Article:

Siciliani, Luigi orcid.org/0000-0003-1739-7289 and Straume, Odd Rune (2019) Competition and Equity in Health Care Markets. Journal of Health Economics. pp. 1-14. ISSN 0167-6296

https://doi.org/10.1016/j.jhealeco.2018.12.002

Competition and Equity in Health Care Markets*

Luigi Siciliani[†] Odd Rune Straume[‡]

September 2018

Abstract

We provide a model where hospitals compete on quality under fixed prices to investigate how hospital competition affects (i) quality differences between hospitals, and as a result, (ii) health inequalities across hospitals and patient severities. The answer to the first question is ambiguous and depends on factors related to both demand and supply of health care. Whether competition increases or reduces health inequalities depends on the type and measure of inequality. Health inequalities due to the postcode lottery are more likely to decrease if the marginal health gains from quality decrease at a higher rate, whereas health inequalities between high- and low-severity patients decrease if patient composition effects are sufficiently small. We also investigate the effect of competition on health inequalities as measured by the Gini and the Generalised Gini coefficients, and highlight differences compared to the simpler dispersion measures.

Keywords: Hospital competition; quality; health inequalities; Gini coefficient.

JEL Classification: I11, I14, L13

*We thank two anonymous reviewers for very helpful suggestions. 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 Related Studies, University of York, Heslington, York YO10 5DD, UK; E-mail: luigi.siciliani@york.ac.uk

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

1 Introduction

Recent and ongoing reforms in several OECD countries aim at stimulating competition and patient choice among publicly-funded hospitals in order to improve quality of care (EXPH, 2015; OECD, 2012). In the U.S. Medicare and Medicaid programmes, hospitals are paid by Diagnosis Related Group (DRG) since 1983. Medicare and Medicaid cover respectively individuals older than 65 years and poor patients. The DRG system involves paying a fixed tariff for every patient treated. In the United Kingdom, under a policy commonly known as 'Payment by Results', hospitals are also paid a tariff for every patient treated, and patients are free to choose the hospital. Hospital competition is also present in other countries such as Denmark, France, Germany, Italy, and Norway. The idea is that hospitals 'compete' on quality to attract patients and are rewarded financially for doing so.

Opponents of hospital competition argue that these policies will harm equity. For example, high-quality hospitals will respond to competition by improving even more, while low-quality hospitals will be left behind. A recent report by the European Commission highlights that despite the extensive literature investigating the effect of competition in the health sector, there is very limited literature focusing on its equity implications (EXPH, 2015). Reduction in health inequalities is an ubiquitous policy objective, and it is surprising that it has received little attention in relation to competition. We contribute to fill this gap in knowledge.

In this study we extend the received theoretical literature by investigating (i) whether competition increases or reduces the gap in quality between high- and low-quality hospitals, and (ii) whether, as a result, competition increases or reduces health inequalities. We focus on two dimensions of (pure) health inequalities (Wagstaff and van Dooerslaer, 2000, Section 5). The first type of health inequalities is what is commonly known, in the hospital context, as inequalities due to 'postcode lottery': a patient living close to a given hospital might receive much poorer quality compared to a patient living close to a good hospital (Dalton, 2014, p.4). We refer to this as *postcode inequality*. The second type of health inequalities relates to disparities in health across patients with different severity: if high-severity patients benefit less from competition than low-severity patients, health inequalities will worsen. The equity concern across severity groups is regularly reflected in sub-group analysis (by severity type) in cost-effectiveness analyses (Sculpher and Gafni, 2001). Given that we have two sources of health inequalities, we also investigate how competition affects the Gini coefficient, a commonly used measure to empirically assess health inequalities within or across countries (Wagstaff and van Dooerslaer, 2000).

Our choice of theoretical framework is a Hotelling model with two hospitals located at the endpoints of a unit line and competing on quality. In this respect we follow the existing theoretical literature, where quality competition is typically analysed within a spatial competition framework. We allow one hospital to have a comparative advantage so that hospitals provide different qualities in equilibrium. We also assume that only a fraction of patients make choices about which hospital to attend for treatment (with the remaining patients being treated as the closest hospital). This fraction represents the degree of *patient choice*, which we use as our measure of competition. This is a highly policy relevant competition measure in hospital markets, since patient choice can be stimulated by the introduction or the enhancement of public reporting of quality indicators (Siciliani et al., 2017).

Our key findings are as follows. Whether competition increases or reduces *quality differences* across hospitals is generally ambiguous, and depends on three key factors related to the demand for health care and the cost of health care provision, namely (i) decreasing marginal health gains from quality, (ii) differences in price-cost margins between high- and low-quality hospitals, and (iii) quality-dependent unit treatment costs. The first and third of these factors contribute in the direction of quality *convergence* as a result of more competition, whereas the second factor contributes in the opposite direction.

Whether competition increases or reduces *health inequalities* depends on the type of inequality, and the effect does not necessarily go in the same direction as the change in hospital quality differences. If health gains are linear in quality, postcode inequalities go hand in hand with quality differences: they increase (decrease) whenever competition induces quality dispersion (convergence). However, if health gains are strictly concave in quality, then health inequalities can reduce even if competition induces quality dispersion. Thus, competition is more likely to reduce postcode inequality if marginal health gains from quality decrease at a higher rate, which also increases the scope for quality convergence as a result of more competition in the first place.

On the other hand, when considering inequality across severity groups, competition generally *reduces* health inequalities between high- and low-severity patients, because high-severity patients benefit more from higher quality than do low-severity patients. However, this reduction can be strengthened or weakened by what we refer to as 'composition effects', which arise when competition induces high-severity patients to exercise choice to a larger extent than low-severity patients, by selecting hospitals with higher quality.

We then derive the effect of competition on aggregate measures of absolute and relative inequality, namely the Generalised Gini and Gini coefficients, respectively. These measures are conceptually distinct from the above-mentioned measures of dispersion across hospitals and severity groups. Consider for example the special case with just one severity group. Even if competition increases differences in health outcomes across hospitals (i.e., an increase in postcode inequality), the Generalised Gini coefficient may still reduce if competition induces more patients to go to the high-quality hospital. Similarly, even if competition has no effect on differences on health outcomes, the Gini coefficient, which measures relative inequality, will still reduce as a result of the overall increase in quality.

In the full model with two severity groups, we identify two key factors that play a crucial role in determining the effect of competition on aggregate (absolute or relative) inequality, namely (i) the distribution of high- and low-severity patients and (ii) the degree to which health benefits are concave in quality. Since competition tends to reduce inequalities between high- and low-severity patients, it also tends to reduce aggregate inequality if the relative shares of these two patient groups are not too unequal, which means that inequality along this dimension has a large weight in the aggregate inequality measure (Gini or Generalised Gini). Furthermore, if marginal health gains decrease at a sufficiently high rate, competition also tends to reduce postcode inequality, as explained above, which further increases the scope for competition to reduce aggregate inequality.

Finally, although our analysis is predominantly positive, we also include a section where we place our analysis in a normative context by specifying a policy objective function that incorporates concerns for health inequalities. Following the approach by Wagstaff (2002), we define a 'health achievement' index that reflects both average health and inequality in the distribution of health, and that is based on an extended Gini coefficient that allows for different degrees of inequality aversion. In this part of the analysis we show that a sufficient (but not necessary) condition for competition to increase overall health achievement is that the market share of the high-quality hospital does not decrease as a result of more competition.

In line with the existing literature, our theoretical model rests on the assumption that hospitals are profit maximisers and suggests that an increase in competition increases quality (Ma and Burgess, 1993; Wolinsky, 1997; Gravelle, 1999; Beitia, 2003; Nuscheler, 2003; Brekke et al., 2006, 2007; Gaynor, 2006; Karlsson, 2007). This result also holds with altruistic providers but only if the degree of altruism is not too high (Brekke et al., 2011, 2012; see also Barigozzi and Burani, 2016).

Equity is not addressed in the existing theoretical literature on hospital competition, though, with the exception of Halonen-Akatwijuka and Propper (2013), who investigate hospital managers' incentive to invest in differential effort for two types of patients. They find that competition could favour the majority group at the expense of the minority one driven by cost substitution across efforts. However, their analysis is different from ours in many respects. First, hospitals are symmetric in costs, therefore ruling out postcode inequalities, which is a key focus for us. Second, patient benefits are linear in effort, so that health inequalities always coincide with quality inequalities. We instead show how increased inequalities in quality are compatible with reduced health inequalities. Third, managers are semi-altruistic and are paid a fixed salary under no competition, while they have a no-profit constraint based on a tariff system under competition; therefore, competition is modelled as a dichotomous variable which introduces a monetary incentive to provide effort. We instead treat the degree of competition as continuous, through a variable which relates to the responsiveness of demand to quality (i.e., the degree of patient choice). Fourth, manager effort varies by patient type, the cost function is quadratic in effort, and efforts can be cost substitutes or complements. We instead assume that quality is common across patient types and adopt a general cost function. Finally, we investigate measures of dispersion such as the Gini index, and the Health Achievement index.

In the empirical literature, the seminal study by Kessler and McClellan (2000) suggests that competition increases quality. This result is also confirmed by Tay (2003), but only partially by Shen (2003), while Gowrinsankaran and Town (2003) find a negative effect. The latest evidence from England suggests that competition, as measured by the introduction of patient choice policies, increases quality under different empirical approaches (Cooper et al., 2011; Gaynor et al., 2013; Bloom et al., 2015). There is only one empirical study which directly tests the effect of competition on equity. Cookson et al. (2013) find that competition did not harm equity, as measured by differences in hip replacement utilisation across socioeconomic status in England. This study is not directly relevant for us given the focus on utilisation as opposed to quality and health outcomes, and the focus on socioeconomic inequalities as opposed to pure health inequalities. Although not focussing on equity, Kessler and Geppert (2005) find that competition improved health for high-severity patients but not for low-severity patients, therefore providing indirect evidence that health inequalities across severity groups reduced.

As previously stressed, our approach is mainly positive rather than normative. Although we could derive the optimal pricing rule set by a welfare maximising regulator, in reality hospital prices are fixed and are set to reflect average treatment costs. We therefore prefer to investigate how competition affects health inequalities under current common financial arrangements. Similarly, although optimal cost regulation (à la Laffont and Tirole, 1993) could be introduced by a regulator to mitigate the implications of cost heterogeneity across hospitals, we are not aware of policy examples where regulators combine price regulations with partial cost reimbursement.[1]

The study is organised as follows. In Section 2, we present the model and derive equilibrium quality. In Section 3, we investigate how competition affects quality differences across hospitals, and in Section 4, how competition affects health inequalities. In Section 5 we place our analysis in a normative context by adopting a policy objective function that incorporates a potential equity-efficiency trade-off in health care provision. Section 6 draws implications for empirical analyses and Section 7 concludes the study.

2 Model

Consider a market for a healthcare treatment (e.g., a coronary bypass or a hip replacement) offered by two different providers (hospitals), located at opposite endpoints of a Hotelling line of length 1. Demand comes from a unit mass of patients who are uniformly distributed on the line. At each point of the line there is a share λ of high-severity patients, denoted by h. The remaining patients have lower severity and are denoted by l. A patient of type k who is treated at Hospital i has the following utility:

$$U_i^k (q_i) = B^k (q_i) - td, \quad k = h, l; \quad i = 1, 2, \tag{1}$$

where $B^k (\cdot)$ is the (expected) health status of a patient with severity k following healthcare treatment; $q_i \geq \underline{q}$ is the quality of treatment at Hospital i; d is the distance travelled by the patient, and t is the marginal cost of travelling. The lower bound $\underline{q}$ on quality represents the minimum treatment quality that the hospitals are allowed to offer, and we can interpret the case of $q_i < \underline{q}$ as malpractice. We assume that: (i) for a given level of treatment quality, the patient with higher severity is in worse health, even after treatment, $B^h (q) < B^l (q)$; and (ii) the patient with higher severity benefits more from a marginal increase in treatment quality, i.e., $\partial B^h/\partial q > \partial B^l/\partial q > 0$ for all q. Thus, for a given level of treatment quality, the difference in post-treatment health status

[1] The Laffont and Tirole (1993) approach has been applied in the health context in several studies (e.g., Jack, 2005; Siciliani, 2006; Choné and Ma, 2011) including quality competition with a Hotelling set-up (Beitia, 2003).

across high- and low-severity patients is smaller the higher the quality of treatment.

We also assume that, at each point on the line, a fraction θ of the patients make utility-maximising choices based on both treatment quality and travelling distance, whereas the remaining fraction $1-\theta$ always attend the closest hospital for treatment, regardless of the treatment qualities offered by the two hospitals. Thus, the parameter θ measures the degree of *patient choice* in the hospital market.[2] For simplicity, we assume that θ is equal for high- and low-severity patients.

Under the assumption of unit demand and full market coverage, utility-maximising behaviour leads to the following demand functions for high- and low-severity patients, respectively, at Hospital i:

$$D_i^h := \lambda\left(\frac{1-\theta}{2}+\theta\left(\frac{1}{2}+\frac{B^h\left(q_i\right)-B^h\left(q_j\right)}{2t}\right)\right), \tag{2}$$

$$D_i^l := (1-\lambda)\left(\frac{1-\theta}{2}+\theta\left(\frac{1}{2}+\frac{B^l\left(q_i\right)-B^l\left(q_j\right)}{2t}\right)\right), \tag{3}$$

where $i=1,2$, $j=1,2$, and $i\neq j$. Total demand for Hospital i is then

$$D_i = D_i^h + D_i^l, \tag{4}$$

while total demand for Hospital j is $D_j = 1-D_i$.

Each hospital is assumed to maximise profits. Under the assumption that the (regulated) price p is the same for both types of patients (e.g., DRG tariff for a coronary bypass)[3], profits of Hospital i are given by

$$\pi_i = \left(p-c_i^h(q_i)\right)D_i^h + \left(p-c_i^l(q_i)\right)D_i^l - C(q_i), \tag{5}$$

where $c_i^k(q_i)$ is the unit cost of treating a patient with severity k, and $C(q_i)$ is the fixed (i.e., output independent) cost of quality (e.g., MRI machines). We assume that the output-independent cost of quality increases with quality at an increasing rate, $\partial C/\partial q_i > 0$ and $\partial^2 C/\partial q_i^2 > 0$, that the unit cost of treatment increases (weakly) with quality, $\partial c_i^k\left(q_i\right)/\partial q_i \geq 0$, and that the cost of treating a high-severity patient is (weakly) higher than the cost of treating a low-severity patient, $c_i^h\left(q_i\right) \geq c_i^l\left(q_i\right)$ for all q_i. We also assume that hospitals differ in unit treatment costs, with Hospital 1 having a

[2] We give more specific interpretations of θ in Section 3.

[3] Our model applies to a given treatment (e.g. hip replacement, coronary bypass) when patients vary in severity and differ in their ability to recover their health due to pre-existing conditions, degree of frailty, pain in motion or at rest, previous heart attacks etc. We treat the DRG tariff as fixed across severity types. In practice, DRG tariffs are in some cases split based on observable patients characteristics, but there are rather crude since DRGs vary within a treatment based on age thresholds (e.g. over 67 years old) and whether the patient has complications (mostly ex-post ones, not before the surgery). Therefore, there remains extensive heterogeneity in severity within a DRG.

cost advantage: $c_1^k(q_1) < c_2^k(q_2)$ and $\partial c_1^k(q_1)/\partial q_1 \leq \partial c_2^k(q_2)/\partial q_2$ for $q_1 = q_2$.

The hospitals simultaneously choose qualities in a non-cooperative one-shot game. We consider an interior-solution Nash equilibrium in which both hospitals choose treatment quality above the minimum level. This Nash equilibrium is implicitly characterised by a pair of first-order conditions, given by[4]

$$\frac{\partial \pi_i}{\partial q_i} = \sum_k \left(p - c_i^k(q_i^*)\right) \frac{\partial D_i^k\left(q_i^*, q_j^*\right)}{\partial q_i} - \sum_k \frac{\partial c_i^k(q_i^*)}{\partial q_i} D_i^k\left(q_i^*, q_j^*\right) - \frac{\partial C(q_i^*)}{\partial q_i} = 0, \tag{6}$$

where

$$\frac{\partial D_i^h}{\partial q_i} = \frac{\lambda\theta}{2t}\frac{\partial B^h}{\partial q_i}; \quad \frac{\partial D_i^l}{\partial q_i} = \frac{(1-\lambda)\theta}{2t}\frac{\partial B^l}{\partial q_i}. \tag{7}$$

Given our assumptions on the hospitals' cost functions, the Nash equilibrium is asymmetric and the hospital with a cost advantage provides a higher quality, $q_1^* > q_2^*$.

3 Competition and quality differences

What is the effect of competition on quality provision? In particular, does fiercer competition reduce or amplify quality differences between the hospitals? We measure the degree of competition by the degree of patient choice, which is given by the parameter θ. A higher degree of patient choice implies that the demand facing each hospital becomes more quality elastic, which yields stronger incentives to compete for patients by providing a higher quality of treatment.

Our competition parameter θ can be given two different interpretations related to relevant policy measures for stimulating hospital competition. First, θ can be interpreted as a measure of the degree of information about quality in the market. More specifically, let θ be the share of patients who are informed about the treatment quality offered by the two hospitals, and thus make choices based on both quality and travelling distance, whereas the remaining patients are uninformed about quality and choose to attend the closest hospital in order to minimise travelling costs. With this interpretation, the degree of patient choice, and thus the degree of competition, is restricted by a lack of information about treatment quality. Consequently, competition can be stimulated by policies that increase the amount of information available to patients, such as public reporting of quality indicators. In our model, such a policy would be captured by an increase in θ.

[4] Second-order and stability conditions are given in the Appendix.

Alternatively, competition could be hampered by administrative restrictions to patient choice, with θ measuring the share of patients who are allowed to choose which hospital to attend for treatment. Thus, a policy of introducing free patient choice in a hospital market (where, previously, all patients were administratively allocated to the closest hospital) would be captured by a discrete increase in θ from 0 to 1.

3.1 Competition and quality provision

In a symmetric model with profit-maximising providers and regulated prices, there is a well-established positive relationship between increased patient choice and equilibrium quality provision (as long as the providers have positive price-cost margins). In our asymmetric setting, however, increased patient choice has additional effects on unilateral quality provision incentives. On the one hand, as in a symmetric model, increased patient choice makes demand more quality elastic, which gives both hospitals an incentive to increase quality. On the other hand, for given quality levels, increased patient choice implies that a larger share of each patient type chooses the high-quality hospital. If unit treatment costs increase with quality, $\partial c_i^k/\partial q_i > 0$, such a reallocation of demand implies higher (lower) marginal cost of quality provision, and therefore weaker (stronger) incentives for quality provision, for the high-quality (low-quality) hospital. However, by applying the first-order conditions, (6), it can be shown (see Appendix) that the former effect dominates the latter, implying that the results from a symmetric model also carry over to an asymmetric one. Increased patient choice leads to higher quality provision in equilibrium for both hospitals: $\partial q_i^*/\partial\theta > 0$, $i = 1, 2$.

3.2 Does competition lead to quality dispersion or quality convergence?

In the present study, we are foremostly interested in whether increased competition amplifies or reduces equilibrium quality differences, defined by $\Delta := q_1^* - q_2^*$. If $\partial\Delta/\partial\theta > 0$, competition leads to quality *dispersion*, whereas, if $\partial\Delta/\partial\theta < 0$, competition leads to quality *convergence*. Using (A5)-(A6) in the Appendix, the effect of increased patient choice on equilibrium quality differences is given by

$$\frac{\partial\Delta}{\partial\theta} = \frac{1}{H}\left[\frac{\partial^2\pi_2}{\partial\theta\partial q_2}\left(\frac{\partial^2\pi_1}{\partial q_1^2} + \frac{\partial^2\pi_1}{\partial q_2\partial q_1}\right) - \frac{\partial^2\pi_1}{\partial\theta\partial q_1}\left(\frac{\partial^2\pi_2}{\partial q_2^2} + \frac{\partial^2\pi_2}{\partial q_1\partial q_2}\right)\right], \tag{8}$$

where $H > 0$, $\partial^2\pi_i/\partial\theta\partial q_i > 0$, $\partial^2\pi_i/\partial q_j\partial q_i \geq 0$ and $\partial^2\pi_i/\partial q_i^2 < 0$. Further details are given in the Appendix.

Proposition 1 *The effect of increased competition on the equilibrium quality difference between the hospitals is generally ambiguous and depends crucially on three different factors: (i) the degree of concavity of the health benefit function, (ii) the difference in equilibrium price-cost margins between the two hospitals, and (iii) the degree to which quality affects unit treatment costs.*

The general ambiguity is proved by the parametric examples given in the next subsection, and the importance of (i)-(iii) in Proposition 1 is established by comparing how increased competition affects (all else equal) the marginal profitability of quality provision for each of the two providers. This comparison is given by

$$
\begin{aligned}
\frac{\partial^2 \pi_1}{\partial \theta \partial q_1} - \frac{\partial^2 \pi_2}{\partial \theta \partial q_2} \quad = \quad & \frac{\lambda}{2t}\left[\left(p - c_1^h\right)\frac{\partial B^h\left(q_1^*\right)}{\partial q_1} - \left(p - c_2^h\right)\frac{\partial B^h\left(q_2^*\right)}{\partial q_2}\right] \\
& + \frac{1-\lambda}{2t}\left[\left(p - c_1^l\right)\frac{\partial B^l\left(q_1^*\right)}{\partial q_1} - \left(p - c_2^l\right)\frac{\partial B^l\left(q_2^*\right)}{\partial q_2}\right] \\
& - \left(\frac{B^h\left(q_1^*\right) - B^h\left(q_2^*\right)}{2t}\right)\left[\lambda\left(\frac{\partial c_2^h}{\partial q_2} + \frac{\partial c_1^h}{\partial q_1}\right) + (1-\lambda)\left(\frac{\partial c_2^l}{\partial q_2} + \frac{\partial c_1^l}{\partial q_1}\right)\right].
\end{aligned}
\tag{9}
$$

This expression consists of three terms. The sign of the first two terms is *a priori* ambiguous, whereas the third term is unambiguously negative, and the overall sign of (9) depends on each of the three factors highlighted in Proposition 1.

(i) *Concavity of the health benefit function.* A strictly concave health benefit function implies that the marginal health gain of quality is higher for patients in the low-quality hospital, which in turn implies that demand responds more strongly to quality for this hospital. Increased competition will therefore lead to a larger increase in the marginal revenue of quality provision for Hospital 2 than for Hospital 1, contributing, all else equal, to quality *convergence* between the two hospitals. Formally, this effect is reflected by $\partial B^k\left(q_1^*\right)/\partial q_1 < \partial B^k\left(q_2^*\right)/\partial q_2$ in the first two terms of (9).

(ii) *Difference in price-cost margins.* Increased patient choice implies that demand becomes more responsive to quality, which increases the marginal revenue of quality and gives both hospitals an incentive to increase quality. If $c_1^k\left(q_1^*\right) < c_2^k\left(q_2^*\right)$, the profit margin is higher for Hospital 1, which implies that the increase in marginal revenue of quality provision, due to more quality-responsive demand, is also higher for Hospital 1, which gives this hospital a stronger incentive to increase quality. This effect contributes, all else equal, to quality *dispersion*. Thus, (ii) counteracts (i) and the sign of each of the first two terms in (9) depends on the relative strength of these two effects.

(iii) *Quality-dependent unit treatment costs.* The increase in demand responsiveness due to increased patient choice also implies that, for given qualities, demand is shifted towards the high-quality hospital (i.e., $\partial D_1/\partial\theta > 0$ and $\partial D_2/\partial\theta < 0$). If unit treatment costs depend on quality, the demand increase (decrease) for the high-quality (low-quality) hospital implies that the marginal cost of quality provision increases (decreases) for the high-quality (low-quality) hospital. All else equal, this gives the high-quality (low-quality) hospital an incentive to reduce (increase) quality. This effect works in the same direction as (i) and contributes, all else equal, to quality *convergence*.

In addition to the difference in the marginal profitability of quality provision between the hospitals, which is determined by the relative strengths of (i)-(iii) as discussed above, the overall effect of competition on the equilibrium quality difference also depends on differences in the curvature of the profit functions (evaluated at the equilibrium point), and on feedback effects related to the strategic interaction between the hospitals, as reflected by the remaining terms in the square brackets of (8).[5] Our conjecture is that these additional effects are likely to be of secondary order. Notice also that the presence of feedback effects relies on the existence of quality-dependent unit treatment costs.[6]

3.3 Parametric examples

In order to further illustrate the main mechanisms at play, we will consider two different parameterisations of the health benefit and unit treatment cost functions. In both examples, we assume that the fixed cost of quality provision is quadratic, $C(q_i) = (k/2)q_i^2$. For expositional simplicity, and without any significant loss of generality, we also assume that unit treatment costs at Hospital i are equal for both severity types, i.e., $c_i^h(q_i) = c_i^l(q_i) = c_i(q_i)$.

3.3.1 Decreasing marginal health gains and constant unit treatment costs

Suppose that unit treatment costs are constant and given by $c_1 < c_2$, and that the health benefit function is given by

$$B^k(q_i) = \alpha_k + \beta_k q_i - \frac{\gamma}{2} q_i^2, \tag{10}$$

[5] The effects that go through changes in the magnitudes of $\partial^2\pi_1/\partial q_1^2$ and $\partial^2\pi_2/\partial q_2^2$ depend on the signs of the third-order derivatives of $B^k(\cdot)$, $c^k(\cdot)$ and $C(\cdot)$, and are thus hard to interpret.

[6] From (A7)-(A8) in the Appendix we see that $\partial^2\pi_i/\partial q_j\partial q_i = 0$ if $\partial c_i^k(q_i)/\partial q_i = 0$.

where $\alpha_l > \alpha_h$ and $\beta_l > \beta_h$. In this case, equilibrium qualities are given by

$$q_i^* = \frac{\theta\overline{\beta}(p - c_i)}{2kt + \theta\gamma(p - c_i)}, \tag{11}$$

where $\overline{\beta} := \lambda\beta_h + (1 - \lambda)\beta_l$. The effect of increased competition on the quality difference is given by

$$\begin{aligned} \frac{\partial\Delta}{\partial\theta} &= \frac{2kt(c_2 - c_1)\overline{\beta}\left(4k^2t^2 - \theta^2\gamma^2(p - c_1)(p - c_2)\right)}{(2kt + \theta\gamma(p - c_1))^2(2kt + \theta\gamma(p - c_2))^2} > (<) 0 \\ \text{if } \gamma &< (>) \frac{2kt}{\theta\sqrt{(p - c_1)(p - c_2)}}. \end{aligned} \tag{12}$$

Under the assumption of constant unit treatment costs, qualities are strategically independent. Thus, with a general health benefit function, the sign of $\partial\Delta/\partial\theta$ will be largely determined by the sign of (9), which in turn is determined by the relative strength of (i) and (ii) as defined in the previous subsection. The result given in (12) confirms this intuition. Increased competition leads to quality convergence if the degree of concavity of the health benefit function (measured by the parameter γ) is sufficiently large, which implies that (i) outweighs (ii). Otherwise, if γ is sufficiently low, increased competition leads to quality dispersion.

3.3.2 Constant marginal health gains and quality-dependent unit treatment costs

Suppose that $\gamma = 0$ in (10), such that marginal health benefits of quality are constant, and suppose also that unit treatment costs are given by $c(q_i) = c_i q_i$, where $c_1 < c_2$. In this case, equilibrium qualities are given by

$$q_i^* = \frac{p\theta^2\overline{\beta}^2(c_i + 2c_j) + t\theta\overline{\beta}(2kp - 3c_ic_j) - 2kt^2c_i}{4kt\left(kt + \theta\overline{\beta}(c_i + c_j)\right) + 3\theta^2\overline{\beta}^2c_ic_j}, \tag{13}$$

and the effect of increased competition on the difference in treatment qualities is given by

$$\begin{aligned} \frac{\partial\Delta}{\partial\theta} &= 4kt(c_2 - c_1)\overline{\beta}\frac{\left(p\theta^2\overline{\beta}^2 - 2kt^2\right)(c_1 + c_2) + t\theta\overline{\beta}(2kp - 3c_1c_2)}{\left(4kt\left(kt + \theta\overline{\beta}(c_1 + c_2)\right) + 3\theta^2\overline{\beta}^2c_1c_2\right)^2} > (<) 0 \\ \text{if } p &> (<) \frac{2kt^2(c_1 + c_2) + 3t\theta\overline{\beta}c_1c_2}{\theta\overline{\beta}\left(\theta\overline{\beta}(c_1 + c_2) + 2kt\right)}. \end{aligned} \tag{14}$$

The assumption of constant marginal health benefits of quality eliminates the first of the three factors identified in Proposition 1, such that the sign of (9) is determined by the relative strength of (ii) and (iii). In addition, the assumption that unit treatment costs depend on quality implies that there are additional feedback effects caused by strategic interaction between the hospitals. In this parametric example we see that increased competition leads to quality dispersion (convergence) if the price p is sufficiently high (low). All else equal, a higher (lower) price increases (reduces) the difference in price-cost margins between the two hospitals, increasing (reducing) the relative strength of (ii), which contributes to quality convergence (dispersion) as a result of more competition. Thus, the result given by (14) is consistent with our general analysis of the main mechanisms identified in Proposition 1.

4 Competition and health inequalities

In the previous section we have identified the main mechanisms that cause competition to induce either a reduction or an increase in inequalities in the level of *quality* across hospitals, which we have referred to as quality convergence and quality dispersion, respectively.

In this section we investigate how competition affects *health* inequalities. In our model we have four groups of patients who differ in severity and the provider from which they receive the treatment, and we answer this question in three steps. First, we look at inequalities in health outcomes across hospitals. These can be thought of as inequalities arising from the 'postcode lottery': some patients will have worse health outcomes than others simply because they live closer to a low-quality hospital, what we refer to as *postcode inequality.* Second, we look at inequalities in health outcomes between patients with high and low severity, and check whether competition increases or reduces the health gap between the two patient groups. Third, we look at aggregate measures of (relative and absolute) health inequality based on the Gini and Generalised Gini coefficients, since these have been commonly used in the health economics empirical literature to measure health inequalities.

4.1 Absolute health inequalities across hospitals (postcode inequality)

When considering health inequalities across hospitals, we restrict attention to inequalities within each patient type. As long as health outcomes (e.g., mortality rates) are risk adjusted, the analysis would be similar in the presence of patients with different severities. For a given level of severity, the difference in health outcomes of patients being treated at Hospital 1 and 2, respectively, is given by

$$\Omega^k := B^k\left(q_1^*\right) - B^k\left(q_2^*\right),\ \ k = h, l, \text{ with } \Omega^h > \Omega^l. \tag{15}$$

The effect of competition on health inequalities is consequently given by

$$\frac{\partial \Omega^k}{\partial \theta} = \frac{\partial B^k}{\partial q_1}\frac{\partial q_1^*}{\partial \theta} - \frac{\partial B^k}{\partial q_2}\frac{\partial q_2^*}{\partial \theta}. \tag{16}$$

If competition induces quality *convergence*, health inequalities across hospitals are also reduced. If the marginal health gain from quality is constant, inequalities are driven by differences in quality. This effect is reinforced if the marginal health gain from quality is decreasing and therefore smaller in the hospital with higher quality. Reductions in inequalities in quality always reduce health inequalities.

If competition induces quality *dispersion*, the effect on health inequalities is instead ambiguous. It is only when the health gain from quality is linear or not too concave that inequalities in levels of quality go hand-in-hand with health inequalities, so that quality dispersion increases postcode inequality. If the marginal health gain from quality is decreasing, the larger quality increase in Hospital 1 arising from competition can be dampened or even offset by the smaller marginal health gain of quality, and quality dispersion can therefore reduce postcode inequality.

Proposition 2 *(i) If competition induces quality convergence, then it reduces health inequalities across hospitals for each severity type. (ii) If competition induces quality dispersion, it increases health inequalities when the health gain from quality is not too concave in quality; it reduces health inequalities when the health gain from quality is sufficiently concave. (iii) If competition has no effect on quality differences across hospitals, it reduces health inequalities if the marginal health gain from quality is decreasing.*

The second part of the Proposition 2 can be illustrated by considering the parameterisation used in Section 3.3.1. Since we are considering the effect of competition on health inequalities for a given severity type, the exposition is simplified by setting $\lambda = 1$. With a quadratic health benefit function, the effect of increased competition on health inequalities is given by

$$\frac{\partial \Omega^h}{\partial \theta} = \frac{4k^2t^2\beta_h^2\left(c_2 - c_1\right)\left[8k^3t^3 - \theta^2\gamma^2\left(p - c_2\right)\left(p - c_1\right)\left(\theta\gamma\left(2p - c_1 - c_2\right) + 6kt\right)\right]}{\left(\theta\gamma\left(p - c_1\right) + 2kt\right)^3\left(\theta\gamma\left(p - c_2\right) + 2kt\right)^3}. \tag{17}$$

It is straightforward to see that the sign of (17) is positive (negative) if γ is sufficiently low (high).

There are two different forces at play here, both of which work in the same direction. For a given increase in quality provision by each hospital, a higher value of γ (which implies a more concave benefit function) contributes directly towards less inequality in health outcomes. In addition, a more concave benefit function also increases the scope for quality convergence as a result of increased patient choice, as shown by (12).

Consider an illustrative numerical example, with $p = k = t = 1$, $c_2 = \theta = \frac{1}{2}$ and $c_1 = \frac{1}{4}$, which yields the following effects of increased patient choice on quality differences and health inequalities:

$$\begin{array}{ll}
\text{(i) } \gamma < 3.26: & \frac{\partial \Delta}{\partial \theta} > 0 \text{ and } \frac{\partial \Omega}{\partial \theta} > 0. \\
\text{(ii) } 3.26 < \gamma < 6.53: & \frac{\partial \Delta}{\partial \theta} > 0 \text{ and } \frac{\partial \Omega}{\partial \theta} < 0. \\
\text{(iii) } \gamma > 6.53: & \frac{\partial \Delta}{\partial \theta} < 0 \text{ and } \frac{\partial \Omega}{\partial \theta} < 0.
\end{array}$$

The interesting case is (ii). When the degree of concavity is in an intermediate range, increased patient choice leads to quality dispersion but simultaneously reduces health inequalities within each severity group, because marginal health gains from quality is decreasing at a sufficiently high rate.

4.2 Absolute health inequalities between high- and low-severity patients

In this sub-section we investigate how competition affects health inequalities across patient severity. These could be due to patients differing in severity within the same condition or across conditions. For example, for patients who had a heart attack (within the same health condition), high severity patients have a history of heart conditions or other comorbidities. Across conditions, we could think of high-severity patients as patients with cancer as opposed to patients in need of a cataract surgery (low-severity patients).[7]

The average (or expected) health outcome for a high-severity patient is given by

$$\overline{B}^h = \frac{1}{\lambda}\left(D_1^h B^h\left(q_1^*\right) + \left(\lambda - D_1^h\right) B^h\left(q_2^*\right)\right), \tag{18}$$

which can be re-written as

$$\overline{B}^h = \frac{B^h\left(q_1^*\right) + B^h\left(q_2^*\right)}{2} + \frac{\theta\left(\Omega^h\right)^2}{2t}. \tag{19}$$

[7] Although our model has only one price, and therefore implicitly considers only one condition, the effects of competition on health inequalities would be similar in a model with more than one condition as long as the price differences across conditions remain constant.

The equivalent expression for a low-severity patient is

$$\overline{B}^l = \frac{B^l(q_1^*) + B^l(q_2^*)}{2} + \frac{\theta\left(\Omega^l\right)^2}{2t}. \tag{20}$$

Health inequalities between patient types can then be defined as $\Phi := \overline{B}^l - \overline{B}^h$. The effect of increased competition is given by

$$\frac{\partial \Phi}{\partial \theta} = -\frac{1}{2}\sum_{i=1}^{2}\left(\frac{\partial B^h}{\partial q_i} - \frac{\partial B^l}{\partial q_i}\right)\frac{\partial q_i^*}{\partial \theta} - \frac{1}{2t}\left[\left(\Omega^h\right)^2 - \left(\Omega^l\right)^2\right] - \frac{\theta}{t}\left(\Omega^h\frac{\partial \Omega^h}{\partial \theta} - \Omega^l\frac{\partial \Omega^l}{\partial \theta}\right). \tag{21}$$

The first term captures the effect of competition on health inequality for given patient allocations. Increased patient choice leads to higher quality provision at both hospitals. Since the marginal health gain of quality is larger for high-severity than for low-severity patients, the inequality in health outcomes between the two patient groups is reduced. Therefore, the first effect is unambiguously negative, and this is regardless of whether more competition induces quality convergence or quality dispersion.

The remaining terms capture the effects of changes in *patient composition* as a result of more competition. An increase in the degree of patient choice makes demand more sensitive to differences in quality between the two hospitals. For given quality levels, an increase in θ implies that a relatively larger share of high-severity patients will choose the high-quality hospital. The resulting effect on health inequality is captured by the second term in (21) and is also unambiguously negative. Once more, since the health gain of having access to higher quality of treatment is larger for high-severity than for low-severity patients, the above described patient reallocation will also reduce inequality in health outcomes across the two patient groups.

The last term in (21) capture the patient composition effects that are related to changes in quality provision as a result of more competition, and the overall sign of these effects is *a priori* indeterminate. Notice, however, that if the marginal benefit of quality is decreasing at a sufficiently low rate, the direction of this effect is uniquely determined by whether competition leads to quality dispersion or quality convergence. To see this, consider the extreme case of linear health benefits, which implies $\partial B^k/\partial q_1 = \partial B^k/\partial q_2 = \partial B^k/\partial q$. The last term in (21) can then be rewritten as

$$-\frac{\theta}{t}\left[\Omega^h\frac{\partial B^h}{\partial q} - \Omega^l\frac{\partial B^l}{\partial q}\right]\frac{\partial \Delta}{\partial \theta}. \tag{22}$$

By the assumption $\partial B^h/\partial q > \partial B^l/\partial q$, the expression in square brackets is positive. This implies that the patient composition effect through changes in quality provision also contributes in the direction of less health inequality if competition leads to quality *dispersion*. Thus, in the case of $\partial\Delta/\partial\theta > 0$, all patient composition effects (given by the second and third terms in (21)) go in the same direction. Since high-severity patients are more responsive than low-severity patients to quality differences, increased competition implies that the share of high-severity patients in the high-quality hospital will increase for given quality levels, and this effect is reinforced if competition leads to quality dispersion. As a result, the health inequality between these two groups of patients is reduced.

This analysis illustrates how increased disparities in quality across hospitals do not necessarily imply increased disparities in health outcomes across patient types. In the above example, with constant marginal health gains, the opposite holds. Since it is the most disadvantaged group, i.e., the high-severity patients, who benefit most from differences in qualities across hospitals, health inequalities are actually reduced. By continuity, this holds also for health benefit functions with a sufficiently low degree of concavity, which allows us to summarise the above derived results as follows:

Proposition 3 *(i) An increase in competition reduces inequalities across patients with different severity if the subsequent changes in patient composition at each hospital are sufficiently small. (ii) If the marginal health gain from quality is constant or decreases slowly with quality, a sufficient condition for increased competition to reduce inequalities across severity types is that competition leads to quality dispersion.*

4.3 Aggregate measures of (absolute and relative) health inequality

In the previous subsections we have studied the effect of competition on health inequalities along two different dimensions: (i) inequalities between patients treated at different hospitals (postcode inequality) and (ii) inequalities between high- and low-severity patients. An aggregate measure of inequality which allows to trade off inequalities along different dimensions is the Gini coefficient, which is also a function of the share of (high/low severity) patients who receive high and low quality. To illustrate the role of the latter we start out with a simplified framework with only one severity level, and then extend to two severity levels.

4.3.1 One severity level

With only one severity level, there are only two patient groups, those receiving high quality (at Hospital 1) and those receiving low quality (at Hospital 2). Using the notational short-hand $B_i := B(q_i^*)$, the Lorenz curve is given by

$$L(x) = \begin{cases} \frac{B_2}{\overline{B}} x & if \quad 0 \leq x \leq 1 - D_1 \\ -\frac{(B_1 - B_2)(1 - D_1)}{\overline{B}} + \frac{B_1}{\overline{B}} x & if \quad 1 - D_1 < x \leq 1 \end{cases}, \tag{23}$$

where $\overline{B} := D_1 B_1 + (1 - D_1) B_2$ is the average health outcome. The Gini coefficient is then given by

$$G = 1 - 2 \int_0^1 L(x)\, dx = 1 - \left(\frac{B_2 + D_1^2 (B_1 - B_2)}{\overline{B}} \right), \tag{24}$$

where

$$\frac{\partial G}{\partial B_1} = \frac{(1 - D_1) D_1 B_2}{\overline{B}^2} > 0, \tag{25}$$

$$\frac{\partial G}{\partial B_2} = -\frac{(1 - D_1) D_1 B_1}{\overline{B}^2} < 0 \tag{26}$$

and

$$\frac{\partial G}{\partial D_1} = -(B_1 - B_2) \frac{B_2 (2D_1 - 1) + D_1^2 (B_1 - B_2)}{\overline{B}^2} < 0. \tag{27}$$

All else equal, a marginal increase in the health outcome of patients at the high-quality (low-quality) hospital will increase (reduce) the Gini coefficient. Furthermore, an increase in the market share of the high-quality hospital – which initially has the larger market share – will reduce the Gini coefficient. Notice also that

$$\frac{\partial G}{\partial B_1} + \frac{\partial G}{\partial B_2} = -\frac{(1 - D_1) D_1}{\overline{B}^2} (B_1 - B_2) < 0. \tag{28}$$

Thus, a marginal increase in health outcome for all patients will, all else equal, reduce the Gini coefficient. This is a reflection of the Gini coefficient being a *relative* measure of inequality, which is reduced when all patients experience an equal absolute increase in health status.

We can convert the Gini coefficient to a measure of *absolute* inequality by multiplying G with the average health outcome, which yields the Generalised Gini coefficient:

$$\widetilde{G} := \overline{B}(q_1^*, q_2^*)\, G = D_1 (1 - D_1)(B_1 - B_2), \tag{29}$$

where

$$\frac{\partial \widetilde{G}}{\partial B_1} = -\frac{\partial \widetilde{G}}{\partial B_2} = D_1 (1 - D_1) > 0 \tag{30}$$

and

$$\frac{\partial \widetilde{G}}{\partial D_1} = -(B_1 - B_2)(2D_1 - 1) < 0. \tag{31}$$

As for the Gini coefficient, a higher market share for the high-quality hospital will also reduce absolute inequality, whereas a marginal improvement in the health status of patients at the high-quality (low-quality) hospital will increase (reduce) absolute inequality. However, for given patient allocations between the two hospitals, an equal absolute increase in the health status of all patients has no effect on absolute inequality (i.e., $\partial \widetilde{G}/\partial B_1 + \partial \widetilde{G}/\partial B_2 = 0$).

The effect of increased competition on *absolute inequality*, as given by the Generalised Gini coefficient, can be expressed as

$$\begin{aligned} \frac{\partial \widetilde{G}}{\partial \theta} &= -(B_1 - B_2)(2D_1 - 1)\left(\frac{1}{2} + \frac{B_1 - B_2}{2t}\right) \\ &\quad + \left[D_1 (1 - D_1) - \frac{\theta}{2t}(2D_1 - 1)(B_1 - B_2)\right] \frac{\partial \Omega}{\partial \theta}. \end{aligned} \tag{32}$$

The overall effect is given by the sum of two effects that potentially go in opposite direction. The first effect is unambiguously negative. For given quality levels, increased patient choice implies a reallocation of patients towards the high-quality hospital. This effect, which is captured by the first line in (32), contributes to lower absolute inequality.

The second effect is related to the effect of competition on postcode inequality (given by $\partial\Omega/\partial\theta$). This effect is captured by the second line in (32) and is *a priori* ambiguous. Suppose that more competition leads to increased postcode inequality ($\partial\Omega/\partial\theta > 0$), which implies a reallocation of patients towards the high-quality hospital ($\partial D_1/\partial\theta > 0$). This has two counteracting effects on the Generalised Gini coefficient, given by the two terms in square brackets in the second line of (32). One the one hand, for given market shares, absolute inequality increases because of increased inequality in health outcomes. However, the reallocation of patients towards the high-quality hospital implies that a lower share of patients experience low quality, which reduces the Generalised Gini coefficient. The relative strength of these two effects depends on the initial quality difference. If the quality difference is small, so that D_1 is close to $\frac{1}{2}$ and B_1 close to B_2, then the first effect dominates and a dispersion in health outcomes increases absolute inequality. On the other hand, if the quality difference is very large, so that D_1 is close to 1, the second effect dominates and further dispersion in health outcomes actually reduces absolute inequality.

The effect of increased competition on *relative inequality*, as measured by the Gini coefficient, can be expressed as

$$\begin{aligned}\frac{\partial G}{\partial \theta} &= -\left(B_1 - B_2\right)\frac{B_2\left(2D_1 - 1\right) + D_1^2\left(B_1 - B_2\right)}{\overline{B}^2}\left(\frac{1}{2} + \frac{B_1 - B_2}{2t}\right) \\ &\quad + \frac{1}{\overline{B}^2}\left[\left(1 - D_1\right)D_1 B_2 - \frac{\theta}{2t}\left(B_1 - B_2\right)\left(B_2\left(2D_1 - 1\right) + D_1^2\left(B_1 - B_2\right)\right)\right]\frac{\partial \Omega}{\partial \theta} \\ &\quad - \frac{\left(1 - D_1\right)D_1\left(B_1 - B_2\right)}{\overline{B}^2}\frac{\partial B_2}{\partial q_2}\frac{\partial q_2}{\partial \theta}.\end{aligned} \tag{33}$$

The first two lines in (33) are completely equivalent to (32) and contain the two effects described above. However, the third line captures an effect that is specific to the Gini coefficient and reflects the fact that G measures *relative* inequality. This is a pure *level effect* and is unambiguously negative. Even if more competition does not lead to any patient reallocations and does not affect the difference in health outcomes across hospitals, such that the sum of the first two effects is zero, the resulting higher quality at both hospitals nevertheless reduces the relative health inequality between the two patient groups.

We summarise the above analysis as follows:

Proposition 4 *Suppose that all patients have the same severity level.*

(i) If competition leads to a dispersion (convergence) of health outcomes between the two hospitals, this will, all else equal, contribute towards an increase (reduction) in absolute and relative inequality if the initial quality difference is sufficiently small, and towards a reduction (increase) in absolute and relative inequality if the initial quality difference is sufficiently large.

(ii) If competition has a sufficiently small effect on the difference in health outcomes between the two hospitals, more competition will reduce both absolute and relative inequality.

4.3.2 Two severity levels

The previous analysis with one severity level can be seen as an approximation of the case where severity differences are small relative to quality differences between the hospitals, such that a patient treated at the high-quality hospital always has a better health outcome than a patient treated at the low-quality hospital, regardless of severity.

Consider now the opposite, that severity differences are large relative to quality differences, in the sense that the health outcome is always better for a low-severity patient than for a high-severity patient, regardless of which hospital the patient is treated at. Thus, and using again the notational short-hand $B_i^k := B^k(q_i^*)$, suppose that $B_1^l > B_2^l > B_1^h > B_2^h$. In this case, the Lorenz curve is given by

$$L(x) = \begin{cases} \frac{B_2^h}{\overline{B}}x & if \quad 0 \le x \le \lambda - D_1^h \\ -\frac{\left(B_1^h - B_2^h\right)\left(\lambda - D_1^h\right)}{\overline{B}} + \frac{B_1^h}{\overline{B}}x & if \quad \lambda - D_1^h < x \le \lambda \\ -\frac{\left(B_1^h - B_2^h\right)\left(\lambda - D_1^h\right) + \left(B_2^l - B_1^h\right)\lambda}{\overline{B}} + \frac{B_2^l}{\overline{B}}x & if \quad \lambda < x \le 1 - D_1^l \\ -\frac{\left(B_1^h - B_2^h\right)\left(\lambda - D_1^h\right) + \left(B_2^l - B_1^h\right)\lambda + \left(B_1^l - B_2^l\right)\left(1 - D_1^l\right)}{\overline{B}} + \frac{B_1^l}{\overline{B}}x & if \quad 1 - D_1^l < x \le 1 \end{cases}, \tag{34}$$

where $\overline{B} := \left(\lambda - D_1^h\right) B_2^h + D_1^h B_1^h + \left(1 - \lambda - D_1^l\right) B_2^l + D_1^l B_1^l$ is average health outcome. The Gini coefficient is given by

$$G = 1 - \left(\frac{B_2^h + \left(1 - \lambda + D_1^h\right)^2 \left(B_1^h - B_2^h\right) + (1 - \lambda)^2 \left(B_2^l - B_1^h\right) + \left(D_1^l\right)^2 \left(B_1^l - B_2^l\right)}{\overline{B}} \right), \tag{35}$$

whereas the Generalised Gini coefficient is given by

$$\widetilde{G} = D_1^l \left(1 - D_1^l\right) \left(B_1^l - B_2^l\right) + D_1^h \left(2\lambda - 1 - D_1^h\right) \left(B_1^h - B_2^h\right) + \lambda (1 - \lambda) \left(B_2^l - B_2^h\right). \tag{36}$$

Competition can affect absolute and relative inequality along three main dimensions:

1. Competition can affect postcode inequalities. For given patient allocations, this effect is described in Proposition 2.

2. Competition can affect inequalities between high- and low-severity patients. This effect is described in Proposition 3.

3. Competition can affect the relative shares of different patient groups, as highlighted by the analysis in the previous subsection, which is summarised in Proposition 4.

For the case of one severity level, the effects along the third dimension listed above are straightforward. If competition leads to patient reallocation towards the high-quality (low-quality) hospital, this will – all else equal – contribute to lower (higher) inequality. For the case of two severity types, which implies four different patient groups, the effects along this dimension are somewhat more complicated. To illustrate this, consider the effect on absolute inequality of patient reallocation towards the high-quality hospital. From (36) we derive

$$\frac{\partial \widetilde{G}}{\partial D_1^h} = -\left(B_1^h - B_2^h\right)\left(2\left(D_1^h - \lambda\right) + 1\right) < 0. \tag{37}$$

and

$$\frac{\partial \widetilde{G}}{\partial D_1^l} = \left(B_1^l - B_2^l\right)\left(1 - 2D_1^l\right) < (>)\, 0 \quad if \quad D_1^l > (<)\, \frac{1}{2}. \tag{38}$$

A reallocation of *high-severity* patients towards the high-quality hospital implies a reallocation of patients from the group with the worst health outcome to the group with the second-worst outcome. This will always reduce inequality. However, a reallocation of *low-severity* patients towards the high-quality hospital, which implies a reallocation of patients from the group with the second-best health outcome to the group with the best health outcome, will reduce inequality only if the latter group constitutes more than half of all patients, which requires that the share of high-severity patients (λ) is very low.

The effects of increased competition on absolute and relative inequality are analytically given by some very involved expressions that yield limited additional insights. It is therefore more illustrative to display the effects by numerical simulations based on our previous parameterisations. Table 1 shows the effects of increased competition based on the parameterisation in Section 3.3.1, with a quadratic health benefit function. The main mechanisms and trade-offs are captured by considering four different numerical configurations, where we vary both the degree of concavity of the health benefit function and the share of high-severity patients.[8] Cases where the marginal health gain decreases at a low rate and at a high rate are shown in Panel A and Panel B, respectively. In each of these cases, we consider both $\lambda = \frac{1}{2}$ and $\lambda = 1$, where the latter assumption implies only one severity level and therefore removes effects related to inequalities between high- and low-severity patients.[9]

Consider first the case of $\lambda = 1$. When all patients have the same severity level, the effect of competition on absolute inequality (as measured by $\widetilde{G}$) is determined by changes in inequality along two different dimensions. On the one hand, higher (lower) postcode inequality contributes to higher (lower) absolute inequality, whereas, on the other hand, increased (reduced) market share

[8] Numerical values are chosen such that the Nash equilibrium is characterised by $B_1^l > B_2^l > B_1^h > B_2^h$, thus satisfying the basic underlying assumption in this subsection.

[9] The case of $\lambda = 1$ is obviously qualitatively equivalent to the case of $\lambda = 0$.

of the high-quality hospital contributes to lower (higher) absolute inequality. These two effects are always counteracting, as discussed in Section 4.3.1.

Table 1: Quadratic health function and constant unit treatment costs

Panel A: Marginal health gains decreasing at a low rate ($\gamma = 1$)

	$\lambda = \frac{1}{2}$						$\lambda = 1$			
	Δ	Ω^h	Ω^l	Φ	G	$\widetilde{G}$	Δ	Ω	G	$\widetilde{G}$
$\theta = \frac{1}{4}$	0.034	0.047	0.031	0.455	0.087	0.119	0.040	0.056	0.012	0.014
$\theta = \frac{1}{2}$	0.058	0.078	0.049	0.415	0.077	0.112	0.070	0.091	0.018	0.023
$\theta = \frac{3}{4}$	0.077	0.097	0.059	0.380	0.069	0.105	0.092	0.112	0.020	0.028
$\theta = 1$	0.091	0.110	0.064	0.348	0.062	0.098	0.109	0.125	0.021	0.031

Panel B: Marginal health gains decreasing at a high rate ($\gamma = 10$)

	$\lambda = \frac{1}{2}$						$\lambda = 1$			
	Δ	Ω^h	Ω^l	Φ	G	$\widetilde{G}$	Δ	Ω	G	$\widetilde{G}$
$\theta = \frac{1}{4}$	0.012	0.012	0.006	0.473	0.092	0.119	0.015	0.013	0.003	0.003
$\theta = \frac{1}{2}$	0.012	0.009	0.003	0.462	0.088	0.116	0.014	0.009	0.002	0.002
$\theta = \frac{3}{4}$	0.011	0.007	0.001	0.457	0.087	0.115	0.013	0.006	0.001	0.001
$\theta = 1$	0.009	0.005	0.001	0.453	0.086	0.114	0.011	0.004	0.001	0.001

Remaining parameter values: $c_1 = \frac{1}{4}$, $c_2 = \frac{1}{2}$, $p = t = k = \alpha_h = \beta_l = 1$, $\beta_h = \alpha_l = \frac{3}{2}$.

If the marginal health benefit of quality decreases at a low rate (Panel A), increased competition leads to quality dispersion ($\partial\Delta/\partial\theta > 0$), with a corresponding increase in postcode inequality ($\partial\Omega/\partial\theta > 0$). This effect is sufficiently strong to outweigh the effect resulting from a lower market share of the high-quality hospital, implying that absolute inequality increases ($\partial\widetilde{G}/\partial\theta > 0$). Increased competition also increases relative inequality (as measured by G), though the increase is relatively smaller for G than for $\widetilde{G}$ because of the aforementioned *level effect*, whereby higher quality in itself reduces relative inequality. However, this effect is not strong enough to prevent an increase also in relative inequality as a result of more competition ($\partial G/\partial\theta > 0$).

However, if the marginal health benefit of quality decreases at a high rate (Panel B), increased competition leads to quality convergence ($\partial\Delta/\partial\theta < 0$). The resulting decrease in postcode inequality ($\partial\Omega/\partial\theta < 0$) is even larger, since a high degree of health benefit concavity strongly reinforces the effect of quality convergence, as discussed in Section 4.1. Even if the market share of the high-quality hospital decreases, which in itself increases relative and absolute inequality, the reduction in postcode inequality is sufficiently strong to cause an overall reduction in both G and $\widetilde{G}$ as a result of more competition.

Consider next the case of $\lambda = \frac{1}{2}$. When patients differ in severity, the effect of competition on absolute and relative inequality is now determined also by changes in inequalities along a third dimension, namely inequalities between high- and low-severity types, as measured by Φ. We see that competition always reduces inequality along this dimension, regardless of whether competition leads to quality dispersion (Panel A) or quality convergence (Panel B). The reason is that high-severity patients benefit more from higher quality than low-severity patients, as discussed in Section 4.2.

If marginal health gains decrease at a low rate (Panel A), more competition has opposite effects on postcode inequality, which increases, and inequality between severity types, which decreases. However, the latter effect is sufficiently strong to cause a reduction in both absolute and relative inequality as a result of more competition. If marginal health gains decrease at a high rate (Panel B), more competition reduces inequality along both dimensions; inequalities in health outcomes decrease both across hospitals and across severity types. Once more, increased competition reduces both absolute and relative inequality.

Overall, these numerical examples suggest that increased competition tends to reduce absolute and relative health inequality, unless marginal health gains decrease at a low rate *and* differences in patient severity is not important (i.e., if λ is close to 1 or 0). An important driving force behind a negative relationship between competition and (aggregate) inequality is that competition reduces health inequalities between high- and low-severity patients, which tends to be a dominant mechanism for reducing absolute and relative inequality as long as the shares of high- and low-severity patients are not too uneven.

5 Inequality aversion

Up to this point, our analysis has been purely positive, in the sense that we have analysed how competition affects various dimensions of health inequality. In this section we bring our analysis into a normative context by introducing a policy objective function that incorporates concerns for health inequalities. This part of our analysis takes inspiration from Wagstaff (2002) who proposes a Health Achievement Index that captures the potential trade-off between mean health and health inequality in a single summary measure. This index is based on the notion of an *extended* Gini coefficient, which highlights the ethical judgements underpinning the Gini coefficient and allows for different degrees of inequality aversion.

5.1 Extended Gini coefficient

Following the approach by Wagstaff (2002), the *extended* Gini coefficient measures health inequalities as a function of inequality aversion,

$$G(v) = 1 - v(v-1) \int_0^1 (1-x)^{v-2} L(x)\, dx, \quad v > 1, \tag{39}$$

where v is a parameter which relates to inequality aversion and $L(x)$ is the Lorenz curve. The approach draws on Yitzhaki (1983) who developed the indicator in the context of income inequalities.[10]

Applying the extended Gini coefficient to our model, the underlying ethical judgements are more explicitly seen by using the following approximation (see O'Donnell et al., 2008, p. 112):

$$G(v) \approx 1 - \sum_{k=h,l;\ i=1,2} \frac{D_i^k B_i^k}{\overline{B}} v(1-R_i^k)^{v-1}, \tag{40}$$

where $D_i^k B_i^k / \overline{B}$ is the share of total health enjoyed by patient group (i,k), while R_i^k is the group's fractional rank (in terms of health status). Therefore, $v(1-R_i^k)^{v-1}$ can be interpreted as the weight given to group (i,k). For $v=1$, this weight is equal to one for all groups, which yields $G(1)=0$, implying that inequalities generate no concerns. For $v>1$, the weight is monotonically decreasing in R, implying that individuals with better health are given a lower weight. For $v=2$, the weight

[10] An alternative to the extended Gini coefficient is the Atkinson index (1970), but this has the disadvantage that it cannot represent the Gini coefficient as a special case. See Le Grand (1987) for an empirical application in relation to health inequalities.

is equal to one for $R = \frac{1}{2}$ and we recover as a special case the Gini coefficient, where patients with health above the median ($R > \frac{1}{2}$) have weights below one and those below the median ($R < \frac{1}{2}$) have weights above one. For values of inequality aversion given by $1 < v < 2$, some patients above the median have also a weight above one (in addition to those below the median), while for $v > 2$ some patients below the median have a weight below one (in addition to those above the median). For $v > 4$, all patients above the median have a weight below 0.5. In this case the extended Gini coefficient heavily discounts the health of patients belonging to the more advantaged groups.

With two severity levels, and under the assumption that $B_1^l > B_2^l > B_1^h > B_2^h$, as in Section 4.3.2, the extended Gini coefficient is given by (see Appendix for an explicit derivation):

$$G(v) = 1 - \left(\frac{B_2^h + \left(1 - \lambda + D_1^h\right)^v \left(B_1^h - B_2^h\right) + (1 - \lambda)^v \left(B_2^l - B_1^h\right) + \left(D_1^l\right)^v \left(B_1^l - B_2^l\right)}{\overline{B}} \right). \quad (41)$$

It is relatively straightforward to see that $G(v)$ is monotonically increasing in v, which is very intuitive, given the interpretation of this parameter. For a given distribution of health status across the four patient groups, a higher degree of inequality aversion (measured by v) implies a higher value of the extended Gini coefficient. It is also straightforward to confirm that (41) coincides with (35) for $v = 2$.

5.2 Health Achievement Index

As argued by Wagstaff (2002), policy makers are likely to care about both the mean health status of the population and the distribution of health across different patient groups, and are also likely willing to trade off increases in health inequality against average health improvements. A policy objective function that captures this potential trade-off is what Wagstaff (2002) defines as an index of 'health achievement', given by

$$I(v) = \overline{B}(1 - G(v)). \quad (42)$$

If $v = 1$, inequality considerations have no bearing on overall health achievement, which is then just given by the mean health status of the population (since $G(1) = 0$). However, for any $v > 1$ health inequalities are incorporated into the overall achievement index.

How does increased competition affect overall health achievement? If competition increases both average health and reduces health inequalities, then $I(v)$ will increase regardless of the inequality aversion parameter. However, if average health increases and health inequalities also increase, then the overall achievement index in (42) gives a tool to assess the trade-off between equity and efficiency. Substituting from (41) into (42), overall health achievement is in our model given by

$$I(v) = B_2^h + \left(1-\lambda+D_1^h\right)^v \left(B_1^h - B_2^h\right) + (1-\lambda)^v \left(B_2^l - B_1^h\right) + \left(D_1^l\right)^v \left(B_1^l - B_2^l\right). \tag{43}$$

Thus, overall health achievement is given by the health status of the patients in the worst-off group (B_2^h) plus the sum of the health differences between each group and the group ranked immediately above, where each of these differences is weighted by the relative share of patients whose health status is at least at the level of the better group in each comparison, raised to the power of v.

This implies that, all else equal, I increases if the distribution of patients is shifted towards groups with a better health status, and more so for a larger degree of inequality aversion (v). Furthermore, for a given patient distribution, $\partial I(v)/\partial B_i^k > 0$ for all i and k. Thus, for a given distribution of patients, an increase in the health status of any patient group will increase I.

The effect of increased competition (patient choice) on overall health achievement is given by

$$\begin{aligned}\frac{\partial I(v)}{\partial\theta} \quad = \quad & \frac{\partial B_2^h}{\partial\theta} + v\left(1-\lambda+D_1^h\right)^{v-1}\frac{\partial D_1^h}{\partial\theta}\left(B_1^h - B_2^h\right) + v\left(D_1^l\right)^{v-1}\frac{\partial D_1^l}{\partial\theta}\left(B_1^l - B_2^l\right) \\ & + \left(1-\lambda+D_1^h\right)^v\frac{\partial\Omega^h}{\partial\theta} + \left(D_1^l\right)^v\frac{\partial\Omega^l}{\partial\theta} + (1-\lambda)^v\frac{\partial\left(B_2^l - B_1^h\right)}{\partial\theta}.\end{aligned} \tag{44}$$

The first term, which is positive, captures the effect that competition increases the health status of the worst-off group. The second and third terms capture the effects of changes in patient composition across hospitals and are positive (negative) if competition leads to a higher (lower) market share for the high-quality hospital. The fourth and fifth terms are positive (negative) if competition leads to more (less) postcode inequality. As we know from Proposition 2, a necessary, but not sufficient, condition for these effects to be positive is that competition leads to quality dispersion. The final term in (44) is ambiguous and depends on whether the health inequality between the second and third ranked group increases or decreases. The term is negative if competition leads to quality dispersion but might be positive if competition leads to quality convergence (if the concavity of B is sufficiently low).

Let us briefly consider the special case of $v = 1$, in which (44) reduces to

$$\frac{\partial I(1)}{\partial \theta} = \left(\lambda - D_1^h\right) \frac{\partial B^h}{\partial q_2}\frac{\partial q_2}{\partial \theta} + D_1^h \frac{\partial B^h}{\partial q_1}\frac{\partial q_1}{\partial \theta} + \left(1 - \lambda - D_1^l\right) \frac{\partial B^l}{\partial q_2}\frac{\partial q_2}{\partial \theta} + D_1^l \frac{\partial B^l}{\partial q_1}\frac{\partial q_1}{\partial \theta} > 0. \tag{45}$$

Even if competition leads to quality convergence, and therefore potentially a reallocation of patients towards the low-quality hospital, the average health always increases. Consequently, in the absence of inequality aversion, increased competition always improves the health achievement.

While the exposition in (44) is useful to illustrate the various effects at play, more clear-cut results can be obtained by rearranging (44) as follows:

$$\begin{aligned} \frac{\partial I(v)}{\partial \theta} = & \left[1 - \left(1 - \lambda + D_1^h\right)^v\right] \frac{\partial B^h}{\partial q_2}\frac{\partial q_2}{\partial \theta} + \left[\left(1 - \lambda + D_1^h\right)^v - (1-\lambda)^v\right] \frac{\partial B^h}{\partial q_1}\frac{\partial q_1}{\partial \theta} \\ & + \left[(1-\lambda)^v - \left(D_1^l\right)^v\right] \frac{\partial B^l}{\partial q_2}\frac{\partial q_2}{\partial \theta} + \left(D_1^l\right)^v \frac{\partial B^l}{\partial q_1}\frac{\partial q_1}{\partial \theta} \\ & + v\left[\left(1 - \lambda + D_1^h\right)^{v-1} \frac{\partial D_1^h}{\partial \theta}\left(B_1^h - B_2^h\right) + \left(D_1^l\right)^{v-1} \frac{\partial D_1^l}{\partial \theta}\left(B_1^l - B_2^l\right)\right] \end{aligned} \tag{46}$$

All but the last term (third line) in (46) are unambiguously positive for all $v \geq 1$, whereas the sign of the last term is uniquely determined by whether competition increases or reduces the market share of the high-quality hospital, which in turn depends, in part, on whether competition increases or reduces postcode inequality. This allows us to arrive at the following general result:

Proposition 5 *(i) Regardless of the degree of inequality aversion, a sufficient, but not necessary, condition for increased competition to improve the overall health achievement is that the market share of the high-quality hospital (for each severity type) does not decrease. (ii) Two necessary, but not sufficient, conditions for competition to reduce overall health achievement is that the market share of the high-quality hospital decreases and that the degree of inequality aversion is sufficiently large.*

As long as the market share of the high-quality hospital does not decrease, competition always leads to a better overall health achievement, regardless of the degree of inequality aversion. The effect is even stronger if competition increases the market share of the high-quality hospital, since this shifts the distribution of patients towards groups with a better health status, which reduces the (extended) Gini coefficient.

What does it take for competition to increase the market share of the high-quality hospital?

From (2)-(3) we derive

$$\frac{\partial D_i^k}{\partial \theta} = \frac{\mu_k}{2t}\left(B_1^k - B_2^k + \theta \frac{\partial \Omega^k}{\partial \theta}\right), \tag{47}$$

where $\mu_k \in \{\lambda, 1-\lambda\}$ is the share of type-k patients. Thus, competition will increase the market share of the high-quality hospital unless postcode inequality is sufficiently reduced. An *increase* in postcode inequality is therefore a sufficient (but far from necessary) condition for competition to improve overall health achievement.

On the other hand, if competition leads to a sufficiently strong *reduction* in postcode inequality, the effect on overall health achievement could potentially be negative, because of a redistribution of patients towards groups with worse health status. But this would require that the aversion towards inequality is sufficiently strong. From (45) we know that $\partial I/\partial \theta > 0$ for $v = 1$. By continuity, this result must hold also for values of v sufficiently close to 1.

6 Implications for empirical analyses

In this section we discuss possible approaches which could be pursued to test empirically how competition affects health inequalities. First, to test for the effect of competition on postcode inequality, researchers could compute measures of dispersion of health outcomes, such as the standard deviation or the coefficient of variation, within a given hospital catchment area and relate them to the degree of patient choice and market structure. For example, future empirical work could test whether in more competitive areas the introduction of patient choice policies lead to an increase or a reduction in AMI mortality dispersion across hospitals.

Second, to test for the effect of competition on health inequalities across severity levels, a subgroup analysis by degree of severity may be appropriate. In line with Kessler and Geppert (2005), high severity could be measured based on the number of previous hospital admissions preceding a health shock (such as AMI). By comparing the effect of competition on mortality for high- and low-severity patients, we can infer the effect on health inequalities across severity groups.

Third, the two types of inequality could be brought together by developing a Generalised Gini or Gini index in a given market area, where patients are ordered by their level of health, i.e., starting with patients with highest severity and lowest hospital quality and ending with patients with lowest severity and highest quality.

Our analysis also illustrates the importance of patient 'composition effects' when measuring the effect of competition due to patients with high and low severity exercising choice to a different degree. Competition affects differentially the health gains for patients with differing severity but also changes the number of patients receiving high and low quality through the composition effect. These will affect both the Gini coefficients and the simple measures of dispersion of health outcomes across hospitals.

The empirical literature which estimates patient choice models as a function of quality and severity tends to confirm that high-severity patients are more likely to choose high-quality hospitals. The elasticity of hospital demand to quality are however generally low and so are the interactions between quality and severity (see Brekke et al., 2014, Section 3.1, for a review of the evidence). We therefore conjecture that overall composition effects are likely to be small in empirical analyses.

Finally, our analysis highlights the importance of distinguishing empirically between quality and health outcomes. Although health outcomes are often used as a proxy of hospital quality, our study highlights how inequalities in qualities do not necessarily go hand-in-hand with inequalities in health outcomes. In relation to postcode inequality, an increase in inequalities in *quality* across hospitals is compatible with a reduction in *health* inequalities across hospitals if the marginal health gain is decreasing, so that patients in high-quality hospitals benefit less from a given quality increase than do patients in low-quality hospitals. Similarly, an increase in quality differences across hospitals is compatible with a reduction in health inequalities across severity types, and this is due to patients with higher severity benefiting more from the increase in quality compared to patients with lower severity.

7 Concluding remarks

Several OECD countries have introduced pro-market policy interventions in the health sector with the aim of stimulating quality of care. Such policies are generally contentious and the subject of an intense political debate. The existing literature has extensively investigated, both theoretically and empirically, the effect of competition on quality but there is very little work on its impact on equity. This is surprising given that reduction in health inequalities is an ubiquitous policy objective. Our study has contributed to fill this gap in knowledge by carefully characterising the conditions under which competition (i) increases or reduces the gap between high-quality and low-quality hospitals and, as a result, (ii) contributes to an increase or reduction in health inequalities.

Our first key finding is that the effect of competition on quality differences between high- and low-quality hospitals is generally ambiguous and depends on three key factors related to both demand and supply of health care: (i) decreasing marginal health gains from quality, which contributes in the direction of quality convergence, (ii) differences in price-cost margins between high- and low-quality hospitals, which contributes in the direction of quality dispersion, and (iii) quality-dependent unit treatment costs, which contributes in the direction of quality convergence. The relative strength of these factors are likely to vary by medical condition, diagnosis and treatment. For example, standardised treatments such as cataract surgery will have treatment costs mildly increasing with quality. This may not be the case for more serious treatments, such as a coronary bypass, where costs will increase more rapidly with quality.

Our second set of key findings is related to the effect of competition on two different types of health inequalities. First, changes in *postcode inequality* go hand in hand with changes in quality differences if health gains are linear in quality. However, these health inequalities are more likely to decrease – regardless of whether competition leads to quality dispersion or quality convergence – if marginal health gains are decreasing. On the other hand, we find that competition generally reduces health inequalities across patients with different *severity*, because high-severity patients benefit more from higher quality than do low-severity patients, although this reduction can be strengthened or weakened by changes in the patient composition at each hospital. Reductions in inequalities across severity types also drive reductions in the Gini and Generalised Gini coefficients, which aggregate different sources of health inequalities both across hospitals and severity types.

Our analysis also reveals that using simple measures of dispersion or aggregate measures of inequality can lead to different conclusions. In other words, when assessing the effect of competition on health inequality, the choice of inequality measure potentially matters. For example, even if competition increases differences in health outcomes across hospitals, the Generalised Gini coefficient may still reduce due to changes in patient composition at each hospital, and the Gini coefficient may reduce further as a result of the overall increase in quality.

In terms of policy implications, our analysis highlights that whether or not competition induces an equity-efficiency trade-off depends on the particular dimension of equity on which policy makers focus. For example, if policy makers focus on equity related to the postcode lottery, then an equity-efficiency trade-off may arise, though it is less likely to be the case if marginal health gains from quality decrease at a high rate. On the other hand, an equity-efficiency trade-off is less likely when considering equity across severity types, if more severe patients tend to benefit more than low-severity ones from increases in quality. If the equity-efficiency trade-off is represented by the Health Achievement Index proposed by Wagstaff (2002), which incorporates both of the aforementioned dimensions of equity, we show that competition always increases the overall health achievement as long as it does not lead to a redistribution of patients from high-quality to low-quality hospitals.

Our study also provides a theoretical framework to guide future empirical work, which should focus not only on testing the effect of competition on quality, but also its equity implications. This can be done, as discussed in Section 6, by developing measures of dispersions in quality and health outcomes within a given hospital catchment or market area, and then by relating these to changes in patient choice through consolidated econometric strategies.

The study assumes that quality is uniform within a hospital across different patients. This is justified because many aspects of quality involve investments in machines (e.g., CT and MRI) and technologies with fixed costs which benefit all patients. However, there may be some aspects of care (e.g., time that doctors spend with patients) which can vary by patient so that quality discrimination is possible (as in Ellis, 1998). We leave this possible extension for future research. Moreover, we assume that health inequalities go hand-in-hand with healthcare inequalities and rule out other sources of inequalities (e.g., due to location in rural versus urban areas, and socioeconomic status). In turn, this extension is likely to affect the empirical implications shifting the focus from *pure* health inequality to socioeconomic (or other) inequalities in health. Future work could include other sources of patient heterogeneity in the model, and derive its empirical implications.

References

[1] Atkinson, A.B., 1970. On the measurement of inequality. Journal of Economic Theory, 2, 244–263.

[2] Barigozzi, F., Burani, N., 2016. Competition and screening with motivated health professionals, Journal of Health Economics, 50, 358–371.

[3] Beitia, A., 2003. Hospital quality choice and market structure in a regulated duopoly. Journal of Health Economics, 22, 1011–1036.

[4] Bloom, N., Propper, C., Seiler, S., Van Reenen, J., 2015. The impact of competition on management quality: evidence from public hospitals. The Review of Economic Studies, 82, 457–489.

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

[6] Brekke, K.R., Nuscheler, R., Straume, O.R., 2007. Gatekeeping in health care. Journal of Health Economics, 26, 149–170.

[7] Brekke, K.R., Siciliani, L., Straume, O.R., 2011. Hospital competition and quality with regulated prices. Scandinavian Journal of Economics, 113, 444–469.

[8] Brekke, K.R., Siciliani, L., Straume, O.R., 2012. Quality competition with profit constraints. Journal of Economic Behavior & Organization, 84, 642–659.

[9] Brekke, K.R., Gravelle, H., Siciliani, L., Straume, O.R., 2014. Patient Choice, Mobility and Competition Among Health Care Providers, Chapter 1, 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.

[10] Calem, P.S., Rizzo, J.A., 1995. Competition and specialization in the hospital industry: an application of Hotelling's location model. Southern Economic Journal, 61, 1182–1198.

[11] Choné, P., C.A. Ma, 2011. Optimal health care contracts under physician agency. Annales d'Economie et de Statistique, 101/102, 229–256.

[12] Cookson, R., Laudicella, M., Li Donni, P., 2013. Does hospital competition harm equity? Evidence from the English National Health Service, Journal of Health Economics, 32, 410–422.

[13] 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.

[14] Dalton, D., 2014. Examining new options and opportunities for providers of NHS care, The Dalton Review, Department of Health and Social Care.

[15] Ellis, R., 1998. Creaming, skimping and dumping: provider competition on the intensive and extensive margins. Journal of Health Economics, 17, 537–555.

[16] EXPH (Expert Panel on Effective Ways of Investing in Health), 2015. Investigating Policy Options regarding Competition among Providers of Health Care Services in EU Member States. European Union, http://ec.europa.eu/health/expert_panel/ index_en.htm.

[17] Gaynor, M., 2006. What do we know about competition and quality in health care markets? Foundations and Trends in Microeconomics, Vol. 2, Issue 6.

[18] Gaynor, M., Moreno-Serra, R., Propper, C., 2013. Death by market power: Reform, competition, and patient outcomes in the National Health Service. American Economic Journal: Economic Policy, 5, 134–166.

[19] Gaynor, M., Town, R., 2011. Competition in health care markets, Handbook of Health Economics, Chapter 9, 499-637.

[20] Gowrisankaran, G., Town, R.J., 1999. Estimating the quality of care in hospitals using instrumental variables. Journal of Health Economics, 18, 747–767.

[21] Gravelle, H., 1999. Capitation contracts: access and quality. Journal of Health Economics, 18, 3, 315–340.

[22] Halonen-Akatwijuka, M., Propper, C., 2013. Competition, equity and quality in healthcare. CEPR Discussion Paper No. 9325.

[23] Jack, W., 2005. Purchasing health care services from providers with unknown altruism. Journal of Health Economics, 24, 73–93.

[24] Karlsson, M., 2007. Quality incentives for GPs in a regulated market. Journal of Health Economics, 26, 699–720.

[25] Kessler, D., McClellan, M., 2000. Is hospital competition socially wasteful? Quarterly Journal of Economics, 115, 577–615.

[26] Kessler, D., Geppert, J. 2005. The effects of competition on variation in the quality and cost of medical care. Journal of Economics and Management Strategy, 14, 575–589.

[27] Laffont, J.J., Tirole, J., 1993. A theory of incentives in procurement and regulation. MIT Press, Cambridge.

[28] Le Grand, J., 1987. Inequalities in health: some international comparisons. European Economic Review, 31, 182–191.

[29] Ma, C.A., Burgess, J.F., 1993. Quality competition, welfare, and regulation. Journal of Economics, 58, 153–173.

[30] O'Donnell, O., van Doorslaer, E., Wagstaff, A., Lindelow, M., 2008. Analyzing health equity using household survey data: A guide to techniques and their implementation. Washington, DC: World Bank.

[31] OECD, 2012. Competition in Hospital Services, OECD Policy Roundtables, Directorate for Financial and Enterprise Affairs Competition Committee.

[32] Sari, N., 2002. Do competition and managed care improve quality? Health Economics, 11, 571–584.

[33] Sculpher, M., Gafni, A., 2001. Recognizing diversity in public preferences: The use of preference sub-groups in cost-effectiveness analysis. Health Economics, 10, 317–324.

[34] Shen, Y.-S., 2003. The effect of financial pressure on the quality of care in hospitals. Journal of Health Economics, 22, 243–269.

[35] Siciliani, 2006. Selection of treatment under prospective payment systems in the hospital sector. Journal of Health Economics, 25, 479–499.

[36] Siciliani, L., Chalkley, M., Gravelle, H., 2017. Competition policy in five European countries. What can be learned for health policy in England, The Health Foundation Working Paper 1, 1–28.

[37] Tay, A., 2003. Assessing competition in hospital care markets: the importance of accounting for quality differentiation. RAND Journal of Economics, 34, 786–814.

[38] Wagstaff, A., 2002. Inequality aversion, health inequalities and health achievement. Journal of Health Economics, 21, 627–641.

[39] Wagstaff, A., Van Doorslaer, E., 2000. Equity in healthcare finance and delivery. Handbook of Health Economics, Vol. 1, 1803–1862.

[40] Wolinsky, A., 1997. Regulation of duopoly: Managed competition vs regulated monopolies. Journal of Economics & Management Strategy, 6, 821–847.

[41] Yitzhaki, S., 1983. On an extension of the Gini index. International Economic Review, 24, 617–628.

Appendix

Second-order and stability conditions

The second-order conditions of the hospitals' profit-maximising problem are given by

$$\frac{\partial^2 \pi_i}{\partial q_i^2} = \sum_k \left(\left(p - c_i^k \right) \frac{\partial^2 D_i^k}{\partial q_i^2} - 2 \frac{\partial c_i^k}{\partial q_i} \frac{\partial D_i^k}{\partial q_i} - \frac{\partial^2 c_i^k}{\partial q_i^2} D_i^k \right) - \frac{\partial^2 C}{\partial q_i^2} < 0, \tag{A1}$$

where

$$\frac{\partial^2 D_i^h}{\partial q_i^2} = \frac{\lambda \theta}{2t} \frac{\partial^2 B^h}{\partial q_i^2}; \quad \frac{\partial^2 D_i^l}{\partial q_i^2} = \frac{(1-\lambda)\,\theta}{2t} \frac{\partial^2 B^l}{\partial q_i^2}. \tag{A2}$$

The Nash equilibrium is stable if

$$H := \frac{\partial^2 \pi_i}{\partial q_i^2} \frac{\partial^2 \pi_j}{\partial q_j^2} - \frac{\partial^2 \pi_i}{\partial q_j \partial q_i} \frac{\partial^2 \pi_j}{\partial q_i \partial q_j} > 0. \tag{A3}$$

Comparative statics

The effects of a marginal change in the degree of patient choice, θ, on equilibrium qualities are given by

$$\begin{bmatrix} \frac{\partial^2 \pi_1}{\partial q_1^2} & \frac{\partial^2 \pi_1}{\partial q_2 \partial q_1} \\ \frac{\partial^2 \pi_2}{\partial q_1 \partial q_2} & \frac{\partial^2 \pi_2}{\partial q_2^2} \end{bmatrix} \begin{bmatrix} dq_1 \\ dq_2 \end{bmatrix} + \begin{bmatrix} \frac{\partial^2 \pi_1}{\partial \theta \partial q_1} \\ \frac{\partial^2 \pi_2}{\partial \theta \partial q_2} \end{bmatrix} d\theta = 0, \tag{A4}$$

which implies

$$\frac{\partial q_1^*}{\partial \theta} = \frac{1}{H} \left[\frac{\partial^2 \pi_2}{\partial \theta \partial q_2} \frac{\partial^2 \pi_1}{\partial q_2 \partial q_1} - \frac{\partial^2 \pi_1}{\partial \theta \partial q_1} \frac{\partial^2 \pi_2}{\partial q_2^2} \right] \tag{A5}$$

and

$$\frac{\partial q_2^*}{\partial \theta} = \frac{1}{H} \left[\frac{\partial^2 \pi_1}{\partial \theta \partial q_1} \frac{\partial^2 \pi_2}{\partial q_1 \partial q_2} - \frac{\partial^2 \pi_2}{\partial \theta \partial q_2} \frac{\partial^2 \pi_1}{\partial q_1^2} \right] \tag{A6}$$

where

$$\frac{\partial^2 \pi_1}{\partial q_2 \partial q_1} = \frac{\theta}{2t}\left(\lambda \frac{\partial c_1^h}{\partial q_1}\frac{\partial B^h}{\partial q_2} + (1-\lambda)\frac{\partial c_1^l}{\partial q_1}\frac{\partial B^l}{\partial q_2}\right) > 0, \tag{A7}$$

$$\frac{\partial^2 \pi_2}{\partial q_1 \partial q_2} = \frac{\theta}{2t}\left(\lambda \frac{\partial c_2^h}{\partial q_2}\frac{\partial B^h}{\partial q_1} + (1-\lambda)\frac{\partial c_2^l}{\partial q_2}\frac{\partial B^l}{\partial q_1}\right) > 0, \tag{A8}$$

$$\begin{aligned}\frac{\partial^2 \pi_1}{\partial \theta \partial q_1} \quad = \quad & \frac{\lambda}{2t}\left[\left(p - c_1^h\right)\frac{\partial B^h}{\partial q_1} - \left(B^h(q_1) - B^h(q_2)\right)\frac{\partial c_1^h}{\partial q_1}\right] \\ & + \frac{1-\lambda}{2t}\left[\left(p - c_1^l\right)\frac{\partial B^l}{\partial q_1} - \left(B^l(q_1) - B^l(q_2)\right)\frac{\partial c_1^l}{\partial q_1}\right],\end{aligned} \tag{A9}$$

and

$$\begin{aligned}\frac{\partial^2 \pi_2}{\partial \theta \partial q_2} \quad = \quad & \frac{\lambda}{2t}\left[\left(p - c_2^h\right)\frac{\partial B^h}{\partial q_2} + \left(B^h(q_1) - B^h(q_2)\right)\frac{\partial c_2^h}{\partial q_2}\right] \\ & + \frac{1-\lambda}{2t}\left[\left(p - c_2^l\right)\frac{\partial B^l}{\partial q_2} + \left(B^l(q_1) - B^l(q_2)\right)\frac{\partial c_2^l}{\partial q_2}\right].\end{aligned} \tag{A10}$$

By applying the first-order conditions, (6), we can simplify and rewrite (A9)-(A10) as

$$\frac{\partial^2 \pi_1}{\partial \theta \partial q_1} = \frac{1}{2\theta}\left(\lambda \frac{\partial c_1^h}{\partial q_1} + (1-\lambda)\frac{\partial c_1^l}{\partial q_1} + 2\frac{\partial C}{\partial q_1}\right) > 0 \tag{A11}$$

and

$$\frac{\partial^2 \pi_2}{\partial \theta \partial q_2} = \frac{1}{2\theta}\left(\lambda \frac{\partial c_2^h}{\partial q_2} + (1-\lambda)\frac{\partial c_2^l}{\partial q_2} + 2\frac{\partial C}{\partial q_2}\right) > 0, \tag{A12}$$

which implies that $\partial q_1^* / \partial \theta > 0$ and $\partial q_2^* / \partial \theta > 0$.

Derivation of the extended Gini coefficient

The extended Gini coefficient is given by (39). Using the expression for $L(x)$ from (34), we have

$$
\begin{aligned}
& \int_0^1 (1-x)^{v-2} L(x)\, dx \\
= \; & \frac{B_2^h}{\overline{B}} \int_0^{\lambda - D_1^h} (1-x)^{v-2} x dx + \frac{B_1^h}{\overline{B}} \int_{\lambda - D_1^h}^{\lambda} (1-x)^{v-2} x dx \\
& + \frac{B_2^l}{\overline{B}} \int_{\lambda}^{1-D_1^l} (1-x)^{v-2} x dx - \frac{\left(B_1^h - B_2^h\right)\left(\lambda - D_1^h\right) + \left(B_2^l - B_1^h\right)\lambda}{\overline{B}} \int_{\lambda}^{1-D_1^l} (1-x)^{v-2} dx \\
& + \frac{B_1^l}{\overline{B}} \int_{1-D_1^l}^{1} (1-x)^{v-2} x dx - \frac{\left(B_1^h - B_2^h\right)\left(\lambda - D_1^h\right)}{\overline{B}} \int_{\lambda - D_1^h}^{\lambda} (1-x)^{v-2} dx \\
& - \frac{\left(B_1^h - B_2^h\right)\left(\lambda - D_1^h\right) + \left(B_2^l - B_1^h\right)\lambda + \left(B_1^l - B_2^l\right)\left(1 - D_1^l\right)}{\overline{B}} \int_{1-D_1^l}^{1} (1-x)^{v-2} dx.
\end{aligned}
\tag{A13}
$$

Solving for the integrals yields

$$
\begin{aligned}
& \int_0^1 (1-x)^{v-2} L(x)\, dx \\
= \; & -\frac{B_2^h}{\overline{B}} \left[\frac{(1-x)^{v-1}((v-1)x+1)}{v(v-1)}\right]_0^{\lambda - D_1^h} - \frac{B_1^h}{\overline{B}} \left[\frac{(1-x)^{v-1}((v-1)x+1)}{v(v-1)}\right]_{\lambda - D_1^h}^{\lambda} \\
& -\frac{B_2^l}{\overline{B}} \left[\frac{(1-x)^{v-1}((v-1)x+1)}{v(v-1)}\right]_{\lambda}^{1-D_1^l} - \frac{B_1^l}{\overline{B}} \left[\frac{(1-x)^{v-1}((v-1)x+1)}{v(v-1)}\right]_{1-D_1^l}^{1} \\
& + \frac{\left(B_1^h - B_2^h\right)\left(\lambda - D_1^h\right)}{\overline{B}} \left[\frac{(1-x)^{v-1}}{v-1}\right]_{\lambda - D_1^h}^{\lambda} \\
& + \frac{\left(B_1^h - B_2^h\right)\left(\lambda - D_1^h\right) + \left(B_2^l - B_1^h\right)\lambda}{\overline{B}} \left[\frac{(1-x)^{v-1}}{v-1}\right]_{\lambda}^{1-D_1^l} \\
& + \frac{\left(B_1^h - B_2^h\right)\left(\lambda - D_1^h\right) + \left(B_2^l - B_1^h\right)\lambda + \left(B_1^l - B_2^l\right)\left(1 - D_1^l\right)}{\overline{B}} \left[\frac{(1-x)^{v-1}}{v-1}\right]_{1-D_1^l}^{1}.
\end{aligned}
\tag{A14}
$$

After evaluating the integrals, we obtain

$$\begin{aligned}
&\int_0^1 (1-x)^{v-2} L(x)\,dx \\
= \quad & -\frac{B_2^h}{\overline{B}}\left(\frac{\left(1-\left(\lambda-D_1^h\right)\right)^{v-1}\left((v-1)\left(\lambda-D_1^h\right)+1\right)-1}{v(v-1)}\right) \\
& -\frac{B_1^h}{\overline{B}}\left(\left(\frac{(1-\lambda)^{v-1}((v-1)\lambda+1)}{v(v-1)}\right)-\left(\frac{\left(1-\left(\lambda-D_1^h\right)\right)^{v-1}\left((v-1)\left(\lambda-D_1^h\right)+1\right)}{v(v-1)}\right)\right) \\
& -\frac{B_2^l}{\overline{B}}\left(\left(\frac{\left(1-\left(1-D_1^l\right)\right)^{v-1}\left((v-1)\left(1-D_1^l\right)+1\right)}{v(v-1)}\right)-\left(\frac{(1-\lambda)^{v-1}((v-1)\lambda+1)}{v(v-1)}\right)\right) \\
& +\frac{B_1^l}{\overline{B}}\left(\frac{\left(1-\left(1-D_1^l\right)\right)^{v-1}\left((v-1)\left(1-D_1^l\right)+1\right)}{v(v-1)}\right) \qquad \text{(A15)} \\
& +\frac{\left(B_1^h-B_2^h\right)\left(\lambda-D_1^h\right)}{\overline{B}}\left(\left(\frac{(1-\lambda)^{v-1}}{v-1}\right)-\left(\frac{\left(1-\left(\lambda-D_1^h\right)\right)^{v-1}}{v-1}\right)\right) \\
& +\frac{\left(B_1^h-B_2^h\right)\left(\lambda-D_1^h\right)+\left(B_2^l-B_1^h\right)\lambda}{\overline{B}}\left(\left(\frac{\left(1-\left(1-D_1^l\right)\right)^{v-1}}{v-1}\right)-\left(\frac{(1-\lambda)^{v-1}}{v-1}\right)\right) \\
& -\frac{\left(B_1^h-B_2^h\right)\left(\lambda-D_1^h\right)+\left(B_2^l-B_1^h\right)\lambda+\left(B_1^l-B_2^l\right)\left(1-D_1^l\right)}{\overline{B}}\left(\frac{\left(1-\left(1-D_1^l\right)\right)^{v-1}}{v-1}\right),
\end{aligned}$$

which reduces to

$$\int_0^1 (1-x)^{v-2} L(x)\,dx = \frac{B_2^h+\left(1-\lambda+D_1^h\right)^v\left(B_1^h-B_2^h\right)+(1-\lambda)^v\left(B_2^l-B_1^h\right)+\left(D_1^l\right)^v\left(B_1^l-B_2^l\right)}{\overline{B}v(v-1)}. \qquad \text{(A16)}$$

By substituting this expression into (39), we obtain the expression given by (41).