Eur J Health Econ (2013) 14:515–526
DOI 10.1007/s10198-012-0396-5

ORIGINAL PAPER

Inter-regional competition and quality in hospital care

Hiroshi Aiura

Received: 22 September 2011 / Accepted: 17 April 2012 / Published online: 22 May 2012

Abstract This study analyzes the effect of episode-of-care payment and patient choice on waiting time and the comprehensive quality of hospital care. The study assumes that two hospitals are located in two cities with different population sizes and compete with each other. We find that the comprehensive quality of hospital care as well as waiting time of both hospitals improve with an increase in payment per episode of care. However, we also find that the extent of these improvements differs according to the population size of the cities where the hospitals are located. Under the realistic assumptions that hospitals involve significant labor-intensive work, we find the improvements in comprehensive quality and waiting time in a hospital located in a small city to be greater than those in a hospital located in a large city. The result implies that regional disparity in the quality of hospital care decreases with an increase in payment per episode of care.

Keywords Patient choice · Waiting time · Hotelling-type spatial competition model · Multi-region model

JEL Classification I18 · L32

Introduction

This study analyzes the effect of inter-regional competition on waiting time and comprehensive quality in hospital care. During the past few decades, several European countries—Norway, Switzerland, the United Kingdom, etc.—have reformed their health care systems; they have introduced free choice in hospitals (so-called *patient choice*) instead of limited or no choice, and an episode-of-care payment to the reimbursement system. These changes aim to improve the quality of health care, especially waiting time, because they provide hospitals with an incentive to compete to acquire patients.[1] Since hospital care is differentiated horizontally by geographical location, hospitals experiencing these changes compete geographically in terms of quality of hospital care.[2]

Several theoretical studies using a Hotelling-type (1929) spatial competition model have attempted to determine the manner in which the incentive to compete for acquiring patients influences the quality of hospital care or waiting time within a health-care system.[3] Gravelle and Masiero (2000), Karlsson (2007), and Brekke et al. (2010) focus on how this incentive influences quality within a health-care system.[4] Although waiting time is modeled implicitly as a part of quality (as mentioned by Brekke et al. 2007), the abovementioned studies do not provide an answer regarding the effect on waiting time alone, that is, separate from quality. On the other hand, while Xavier (2003), Siciliani (2005), and Brekke et al. (2008) focus on the impact of this

H. Aiura (✉)
Faculty of Economics, Oita University,
700, Dannoharu, Oita 870-1192, Japan
e-mail: h-aiura@oita-u.ac.jp; office@aiura.info

[1] Gravelle and Sivey (2010) examine whether better information provides an incentive for hospitals to improve quality when patient choice is introduced.

[2] Tay (2003) shows empirically that the demand responses to both distance and quality are substantial.

[3] With regard to empirical studies, Gowrisankaran and Town (2003) estimate the effects of competition on the quality decisions of hospitals in southern California. Dawson et al. (2007) estimate the impact on waiting time for ophthalmology in London.

[4] For the Hotelling-type spatial competition, see also Alonso (1964) and d'Aspremont et al. (1979). Nuscheler (2003), Montefiori (2005), and Sanjo (2009) deal with medical service quality by using the spatial competition model.

incentive on waiting time, they do not consider comprehensive quality.

Moreover, these studies do not assume that patients undertake trips across provinces with different population sizes. If geographical constraints did not exist, patients would prefer hospitals located in large cities to those in small cities, because the former seems to have high-quality health care that is based on economies of scale, due to the sharing of technology and health-care expertise.[5] Therefore, when patient choice is introduced, a certain number of patients flow into hospitals located in large cities in preference to those from small cities; the impact of this, primarily on waiting time and secondarily on comprehensive quality, is different for hospitals in both types of cities. Aiura and Sanjo (2010) consider this flow of patients and derive the competitive equilibrium qualities of local public hospitals located in two regions with different population sizes. However, economies of scale and scope do not work in their model, and a counterintuitive result—that a rural public hospital always offers higher quality of hospital care than an urban public hospital—is derived. Furthermore, they do not analyze the effect of the incentive to compete.

In this study, we assume that the comprehensive quality of hospital care is reflected by several factors, such as waiting time, medical technology, and the skills and experience of medical staff. The model in this study divides these factors into two categories: waiting time and factors irrelevant to waiting time. By this division, we can study the effect not only on the comprehensive quality of hospital care but also on waiting time alone; this was not addressed in the aforementioned studies. Further, we note the difference in the influence of increased demand between waiting time and the factors irrelevant to it. When the demand on hospitals increases, hospitals become crowded, and additional resources are needed to reduce congestion and maintain a certain length of waiting time. Thus, we assume that the costs of maintaining a certain waiting time length would increase with the demand on hospitals. On the other hand, we assume that the quality of hospital care as wrought by factors irrelevant to waiting time does not worsen even if the demands on that hospital increase. For example, the medical technology and the skills and experience of medical staff of a hospital remain at the same quality level, regardless of its number of patients (although improvement in these factors may increase its waiting time by attracting more patients).[6] Accordingly, we assume that there are economies of scale in the costs of improving the factors irrelevant to waiting time; this assumption was not made by Aiura and Sanjo (2010).

When these assumptions regarding the comprehensive quality of hospitals and costs as well as patient choice across cities are introduced, we derive the competitive equilibrium qualities of hospitals located in two cities with different population sizes. When the difference in the population between large and small cities is sufficiently great, there exists an equilibrium in which a hospital in a large city is superior to a hospital in a small city in terms of comprehensive quality. This equilibrium is intuitive and is not shown by Aiura and Sanjo (2010). Under the conditions in which this equilibrium exists, we analyze the effect of episode-of-care payment and find that, with an increase in this payment, the hospitals in the two cities improve not only in terms of comprehensive quality but also in terms of waiting time; however, the extent of these improvements differs between hospitals. In an actual situation—in which hospitals involve significant labor-intensive work—these improvements are found to be greater for the hospital in the small city than for the hospital in the large city. This result implies that regional disparity in the quality of hospital care decreases with an increase in episode-of-care payment. Since the costs required by a hospital that accepts only a few patients in exchange for a certain decrease in waiting time are lower than those required by a hospital that accepts many patients, the rationale behind this result is that the hospital in the small city, which has a relatively small demand, has a cost advantage in terms of improving in waiting time. Further, on the basis of these assumptions, we can also infer that the reduction in the disparity in waiting time between the two hospitals in the large and small cities is greater than that in comprehensive quality. In other words, when patients are given a free choice of hospitals and episode-of-care payment to hospitals is adequate, regional disparity in waiting time appears to be smaller than that in comprehensive quality.

This result within the present study can be interpreted thus: regional disparity in the quality of hospital care would decrease with increasing intensity of competition among regions, because an increase in episode-of-care payment intensifies competition among hospitals for acquiring patients. This implication is supported by the findings in *OECD Regions at a Glance* 2009, which shows that Japan

[5] Aletras (1999) suggests that, apparently, economies of scale work effectively in acute care hospitals with 100–200 beds. Preyra and Pink (2006) show that economies of scale and scope through hospital consolidations are almost certainly possible.

[6] An increase in the number of patients may put more pressure on medical staff, and their overall level of skill may therefore decrease. However, this impact would be sufficiently smaller than the impact on waiting time as wrought by an increase in the number of patients. Therefore, we neglect the impact on factors irrelevant to waiting time as wrought by an increase in the number of patients.

has a more balanced regional distribution of physicians than most European countries. The report also shows that the number of physicians in the urban regions of each European country is correlated positively with population share, whereas the number of physicians in the urban regions of Japan is correlated negatively with population share. These findings imply that the disparity in quality of care would be small between urban and rural regions in Japan; these findings can be explained as follows. Japan is geographically small in size and people in Japan—a country that has been permitting patient choice in hospitals since the 1960s—are accustomed to exercising patient choice; thus, Japanese hospitals seem to be more competitive than European countries. In terms of the implications of the present study, this Japanese feature suggests that regional disparity in Japan appears to be small compared to that in European countries.

The remainder of this study is organized as follows. "Model" presents the model. "First-best quality of hospital care" shows the properties of the first-best quality of hospital care in maximizing social welfare. "Inter-regional competition among hospitals" derives the equilibrium at which hospitals compete on quality and investigates how this equilibrium changes with an increase in payment per episode of care. "Numerical analysis" uses numerical analysis to support the implications of "Inter-regional competition among hospitals". "Conclusion" presents concluding remarks.

Model

In this study, we consider an economy extended over a linear segment with length 1. Two cities, city 1 and city 2, are located at the two endpoints of this segment. Geographically speaking, the measure of each city is 0; that is, each city is regarded as a point on the segment.[7] The area between the two cites is assumed to be agricultural, and the population in this agricultural area is distributed uniformly. Hereafter, the agricultural area is referred to as the "village." We indicate the populations in city 1, city 2, and the village as N_1, N_2, and 1, respectively. Further, we assume that $N_1 > N_2 > 1$, which implies that the populations of each of the two cities is larger than that of the village, and that the population of city 1 is larger than that of city 2. Only the cities have hospitals; the village does not have a hospital because there is not sufficient demand in the village. Therefore, the people in the village need to travel to either of the two cities in order to receive hospital care.

[7] Takahashi (2004) also considers a similar spatial economy. Even if the people in cities are spread over a segment with a certain length, the results do not change within the parameter domain of this study.

Residents

Residents are endowed with a utility function separable in money and benefits derived from the public goods that government provides and hospital care. Every resident earns the same income, y, pays the same head tax, h, and demands one episode of hospital care. Public goods give each resident benefits equal to $g(z)$, which is an increasing function of government expenditure, z. When a resident takes one episode of hospital care available in city i, he/she gains benefits equal to $q(w_i, \theta_i)$, which is a function of two factors: (1) waiting time, denoted by w_i, and (2) the amount of such resources that yield benefits of hospital care but do not influence waiting time, denoted by θ_i. These two factors, w_i and θ_i, are substitutable, but not perfectly. As an extreme example, we would definitely not want a hospital in which the waiting time exceeds 10 years, even if it had the best medical technology in the world. Therefore, we assume that

$$q(w_i, \theta_i) = A w_i^{-\alpha} \theta_i^{\beta},$$

where A, α, and β are constant and greater than 0. The elasticity of the comprehensive quality with regard to waiting time is equal to α; that is, waiting time increases in worth for patients as α increases. Since most patients consult their general practitioners (GPs) before accessing hospital care, they are well informed by their GPs about the hospitals they will access; thus, we assume that the residents know of the benefits gained from hospital care before receiving them.[8] Additionally, we identify the benefits, $q(w_i, \theta_i)$, with the comprehensive quality of hospital care. The residents of the cities and the village who receive hospital care from another city incur transportation costs for traveling from their homes to the city that provides hospital care. When a resident residing at $x \in [0, 1]$ receives hospital care in city 1, we assume that he/she incurs tx as transportation costs, where t is a constant and greater than 0. Similarly, when the resident receives hospital care in city 2, we assume that he/she incurs $t(1 - x)$ in transportation costs.[9] Therefore, a resident residing at x, who consumes one unit of hospital care available in city i (= 1, 2) gains a utility—denoted by $u_i(x)$—that is equal to

$$u_1(x) = y - h + g(z) + q(w_1, \theta_1) - sp - tx,$$
$$u_2(x) = y - h + g(z) + q(w_2, \theta_2) - sp - t(1 - x),$$

[8] If the residents do not know of these benefits beforehand, but the errors that patients make with regard to the information of hospitals are distributed identically and independently, then the results presented in this study would hold qualitatively; however, the effect of episode-of-care payment would weaken.

[9] Even if we assume quadratic transportation costs, the results remain unchanged.

where p and s denote the payment per episode of care to the hospital and the co-payment rate of the patient, respectively, in the medical security system that the government constructs.[10] Accordingly, residents go to the city that offers a higher utility and receive hospital care that is available in that city. Let X denote the location of a resident who receives the same amount of surplus from both cities. Thus, we obtain the following equation:

$$q(w_1, \theta_1) - tX = q(w_2, \theta_2) - t(1 - X),$$

which yields

$$X = \frac{1}{2} + \frac{q(w_1, \theta_1) - q(w_2, \theta_2)}{2t}.$$

The residents on the left side of X consume one episode of hospital care available in city 1 and gain u_1, whereas those on the right side of X consume one episode of hospital care available in city 2 and gain u_2.[11]

Accordingly, we obtain the demand for hospital care in city i (=1, 2), D_i, as follows[12]:

$$D_1(q(w_1, \theta_1), q(w_2, \theta_2)) = \begin{cases} 0, & \text{if } X < 0 \\ N_1 + X, & \text{if } 0 \leq X \leq 1 \\ N_1 + N_2 + 1, & \text{if } X > 1 \end{cases},$$

$$D_2(q(w_1, \theta_1), q(w_2, \theta_2)) = (N_1 + N_2 + 1) - D_1(q(w_1, \theta_1), q(w_2, \theta_2)).$$

Hospitals

Hospitals are financed by the government, which offers a lump-sum transfer T_i $(i = 1, 2)$ and a payment based on the number of episodes of care. We assume that hospitals consider not only their profits but also the benefits of the residents that live in the city to which each hospital belongs, and the objective function of the hospital in city i is assumed to be given by

$$\pi_i = T_i + pD_i(q_1, q_2) - c(\cdot) + \delta N_i u_i(i - 1), \qquad (1)$$

where δ is constant and greater than 0, and $c(\cdot)$ denotes a total costs function. In this objective function, the profits and the total benefits of the residents living in city i are given by $T_i + pD_i(q_1, q_2) - c(\cdot)$ and $N_i u_i(i - 1)$, respectively, while δ captures the relative weight attached to the benefits.[13] The payment per episode of care is regulated by the government; thus, the hospital in city i decides its waiting time (w_i) and the amount of resources that do not influence waiting time (θ_i) in order to maximize its own objective function. The revenue of hospital i is pD_i, which depends on the w_i and θ_i of the hospital in city i. On the other hand, the total costs of the hospital in city i, $c(\cdot)$, depend on not only w_i and θ_i but also on the number of episodes of care (which is equal to D_i), because the hospital in city i that decided a certain waiting time requires resources in proportion to the number of episodes that it has, in order to maintain its waiting time. For example, if the hospital handled k times the demand without increasing its required resources, the waiting time would become k times longer; thus, if the hospital handled k times the demand without increasing waiting time, a k-fold increase in resources would be required. Therefore, we assume the costs to provide D_i episodes of hospital care at its decided waiting time (which is w_i) as $r_i HD_i/w_i$, where r_i and D_i/w_i denote the price and the quantity of the resources in city i, respectively, to provide one episode of hospital care per unit of time, and H denotes the total operating hours of the hospital. In addition, the hospital in city i also requires θ_i (which denotes the amount of hospital-care resources that do not influence waiting time); thus, the cost function is

$$c(\theta_i, w_i, D_i) = r_i H \frac{D_i}{w_i} + \hat{r}_i \theta_i, \qquad (2)$$

where $\hat{r}_i$ denotes the unit price of θ_i. We assume that all resources to provide hospital care are mobile across regions and sectors, and that their prices are determined in a competitive marketplace. Therefore, their prices are the same in cities 1 and 2, and we set $\hat{r}_1 = \hat{r}_2 = 1$ as the numéraire; we also set $r_1 H = r_2 H = r$.

Since the demands of both hospitals remains unchanged unless their comprehensive qualities change, the hospitals face the following cost minimization problem in providing hospital care at the comprehensive quality level q_i:

$$\min_{w_i, \theta_i} r \frac{D_i(q_1, q_2)}{w_i} + \theta_i \quad \text{s.t. } q(w_i, \theta_i) = q_i \ (i = 1, 2).$$

The first-order conditions for the minimization problem yield

[10] This study omits private medical insurance, which is considered by Barros and Martinez-Giralt (2002). We omit this, because we focus solely on the episode-of-care effect.

[11] Note that the co-payment of hospital care and the benefits derived from public goods do not both affect the residents' choice between two cities, because these two factors are the same, regardless of their choice.

[12] Regarding this derivation of demand functions, we basically follow the model presented by Montefiori (2005) and Sanjo (2009).

[13] This formulation is based on Brekke et al. (2008), Chalkley and Malcomson (1998), and Jack (2005), but it differs from these previous studies in that the formulation in the current study does not include the benefits of the residents that visit the hospital from outside cities (whereas the formulation in these previous studies includes the benefits of every resident that visits the hospital). We apply this formulation because it is more tractable for symbolic analysis, but the results would not be qualitatively changed, even if we were to apply the formulation—as do previous studies—because the benefits of the residents that visit the hospital from outside the city are small compared to those of the residents inside the city.

$$w_i(q_1,q_2) = A^{\gamma}\left(\frac{\beta}{\alpha}r\right)^{\beta\gamma} D_i(q_1,q_2)^{\beta\gamma} q_i^{-\gamma}, \tag{3}$$

$$\theta_i(q_1,q_2) = \frac{1}{A^{\gamma}}\left(\frac{\beta}{\alpha}r\right)^{\alpha\gamma} D_i(q_1,q_2)^{\alpha\gamma} q_i^{\gamma}, \tag{4}$$

where $\gamma = 1/(\alpha + \beta)$. Equations 3 and 4 indicate the levels of w_i and θ_i that hospital i, which is rational, chooses at the level of q_i. Substituting Eqs. 3 and 4 into Eq. 2, we obtain the cost function that is minimal cost at the level of q_i as

$$c_i(q_1,q_2) = \hat{A}D_i(q_1,q_2)^{\alpha\gamma} q_i^{\gamma},$$
$$\text{where } \hat{A} = \frac{1}{A^{\gamma}}\left[\left(\frac{\beta}{\alpha}\right)^{\alpha\gamma} + \left(\frac{\alpha}{\beta}\right)^{\beta\gamma}\right] r^{\alpha\gamma}. \tag{5}$$

Note that $(rD_i/w_i)/\theta_i = \alpha/\beta$, which means that $\alpha\gamma$ denotes the share of costs with improvement in waiting time in total costs and $\beta\gamma$ denotes the share of costs without improvement in waiting time in total costs. The intuition behind this result is that if α is larger (smaller) than β, which implies that patients put more (less) value on waiting time than on other factors except for waiting time, then costs with (without) improvement in waiting time required to attract patients will be more.

The average cost is given by

$$c_i(q_1,q_2)/D_i = \hat{A}D_i(q_1,q_2)^{-\beta\gamma} q_i^{\gamma}, \tag{6}$$

which shows that the average cost decreases with an increase in demand (i.e., economies of scale work) if the comprehensive quality of hospital care is fixed. However, the amount of demand depends on, and is in proportion to, the height of the comprehensive quality. Therefore, economies of scale appear for a low amount of the demand (a low number of episodes of care), then decreasing returns to scale appear, if $\gamma > 1$. This change is consistent with the empirical results indicated by Aletras (1999). Equation 6 also implies that a decrease in $\beta\gamma$ and increase in γ weaken economies of scale; in other words, economies of scale weaken as the share of the costs without improvement in waiting time in total costs decreases or as the cost function becomes more convex with regard to the comprehensive quality of hospital care. Accordingly, substituting Eq. 5 into Eq. 1, we obtain the objective function that depends on q_1, q_2 as

$$\pi(q_i) = T_i + pD_i(q_1,q_2) - \hat{A}D_i(q_1,q_2)^{\alpha\gamma} q_i^{\gamma} + \delta N_1[y - h + g(z) + q_1 - sp].$$

Governments

The government spends its tax revenue on the payments to hospital and the provision of public goods. In addition, taxation and redistribution cause shadow costs equal to ζh $(0<\zeta<1)$. Since a lump-sum transfer T_i $(i = 1, 2)$ and a payment based on the number of episodes of care that the government pay to hospitals are $T_1 + T_2$ and $(N_1 + N_2 + 1)(1 - s)p$, respectively, the expenditure on the provision of public goods is $z = (N_1 + N_2 + 1)[(1-\zeta)h - (1-s)p] - (T_1 + T_2)$.

First-best quality of hospital care

Although an increase in episode-of-care payment would provide hospitals with an incentive to compete for acquiring patients and lead them to improve their comprehensive quality of care and waiting time, the central government would not adjust a payment per episode of care to such a high value that its costs would be significantly higher than its benefits. In this section, to determine a reasonable range for episode-of-care payments, we derive the property of the first-best quality of hospital care, in a case where social welfare is at a maximum value.

When the comprehensive quality of every hospital is first-best, then the marginal benefits and marginal costs of a marginal improvement in a comprehensive quality of each hospital are equal from the viewpoint of an entire economy. The total utility in the economy, TU, is given by

$$\begin{aligned} TU &= N_1u_1(0) + \int_0^X u_1(x)dx + \int_X^1 u_2(x)dx + N_2u_2(1) \\ &= (y - h + g(z) - sp)(N_1 + N_2 + 1) \\ &\quad - t\left(\int_0^X xdx + \int_X^1 (1-x)dx\right) + q_1D_1(q_1,q_2) \\ &\quad + q_2D_2(q_1,q_2), \end{aligned}$$

and we obtain the marginal benefits of a marginal improvement in a comprehensive quality of each hospital, MB_i $(i = 1, 2)$, as

$$MB_i = \frac{\partial TU}{\partial q_i} = D_i. \tag{7}$$

On the other hand, we obtain the marginal cost of a marginal improvement in a comprehensive quality of each hospital, MC_i (i = 1, 2), as follows[14]:

$$\begin{aligned} MC_i &= \frac{\partial(c_1(q_1,q_2) + c_2(q_1,q_2))}{\partial q_i} \\ &= \hat{A}\gamma\left[\alpha(D_i^{-\beta\gamma}q_i^{\gamma} - D_{-i}^{-\beta\gamma}q_{-i}^{\gamma})\frac{dD_i}{dq_i} + D_i^{\alpha\gamma}q_i^{\gamma-1}\right], \end{aligned} \tag{8}$$

where $i \neq -i$. From Eqs. 7 and 8, we obtain the following lemmas.

[14] Note that the shadow costs are sunk (independent of the level of comprehensive quality).

Lemma 1. *If $\gamma < 1$, the optimal level of comprehensive quality does not exist in each hospital.*

Proof See Appendix 1. □

From Lemma 1, we assume $\gamma \geq 1$ to omit the cases in which the optimal level of the comprehensive quality does not exist. Moreover, we consider only those cases in which every hospital has a positive demand at the first-best quality, because it is unrealistic to expect that only one facility would provide hospital care to all residents in the two cities and one village. In this consideration, $0 < X < 1$ is satisfied, and the properties of the optimal level of comprehensive quality are shown in the following Lemma.

Lemma 2. *Let us denote q_1^{OPT} and q_2^{OPT} as the optimal level of the comprehensive quality in each hospital. When each hospital has a positive demand at $q_1 = q_1^{OPT}$ and $q_2 = q_2^{OPT}$, the following conditions regarding q_1^{OPT} and q_2^{OPT} are satisfied:*

$$2tD_i - \beta q_i^{OPT} > 0 (\mathrm{i} = 1, 2), q_1^{OPT} > q_2^{OPT}, \quad \text{and} \quad w_1(q_1^{OPT}, q_2^{OPT}) < w_2(q_1^{OPT}, q_2^{OPT}). \tag{9}$$

Proof See Appendix 2. □

Lemma 2 implies that the hospital in city 1 (large city) is better than the hospital in city 2 (small city), in terms of both quality and waiting time. The qualities based on resources that do not influence waiting time are independent of demand; for example, an increase in the number of patients does not influence the quality of the medical technology employed. Therefore, the costs per unit of demand with regard to the resources that do not influence waiting time decrease with an increase in the amount of demand; thus, it is cost-effective to invest additional resources in a hospital located in a large city, which has a large demand, rather than in a hospital located in a small city.[15] In addition, in order to utilize these resources, it is better for a hospital located in a large city to perform more operations per amount of time by reducing its waiting time, compared to a hospital in a small city. Accordingly, the comprehensive quality of the large city's hospital is higher than that of the small city's hospital, and the waiting time at the large city's hospital is shorter than that at the small city's hospital.

Inter-regional competition among hospitals

In this section, we derive the equilibrium properties under which hospitals compete in terms of quality. For the same reason indicated in "First-best quality of hospital care", we assume that $0 < X < 1$ is satisfied under equilibrium. Further we assume that $\gamma \geq 1$ which is the condition in which the first-best quality exists. Since the central government would not adjust an episode-of-care payment to such a high level that it would cost significantly more than its benefits, we consider only the equilibrium condition in which the quality of hospital care in each city does not differ significantly from the quality of hospital care in each city at the first-best quality.[16]

Based on this consideration and condition (9), at equilibrium, the quality of each hospital satisfies that $2tD_i - \beta q_i > 0\,(i = 1, 2)$, which is indicated in the gray area of Fig. 1.

The hospital in city i determines the quality of its care in order to maximize its objective function, and the first-order conditions for maximizing π_1 and π_2 are given by

$$R_1 = \frac{\partial \pi_1}{\partial q_1} = \delta N_1 + \frac{p}{2t} - \hat{A}\gamma\left(\frac{\alpha}{2t} D_1^{-\beta\gamma} q_1^{\gamma} + D_1^{\alpha\gamma} q_1^{\gamma-1}\right) = 0 \tag{10}$$

and

$$R_2 = \frac{\partial \pi_2}{\partial q_2} = \delta N_2 - \frac{p}{2t} - \hat{A}\gamma\left(\frac{\alpha}{2t} D_2^{-\beta\gamma} q_2^{\gamma} + D_2^{\alpha\gamma} q_2^{\gamma-1}\right) = 0, \tag{11}$$

respectively.

We obtain the Nash equilibrium by identifying the intersection of the two response functions, $R_1(q_1, q_2) = 0$ and $R_2(q_1, q_2) = 0$. The following Lemma proposes the properties of these response functions.

Lemma 3 *The slope of $R_1(q_1, q_2) = 0$, $dq_1/dq_2|_{R1(q1,\, q2)=0}$ is greater than 2, and the slope of $R_2(q_1, q_2) = 0$, $dq_1/dq_2|_{R2(q1,\, q2)\,=\,0}$, is between 0 and 1/2. The q_1-coordinate*

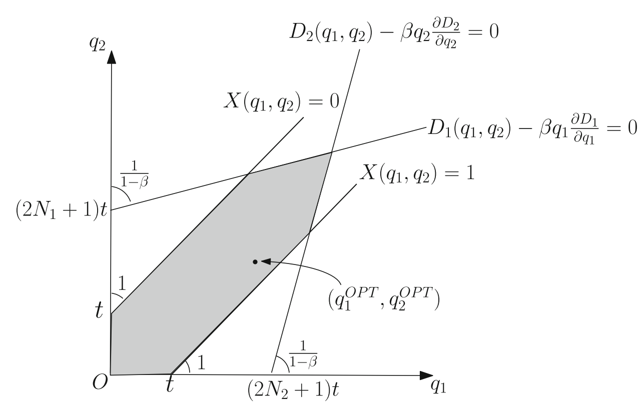

Fig. 1 Area of analysis considered in "Inter-regional competition among hospitals"

[15] If $\beta = 0$, this cost-effectiveness disappears, and we obtain $q_1^{OPT} = q_2^{OPT}$; this is consistent with the result of Aiura and Sanjo (2010).

[16] The central government, which considers the positive externalities of health care, would adjust an episode-of-care payment whose costs reasonably outweigh its direct benefits to residents.

of the intersection of the two lines $R_1(q_1, q_2) = 0$ *and* $D_1(q_1, q_2) = D_2(q_1, q_2)$ *is larger than the* q_2*-coordinate of the intersection of the two lines* $R_2(q_1, q_2) = 0$ *and* $D_1(q_1, q_2) = D_2(q_1, q_2)$.

Proof See Appendix 3. □

Figure 2 illustrates the two response functions, $R_1 = 0$ and $R_2 = 0$, that satisfy the properties in Lemma 3, and leads to the following propositions.

Lemma 4 *At the Nash equilibrium in the gray area of Fig. 1, the demand for hospital care in city* 1 *is higher than that in city* 2*—that is,* $D_1(q_1^*, q_2^*) > D_2(q_1^*, q_2^*)$, *where* q_1^* *and* q_2^* *denote the comprehensive quality levels at the equilibrium.*

Proof Figure 2 shows that the intersection of the two lines $R_1(q_1, q_2) = 0$ and $D_1(q_1, q_2) = D_1(q_1, q_2)$ is located at the upper right of the intersection of the two lines $R_2(q_1, q_2) = 0$ and $D_2(q_1, q_2) = D_1(q_1, q_2)$. Moreover, the slope of $R_1(q_1, q_2) = 0$ is greater than 2, and the slope of $R_2(q_1, q_2) = 0$ is between 0 and 1/2. Therefore, we ensure that the intersection of the two response functions, $R_1 = 0$ and $R_2 = 0$, is located in an area lower than the line $D_1(q_1, q_2) = D_1(q_1, q_2)$; this means that $D_1(q_1^*, q_2^*) > D_2(q_1^*, q_2^*)$. □

As mentioned in "First-best quality of hospital care", the hospital in the large city—which has a large demand—is cost-effective in terms of investing resources that do not influence waiting time. Lemma 4 implies that this cost-effectiveness of the large city's hospital holds, even at equilibrium. However, Lemma 4 does not point to which hospital—large city's hospital or small city's hospital—is superior in terms of comprehensive quality. The following Lemma proposes the condition in which the comprehensive quality of the large city's hospital is greater than that of the small city's hospital.

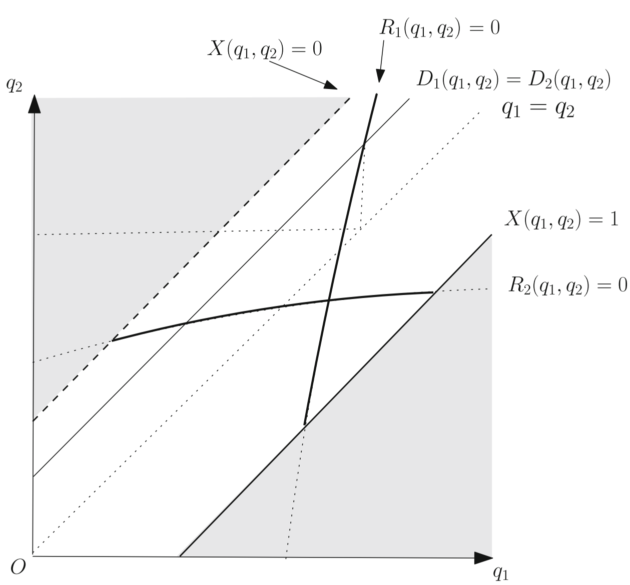

Fig. 2 Response functions

Lemma 5 *When* (1) *the factors other than waiting time are sufficiently important for residents with regard to the quality of hospital care as compared to the waiting time, or* (2) *the population of city* 1 *is sufficiently larger than that of city* 2, *the comprehensive quality of the hospital in city 1 is greater than that of city 2 at the equilibrium—that is,* $q_1^* > q_2^*$ *if* $\beta\gamma \to 1$ *or* $N_1 \to \infty$.

Proof See Appendix 4. □

The conditions in Lemma 5 are the same as those needed to make the large city's hospital more cost-effective in terms of investing in resources that do not influence waiting time. Needless to say, the higher the level of cost-effectiveness, the greater the incentive to improve quality. Therefore, if either of these conditions is satisfied, the comprehensive quality of the large city's hospital will be greater than that of the small city's hospital.[17] Since the large city's hospital is usually superior to the small city's hospital, hereafter, we assume that the difference of population between the large and small cities is sufficiently large in order to secure it.

Our main focus in the present study is to analyze the effect of episode-of-care payments on waiting time and comprehensive quality in hospital care; we propose the following propositions.

Proposition 1 *In the competitive equilibrium among hospitals, the comprehensive quality of hospital care in both hospitals increases with an increase in payment per episode of care.*

Proof Both $R_1(q_1, q_2) = 0$ and $R_2(q_1, q_2) = 0$ move away from the origin as p increases with respect to q_i, because $D_i^{-\beta\gamma} q_i^{\gamma}$ and $D_i^{\alpha\gamma} q_i^{\gamma-1}$ are increasing functions of q_i. Therefore, the intersection of $R_1(q_1, q_2) = 0$ and $R_2(q_1, q_2) = 0$ move up and right as p increases, which ensure Proposition 1. □

Proposition 2 *In the competitive equilibrium among hospitals, waiting time shortens with an increase in payment per episode of care.*

Proof See Appendix 5. □

If an episode-of-care payment is sufficient, it would serve as an incentive for hospitals to provide the residents of the village with hospital care. Therefore, as the episode-of-care payment increases, the hospitals in each city will

[17] On the other hand, if the difference in population in the two cities is small and the waiting time is sufficiently important for residents with regard to hospital quality, the cost-effectiveness weakens; thus, the hospital in the large city is not superior to that in the small city, in terms of comprehensive quality at equilibrium.

increase their comprehensive quality of health care in order to acquire patients from the village. Since an increase in the comprehensive quality requires a balanced improvement between waiting time and the factors other than waiting time, the waiting time in each hospital will also improve with an increase in the payment per episode of care, as shown in Proposition 2.

However, the degree of improvement is different between the hospitals in the large and small cities. We derive dq_1^*/dp and dq_2^*/dp by solving

$$\begin{pmatrix} \partial R_1/\partial q_1 & \partial R_1/\partial q_2 \\ \partial R_2/\partial q_1 & \partial R_2/\partial q_2 \end{pmatrix} \begin{pmatrix} dq_1^*/dp \\ dq_2^*/dp \end{pmatrix} = -\begin{pmatrix} \partial R_1/\partial p \\ \partial R_2/\partial p \end{pmatrix},$$

which yields

$$\begin{aligned} \frac{dq_1^*}{dp} - \frac{dq_2^*}{dp} &= \frac{(R_1/\partial q_1 + R_1/\partial q_2)(\partial R_2/\partial p) - (R_2/\partial q_1 + R_2/\partial q_2)(\partial R_1/\partial p)}{(R_1/\partial q_1)(R_2/\partial q_2) - (R_1/\partial q_2)(R_2/\partial q_1)} \\ &= \frac{1}{\Delta}\frac{\gamma}{2t}\left[\frac{\alpha\gamma}{2t}\left(D_2^{-\beta\gamma} q_2^{\gamma-1} - D_1^{-\beta\gamma} q_1^{\gamma-1}\right) + (\gamma-1)\left(D_2^{\alpha\gamma} q_2^{\gamma-2} - D_1^{\alpha\gamma} q_1^{\gamma-2}\right)\right], \end{aligned} \tag{12}$$

where $\Delta = (R_1/\partial q_1)(R_2/\partial q_2) - (R_1/\partial q_2)(R_2/\partial q_1) > 0$.

Since we assume the difference in population between the large and small cities to be sufficient, $q_1 > q_2$ and $D_1 > D_2$; thus, the sign of $dq_1^*/dp - dq_2^*/dp$ depends on α, β, and γ (=$1/(\alpha + \beta)$). The following Lemma presents the relationships between the sign of $dq_1^*/dp - dq_2^*/dp$ and the conditions of α, β and γ.

Lemma 6 *When $\gamma \geq 2$ and α is sufficiently large compared to β, the sign of $dq_1^*/dp - dq_2^*/dp$ is negative—that is, $dq_1^*/dp < dq_2^*/dp$. When $1 < \gamma < 2$ and β is sufficiently large compared to α, the sign of $dq_1^*/dp - dq_2^*/dp$ is positive—that is, $dq_1^*/dp > dq_2^*/dp$*

Proof See Appendix 6. □

As mentioned in "Model", a decrease in $\beta\gamma$ and an increase in γ weaken economies of scale; thus, Lemma 6 implies that the strength of economies of scale determines how a disparity in comprehensive quality between the hospitals in large and small cities changes as a result of an increase in payment per episode of care. The hospital industry is generally categorized as a labor-intensive industry, and so economies of scale do not generally work well there. Accordingly, we rewrite Lemma 6 as the following proposition.

Proposition 3 *In the competitive equilibrium among labor-intensive hospitals, an increase in payment per episode of care reduces the difference in the comprehensive quality of hospital care between the two hospitals.*

Since the amount of labor (such as doctors and nurses) deeply influences waiting time, the costs relevant to waiting time are large compared to other costs in a labor-intensive hospital; the effect of this cost on comprehensive quality is also large. Moreover, the additional cost related to improving the waiting time is larger in the large city's hospital, which has many more episodes of care than does the small city's hospital. Therefore, compared to the large city's hospital, the small city's hospital can increase the comprehensive quality of its hospital care at a lower cost; thus, the small city's hospital has a cost advantage over the large city's hospital in terms of competing to acquire patients from the village. In addition, an increase in payment per episode of care makes the hospitals in both cities more enthusiastic vis-a-vis the acquisition of patients from the village. As a result, an increase in payment per episode of care makes the small city's hospital, which has a cost advantage, more enthusiastic than the large city's hospital with respect to acquiring patients from the village; thus, the difference in the comprehensive quality of hospital care between the two cities decreases.

As described, when the costs associated with improvements in waiting time occupy a large proportion of the total costs, the improvement in waiting time plays a major role in reducing the difference in quality of hospital care between the two cities. However, the comprehensive quality of hospital care is partially influenced by resources that do not influence waiting time, which create economies of scale and weaken the reduction of the difference in the comprehensive quality of hospital care, owing to improvements in waiting time. Accordingly, the reduction of the difference in comprehensive quality of hospital care between the hospitals in the two cities is less than that between their waiting times. On the contrary, the reduction in the difference in waiting time between the hospitals in the two cities is more than that in their comprehensive quality of hospital care. Proposition 4 presents its extreme case.

Proposition 4 *The equilibrium at which the comprehensive quality of hospital care in city 1 (the large city) is higher than that in city 2 (the small city) (i.e., $q_1^* > q_2^*$) does not guarantee that the waiting time for the hospital in city 1* *is shorter than that for the hospital in city* 2. [*i.e.*, $w_1(q_1^*, q_2^*) < w_2(q_1^*, q_2^*)$].

Proof See Appendix 7. □

Proposition 4 implies that the hospital whose waiting time is the shorter of the two hospitals is not always superior to the other hospital in terms of comprehensive quality.

Numerical analysis

The previous section abstractly presented the effect of episode-of-care payment by symbolic analysis, but did not demonstrate a specific equilibrium. Therefore, in this section, we show that the implications mentioned in the previous section remain valid within realistic parameters; we do by using numerical analysis.

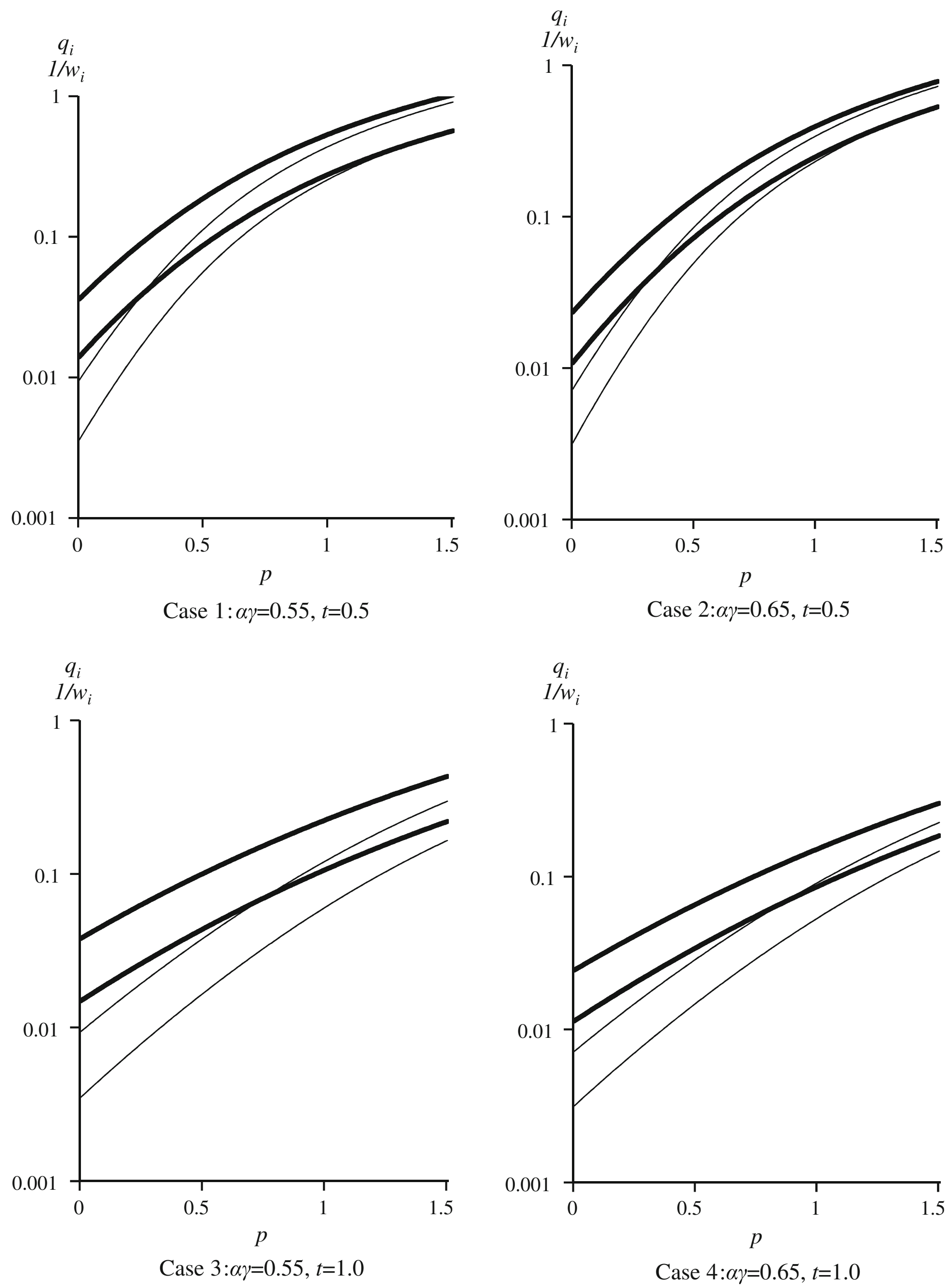

Fig. 3 Changes in q_i and $1/w_i$, with p. (Note. *Upper bold line*: q_1; *lower bold line*: $1/w_1$; *upper thin solid line*: q_2; *lower thin solid line*: $1/w_2$. q_i and $1/w_i$ are logarithmically scaled on the *left axis*. The other parameters are same for any case: $N_1 = 2$, $N_2 = 1.5$, $\delta = 1$, $A = 1.25$, $r = 1$, $1/\gamma = 0.9$, and $T_1 = T_2 = 0$.)

Generally speaking, the share of labor costs in total costs of a hospital is between 50 and 70%.[18] Since health care is a service that doctors and nurses provide by spending time, labor is a resource that influences waiting time; thus, we set $\alpha\gamma$ (=$1-\beta\gamma$) to either 0.55 or 0.65. In order to encourage the effect of economies of scale, we set $1/\gamma = \beta + \alpha = 0.9$. Further, we assume two cases for t, $t = 0.5$ and $t = 1$, in order to investigate the effect of transport improvements. Regarding the other parameters, we set $N_1 = 2$, $N_2 = 1.5$, $\delta = 1$, $A = 1.25$, $r = 1$, and $T_1 = T_2 = 0$.[19]

Figure 3 depicts the change of q_i and $1/w_i$ with p for four cases with different parameter values. In all cases, the hospital in city 1 is superior to the hospital in city 2 in terms of comprehensive quality (q) and shortness of waiting time (w) when the payment per episode of care (p) is small; these disparities decrease with an increase in payment per episode of care. Further, in all cases, the waiting time for hospital care is more effective than the comprehensive quality of hospital care, in terms of reducing of these disparities. These results correspond to and support the discussions presented in "Inter-regional competition among hospitals".

When we compare the results between the cases of $\alpha\gamma = 0.55$ and $\alpha\gamma = 0.65$, the reduction of these disparities is faster when $\alpha\gamma = 0.65$ than when $\alpha\gamma = 0.55$. The economies of scale in hospitals weaken as hospitals become more labor-intensive. Therefore, the higher the share of costs relating to improvements in waiting time in the total costs, the faster the reduction of disparities as a result of an increase in payment per episode of care between the two hospitals. Further, when we compare the results between the cases of $t = 0.5$ and $t = 1$, the reduction of disparities between the two hospitals occurs more quickly in the case of $t = 0.5$ than in the case of $t = 1$, because the reduction of t encourages competition between hospitals.

Conclusion

This study illustrates the effect of episode-of-care payment on the comprehensive quality of hospital care and waiting time between two hospitals that are located in two different cities and compete for acquiring patients, under the assumption that the population is sufficiently different between the two cities. The results show that labor-intensive hospitals improve their comprehensive quality of hospital care and waiting time as a result of an increase in payment per episode of care, but the extent of these improvements differs according to the population size of the cities: the hospital in the small city shows greater improvements in the comprehensive quality of hospital care and waiting time than the hospital in the large city. This result implies that regional disparity in the quality of hospital care decreases with an increase in payment per episode of care. Further, we show that a reduction in the disparity of waiting time between the two hospitals is greater than that of the comprehensive quality of hospital care; we then present the case in which the hospital with the longer waiting time is superior in terms of comprehensive quality. These results cautions us: although we often focus on the ratio of doctors to the population using public censuses, in order to determine the difference in the health-care environment between regions, we may make faulty decisions if we rely only on this metric. For example, OECD Regions at a Glance 2009 shows that Japan has a more balanced regional distribution of physicians than most European countries. Although this implies that access to health care in Japan is even between rural and urban regions, it does not imply that the comprehensive quality of health care in the rural regions of Japan is comparable to that in urban regions.

In summary, an increase in episode-of-care payment gives an incentive to every hospital, but the impact of incentives for hospitals to compete for acquiring patients on the basis of their comprehensive quality and waiting time depends on the population size of the city in which the hospitals are located. Therefore, we must pay attention to the population surrounding the hospitals while theoretically and empirically analyzing the effect of policies that induce hospitals to compete with each other.

Acknowledgments We would like to thank William M. Doerner, Tomoya Ida, Toshiharu Ishikawa, Akio Kawasaki, Yuichi Kimura, Tatsuaki Kuroda, Noriaki Matsushima, Kenichi Mizuta, Toru Naito, Hikaru Ogawa, Hiroshi Ono, Sasuo Sanjo, Yasuhiro Sato, Norio Shimoda, Luigi Siciliani, Hiroyuki Takami, Chris White and two anonymous referees for comments. In particular, the extensions provided in Section 4 are motivated by the referees' very useful comments and constructive suggestions. Any mistakes herein are, of course, our own. I acknowledge the financial support from KAKENHI (22330095).

Appendix 1: Proof of Lemma 1

When we let q_1 go to infinity with q_2 fixed, $D_1 = N_1 + N_2 + 1$, $D_2 = 0$ and $dD_i/dq_i = 0$ (i = 1,2). Therefore, we obtain

$$\lim_{q_1\to\infty} MB_1 = N_1 + N_2 + 1 > 0 \tag{13}$$

and

$$\lim_{q_1\to\infty} MC_1 = \hat{A}\gamma(N_1 + N_2 + 1)^{\alpha\gamma} q_i^{\gamma-1}. \tag{14}$$

When $\gamma < 1$, $\lim_{q_1\to\infty} MB_1 > \lim_{q_1\to\infty} MC_1 = 0$, which means that social welfare is divergent as q_1 goes to infinity.

[18] For example, the share of labor costs in total costs was 57.1% in 2008 in the United Kingdom (as shown by the UK Centre for the Measurement of Government Activity, 2010) and 67.9% between 1997 and 2001 in Switzerland (as shown by Farsi and Filippini, 2006).

[19] This analysis is not affected by y, h, and s.

Accordingly, if $\gamma < 1$, the optimal level of the comprehensive quality does not exist.

Appendix 2: Proof of Lemma 2

Proof of $2tD_i - \beta q_i^{OPT} > 0(\mathrm{i}=1,2)$. From (7) and (8) we obtain

$$MB_i - MC_i = \left[1 - \hat{A}\gamma\frac{1}{q_i}D_i^{-\beta\gamma}q_i^{\gamma}\right]D_i - \hat{A}\gamma\left[\frac{\alpha}{2t}\left(D_i^{-\beta\gamma}q_i^{\gamma} - D_{-i}^{-\beta\gamma}q_{-i}^{\gamma}\right)\right].$$

Therefore, the second-order conditions for the maximization of social welfare at $q_1 = q_1^{OPT}$ and $q_2 = q_2^{OPT}$ would require

$$\frac{\partial\left(D_i^{-\beta\gamma}q_i^{OPT\gamma}\right)}{\partial q_i} = \frac{\gamma}{2t}D_i^{-\beta\gamma-1}\left(q_i^{OPT}\right)^{\gamma-1}\left(2tD_i - \beta q_i^{OPT}\right) > 0,$$

which derives $2tD_i - \beta q_i^{OPT} > 0$ (i = 1, 2).

Proof of $q_1^{OPT} > q_2^{OPT}$ and $w_1(q_1^{OPT}, q_2^{OPT}) < w_2(q_1^{OPT}, q_2^{OPT})$ First, we will confirm that $D_1(q_1^{OPT}, q_2^{OPT}) \geq D_2(q_1^{OPT}, q_2^{OPT})$. q_1 and q_2 that satisfies that $D_1 = D_2 = (N_1 + N_2 + 1)/2 = \bar{N}$ are denoted by $\bar{q}_1$ and $\bar{q}_2$ respectively. Since $N_1 > N_2$, $\bar{q}_1 < \bar{q}_2$ is satisfied. Moreover, $TU|_{q_1=\bar{q}_1+\epsilon \text{and } q_2=q_2-\epsilon} = TU|_{q_1=\bar{q}_1-\epsilon \text{ and } q_2=q_2+\epsilon}$ and

$$\begin{aligned}(c_1 + c_2)|_{q_1=\bar{q}_1+\epsilon \text{ and } q_2=\bar{q}_2-\epsilon} - (c_1 + c_2)|_{q_1=\bar{q}_1-\epsilon \text{ and } q_2=q_2+\epsilon} \\ = \hat{A}\{(\bar{N}+\epsilon)^{\alpha\gamma}[(\bar{q}_2+\epsilon)^{\gamma} - (\bar{q}_1+\epsilon)^{\gamma}] - (\bar{N}-\epsilon)^{\alpha\gamma} \\ [(\bar{q}_2-\epsilon)^{\gamma} - (\bar{q}_1-\epsilon)^{\gamma}]\}.\end{aligned} \tag{15}$$

Since $\bar{q}_1 < \bar{q}_2$, $\alpha > 0$, and $\gamma \geq 1$, the right side of (15) is more than 0 if $\epsilon > 0$; thus, social welfare is greater at $q_1 = \bar{q}_1 + \epsilon$ and $q_2 = \bar{q}_2 - \epsilon$ than at $q_1 = \bar{q}_1 - \epsilon$ and $q_2 = \bar{q}_2 + \epsilon$, which confirms that $D_1(q_1^{OPT}, q_2^{OPT}) \geq D_2(q_1^{OPT}, q_2^{OPT})$.

Using (5), we transform $MB_i = MC_i$ $(i = 1, 2)$ to

$$\begin{pmatrix} \frac{\alpha\gamma}{2tD_1} + \frac{\gamma}{q_1} & -\frac{\alpha\gamma}{2tD_2} \\ -\frac{\alpha\gamma}{2tD_1} & \frac{\alpha\gamma}{2tD_2} + \frac{\gamma}{q_2} \end{pmatrix}\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} D_1 \\ D_2 \end{pmatrix}. \tag{16}$$

The solution of (16) with respect to c_1 and c_2 is

$$\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \frac{1}{\bar{A}}\begin{pmatrix} \left(\frac{\alpha\gamma}{2tD_2} + \frac{\gamma}{q_2}\right)D_1 + \frac{\alpha\gamma}{2t} \\ \frac{\alpha\gamma}{2t} + \left(\frac{\alpha\gamma}{2tD_1} + \frac{\gamma}{q_1}\right)D_2 \end{pmatrix}, \tag{17}$$

where

$$\bar{A} = \left(\frac{\alpha\gamma}{2tD_1} + \frac{\gamma}{q_1}\right)\left(\frac{\alpha\gamma}{2tD_2} + \frac{\gamma}{q_2}\right) - \left(\frac{\alpha\gamma}{2t}\right)^2\frac{1}{D_1D_2} > 0.$$

(17) is transformed to

$$\begin{aligned}\frac{c_1}{c_2} &= \frac{D_1q_1[(\alpha\gamma q_2 + 2t\gamma D_2)D_1 + \alpha\gamma q_2D_2]}{D_2q_2[(\alpha\gamma q_1 + 2t\gamma D_1)D_2 + \alpha\gamma q_1D_1]} \\ &= \frac{D_1q_1(\bar{N}\alpha q_2 + tD_1D_2)}{D_2q_2(\bar{N}\alpha q_1 + tD_1D_2)}.\end{aligned} \tag{18}$$

Substituting (5) into the left side of (18), we have

$$\left(\frac{D_2}{D_1}\right)^{\beta\gamma}\left(\frac{q_1}{q_2}\right)^{\gamma-1} = \frac{\bar{N}\alpha q_2 + tD_1D_2}{\bar{N}\alpha q_1 + tD_1D_2}. \tag{19}$$

To satisfy (19), $q_1 > q_2$ must hold, because $\beta > 0$, $\gamma > 1$, and $D_1 \geq D_2$. Therefore, $q_1^{OPT} > q_2^{OPT}$ is satisfied.

Further,

$$\frac{w_2(q_1^{OPT}, q_2^{OPT})}{w_1(q_1^{OPT}, q_2^{OPT})} = \frac{q_1(\bar{N}\alpha q_2 + tD_1D_2)}{q_2(\bar{N}\alpha q_1 + tD_1D_2)} > \frac{q_1\bar{N}\alpha q_2}{q_2\bar{N}\alpha q_1} = 1,$$

which derives $w_1(q_1^{OPT}, q_2^{OPT}) < w_2(q_1^{OPT}, q_2^{OPT})$.

Appendix 3: Proof of Lemma 3

From (10) and (11), we obtain

$$\left.\frac{dq_2}{dq_1}\right|_{R_1(q_1,q_2)=0} = -\frac{\partial R_1/\partial q_1}{\partial R_1/\partial q_2} = 2 + \frac{\beta\alpha\gamma q_1^2 + (\gamma-1)(2tD_1)^2}{\alpha\gamma(2tD_1 - \beta q_1)q_1}$$

and

$$\left.\frac{dq_2}{dq_1}\right|_{R_2(q_1,q_2)=0} = -\frac{\partial R_2/\partial q_1}{\partial R_2/\partial q_2} = 1\bigg/\left(2 + \frac{\beta\alpha\gamma q_2^2 + (\gamma-1)(2tD_2)^2}{\alpha\gamma(2tD_2 - \beta q_2)q_2}\right).$$

Since $2tD_i - \beta q_i > 0$ is satisfied, the slope of $R_1(q_1, q_2) = 0$ is greater than 2, and the slope of $R_2(q_1, q_2) = 0$ is between 0 and 1/2.

Appendix 4: Proof of Lemma 5

Proof of that $q_1^* > q_2^*$ *if* $\beta\gamma \to 1$ Since $\lim_{\beta\gamma\to1}\alpha\gamma = 0$ and $D_1(q_1^*, q_2^*) > D_2(q_1^*, q_2^*)$, we obtain $\lim_{\beta\gamma\to1}R_1(q_1^*, q_2^*) > \lim_{\beta\gamma\to1}R_2(q_1^*, q_2^*)$ if $q_1^* \leq q_2^*$. Therefore, the equilibrium satisfying $q_1^* \leq q_2^*$ does not exist; thus, $q_1^* > q_2^*$ is satisfied if $\beta\gamma \to 1$.

Proof of that $q_1^* > q_2^*$ *if* $N_1 \to \infty$ The line $R_1 = 0$ in Fig. 2 shifts right as N_1 increases. Therefore, if N_1 is sufficiently large, $q_1^* > q_2^*$ is satisfied.

Appendix 5: Proof of Proposition 2

$$\frac{dw_i(q_1^*, q_2^*)}{dp} = -a^{\gamma}\left(\frac{\beta}{\alpha}r\right)^{\beta\gamma} \times \left[\frac{\gamma}{2t}\left(2tD_i - \beta q_i^*\right)D_i^{-\alpha\gamma}q_i^{*-(\gamma+1)}\frac{dq_i^*}{dp} - \frac{\beta\gamma}{2t}D_i^{-\alpha\gamma}q_i^{*-\gamma}\frac{dq_{-i}^*}{dp}\right].$$

Here, $i \neq -i \in \{1,2\}$. $2tD_i - \beta q_i > 0$ and $\partial D_i/\partial q_{-i} < 0$ are satisfied, and Proposition 1 shows that $dq_i^*/dp > 0$ and $dq_{-i}^*/dp > 0$. Therefore, we derive that $dw_i(q_1^*, q_2^*)/dp < 0$.

Appendix 6: Proof of Lemma 6

Since $\lim_{\beta/\alpha\to\infty}\alpha\gamma = 0$ and $\lim_{\alpha/\beta\to\infty}\beta\gamma = 0$, when $\gamma \geq 2$,

$$\lim_{\alpha/\beta\to\infty} \operatorname{sgn}\left[\frac{dq_1^*}{dp} - \frac{dq_2^*}{dp}\right] = \operatorname{sgn}\left[\frac{1}{2t}(q_2^{\gamma-1} - q_1^{\gamma-1}) + \left(\gamma-1\right)(D_2q_2^{\gamma-2} - D_1q_1^{\gamma-2})\right] < 0.$$

When $1 < \gamma < 2$,

$$\lim_{\beta/\alpha\to\infty} \operatorname{sgn}\left[\frac{dq_1^*}{dp} - \frac{dq_2^*}{dp}\right] = \operatorname{sgn}\left[(\gamma-1)(q_2^{\gamma-2} - q_1^{\gamma-2})\right] > 0.$$

These equations confirm Lemma 6.

Appendix 7: Proof of Proposition 4

If $q_1 \geq q_2$, $D_1 > D_2$ is satisfied. Therefore,

$$\lim_{q_1\to q_2}(w_1(q_1,q_2) - w_2(q_1,q_2)) = \frac{1}{a^{\gamma}}\left(\frac{\beta}{\alpha}r\right)^{\beta\gamma}\left(D_1^{\beta\gamma} - D_2^{\beta\gamma}\right)q_2^{-\gamma} > 0,$$

which implies that $w_1(q_1^*, q_2^*) > w_2(q_1^*, q_2^*)$ if q_1^* is slightly more than q_2^*.

References

Aiura, H., Sanjo, Y.: Privatization of local public hospitals: Effect on budget, medical service quality, and social welfare. International Journal of Health Care Finance and Economics **10**, 275–299 (2010)

Alonso, W.: Location theory. In: Friedman, J., Alonso, W. (eds) Regional Development and Planning, pp. 78–106. MIT, Cambridge, MA (1964)

Aletras, V.H.: A comparison of hospital scale effects in short-run and long-run cost functions. Health Econ. **8**, 521–530 (1999)

Barros, P.P., Martinez-Giralt, X.: Public and private provision of health care. J. Econ. Manage. Strat. **11**, 109–133 (2002)

Brekke, K.R., Cellini, R., Siciliani, L., Straume, O.R.: Competition and quality in health care markets: a differential-game approach. J. Health Econ. **29**, 508–523 (2010)

Brekke, K.R., Nuscheler, R., Straume, O.R.: Gatekeeping in health care. J. Health Econ. **26**, 149–170 (2007)

Brekke, K.R., Siciliani, L., Straume, O.R.: Competition and waiting times in hospital markets. J. Publ. Econ. **92**, 1607–1628 (2008)

Chalkley, M., Malcomson, J.M.: Contracting for health services when patient demand does not reflect quality. J. Health Econ. **17**, 1–19 (1998)

d'Aspremont, C., Gabszewicz, J.J., & Thisse, J.-F.: On hotelling's 'stability in competition'. Econometrica **47**, 1145–1150 (1979)

Dawson, D., Gravelle, H., Jacobs, R., Martin, S., Smith, P.C.: The Effects of expanding patient choice of provider on waiting times: Evidence from a policy experiment. Health Econ. **16**, 113–128 (2007)

Farsi, M., Filippini, M.: An analysis of efficiency and productivity in Swiss hospitals. Swiss J. Econ. Stat. **142**, 1–37 (2006)

Gowrisankaran, G., Town, R.J.: Competition, payers, and hospital quality. Health Serv. Res. **38**, 1403–1422 (2003)

Gravelle, H., Masiero, G.: Quality incentives in a regulated market with imperfect information and switching costs: capitation in general practice. J. Health Econ. **19**, 1067–1088 (2000)

Gravelle, H., Sivey, P.: Imperfect information in a quality-competitive hospital market. J. Health Econ. **29**, 524–535 (2010)

Hotelling, H.: Stability in competition. Econ. J. **39**, 41–57 (1929)

Jack, W.: Purchasing health care services from providers with unknown altruism. J. Health Econ. **24**, 73–93 (2005)

Karlsson, M.: Quality incentives for GPs in a regulated market. J. Health Econ. **26**, 699–720 (2007)

Montefiori, M.: Spatial competition for quality in the market for hospital care. Eur. J. Health Econ. **6**, 131–135 (2005)

Nuscheler, R.: Physician reimbursement, time consistency, and the quality of care. J. Inst. Theor. Econ.**159**, 302–322 (2003)

OECD: OECD Regions at a Glance 2009. OECD, Paris (2009)

Preyra, C., Pink, G.: Scale and scope efficiencies through hospital consolidations. J. Health Econ. **25**, 1049–1068 (2006)

Sanjo, Y.: Quality choice in a health care market: A mixed duopoly approach. Eur. J. Health Econ. **10**, 207–215 (2009)

Siciliani, L.: Does more choice reduce waiting times?. Health Econ. **14**, 17–23 (2005)

Takahashi, T.: Spatial competition of governments in the investment on public facilities. Reg. Sci. Urban Econ. **34**, 455–488 (2004)

Tay, A.: Assessing competition in hospital care markets: the importance of accounting for quality differentiation. RAND J. Econ. **34**, 786–814 (2003)

UK Centre for the Measurement of Government Activity, 2010. Public service output, inputs and productivity: Healthcare, The Office for National Statistics, the UK

Xavier, A.: Hospital competition, GP fundholders and waiting times in the UK internal market: the case of elective surgery. Int. J. Health Care Finance Econ. **3**, 25–51 (2003)