m^*(q^*(p_m, p_r)) + p_r r^*(q^*(p_m, p_r)) - \frac{p}{2}. \tag{33}$$

The effects of (a marginal increase in) P4P penalties on purchaser utility are given by

$$\frac{\partial \widehat{W}}{\partial p_m} = \frac{\partial \widetilde{W}}{\partial p_m} + \overline{m}, \tag{34}$$

$$\frac{\partial \widehat{W}}{\partial p_r} = \frac{\partial \widetilde{W}}{\partial p_r} + \overline{r}. \tag{35}$$

Thus, from a purchaser's perspective, penalties are more likely to be welfare improving than bonuses. For example, if we consider the introduction of a small penalty, a sufficient condition for this to increase purchaser utility is that quality is not negatively affected by the penalty. Even if the

penalty introduction has no effect on equilibrium quality provision ($\partial q^*/\partial p_m = 0$ or $\partial q^*/p_r = 0$), purchaser utility nevertheless increases because of the revenues generated by the P4P penalties. However, when basing the P4P scheme on penalties instead of bonuses, the purchaser needs to ensure that the providers' participation constraints are satisfied, since $\partial\Omega\left(q^*\right)/\partial p_m = -m^*(q^*) < 0$ and $\partial\Omega\left(q^*\right)/\partial p_r = -r^*(q^*) < 0$.

5.1.2 Welfare versus purchaser perspective

In this sub-section we ask whether the effect of the P4P scheme is higher under the *welfare* perspective or the *purchaser* perspective. The distinction between the two perspectives is potentially important for empirical analyses which evaluate the benefits and costs of P4P schemes, where the *cost* for the purchaser is given by the transfer to the provider. Therefore, we would like to know which of the two perspectives provides a stricter test for evaluating the effectiveness of a given scheme.

We only provide the comparison in relation to mortality, since a similar analysis holds for readmission rates. In case of P4P *bonuses*, the difference between the marginal effects on welfare and purchaser utility is given by

$$\frac{\partial W}{\partial p_m} - \frac{\partial \widetilde{W}}{\partial p_m} = [\overline{m} - m^*(q^*)] - \left[p_m \frac{\partial m^*}{\partial q^*} + p_r \frac{\partial r^*}{\partial q^*}\right] \frac{\partial q^*}{\partial p_m} - C_q\left(1/2, q^*\right) \frac{\partial q^*}{\partial p_m}. \tag{36}$$

Using (16), this difference can alternatively be re-written as

$$\frac{\partial W}{\partial p_m} - \frac{\partial \widetilde{W}}{\partial p_m} = [\overline{m} - m^*(q^*)] - \left[\left(p - C_D\right) D_q\left(q^*\right) + \alpha B_q\left(q^*\right)\right] \frac{\partial q^*}{\partial p_m}. \tag{37}$$

Patient benefits from P4P are taken into account under both the welfare and the purchaser perspective. However, the two perspectives differ with respect to the perceived costs of the P4P scheme. The first two terms in (36) reflect the marginal cost of a higher P4P bonus under the purchaser perspective. If we consider the introduction of a small bonus, this effect is unambiguously positive, thus contributing to making the P4P scheme more effective from a welfare perspective than from a purchaser perspective. However, the third term in (36) pulls in the opposite direction. This term reflects the marginal cost of the P4P bonus from the welfare perspective, which takes the cost of quality provision into account. The size of this latter effect depends on the providers' incentives for quality provision. From (37) we see that, if demand responsiveness to quality and the degree

of provider altruism are both sufficiently low, the marginal cost of P4P from a welfare perspective is smaller than the corresponding marginal cost from a purchaser perspective. In this case, the introduction of a P4P scheme is more likely to be deemed successful from a welfare perspective than from a purchaser perspective.

The above analysis is markedly different if we consider a P4P scheme with penalties instead of bonuses. In this case, the expression equivalent to (37) is given by

$$\frac{\partial W}{\partial p_m} - \frac{\partial \widehat{W}}{\partial p_m} = -m^*(q^*) - \left[(p - C_D)\, D_q\left(q^*\right) + \alpha B_q\left(q^*\right)\right] \frac{\partial q^*}{\partial p_m}. \tag{38}$$

Since, from a purchaser perspective, a P4P scheme with penalties brings revenues instead of costs, a sufficient (but not necessary) condition for the marginal gains of P4P to be higher from a purchaser than from a welfare perspective, is that a higher penalty does not lead to lower quality provision in equilibrium.

5.2 Budget neutral P4P scheme

Let us once more return to the case of P4P bonuses, but suppose that the purchaser is resource constrained so that any payment to the providers from the P4P scheme is financed through a reduction in the DRG tariff. This is in line with some existing schemes implemented in recent years. For example, the English NHS has introduced several *Best Practice Tariffs* which reward higher quality within a given DRG and are combined with a reduction of the basic DRG tariff. To keep the exposition simple, we set $p_r = 0$. Then the following constraint holds:

$$\frac{p}{2} + p_m[\overline{m} - m^*(q^*(p_m, p))] = F, \tag{39}$$

where F is a fixed budget. By totally differentiating (39) we obtain

$$\frac{dp}{dp_m} = -\frac{(\overline{m} - m^*) - p_m\left(\partial m^*/\partial q^*\right)\left(\partial q^*/\partial p_m\right)}{1/2 - p_m\left(\partial m^*/\partial q^*\right)\left(\partial q^*/\partial p\right)}. \tag{40}$$

We can write welfare as

$$W(q^*(p_m, p(p_m))) = \overline{V}\left(q^*(p_m, p(p_m))\right) - C\left(1/2, q^*(p_m, p(p_m))\right). \tag{41}$$

The marginal effect of P4P on welfare is then given by

$$\frac{\partial W}{\partial p_m} = \left[\overline{V}_q\left(q^*\right) - C_q\left(1/2, q^*\right)\right]\left(\frac{\partial q^*}{\partial p_m} + \frac{\partial q^*}{\partial p}\frac{dp}{dp_m}\right), \tag{42}$$

where

$$\frac{\partial q^*}{\partial p_m} + \frac{\partial q^*}{\partial p}\frac{dp}{dp_m} = \frac{(1/2)(\partial q^*/\partial p_m) - (\overline{m} - m^*)\,\partial q^*/\partial p}{1/2 - p_m\left(\partial m^*/\partial q^*\right)\left(\partial q^*/\partial p\right)}. \tag{43}$$

Suppose that equilibrium quality is suboptimally low (i.e., $\overline{V}_q\left(q^*\right) > C_q\left(1/2, q^*\right)$), and suppose that the introduction of P4P stimulates quality provision (i.e., $\partial q^*/\partial p_m > 0$). In this case, nevertheless, the marginal welfare effect of P4P is in principle indeterminate. On the one hand, the P4P bonus induces an increase in quality. On the other hand, the higher bonus involves a reduction in the DRG tariff, which in turn implies a reduction in quality. Recall that

$$\frac{\partial q^*}{\partial p_m} = \frac{\partial m_i\left(q^*\right)/\partial q_i}{\partial \widehat{\Omega}/\partial q_i} \text{ and } \frac{\partial q^*}{\partial p} = -\frac{\partial D_i\left(q^*\right)/\partial q_i}{\partial \widehat{\Omega}/\partial q_i}. \tag{44}$$

By substitution, we obtain

$$\frac{\partial q^*}{\partial p_m} + \frac{\partial q^*}{\partial p}\frac{dp}{dp_m} = \frac{-\left(\partial m_i\left(q^*\right)/\partial q_i\right) - 2\left(\overline{m} - m^*\right)\left(\partial D_i\left(q^*\right)/\partial q_i\right)}{-\left(\partial \widehat{\Omega}/\partial q_i + 2p_m\left(\partial m^*/\partial q^*\right)\left(\partial D_i\left(q^*\right)/\partial q_i\right)\right)} \tag{45}$$

The denominator in (45) is positive, so the sign of (45) is given by the sign of the numerator. The first term in the numerator reflects the (assumed positive) effect of a higher bonus on equilibrium quality provision, which depends on the effect of a unilateral quality increase on the mortality rate. The second term reflects the (negative) effect of the DRG price adjustment on equilibrium quality provision, which depends on the demand responsiveness to quality. Intuitively, as long as the demand responsiveness to quality is sufficiently low (which is a potential reason for introducing P4P in the first place), then the introduction of P4P is welfare improving despite the weaker incentive for quality provision induced by the reduction in the DRG tariff.

Finally, we can also compare the results under the *purchaser* perspective. The purchaser objective is now simply defined by $\widetilde{W}(q^*(p_m, p_r)) = \overline{V}\left(q^*(p_m, p_r)\right) - F$. Therefore, any P4P scheme which improves quality increases the purchaser's utility.

6 Extension: public reporting

In our main analysis, patient choice is assumed to be based, in part, on the quality of treatment offered by the hospitals. The underlying assumption is that quality is observable but not directly contractible, implying that hospital payments cannot be made directly contingent on quality, but must be based on some proxies for treatment quality, such as mortality and readmission rates.

In this section we relax the assumption of observability (e.g., Gravelle and Sivey, 2010) and assume that quality is neither observable nor directly contractible. Patients only have access to a publicly reported measure of quality for each hospital, denoted by ρ_i, where a higher value of ρ_i indicates a higher level of treatment quality at Hospital i.[19] To facilitate a comparison with the results from our main model, suppose that the indifferent patient of type s is located at

$$x(s) = \frac{1}{2} + \frac{\theta^s \left(\rho_i - \rho_j\right)}{2t}, \tag{46}$$

where the parameter θ^s measures how strongly the demand from patients of severity type s responds to changes in the reported measures of quality. For the sake of brevity of analysis, we will here only consider the case where demand responses are stronger from high-severity patients; i.e., $\theta^H > \theta^L$. Maintaining all other assumptions from the main model, total demand for Hospital i is then given by

$$D_i(q_i, q_j) = \frac{1}{2} + \frac{\gamma\theta^L + (1-\gamma)\theta^H}{2t}\left(\rho_i - \rho_j\right). \tag{47}$$

In the following we will consider the cases where patient choice is determined either by reported mortality rates or by reported readmission rates. In each of these cases, we analyse the effects (on mortality or readmission rates) of a marginal and unilateral increase in quality, evaluated at a symmetric starting point. As we have shown in Sections 3-4, the direction of these effects determines (qualitatively) the equilibrium effects of P4P bonuses.

[19] Despite the intuitive role of public reporting in reducing patients' information asymmetries and providing incentives to improve quality, the literature on the effects of quality reporting in health care shows mixed results. Dranove et al. (2003) find that public disclosure of patient health outcomes (CABG mortality rate) at the hospital level in New York and Pennsylvania led to worse health outcomes, especially for sicker patients. Studying the relationship between publicly available quality indicators and patient choice in the Netherlands, Varkevisser et al. (2012) conclude that (not fully risk-adjusted) readmission rates may not provide a correct signal of hospital quality and, as such, may lead to suboptimal choices and risk selection by hospitals. On the other hand, Kolstad (2013) finds that the public release of information on hospital performance (CABG mortality rate) in Pennsylvania led to an intrinsic quality response significantly larger than the response to profit incentives, suggesting that quality reporting can increase the accountability of providers irrespective of the effect mediated through changes in demand.

6.1 Patient choices based on mortality rates

Suppose that patients use publicly reported mortality rates as a source of information about actual treatment quality when choosing provider; i.e., $\rho_i = -m_i$. Because of patient composition effects, the mortality rate of each hospital depends on the treatment qualities at both hospitals. Thus, for given values of q_i and q_j, the mortality rates at the two hospitals are simultaneously determined by the following system:[20]

$$m_i = \frac{\gamma x(L)}{D_i} M(L) + \frac{(1-\gamma) x(H)}{D_i} M(H), \tag{48}$$

$$m_j = \frac{\gamma (1 - x(L))}{1 - D_i} M(L) + \frac{(1-\gamma)(1 - x(H))}{1 - D_i} M(H), \tag{49}$$

where $x(s)$ and D_i are given by (46) and (47), respectively, with $\rho_i = -m_i$.

We can evaluate the effect of a marginal increase in quality by Hospital i by differentiating (48)-(49) with respect to m_i, m_j and q_i, and applying Cramer's Rule. Evaluated at a symmetric steady state, where $q_i = q_j$ and $m_i = m_j$, this effect is given by (see the Appendix for details):

$$\left.\frac{\partial m_i}{\partial q_i}\right|_{q_i = q_j} = \frac{\left[\gamma M_{q_i}(L) + (1-\gamma) M_{q_i}(H)\right]\left[\frac{1}{2} + \frac{1}{2t}(1-\gamma)\gamma\left(\theta^H - \theta^L\right)\Delta M(q_i)\right]}{\frac{1}{2} + \frac{1}{t}(1-\gamma)\gamma\left(\theta^H - \theta^L\right)\Delta M(q_i)} < 0. \tag{50}$$

Thus, a unilateral increase in quality by Hospital i will ultimately lead to a lower mortality rate. As in the main model, the direct effect on mortality is dampened by a patient composition effect, which is stronger the higher the difference in demand responsiveness between the high- and low-severity patients.[21] However, in contrast to the main model, the patient composition effect through demand is here only a second-order effect that never dominates the direct effect. The reason is that, in this extended version of the model, demand does not respond directly to quality changes. Instead, patient responses are (second-order) feedback responses to the direct effect of quality on mortality probabilities.

[20] The recursiveness of the interaction between mortality rates and patient choices implies that we should think of the solution to (48)-(49) as the steady state of a dynamic process where, for given quality levels, demand responds to changes in reported mortality rates, and *vice versa*, over time.

[21] It is straightforward to verity that the absolute value of (50) is increasing in $\left(\theta^H - \theta^L\right)$.

6.2 Patient choices based on readmission rates

Suppose instead that patients use publicly reported information about readmission rates when making choices about their preferred provider; i.e., suppose that $\rho_i = -r_i$. As for the case of mortality rates, the readmission rate of each hospital depends, through patient choices, on the treatment quality at each hospital. Thus, the two readmission rates are simultaneously determined by the following system:

$$r_i = \frac{\gamma x(L) S(L) R(L)}{\gamma x(L) S(L) + (1-\gamma) x(H) S(H)} + \frac{(1-\gamma) x(H) S(H) R(H)}{\gamma x(L) S(L) + (1-\gamma) x(H) S(H)}, \tag{51}$$

$$r_j = \frac{\gamma (1-x(L)) S(L) R(L)}{\gamma (1-x(L)) S(L) + (1-\gamma)(1-x(H)) S(H)} + \frac{(1-\gamma)(1-x(H)) S(H) R(H)}{\gamma (1-x(L)) S(L) + (1-\gamma)(1-x(H)) S(H)}, \tag{52}$$

where $x(s)$ is given by (46) with $\rho_i = -r_i$.

The effect of a marginal quality increase by Hospital i, evaluated at a symmetric steady state, is given by

$$\left.\frac{\partial r_i}{\partial q_i}\right|_{q_i=q_j} = \frac{\left[\gamma S(L) R_{q_i}(L) + (1-\gamma) S(H) R_{q_i}(H) + \frac{\gamma(1-\gamma)\Delta R(q_i)}{\gamma S(L)+(1-\gamma)S(L)}\Lambda\right]\Psi}{\frac{1}{2}(\gamma S(L) + (1-\gamma) S(H))^2 + \gamma S(L) S(H) \Delta R(q_i) \frac{(\theta^H - \theta^L)(1-\gamma)}{t}}, \tag{53}$$

where

$$\Lambda := S_{q_i}(H) S(L) - S_{q_i}(L) S(H) > 0 \tag{54}$$

and

$$\Psi := \frac{1}{2}(\gamma S(L) + (1-\gamma) S(H)) + \frac{\gamma(1-\gamma)\left(\theta^H - \theta^L\right) S(L) S(H) \Delta R(q_i)}{2t(\gamma S(L) + (1-\gamma) S(H))} > 0. \tag{55}$$

The sign of $\partial r_i / \partial q_i$ is determined by the sign of the expression in the square brackets in the numerator of (53), which is a priori ambiguous. As in the main model, there are three different effects. (i) The direct effect, which is represented by the first and second terms in the square brackets, is obviously negative. (ii) The direct effect is dampened by a *Patient Composition Effect through Demand*, which is represented by the second term in the denominator. However, as for the case of patient choices based on reported mortality rates, this effect is here only a second-order effect that never dominates the direct effect. (iii) The final effect is the *Patient Composition Effect through Mortality*, which is the same as in the main model and is represented by the last term in the square brackets. This effect might dominate the direct effect, leading to a negative relationship

between treatment quality and readmission rates, if the direct effect of quality on mortality is large relative to the direct effect on readmissions.

Overall, we see that, compared with the results from our main model, the effects of unilateral quality changes on mortality and readmission rates are very similar under these alternative assumptions regarding patient choices. Each of the mechanisms identified in our main analysis are also present here, and each of them go in the same direction. The only difference is that the *Patient Composition Effect through Demand* is strictly a second-order effect when demand does not respond directly to quality changes. The implications for the potential equilibrium effects of P4P bonuses are also similar. In both cases, the scope for counterproductive effects of P4P schemes is larger when the financial incentives are tied to readmission rates instead of mortality rates.

7 Conclusions

This study has investigated how P4P schemes which reward common health outcome indicators, such as mortality and readmission rates, affect the incentives to provide quality. We have done so in a rich and realistic environment when risk-adjustment is not fully accounted for and unobserved dimensions of severity reduce the reliability of mortality and readmission rates. We show that this source of selection bias can in principle weaken or strengthen quality incentives. The direction depends on whether demand responsiveness to quality is stronger for high- or low-severity patients, which we show is *a priori* ambiguous: for example, while marginal utility from quality of high-severity patients is higher, their post-treatment utility is lower, generating opposite incentives for patients.

Empirically, some studies that model patient choice as a function of quality suggest that demand responsiveness is higher for high-severity patients. For example, Gaynor et al. (2016) find that patients in need of a coronary bypass in England are less likely to choose hospitals with higher mortality and this effect is stronger for patients with higher severity, as measured by the Charlson index. Tay (2003) shows that more-severely ill patients in need of cardiovascular procedures in the US are more likely to choose high-volume hospitals (which are likely to provide higher quality due to learning-by-doing effects). Therefore, if more severe patients respond more to quality, providers will have weaker incentives for quality provision when selection bias through patient choice is important. Indeed, there is a point where the selection bias is so large that the introduction of a P4P scheme may reduce rather than increase quality. But the opposite is also possible. Gutacker et al. (2016)

show that hip replacement patients are more willing to travel if they are less severe, where severity is accurately measured by pre-operative Oxford Hip Score, which reflects patient mobility and degree of pain before the surgery takes place. In this case, providers will have stronger, rather than weaker, incentives for quality provision.

Moreover, we show that since readmission rates suffer from selection bias through mortality, which affects the risk distribution of survived patients and therefore the readmission rate, this bias always weakens providers' quality incentives and increases the scope for counterproductive effects of P4P. This result might contribute to explain recent (if still limited) empirical evidence on the effects of the HRRP in the US on Medicare readmission rates. Mellor et al. (2017) find that, while the HRRP seems to have reduced readmission rates for AMI patients, it has been ineffective in reducing other readmission rates targeted by the programme.

We also show that selection effects through patient choice are more important in more competitive areas. This is in line with some empirical evidence suggesting that P4P has a weaker effect on performance in more competitive market structures (Werner et al., 2011).[22] P4P and competition policies may therefore be substitutes: while competition tends to increase quality, it weakens the potential effect of P4P schemes.

Finally, we show that the welfare implications of P4P schemes depend on how welfare is defined, whether a bonus or a penalty is introduced, and other policy instruments at the regulator's disposal. As long as higher quality translates into some reduction in mortality and readmission rates, the introduction of a bonus will increase welfare if patient benefits are higher than the costs, and if quality pre-policy intervention is below the first best level. If welfare is more narrowly defined from the purchaser perspective, as benefits minus transfers to providers, the (perceived) costs of a P4P scheme are related to bonus payments rather than the induced changes in the costs of quality provision. The former is larger than the latter, implying that the scope for P4P based on bonuses is smaller from a purchaser perspective than from a welfare perspective, if demand responsiveness to quality and the degree of provider altruism are both sufficiently low. Whether a bonus or a penalty is introduced does not affect welfare, defined as benefits minus costs. The purchaser instead has a clear preference for a penalty: it has a similar effect on quality but increases rather than reduces

[22] Werner et al. (2011) suggest that "Pay-for-performance had a larger effect on hospital performance in markets with less competition. After stratifying by level of market competition, we found that pay-for-performance had a larger effect on performance improvement than public reporting alone did among hospitals in the least competitive—sometimes called the most concentrated—markets (Appendix Figure 2, Panel A). There were only modest differences in improvement between pay-for-performance and control hospitals in competitive markets ...".

the transfers. However, a penalty can only be introduced if the provider participation constraint is not binding.

Some health systems have introduced P4P schemes which are budget neutral, so that a bonus is financed through a reduction of the basic DRG tariff. We show that P4P is welfare improving if the higher quality induced by P4P is higher than the lower quality induced by a lower price mark-up, as long as the initial quality is below the first-best level (where marginal patient benefit is equal to provider marginal cost). This is likely to hold when demand responsiveness to quality is low.

Overall, our results contribute to the current debate on the optimal design of pay-for-performance schemes in the health sector by highlighting scenarios under which P4P will lead to the desired policy outcomes.

References

[1] Beitia, A. (2003), "Hospital quality choice and market structure in a regulated duopoly", *Journal of Health Economics* 22: 1011–1036.

[2] Bardey, B., Canta, C., Lozachmeur, J.M. (2012) "The regulation of health care providers' payments when horizontal and vertical differentiation matter", *Journal of Health Economics* 31(5), 691-704.

[3] Berta, P., Martini, G., Moscone, F., Vittadini, G. (2016) "The association between asymmetric information, hospital competition and quality of healthcare: evidence from Italy", *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 179(4): 907-926.

[4] Berenson, B., Pronovost, P., Krumholz, H. (2013), *Achieving the potential of health care performance measures*. Washington, DC: Urban Institute.

[5] Brekke, K.R., Cellini, R., Siciliani, L., Straume, O.R. (2010), "Competition and quality in health care markets: A differential–game approach", *Journal of Health Economics* 29: 508–523.

[6] Brekke, K.R., Nuscheler, R., Straume, O.R. (2006), "Quality and location choices under price regulation", *Journal of Economics & Management Strategy* 15: 207–227.

[7] Brekke, K.R., Nuscheler, R., Straume, O.R. (2007), "Gatekeeping in health care", *Journal of Health Economics* 26: 149–170.

[8] Brekke, K.R., Siciliani, L., Straume, O.R. (2008), "Competition and waiting times in hospital markets", *Journal of Public Economics* 92: 1607–1628.

[9] Brekke, K.R., Siciliani, L., Straume, O.R. (2011), "Hospital Competition and Quality with Regulated Prices", *The Scandinavian Journal of Economics* 113: 444–469.

[10] Brekke, K., Gravelle, H., Siciliani, L., Straume, O.R. (2014), "Patient Choice, Mobility and Competition Among Health Care Providers", Chapter 1, pages 1-26, in R. Levaggi and M. Montefiori (eds.), Health Care Provision and Patient Mobility, Developments in Health Economics and Public Policy, 12, Springer-Verlag.

[11] Busse, R., Geissler, A., Quentin, W., Wiley, M. (2011), *Diagnosis-Related Groups in Europe: Moving towards transparency, efficiency and quality in hospitals*. Open University Press.

[12] Busse, R., Nimptsch, U., Mansky, T. (2009), "Measuring, monitoring, and managing quality in Germany's hospitals", *Health Affairs* 28(2): w294-w304.

[13] Cashin, C., Chi, Y.-L., Smith, P., Borowitz, M., Thomson, S. (2014), *Paying for Performance in Healthcare: Implications for Health System Performance and Accountability*. Open University Press.

[14] Cavalieri, M., Gitto, L., Guccio, C. (2013), "Reimbursement systems and quality of hospital care: an empirical analysis for Italy", *Health Policy* 111(3): 273-289.

[15] Chalkley, M., Malcomson, J.M., (1998a), "Contracting for health services with unmonitored quality". *The Economic Journal* 108: 1093-1110.

[16] Chalkley, M., Malcomson, J.M., (1998b), "Contracting for health services when patient demand does not reflect quality". *Journal of Health Economics* 17: 1–19.

[17] Chua, C.L., Palangkaraya, A., Yong, J. (2010), "A two-stage estimation of hospital quality using mortality outcome measures: an application using hospital administrative data", *Health Economics* 19(12): 1404–1424.

[18] Conrad, D. A. (2015), "The Theory of Value-Based Payment Incentives and Their Application to Health Care", *Health Services Research* 50(S2): 2057-2089.

[19] Cooper, Z., Gibbons, S., Jones, S., McGuire, A. (2011), "Does hospital competition save lives? Evidence from the English NHS patient choice reforms", *The Economic Journal* 121(554): F228–F260.

[20] Cox, J. C., Sadiraj, V., Schnier, K. E., Sweeney, J. F. (2016), "Incentivizing cost-effective reductions in hospital readmission rates", *Journal of Economic Behavior & Organization* 131: 24-35.

[21] Custers, T., Arah, O.A., Klazinga, N.S. (2007), "Is there a business case for quality in The Netherlands?: A critical analysis of the recent reforms of the health care system", *Health Policy* 82(2): 226-239.

[22] Dafny, L.S. (2005), "How do hospitals respond to price changes?", *The American Economic Review* 95: 1525–1547.

[23] Department of Health (2011), *Payment by Results Guidance for 2011-12*. https://www.gov.uk/government/collections/payment-by-results-pbr-in-the-nhs.

[24] Dranove, D., Kessler, D., McClellan, M., Satterthwaite, M. (2003), "Is more information better? The effects of "report cards" on health care providers", *Journal of Political Economy* 111(3): 555-588.

[25] Eggleston, K. (2005), "Multitasking and mixed systems for provider payment", *Journal of Health Economics* 24: 211-223.

[26] Eijkenaar, F. (2013), "Key issues in the design of pay for performance programs", *The European Journal of Health Economics* 14(1): 117-131.

[27] Eijkenaar, F., Emmert, M., Scheppach, M., Schöffski, O. (2013), "Effects of pay for performance in health care: a systematic review of systematic reviews", *Health Policy* 110(2): 115-130.

[28] Ellis, R.P., McGuire, T. (1986), "Provider behavior under prospective reimbursement: Cost sharing and supply", *Journal of Health Economics* 5: 129–151.

[29] Epstein, A.M. (2012), "Will pay for performance improve quality of care? The answer is in the details", *New England Journal of Medicine* 367(19):1852-1853.

[30] Fischer, C., Lingsma, H.F., Marang-van de Mheen, P.J., Kringos, D.S., Klazinga, N.S., Steyerberg, E.W. (2014), "Is the readmission rate a valid quality indicator? A review of the evidence", *PLoS One* 9(11): e112282.

[31] Flodgren, G., Eccles, M.P., Shepperd, S., Scott, A., Parmelli, E., Beyer, F.R. (2011), "An overview of reviews evaluating the effectiveness of financial incentives in changing healthcare professional behaviours and patient outcomes", *Cochrane Database of Systematic Reviews* 7: CD009255.

[32] Gaynor, M. (2006), "What do we know about competition and quality in health care markets?", *NBER Working Paper* No. 12301.

[33] Gaynor, M., Propper, C., Seiler, S. (2016), "Free to choose? Reform, choice, and consideration sets in the English National Health Service", *The American Economic Review* 106(11): 3521-3557.

[34] Geissler, A., Scheller-Kreinsen, D., Busse, R. (2011), "Germany: Understanding G-DRGs". In: Busse R., Geissler A., Quentin W., Wiley M. (Eds), *Diagnosis Related Groups in Europe. Moving Towards Transparency, Efficiency and Quality in Hospitals*. Open University Press.

[35] Gravelle, H., Santos, R., Siciliani, L. (2014), "Does a hospital's quality depend on the quality of other hospitals? A spatial econometrics approach", *Regional Science and Urban Economics* 49: 203–216.

[36] Gravelle, H., Sivey, P. (2010), "Imperfect information in a quality-competitive hospital market", *Journal of Health Economics* 29(4): 524-535.

[37] Gravelle, H., Sutton, M., Ma, A. (2010), "Doctor behaviour under a pay for performance contract: treating, cheating and case finding?", *The Economic Journal* 120(542): F129–F156.

[38] Guccio, C., Lisi, D., Pignataro, G. (2016), "Readmission and Hospital Quality under Different Payment Regimes", *FinanzArchiv/Public Finance Analysis* 72(4): 453-474.

[39] Gutacker, N., Siciliani, L., Moscelli, G., Gravelle, H., (2016), "Choice of hospital: which type of quality matters?", *Journal of Health Economics* 50: 230-246.

[40] Holmstrom, B., Milgrom, P. (1991), "Multitask principal-agent analyses: incentive contracts, asset ownership, and job design", *Journal of Law, Economics and Organization* 7: 24–52.

[41] Kaarboe, O., Siciliani, L. (2011), "Quality, Multitasking and Pay for Performance", *Health Economics* 2: 225-238.

[42] Kessler, D.P., McClellan, M.B. (2000), "Is hospital competition socially wasteful?", *The Quarterly Journal of Economics* 115(2): 577-615.

[43] Kolstad, J. T. (2013), "Information and quality when motivation is intrinsic: Evidence from surgeon report cards", *The American Economic Review* 103(7): 2875-2910.

[44] Kristensen, S.-R., Bech, M., Quentin, W. (2015), "A roadmap for comparing readmission policies with application to Denmark, England, Germany and the United States", *Health Policy* 119(3): 264-273.

[45] Kristensen, S.-R., Siciliani, L., Sutton, M. (2016), "Optimal price-setting in pay for performance schemes in health care", *Journal of Economic Behavior & Organization* 123: 55–77.

[46] Kuhn, M., Siciliani, L. (2009), "Performance indicators for quality with costly falsification", *Journal of Economics and Management Strategy* 18: 1137-1154.

[47] Kuhn, M., Siciliani, L. (2013), "Manipulation and Auditing of Public Sector Contracts", *European Journal of Political Economy* 32: 251-267

[48] La Forgia, G.M., Couttolenc, B.F. (2008), *Hospital performance in Brazil – the search for excellence*. Washington, DC: World Bank.

[49] Laudicella, M., Li Donni, P., Smith, P.C. (2013), "Hospital readmission rates: Signal of failure or success?", *Journal of Health Economics* 32: 909–921.

[50] Lindenauer, P.K., Remus, D., Roman, S., Rothberg, M.B., Benjamin, E.M., Ma, A., Bratzler, D.W. (2007), "Public reporting and pay for performance in hospital quality improvement", *New England Journal of Medicine* 356(5): 486-496.

[51] Lisi, D., Moscone, F., Tosetti, E., Vinciotti, V. (2017), "Hospital interdependence in a competitive institutional environment: Evidence from Italy", *HEDG Working Paper* No. 17/07.

[52] Ma, C-T. A. (1994), "Health Care Payment Systems: Cost and Quality Incentives", *Journal of Economics & Management Strategy* 3: 93–112.

[53] Ma, C-T. A., Mak, H. Y. (2015), "Information disclosure and the equivalence of prospective payment and cost reimbursement", *Journal of Economic Behavior & Organization* 117: 439–452.

[54] Mak, H. Y. (2018), "Managing imperfect competition by pay for performance and reference pricing", *Journal of Health Economics* 57: 131-146.

[55] McClellan, M., Staiger, D. (1999), "The Quality of Health Care Providers", *NBER Working Paper* No. 7327.

[56] Mellor, J., Daly, M., Smith, M. (2017), "Does it pay to penalize hospitals for excess readmissions? Intended and unintended consequences of Medicare's Hospital Readmissions Reductions Program", *Health Economics* 26(8): 1037-1051.

[57] Milstein, R., Schreyoegg, J. (2016), "Pay for performance in the inpatient sector: A review of 34 P4P programs in 14 OECD countries", *Health Policy* 120(10): 1125-1140.

[58] Mohammed, M.A., Deeks, J.J., Girling, A., Rudge, G., Carmalt, M., Stevens, A.J., Lilford, R.J. (2009), "Evidence of methodological bias in hospital standardised mortality ratios: retrospective database study of English hospitals", *British Medical Journal* 338: b780.

[59] Mullen, K. J., Frank, R. G., Rosenthal, M. B. (2010), "Can you get what you pay for? Pay-for-performance and the quality of healthcare providers", *The RAND Journal of Economics* 41(1): 64-91.

[60] Papanicolas, I., McGuire, A. (2017), "Measuring and forecasting quality in English hospitals", *Journal of the Royal Statistical Society: Series A* 180(2): 409-432.

[61] Propper, C., Burgess, S., Gossage, D. (2008), "Competition and Quality: Evidence from the NHS Internal Market", *The Economic Journal* 118 (525): 138–170.

[62] Propper, C., Burgess, S., Green, K. (2004), "Does competition between hospitals improve the quality of care?: Hospital death rates and the NHS internal market", *Journal of Public Economics* 88 (7-8): 1247–1272.

[63] Schreyogg, J., Stargardt, T. (2010), "The trade-off between costs and outcomes: the case of acute myocardial infarction", *Health Services Research* 45: 1585–1601.

[64] Shahian, D.M., Wolf, R.E., Iezzoni, L.I., Kirle, L., Normand, S.L. (2010), "Variability in the measurement of hospital-wide mortality rates", *New England Journal of Medicine* 363(26): 2530-2539.

[65] Sherry, T. B. (2016), "A Note on the Comparative Statics of Pay-for-Performance in Health Care", *Health Economics* 25(5): 637-644.

[66] Siciliani, L., Straume, O.R., Cellini, R. (2013), "Quality competition with motivated providers and sluggish demand", *Journal of Economic Dynamics and Control* 37: 2041–2061.

[67] Sutton, M., Nikolova, S., Boaden, R., Lester, H., McDonald, R., Roland, M. (2012), "Reduced mortality with hospital pay for performance in England", *New England Journal of Medicine* 367(19): 1821-1828.

[68] Tay, A. (2003), "Assessing Competition in Hospital Care Markets: the Importance of Accounting for Quality Di erentiation." *RAND Journal of Economics* 34(4): 786-814.

[69] Van Herck, P., De Smedt, D., Annemans, L., Remmen, R., Rosenthal, M.B., Sermeus, W. (2010), "Systematic review: effects, design choices, and context of pay-for-performance in health care", *BMC health services research* 10(1): 247.

[70] Varkevisser, M., van der Geest, S. A., Schut, F. T. (2012), "Do patients choose hospitals with high quality ratings? Empirical evidence from the market for angioplasty in the Netherlands", *Journal of Health Economics* 31(2): 371-378.

[71] Wennberg, J.E., Staiger, D.O., Sharp, S.M., Gottlieb, D.J., Bevan, G., McPherson, K., Welch, H.G. (2013), "Observational intensity bias associated with illness adjustment: cross sectional analysis of insurance claims", *British Medical Journal* 346: f549.

[72] Werner, R. M., Kolstad, J. T., Stuart, E. A., Polsky, D. (2011), "The effect of pay-for-performance in hospitals: lessons for quality improvement", *Health Affairs* 30(4): 690-698.

[73] Zhang, D. J., Gurvich, I., Van Mieghem, J. A., Park, E., Young, R. S., Williams, M. V. (2016), "Hospital readmissions reduction program: An economic and operational analysis", *Management Science* 62(11): 3351-3371.

Appendix

Below we present the underlying calculations for the results given in Section 6.

Patient choices based on mortality rates

For given quality levels, equilibrium mortality rates at the two hospitals are implicitly given by the system (48)-(49), which we re-write as

$$\Theta_i := m_i D_i - \gamma x\left(L\right) M\left(L, q_i\right) - \left(1-\gamma\right) x\left(H\right) M\left(H, q_i\right) = 0, \tag{A1}$$

$$\Theta_j := m_j\left(1-D_i\right) - \gamma\left(1-x\left(L\right)\right) M\left(L, q_j\right) - \left(1-\gamma\right)\left(1-x\left(H\right)\right) M\left(H, q_j\right) = 0. \tag{A2}$$

Differentiating (A1)-(A2) with respect to m_i, m_j and q_i yields

$$\left[\begin{array}{cc} \frac{\partial \Theta_i}{\partial m_i} & \frac{\partial \Theta_i}{\partial m_j} \\ \frac{\partial \Theta_j}{\partial m_i} & \frac{\partial \Theta_j}{\partial m_j} \end{array}\right]\left[\begin{array}{c} dm_i \\ dm_j \end{array}\right] + \left[\begin{array}{c} \frac{\partial \Theta_i}{\partial q_i} \\ \frac{\partial \Theta_j}{\partial q_i} \end{array}\right] dq_i = 0, \tag{A3}$$

where

$$\frac{\partial \Theta_i}{\partial m_i} = \frac{1}{2} - \frac{\gamma\theta^L + \left(1-\gamma\right)\theta^H}{2t}\left(2m_i - m_j\right) + \gamma\frac{\theta^L}{2t} M\left(L, q_i\right) + \left(1-\gamma\right)\frac{\theta^H}{2t} M\left(H, q_i\right), \tag{A4}$$

$$\frac{\partial \Theta_i}{\partial m_j} = \frac{\gamma\theta^L + \left(1-\gamma\right)\theta^H}{2t} m_i - \gamma\frac{\theta^L}{2t} M\left(L, q_i\right) - \left(1-\gamma\right)\frac{\theta^H}{2t} M\left(H, q_i\right), \tag{A5}$$

$$\frac{\partial \Theta_j}{\partial m_i} = \frac{\gamma\theta^L + \left(1-\gamma\right)\theta^H}{2t} m_j - \gamma\frac{\theta^L}{2t} M\left(L, q_j\right) - \left(1-\gamma\right)\frac{\theta^H}{2t} M\left(H, q_j\right), \tag{A6}$$

$$\frac{\partial \Theta_j}{\partial m_j} = \frac{1}{2} - \frac{\gamma\theta^L + \left(1-\gamma\right)\theta^H}{2t}\left(2m_j - m_i\right) + \gamma\frac{\theta^L}{2t} M\left(L, q_j\right) + \left(1-\gamma\right)\frac{\theta^H}{2t} M\left(H, q_j\right), \tag{A7}$$

$$\frac{\partial \Theta_i}{\partial q_i} = -\gamma\left(\frac{1}{2} - \frac{\theta^L\left(m_i - m_j\right)}{2t}\right) M_{q_i}\left(L, q_i\right) - \left(1-\gamma\right)\left(\frac{1}{2} - \frac{\theta^H\left(m_i - m_j\right)}{2t}\right) M_{q_i}\left(H, q_i\right), \tag{A8}$$

$$\frac{\partial \Theta_j}{\partial q_i} = 0. \tag{A9}$$

By imposing symmetry, $q_i = q_j$, which implies

$$m_i = m_j = \gamma M\left(L\right) + \left(1-\gamma\right) M\left(H\right), \tag{A10}$$

and by applying the expressions for $x(s)$ and D_i, (A4)-(A8) can be re-written as

$$\frac{\partial \Theta_i}{\partial m_i} = \frac{\partial \Theta_j}{\partial m_j} = \frac{1}{2} + \frac{1}{2t}(1-\gamma)\gamma\left(\theta^H - \theta^L\right)\Delta M(q_i) > 0, \tag{A11}$$

$$\frac{\partial \Theta_i}{\partial m_j} = \frac{\partial \Theta_j}{\partial m_i} = -\frac{1}{2t}\gamma(1-\gamma)\left(\theta^H - \theta^L\right)\Delta M(q_i) < 0, \tag{A12}$$

$$\frac{\partial \Theta_i}{\partial q_i} = -\frac{1}{2}\left(\gamma M_{q_i}(L) + (1-\gamma) M_{q_i}(H)\right) > 0. \tag{A13}$$

Using Cramer's Rule, the effect of a marginal (unilateral) increase in q_i on the mortality rate of Hospital i is given by

$$\left.\frac{\partial m_i}{\partial q_i}\right|_{q_i=q_j} = \frac{\begin{vmatrix} -\frac{\partial \Theta_i}{\partial q_i} & \frac{\partial \Theta_i}{\partial m_j} \\ -\frac{\partial \Theta_j}{\partial q_i} & \frac{\partial \Theta_j}{\partial m_j} \end{vmatrix}}{\begin{vmatrix} \frac{\partial \Theta_i}{\partial m_i} & \frac{\partial \Theta_i}{\partial m_j} \\ \frac{\partial \Theta_j}{\partial m_i} & \frac{\partial \Theta_j}{\partial m_j} \end{vmatrix}} = \frac{-\frac{\partial \Theta_i}{\partial q_i}\frac{\partial \Theta_j}{\partial m_j}}{\frac{\partial \Theta_i}{\partial m_i}\frac{\partial \Theta_j}{\partial m_j} - \frac{\partial \Theta_j}{\partial m_i}\frac{\partial \Theta_i}{\partial m_j}}, \tag{A14}$$

which, using (A11)-(A13), is given by (50) in Section 6.

Patient choices based on readmission rates

For given quality levels, equilibrium readmission rates at the two hospitals are implicitly given by the system (51)-(52), which we re-write as

$$\Phi_i := \begin{bmatrix} r_i\left(\gamma x(L) S(L,q_i) + (1-\gamma) x(H) S(H,q_i)\right) \\ -\gamma x(L) S(L,q_i) R(L,q_i) - (1-\gamma) x(H) S(H,q_i) R(H,q_i) \end{bmatrix} = 0, \tag{A15}$$

$$\Phi_j := \begin{bmatrix} r_j\left(\gamma\left(1-x(L)\right) S(L,q_j) + (1-\gamma)\left(1-x(H)\right) S(H,q_j)\right) \\ -\gamma\left(1-x(L)\right) S(L,q_j) R(L,q_j) - (1-\gamma)\left(1-x(H)\right) S(H,q_j) R(H,q_j) \end{bmatrix} = 0. \tag{A16}$$

Differentiating (A15)-(A16) with respect to r_i, r_j and q_i yields

$$\begin{bmatrix} \frac{\partial \Phi_i}{\partial r_i} & \frac{\partial \Phi_i}{\partial r_j} \\ \frac{\partial \Phi_j}{\partial r_i} & \frac{\partial \Phi_j}{\partial r_j} \end{bmatrix} \begin{bmatrix} dr_i \\ dr_j \end{bmatrix} + \begin{bmatrix} \frac{\partial \Phi_i}{\partial q_i} \\ \frac{\partial \Phi_j}{\partial q_i} \end{bmatrix} dq_i = 0, \tag{A17}$$

where

$$\begin{aligned}\frac{\partial \Phi_i}{\partial r_i} &= \gamma\left(\frac{1}{2}-\frac{\theta^L\left(2r_i-r_j\right)}{2t}\right)S\left(L,q_i\right)+\left(1-\gamma\right)\left(\frac{1}{2}-\frac{\theta^H\left(2r_i-r_j\right)}{2t}\right)S\left(H,q_i\right)\\ &\quad+\gamma\frac{\theta^L}{2t}S\left(L,q_i\right)R\left(L,q_i\right)+\left(1-\gamma\right)\frac{\theta^H}{2t}S\left(H,q_i\right)R\left(H,q_i\right),\end{aligned} \tag{A18}$$

$$\begin{aligned}\frac{\partial \Phi_i}{\partial r_j} &= r_i\left(\gamma\frac{\theta^L}{2t}S\left(L,q_i\right)+\left(1-\gamma\right)\frac{\theta^H}{2t}S\left(H,q_i\right)\right)\\ &\quad-\gamma\frac{\theta^L}{2t}S\left(L,q_i\right)R\left(L,q_i\right)-\left(1-\gamma\right)\frac{\theta^H}{2t}S\left(H,q_i\right)R\left(H,q_i\right),\end{aligned} \tag{A19}$$

$$\begin{aligned}\frac{\partial \Phi_j}{\partial r_i} &= r_j\left(\gamma\frac{\theta^L}{2t}S\left(L,q_j\right)+\left(1-\gamma\right)\frac{\theta^H}{2t}S\left(H,q_j\right)\right)\\ &\quad-\gamma\frac{\theta^L}{2t}S\left(L,q_j\right)R\left(L,q_j\right)-\left(1-\gamma\right)\frac{\theta^H}{2t}S\left(H,q_j\right)R\left(H,q_j\right),\end{aligned} \tag{A20}$$

$$\begin{aligned}\frac{\partial \Phi_j}{\partial r_j} &= \gamma\left(\frac{1}{2}-\frac{\theta^L\left(2r_j-r_i\right)}{2t}\right)S\left(L,q_j\right)+\left(1-\gamma\right)\left(\frac{1}{2}-\frac{\theta^H\left(2r_j-r_i\right)}{2t}\right)S\left(H,q_j\right)\\ &\quad+\gamma\frac{\theta^L}{2t}S\left(L,q_j\right)R\left(L,q_j\right)+\left(1-\gamma\right)\frac{\theta^H}{2t}S\left(H,q_j\right)R\left(H,q_j\right),\end{aligned} \tag{A21}$$

$$\begin{aligned}\frac{\partial \Phi_i}{\partial q_i} &= r_i\left(\gamma\left(\frac{1}{2}-\frac{\theta^L\left(r_i-r_j\right)}{2t}\right)S_q\left(L,q_i\right)+\left(1-\gamma\right)\left(\frac{1}{2}-\frac{\theta^H\left(r_i-r_j\right)}{2t}\right)S_q\left(H,q_i\right)\right)\\ &\quad-\gamma\left(\frac{1}{2}-\frac{\theta^L\left(r_i-r_j\right)}{2t}\right)\left(S\left(L,q_i\right)R_{q_i}\left(L,q_i\right)+S_{q_i}\left(L,q_i\right)R\left(L,q_i\right)\right)\\ &\quad-\left(1-\gamma\right)\left(\frac{1}{2}-\frac{\theta^H\left(r_i-r_j\right)}{2t}\right)\left(S\left(H,q_i\right)R_{q_i}\left(H,q_i\right)+S_{q_i}\left(H,q_i\right)R\left(H,q_i\right)\right),\end{aligned} \tag{A22}$$

$$\frac{\partial \Phi_j}{\partial q_i}=0. \tag{A23}$$

By imposing symmetry, $q_i=q_j$, which implies

$$r_i=r_j=\frac{\gamma S\left(L\right)R\left(L\right)+\left(1-\gamma\right)S\left(H\right)R\left(H\right)}{\gamma S\left(L\right)+\left(1-\gamma\right)S\left(H\right)}, \tag{A24}$$

and by applying the expressions for $x(s)$ and D_i, (A18)-(A22) can be re-written as

$$\frac{\partial \Phi_i}{\partial r_i} = \frac{\partial \Phi_j}{\partial r_j} = \frac{1}{2}\left(\gamma S(L) + (1-\gamma) S(H)\right) + \frac{\gamma(1-\gamma)\left(\theta^H - \theta^L\right) S(L) S(H) \Delta R(q_i)}{2t\left(\gamma S(L) + (1-\gamma) S(H)\right)} > 0, \quad \text{(A25)}$$

$$\frac{\partial \Phi_i}{\partial r_j} = \frac{\partial \Phi_j}{\partial r_i} = -\frac{\gamma(1-\gamma)\left(\theta^H - \theta^L\right) S(L) S(H) \Delta R(q_i)}{2t\left(\gamma S(L) + (1-\gamma) S(H)\right)} < 0, \quad \text{(A26)}$$

$$\frac{\partial \Phi_i}{\partial q_i} = -\frac{1}{2}\left[\begin{array}{c} \left(\gamma S(L) R_{q_i}(L) + (1-\gamma) S(H) R_{q_i}(H)\right) \\ + \frac{\gamma(1-\gamma)\Delta R(q_i)}{\gamma S(L) + (1-\gamma) S(L)} \left(S_{q_i}(H) S(L) - S_{q_i}(L) S(H)\right) \end{array}\right] \gtrless 0. \quad \text{(A27)}$$

Using Cramer's Rule, the effect of a marginal (unilateral) increase in q_i on the mortality rate of Hospital i is given by

$$\left.\frac{\partial r_i}{\partial q_i}\right|_{q_i = q_j} = \frac{\left|\begin{array}{cc} -\frac{\partial \Phi_i}{\partial q_i} & \frac{\partial \Phi_i}{\partial r_j} \\ -\frac{\partial \Phi_j}{\partial q_i} & \frac{\partial \Phi_j}{\partial r_j} \end{array}\right|}{\left|\begin{array}{cc} \frac{\partial \Phi_i}{\partial r_i} & \frac{\partial \Phi_i}{\partial r_j} \\ \frac{\partial \Phi_j}{\partial r_i} & \frac{\partial \Phi_j}{\partial r_j} \end{array}\right|} = \frac{-\frac{\partial \Phi_i}{\partial q_i}\frac{\partial \Phi_j}{\partial r_j}}{\frac{\partial \Phi_i}{\partial r_i}\frac{\partial \Phi_j}{\partial r_j} - \frac{\partial \Phi_i}{\partial r_i}\frac{\partial \Phi_j}{\partial r_j}}, \quad \text{(A28)}$$

which, using (A25)-(A27), is given by (53) in Section 6.