Article:
Brekke, Kurt R., Levaggi, Rosella, Siciliani, Luigi orcid.org/0000-0003-1739-7289 et al. (1 more author) (2016) Patient Mobility and Health Care Quality when Regions and Patients Differ in Income. Journal of Health Economics. pp. 372-387. ISSN: 0167-6296

https://doi.org/10.1016/j.jhealeco.2016.05.003

Patient Mobility and Health Care Quality when Regions and Patients Differ in Income*

Kurt R. Brekke[†] Rosella Levaggi[‡] Luigi Siciliani[§] Odd Rune Straume[¶]

16 May 2016

Abstract

We study the effects of cross-border patient mobility on health care quality and welfare when income varies across and within regions. We use a Salop model with a high-, middle-, and low-income region. In each region, a policy maker chooses health care quality to maximise the utility of its residents when health care costs are financed by general income taxation. In equilibrium, regions with higher income offer better quality, which creates an incentive for patient mobility from lower- to higher-income regions. Assuming a prospective payment scheme based on DRG-pricing, we find that lower non-monetary (administrative) mobility costs have (i) no effect on quality or welfare in the high-income region; (ii) a negative effect on quality but a positive effect on welfare for the middle-income region; and (iii) ambiguous effects on quality and welfare for the low-income region. Lower monetary mobility costs (copayments) might reduce welfare in both the middle- and low-income region. Thus, health policies that stimulate cross-border patient mobility can be counterproductive when regions differ in income.

Keywords: Patient mobility; Health care quality; Income inequalities; Regional welfare.

JEL Classification: H51; H73; I11; I18

*This work was carried out with funding from COMPETE reference n^o POCI-01-0145-FEDER-006683, with the FCT/MEC's financial support through national funding and by the ERDF through the Operational Programme on "Competitiveness and Internationalization – COMPETE 2020 under the PT2020 Partnership Agreement. The paper has benefited from being presented at the European Health Economics Workshop 2015 in Toulouse. We are grateful for comments by Christoph Schottmüller, Helmuth Cremer, and three anonymous referees.

†Department of Economics, Norwegian School of Economics, Helleveien 30, N-5045 Bergen, Norway. E-mail: kurt.brekke@nhh.no.

‡Department of Economics, University of Brescia, Via San Faustino 74b, 25100 Brescia, Italy. E-mail: rosella.levaggi@unibs.it.

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

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

1 Introduction

Cross-border patient mobility is currently a key issue for health policy. In the European Union (EU), patient mobility across member states has been high on the political agenda for many years, despite the fact that the free movement principles do not apply to health care provision. A key example is the new directive adopted by the European Parliament and the Council in 2011, which gives patients the right to choose among health care providers across all EU member states.[1] On the 25th of October 2013, when the directive came into force, the Health Commissioner Tonio Borg said:[2]

> "Today is an important day for patients across the European Union. As of today, EU law in force enshrines citizens' right to go to another EU country for treatment and get reimbursed for it (. . .). For patients, this Directive means empowerment: greater choice of healthcare, more information, easier recognition of prescriptions across-borders."

The actual patient flows across EU member states are small. In 2011, when the new directive was implemented, the European Commission estimated the demand for cross-border health care to €10 billion, i.e., 1% of public health-care spending in the EU. In a recent Eurobarometer survey, only 5% of the respondents reported that they had received medical treatment in another EU country.[3] However, according to the same survey, 49% said they would be willing to travel to another EU country to receive medical treatment with the two main reasons being to receive treatment not available in home country (71%) and to receive better quality treatment (53%). These figures suggest a large potential for cross-border health care within the EU.

Cross-border patient mobility is also an important policy issue for countries with regional health-care provision. Sweden implemented a 'free choice' reform in 2003 that allowed patients to demand health care outside their home county and ensured cost reimbursement by specifying transfer payments for cross-border care. A similar system is in place in Italy, where many patients migrate from the south to the north to obtain better medical care. However, in Canada, patient mobility across provinces is generally limited to emergency and sudden illness or allowed only in special circum-

[1] Directive 2011/24/EU of the European Parliament and of the Council of 9 March 2011 on the application of patients' rights in cross-border healthcare.

[2] The full statement of the Health Commissioner can be found here: http://europa.eu/rapid/press-release_MEMO-13-932_en.htm

[3] The survey, Special Eurobarometer 425 "Patients' rights in crossborder healthcare in the European Union", was conducted by TNS Opinion & Social for the European Commission in 2014; http://ec.europa.eu/public_opinion/index_en.htm.

stances. In the US, state-specific regulations restrict individuals from purchasing health insurance outside their home state, which limits patient mobility across state lines.[4]

These real-world observations highlight the importance of understanding the effects of cross-border patient mobility. Would the new EU directive improve access to better health care for patients, or is the reform beneficial only for some countries while others are worse off? Would cross-border patient mobility in the EU stimulate quality provision among member states, or can the reform have adverse effects on the quality of care? Should Canada do as Sweden and Italy by allowing for patient mobility across regions, or are they better off with the current system?

In this paper, we study the impact of cross-border patient mobility on health-care quality and welfare, and discuss the implications for health policy. To do so, we develop a spatial model á la Salop (1979) with three regions that differ in income distribution, i.e., a high-income, a middle-income, and a low-income region. In the context of EU, we can interpret the high-income region as Northern European countries, the middle-income region as the Southern European countries, and the low-income region as the new member states in Eastern Europe.

The policy maker in each region chooses quality to maximise the utility of its own residents subject to a budget constraint, where the total cost of health care provision is financed by general income taxation. To allow for income effects, we assume individuals have decreasing marginal utility of income, which implies that the marginal cost of raising tax revenues decreases with average income. Consequently, health care quality is increasing in the regions' income level, inducing cross-border patient flows from poorer to richer regions. We focus on the equilibrium where the high-income region only imports patients, the low-income region only exports patients, whereas the middle-income region both exports and imports patients.

In the analysis, a key assumption is that the regions apply a prospective payment scheme where the health care providers receive a fixed price equal to the treatment cost per patient. This payment scheme is consistent with Diagnosis-Related Group (DRG) pricing that has been adopted by almost every European country, and is now by far the most commonly used payment scheme in the EU (Busse et al., 2011). As the treatment costs are arguably higher in richer regions with better health care quality, DRG-pricing implies that the exporting regions are facing a net financial loss related to cross-border patient mobility, whereas the importing regions are fully compensated

[4]During the debate over Obamacare, the Republicans promoted an alternative approach that involved allowing individuals to purchase health insurance across state lines. See, for instance, http://www.forbes.com/sites/theapothecary/2012/05/11/will-buying-health-insurance-across-state-lines-reduce-costs/

for the treatment of migrating patients. Another key assumption is that we allow for the exporting regions to possibly charge a copayment from the migrating patients to cover the difference in the treatment costs. This assumption is in line with the new EU directive that entitles patients that demand cross-border care within the EU to cost reimbursement equal to that of their home country.[5]

Based on this framework, we derive a rich set of results regarding the regional effects of cross-border patient mobility on quality provision and welfare. First, a reduction in *non-monetary* mobility costs (e.g., a simplification of administrative procedures) has no quality or welfare effects in the high-income region, since the costs of treating migrating patients are fully compensated. However, for the two other regions, a reduction in *non-monetary* mobility costs tends to reduce quality. In both the middle-income and the low-income regions, the direct effects of lower (non-monetary) mobility costs on quality provision are unambiguously negative. Increased patient mobility reduces the marginal benefit of quality provision, because fewer patients will be treated in their home region. In addition, since patient export has monetary costs – both for the patients who migrate and for the tax payers – higher mobility needs to be financed by a higher income tax rate for exporting regions. For the low-income region, we also identify an indirect effect related to quality provision in this region being a strategic substitute to the quality provision in the middle-income region, which counteracts the direct effects mentioned above.

Second, the effects of reducing *monetary* costs of patient mobility (i.e., patient copayments) are qualitatively similar to the effects of reducing non-monetary costs. However, there is an additional budget effect that makes the overall effect generally indeterminate. A lower copayment implies that a larger share of the costs of patient export needs to be financed by the exporting regions' tax payers, which in turn implies a tightening of the government's budget constraint. This gives the exporting (low- and middle-income) regions an incentive to increase quality in order to mitigate the increase in mobility caused by lower patient copayments.

Third, the effects of cross-border health care on *regional welfare* are mixed with winners and losers from such a policy. A reduction in *non-monetary* mobility costs has (i) no welfare effect in the high-income region, (ii) a positive welfare effect in the middle-income region, and (iii) an indeterminate welfare effect in the low-income region. The middle-income region unambiguously benefits because of the cost reduction for the patients who seek treatment in the high-income region. A similar effect also applies to the low-income region. However, in this region there is a potentially

[5] See Directive 2011/24/EU on the application of patients' rights in cross-border healthcare.

counteracting welfare effect due to the quality reduction in the middle-income region, which harms the migrating patients but benefits the tax payers. If, on the other hand, cross-border mobility is stimulated by a reduction in *monetary* mobility costs (lower copayment), the welfare effects in the middle- and low-income regions are generally ambiguous.

Finally, we analyse the effects of more income dispersion both across and within regions. Larger *inter-regional* income inequality leads to higher quality in the high-income region, lower quality in the middle-income region, whereas the effect in the low-income region is ambiguous. The effects of higher *intra-regional* income inequality, on the other hand, depend on the region in which income dispersion increases. Higher income inequality in the high-income region leads to higher quality in that region and lower quality in the other two regions, while higher income inequality in either the middle-income or the low-income region has no effect on quality provision in the high-income region and indeterminate effects on the other two regions. Thus, allowing for cross-border patient mobility can create negative spillover effects of higher income inequality in the form of lower quality of health care in neighbouring regions.

The literature on cross-border patient mobility is limited but growing.[6] The recent papers by Andritsos and Tang (2013, 2014) use a queueing framework to analyse the effect of cross-border patient mobility on waiting times and reimbursement policies.[7] Andritsos and Tang (2013) find that patient mobility can increase patient welfare due to increased access to care. However, the effects on waiting times and reimbursement rates are mixed, and the additional costs of mobility are disproportionately shared between the participating countries. Andritsos and Tang (2014) find that patient mobility can be beneficial to public health-care systems (NHS), as health-care funders can reduce their costs without increasing the patients' waiting time. In border regions, where the cost of crossing the border is low, 'outsourcing' the high-cost country's elective care services to the low-cost country is a viable strategy from which both countries can benefit. Despite similarities, these studies do not consider the effect of patient mobility on health-care quality nor the role of differences in income distribution across and within regions, which is the key focus of our paper.

The closest paper to ours is Brekke et al. (2014b) who consider a Hotelling model with two regions that differ in health-care technology, where the region with more efficient technology offers

[6] See, for instance, the review by Brekke et al. (2014a).

[7] There is also a paper by Petretto (2000) that looks at regionalisation of a National Health Service. It provides conditions for establishing whether devolution for health care expenditure is desirable. Variations in health expenditure will depend on its marginal benefit and the marginal cost of public funds, including higher or lower transfers originating from mobility. However, this paper has no explicit spatial dimension and it is not concerned with the quality of care. It is thus very different from ours.

higher health-care quality and attracts patients from the region with less efficient technology. A key finding is that the effects of patient mobility depend on the transfer payment scheme. If the transfer payment is below marginal cost, mobility leads to a 'race-to-the-bottom' in quality and lower welfare in both regions. Thus, patient mobility can have adverse effects on quality provision and welfare unless an appropriate transfer payment scheme is implemented.

In the current paper, we take a different approach by focusing on differences in the income distribution across regions as the key source of cross-border patient mobility.[8] This is an issue that has received much attention in the EU debate, especially after the recent extension of the EU to Eastern Europe. To capture this, we therefore apply a model with three (rather than two) regions. This also implies that the same region can be both importing and exporting patients, which cannot arise with the two-region set up in Brekke et al. (2014b). Finally, we use more general cost functions, allow for copayments when patients demand care outside their region, and allow for heterogeneity in income within countries/regions (with richer patients more likely to move). Critically, we introduce income effects through decreasing marginal utility of income, which implies that qualities are in most reasonable scenarios strategic substitutes, and this is an important driver of some key results. We investigate the effect of policy-relevant parameters such as patients' copayments and inter- and intra-regional income dispersion. Thus, our paper is significantly different from Brekke et al. (2014b).

Our paper also relates to the broader health economics literature on provider competition and quality incentives. A key finding from this literature is that with regulated prices, competition increases health-care quality if providers are profit-maximisers, whereas the relationship between competition and quality is generally ambiguous if providers are (partly) altruistic.[9] Despite some similarities, our study differs from this literature as we consider competition between regions (rather than providers), where health-care quality is set by policy makers that maximise regional welfare financed through taxation. Moreover, the income distribution across and within regions is central to our study, but not a part of the previously cited papers.[10] Thus, the competitive mechanisms in our model are clearly different from the more general literature on provider competition and

[8] Analytically, differences in quality in the current paper are driven by differences in income. In Brekke et al. (2014b) countries differ in the marginal cost of quality, ie a country has a technology advantage.

[9] See for example Gravelle (1999), Gravelle and Sivey (2010), Brekke, Nuscheler and Straume (2006), and Brekke, Siciliani and Straume (2011). See Gaynor (2006) for an excellent review of the literature on competition and quality in health care markets.

[10] There is a paper by Aiura and Sanjo (2010) that uses a Hotelling model with two regions that differ in their population density to study incentives for health care quality. While this paper shares some similarities in the demand structure, the focus is very different as they study the impact of privatisation of local public hospitals.

quality incentives.[11]

The rest of the paper is organised as follows. In Section 2 we present our model. Equilibrium quality provision is derived and presented in Section 3, whereas, in Section 4, we describe the strategic relationship between regional quality choices. In Section 5 we analyse the effects of policies to stimulate cross-border patient mobility – a reduction of either monetary or non-monetary mobility costs – on regional quality provision and welfare. In Section 6 we explore the effects of (inter-regional or intra-regional) income inequality on regional quality provision and analyse how these effects depend on cross-border patient mobility. Finally, Section 7 concludes the paper.

2 Model

Consider a market for health care where patients are uniformly distributed on a circle with circumference equal to 1 and the total patient mass normalised to 1. The market consists of three different regions, which can be interpreted either as neighbouring countries or neighbouring regions within the same country. The three regions, indexed by $i = L, M, H$, are of equal size, each covering 1/3 of the circle. The index i denotes whether the region has Low, Middle or High average income. The market is served by three health care providers (hospitals), one in each region, where the provider in Region i is located at x_i. We assume that each provider is located at the center of its region, implying that the residents of Region i are located on the line segment $[x_i - \frac{1}{6}, x_i + \frac{1}{6}]$. Each patient demands one unit of health care (one treatment) from the most preferred provider. We assume that health care provision is publicly funded through general income taxation and is free at the point of consumption (at least for patients who seek treatment in their own region).[12]

The net utility of a patient located at z and receiving health care from the provider in Region i, located at x_i, is given by

$$U(z, x_i) = v + bq_i - t\,|z - x_i| + u\left(Y_i^k\right) \tag{1}$$

[11] Our paper also relates to the economic literature on fiscal federalism and interregional competition, in particular the part of this literature concerned with cross-border shopping. However, this literature is mainly concerned with taxation rather than health-care quality as an incentive for cross-border mobility. See, for instance, Kanbur and Keen (1993), Trandel (1994), Wang (1999), and Nielsen (2001).

[12] We therefore do not allow for the presence of a private sector alternative. Adding this additional choice would make the presentation of the model much more complicated without gaining additional insights (see Barros and Siciliani (2011) for a detailed review of the literature). Moreover, note that in our model patients may have to pay a copayment when choosing treatment in a different region, which has some analogies with modelling public versus private patient's choice.

if the patient receives treatment in the region to which she resides. $v > 0$ is the patient's gross utility of being treated, $q_i \geq \underline{q}$ is the quality offered by the provider in Region i (with $b > 0$ measuring the marginal utility of quality)[13] and t is the marginal disutility of travelling.[14] The utility of income is measured by a strictly concave utility function $u(\cdot)$. We assume that patients are heterogeneous in income y^k with $k = P, R$ and with $y^R > y^P$, i.e., we allow for high income (Rich) and low income (Poor) patients. Assuming a proportional income tax rate (or social security contribution), $\tau_i > 0$, set by the government of Region i, the net income of a type-k patient in Region i is given by

$$Y_i^k := y^k (1 - \tau_i). \tag{2}$$

The proportion of high-income residents, λ_i, is assumed to differ across regions, with $\lambda_H > \lambda_M > \lambda_L > 0$. For later reference, it is useful to define the average gross income in Region i as

$$\overline{y}_i := \lambda_i y^R + (1 - \lambda_i) y^P. \tag{3}$$

We also define the average utility gain of a marginal reduction in the income tax rate in Region i (when all residents in the region seek treatment in their own region) as

$$\overline{u}_{\tau_i} := \lambda_i u' \left(Y_i^R\right) y^R + (1 - \lambda_i) u' \left(Y_i^P\right) y^P. \tag{4}$$

The net utility of a patient located at z and receiving health care from the provider in a neighbouring Region j (different from where the patient resides), located at x_j, is given by

$$U(z, x_j) = v + bq_j - t |z - x_j| + u \left(\widehat{Y}_i^k\right) - F, \tag{5}$$

where F is a non-monetary mobility cost (disutility) of seeking care in a different region (because of different administrative rules and language barriers, for example). We also assume that a patient who receives care in a different region (with higher treatment costs) must pay a copayment π, such that the net income of a type-k patient in Region i who seeks care in a neighbouring region is given

[13] The lower bound $\underline{q}$ represents the lowest possible quality the providers can offer without being charged with malpractice and is, for simplicity, set to 0.

[14] For tractability reasons, we make the standard assumption that utility is separable in quality, distance and consumption. We also assume, without much loss of generality, that utility is linear in quality and distance. There is strong empirical evidence showing that distance is a major predictor of patients' choice of hospital, see, e.g., Tay (2003) and Beckert et al (2012). We also assume that utility is separable in quality and consumption.

by[15]

$$\widehat{Y}_i^k := y^k(1-\tau_i) - \pi. \tag{6}$$

Assuming that each patient makes a utility-maximising choice of provider, and assuming that v is sufficiently large to ensure full market coverage in equilibrium, patients of type k who travel from Region i to Region j for treatment are located on a line segment of length $\max\{0, \phi_{ij}^k\}$, where

$$\phi_{ij}^k := \frac{1}{2t}\left(b\,(q_j - q_i) + u\left(\widehat{Y}_i^k\right) - u\left(Y_i^k\right) - F\right). \tag{7}$$

Notice that

$$\frac{\partial \phi_{ij}^k}{\partial y_k} = \left(\frac{1-\tau_i}{2t}\right)\left(u'\left(\widehat{Y}_i^k\right) - u'\left(Y_i^k\right)\right) > 0 \text{ if } \pi > 0. \tag{8}$$

Because of decreasing marginal utility of income, richer patients have a lower disutility of paying the copayment and are therefore more prone to choose cross-border health care (as long as $\pi > 0$).

The total number of patients travelling from Region i to Region j is then given by $\max\{0, \Phi_{ij}\}$, where

$$\Phi_{ij} := \lambda_i \phi_{ij}^R + (1-\lambda_i)\,\phi_{ij}^P. \tag{9}$$

Notice here that $\partial \Phi_{ij}/\partial q_j = -\left(\partial \Phi_{ij}/\partial q_i\right) = b/2t$.

Since utility is assumed to be strictly concave in income, the marginal cost of raising tax revenues decreases with average income, implying that the optimally chosen health care quality will be higher in richer regions (see section 3.2 and Appendix B). This creates an incentive for patient migration from poorer to richer regions and we will look for an equilibrium that has this direction of patient flows. The total demand for health care in each region is then given by

$$D_L = \frac{1}{3} - \Phi_{LH} - \Phi_{LM}, \tag{10}$$

$$D_M = \frac{1}{3} - \Phi_{MH} + \Phi_{LM}, \tag{11}$$

$$D_H = \frac{1}{3} + \Phi_{MH} + \Phi_{LH}. \tag{12}$$

[15] The paremeter π could alternatively be interpreted as including other monetary costs of seeking cross-border health care. This would have some minor implications for the way we specify the budget restriction of each region (see Section 3) but would not qualitatively affect the main results of our analysis.

The provider in Region i is assumed to have the following the cost function:

$$C(D_i, q_i) = c_i D_i + K(q_i), \tag{13}$$

where K is increasing and strictly convex in quality, and where $c_H \geq c_M \geq c_L$. We make the simplifying assumption that quality costs are independent of the number of patients treated, implying that quality can be characterised as a local public good at each hospital. A typical example would be investment in medical machinery and other treatment specific hospital facilities whose running cost, being mostly labour, is rather similar. We also make the assumption that marginal treatment costs (might) differ across the three regions, where we postulate a positive relationship between income level and treatment costs, motivated by the fact that labour (including health care personnel) tends to be more expensive in high-income countries.

3 Optimal quality provision

The policy maker in each region chooses quality to maximise the utility of its own residents subject to a budget constraint, where the total cost of health care provision is financed by general income taxation with a tax rate τ_i. For patients travelling from Region i to Region j to receive treatment, the provider in Region j receives a payment p_j per patient. We assume that part of this payment, π, is paid by the patient himself, whereas the remaining part, $p_j - \pi$, is paid by the tax payers in Region i. As previously mentioned, we will consider an (interior solution) equilibrium in which quality is highest in the high-income region and lowest in the low-income region. This implies that the high-income region will attract (import) some patients, and the low-income region will export some patients. The middle-income region both imports patients from the low-income region and exports some patients to high-income region.

Let aggregate utility (i.e., total welfare) in Region i be denoted by W_i.[16] The problem of the policy maker in Region i is then to maximise W_i with respect to q_i, subject to a budget constraint. We assume that this problem is well-behaved.[17] The budget constraints in the high-, middle- and low-income regions, respectively, are given by

$$\tau_H \frac{\overline{y}_H}{3} = K(q_H) + \frac{c_H}{3} - (p_H - c_H)(\Phi_{LH} + \Phi_{MH}), \tag{14}$$

[16] See Appendix A for the full welfare expressions for each region.

[17] The second-order conditions are presented in Appendix A.

$$\tau_M \frac{\overline{y}_M}{3} = K(q_M) + \frac{c_M}{3} + (p_H - \pi - c_M)\,\Phi_{MH} - (p_M - c_M)\,\Phi_{LM}, \tag{15}$$

$$\tau_L \frac{\overline{y}_L}{3} = K(q_L) + \frac{c_L}{3} + (p_H - \pi - c_L)\,\Phi_{LH} + (p_M - \pi - c_L)\,\Phi_{LM}. \tag{16}$$

The left-hand sides of (14)-(16) are the tax revenues collected. The first term on the right hand sides is the cost of quality provision, whereas the second term is the cost of treating all of the region's own residents. The remaining terms are the net fiscal losses related to patient migration. All else equal, if patient migration is profitable for a particular region, a lower tax rate is needed to finance a given quality provision. The high-income region only imports patients, and these patients bring a net gain (loss) if $p_H > (<)\, c_H$. On the other hand, the low-income region only exports patients, and this patient emigration has a positive (negative) effect on the government budget if $p_H + p_M - 2\pi < (>)\, 2c_L$. Finally, the middle-income region is both an exporter and importer of patients. For this region, patient immigration is profitable (unprofitable) if $p_M > (<)\, c_M$, whereas patient emigration is profitable (unprofitable) if $p_H - \pi < (>)\, c_M$.

Notice that, as long as patient migration is not budget neutral, a change in quality provision in one region will have budgetary effects in all regions due to changes in patient flows across regions, thereby affecting the tax rates necessary to finance a given provision of health care.[18] This will, in turn, affect the incentives for quality provision.

3.1 First-order conditions

We derive the first-order condition for optimal quality provision in Region i by substituting the budget constraint into the welfare function and maximising it with respect to quality. Thus, on general form, the first-order condition for optimal quality provision in Region i is given by

$$\frac{dW_i}{dq_i} = \frac{\partial W_i}{\partial q_i} + \frac{\partial W_i}{\partial \tau_i}\frac{\partial \tau_i}{\partial q_i} = 0, \quad i = H, M, L. \tag{17}$$

Using the expressions for welfare and the budgetary effects of quality changes detailed in Appendix A, we can characterise the optimal quality provision in each region more specifically.

In the *high-income region*, the first-order condition for optimal quality provision is given by

$$\frac{dW_H}{dq_H} = \frac{b}{3} - \frac{\overline{u}_{\tau_H}}{\overline{y}_H}\left(K'(q_H) - (p_H - c_H)\,\frac{b}{t}\right) = 0. \tag{18}$$

[18] See Appendix A for an explicit derivation of the effects of quality provision on tax rates in each of the three regions.

The first term is the marginal utility of health care quality in Region H. Notice that all patients in Region H seek treatment in their own region and therefore all benefit from an increase in q_H. The second term is the marginal cost of health care quality, which is the higher income tax rate necessary to finance a marginal quality improvement, times the utility loss of higher taxes. Higher quality will attract more patients from neighbouring regions. Thus, the amount of tax revenues that need to be raised in order to finance higher health care quality depends partly on the profitability of treating such patients. If $p_H > (<)\, c_H$, less (more) tax revenues need to be raised when patient immigration increases.

In the *low-income region*, the first-order condition for optimal quality provision is given by

$$\begin{aligned} \frac{dW_L}{dq_L} &= b\left(\frac{1}{3} - \left(\Phi_{LH} + \Phi_{LM}\right)\right) - \frac{\left[K'\left(q_L\right) - \left(p_H + p_M - 2\pi - 2c_L\right)\frac{b}{2t}\right]}{\overline{y}_L\left(1+\psi_L\right)} \\ &\quad \times \begin{bmatrix} \overline{u}_{\tau_L} + 3\lambda_L\left(\phi^R_{LH} + \phi^R_{LM}\right) y_R \left(u'\left(\widehat{Y}^R_L\right) - u'\left(Y^R_L\right)\right) \\ +3\left(1-\lambda_L\right)\left(\phi^P_{LH} + \phi^P_{LM}\right) y_P \left(u'\left(\widehat{Y}^P_L\right) - u'\left(Y^P_L\right)\right) \end{bmatrix} = 0, \end{aligned} \tag{19}$$

where ψ_L is defined by equation (A7) in Appendix A. The first term in (19) is the marginal utility of higher quality for patients in Region L. These benefits accrue only to the patients who seek treatment in their own region. Because of patient emigration, these are lower than the corresponding marginal benefits in the high-income region. The second term is the marginal cost of health care quality, which is the utility loss of higher taxes. The second and third term in the square brackets are the extra (per patient) utility loss of higher taxes for rich and poor patients, respectively, who travel out of the region for treatment. If the copayment is positive ($\pi > 0$), the net income is higher for patients who stay than for patients who go, implying an extra utility loss of higher taxes (because of decreasing marginal utility of income) for patients who are treated in other regions.

Finally, in the *middle-income region*, the first-order condition for optimal quality provision is given by

$$\begin{aligned} \frac{dW_M}{dq_M} &= b\left(\frac{1}{3} - \Phi_{MH}\right) - \frac{\left[K'\left(q_M\right) - \left(p_H - \pi - c_M\right)\frac{b}{2t}\right]}{\overline{y}_M\left(1+\psi_M\right)} \\ &\quad \times \begin{bmatrix} \overline{u}_{\tau_M} + 3\lambda_M \phi^R_{MH} y_R \left(u'\left(\widehat{Y}^R_M\right) - u'\left(Y^R_M\right)\right) \\ +3\left(1-\lambda_M\right)\phi^P_{MH} y_P \left(u'\left(\widehat{Y}^P_M\right) - u'\left(Y^P_M\right)\right) \end{bmatrix} = 0, \end{aligned} \tag{20}$$

where ψ_M is defined by equation (A6) in Appendix A. The interpretation is equivalent to the interpretation for optimal quality in Region L, where the differences only account for differences in cross-regional patient flows.

3.2 Nash equilibrium

The (interior solution) Nash equilibrium is given by the triple (q_H^*, q_M^*, q_L^*) that solves the system of equations given by (18)-(20). For the remainder of the analysis, we will assume that prices are set equal to the marginal treatment costs, i.e., $p_H = c_H$ and $p_M = c_M$. This is in line with current DRG pricing in several countries where fixed costs are not included in the tariff. This is the case in Denmark, Germany, Ireland, Norway, Portugal, Spain and Switzerland (Levaggi et al., 2014).[19] Moreover, Levaggi et al. (2012) show that it is optimal to pay the provider based on a marginal cost pricing rule and fund capital costs separately; otherwise the price is too high and undertreatment arises.

We also make the realistic assumption that the patient copayment π is such that a patient from Region j seeking treatment in Region i never pays more than the difference between the price charged by the importing region and the price reimbursed by the exporting region; i.e., $\pi \leq c_i - c_j$. As mentioned in the Introduction, this is in line with the new EU directive. In Appendix B we provide sufficient conditions for the Nash equilibrium to exist under these assumptions.

[Figures 1 and 2 here]

The Nash equilibrium outcome is illustrated in Figures 1 and 2. In Figure 1 (Figure 2) the blue (red) dots on the circle indicate the location of rich (poor) patients who are indifferent between seeking health care in their home region and being treated in the nearest neighbouring region. Total migration is constituted by all patients (rich and poor) who are located on the intervals between the dots (blue and red, respectively) and the nearest regional border (indicated by a blue square in the figures). The relative positioning of the red and blue dots indicate that rich patients are more prone to seek cross border health care. In other words, the share of rich people is higher among the migrating patients than in the general population. Only in the special case of $\pi = 0$ do the red and blue dots coincide on the circle.

[19] Other countries, such as Austria, Finland and the Netherlands, fund capital costs through the DRG price, whereas other countries, such as France, Italy and Poland, use either one or the other pricing rule depending on the treatment.

4 Strategic interaction between regions

In order to understand the main mechanisms involved in the model, it is instructive to study the nature of the strategic interaction between the different regions. In order to facilitate this analysis, and the comparative statics analysis in the subsequent sections, we also make the additional simplifying assumption of $u''' (\cdot) = 0$, which implies that the marginal utility of income is decreasing at a constant rate. This allows us also to adopt a more compact notation. Let

$$\Delta u_Y := u' \left(\widehat{Y}_i^k \right) - u' \left(Y_i^k \right) > 0 \tag{21}$$

denote the difference in the marginal utility of income for patients who move and pay a copayment and those who do not. Under the assumption $u''' (\cdot) = 0$, notice that Δu_Y is the same for all individuals, regardless of whether they are rich or poor, and regardless of which region they move from and to. Let

$$\overline{u}_{\tau_i \tau_i} := u'' (\cdot) \left[\lambda_i \left(y^R \right)^2 + (1 - \lambda_i) \left(y^P \right)^2 \right] < 0 \tag{22}$$

denote the expected degree of concavity (across individuals with different income) of utility with respect to the tax rate; let

$$u_{\pi_i} := \lambda_i u' \left(\widehat{Y}_i^R \right) + (1 - \lambda_i) u' \left(\widehat{Y}_i^P \right) > 0 \tag{23}$$

denote the marginal utility of income due to a reduction in copayment; and, finally, let

$$\widetilde{K}_M := \frac{K' (q_M) - (p_H - \pi - c_M) b/2t}{\overline{y}_M (1 + \psi_M)} > 0 \tag{24}$$

and

$$\widetilde{K}_L := \frac{K' (q_L) - (p_H + p_M - 2\pi - 2c_L) b/2t}{\overline{y}_L (1 + \psi_L)} > 0 \tag{25}$$

denote the marginal costs of quality in Region M and Region L, respectively, adjusted for budgetary effects of patient migration.

If quality provision increases in one region, what is the optimal response by the policy maker in a neighbouring region? This depends crucially on the direction of patient flow between these two regions. Suppose that the patient flow in equilibrium is from Region j to Region i (which implies that $i = H, M,\ j = M, L,\ i \neq j$). Our assumptions on prices and copayments imply that the budgetary effects of quality provision (see the general expressions in Appendix A) are characterised by $\partial\tau_j/\partial q_i > 0$ and $\partial\tau_i/\partial q_j = 0$. In other words, patient immigration is budget neutral for the host region, while there is a negative budget effect in the region from which patients emigrate, as long as the migrating patients do not cover the entire difference in treatment costs themselves. The budget neutrality of patient immigration for the host region implies that $d^2W_i/dq_i dq_j = 0$; i.e., a quality increase in a patient-exporting region does not affect the optimal quality provision in a patient-importing region.

However, a quality increase in a host region will affect the incentives for quality provision in regions from which patients are emigrating. The policy response of Region j to a quality increase in Region i (i.e., the policy response to a quality increase in a patient-importing region), is given by

$$\begin{aligned} \frac{d^2W_j}{dq_j dq_i} &= -\frac{b}{2t}\left(b - \Delta u_Y \overline{y}_j \frac{\partial \tau_j}{\partial q_i}\right) \\ &\quad -\widetilde{K}_j\left[\frac{3b}{2t}\Delta u_Y \overline{y}_j - \frac{\partial \tau_j}{\partial q_i}\gamma_j \left(\Delta u_Y\right)^2 \left(\lambda_j \left(y^R\right)^2 + (1-\lambda_j)\left(y^P\right)^2\right) - \frac{\partial \tau_j}{\partial q_i}\overline{u}_{\tau_j\tau_j}\right], \quad (26) \\ i &= H, M, \quad j = M, L, \quad i \neq j, \end{aligned}$$

where

$$\gamma_j = \begin{cases} \frac{3}{t} & if \quad j = L \\ \frac{3}{2t} & if \quad j = M \end{cases}.$$

The optimal response can be decomposed into three different effects: (i) a direct utility effect, (ii) an income effect, and (iii) indirect effects through the budget constraint. The first two effects are unambiguously negative and the third effect is also likely negative. Below we will explain each of these three effects in detail.

(i) An increase in quality of a neighbouring (patient-importing) region reduces the number of patients who seek care in the home region. This reduces the marginal benefit of the exporting region to provide quality. This effect is captured by the first term in (26).

(ii) If patients who travel to a neighbouring region pay a copay ($\pi > 0$), their net income is lower and the marginal utility from income higher, which effectively increases the marginal cost of providing quality in the exporting region. This effect is captured by the first term on the second line of (26).

(iii) Since patient mobility is unprofitable for the exporting region, which is the case for $\pi <$

$c_i - c_j$, higher quality in a neighbouring region requires that more tax revenues need to be raised in order to maintain the same level of quality provision (i.e., $\partial\tau_j/\partial q_i > 0$). This budget effect is captured by the sum of the terms including $\partial\tau_j/\partial q_i$ in (26) and has one first-order and two second-order effects on the incentives for quality provision. The first-order effect, given by the last term in the square brackets of (26), is that the tax increase will increase the marginal cost of quality provision (because of decreasing marginal utility of income). This effect is dampened by two second-order effects (given by the two other terms in (26) that include $\partial\tau_j/\partial q_i$). Because of copayments ($\pi > 0$), and since $u''(\cdot) < 0$, a higher tax rate will make it relatively more costly for patients to travel out of the region, which dampens mobility and therefore increases (reduces) the marginal benefit (cost) of quality provision.

Thus, given that the two second-order effects described above do not outweigh the first-order effect, the third effect reinforces the first two effects. We conduct the remainder of the analysis under this highly likely assumption, implying that quality in an exporting region is a strategic substitute to quality in an importing region; i.e., $d^2W_M/dq_M dq_H < 0$, $d^2W_L/dq_L dq_H < 0$ and $d^2W_L/dq_L dq_M < 0$.

5 Cross-border health care

In this section, we analyse how policies intended to increase cross-border patient mobility is likely to affect quality provision and social welfare in each of the three regions. We do so by conducting comparative statics with respect to each of the two mobility cost parameters; the non-monetary cost F and the monetary cost π. A reduction in the patient copayment, π, has a straightforward policy interpretation, while a reduction in F can be interpreted as a policy to reduce the 'red tape' costs of seeking treatment in another region.

5.1 Administrative mobility costs

Suppose policymakers reduce the complexity of administrative procedures to obtain health care in a different region, in order to encourage mobility. In our model, this policy corresponds to a reduction in F. For given quality levels, lower non-monetary mobility costs increase patient mobility and thus the demand for cross-border health care. How will this affect regional health care quality? Applying Cramer's rule, we obtain

$$
\begin{aligned}
\frac{dq_H^*}{dF} &= 0, & (27)\\
\frac{dq_M^*}{dF} &= \left(-\frac{d^2W_M}{dq_M^2}\right)^{-1}\frac{d^2W_M}{dq_M dF}, & (28)\\
\frac{dq_L^*}{dF} &= \left(-\frac{d^2W_L}{dq_L^2}\right)^{-1}\frac{d^2W_L}{dq_L dF}+\frac{dq_M^*}{dF}\frac{d^2W_L}{dq_L dq_M}, & (29)
\end{aligned}
$$

where

$$
\frac{d^2W_M}{dq_M dF}=\frac{b}{2t}\left(1-\rho_M\frac{\partial\tau_M}{\partial F}\right)+\widetilde{K}_M\left[\frac{3\rho_M}{2t}-\overline{u}_{\tau_M\tau_M}\left(\frac{3\left(\Delta u_Y\right)^2}{u''\left(\cdot\right)t}-1\right)\frac{\partial\tau_M}{\partial F}\right]>0, \tag{30}
$$

$$
\frac{d^2W_L}{dq_L dF}=\frac{b}{t}\left(1-\rho_L\frac{\partial\tau_L}{\partial F}\right)+\widetilde{K}_L\left[\frac{3\rho_L}{t}-\overline{u}_{\tau_L\tau_L}\left(\frac{3\left(\Delta u_Y\right)^2}{u''\left(\cdot\right)t}-1\right)\frac{\partial\tau_L}{\partial F}\right]>0, \tag{31}
$$

and

$$
\begin{aligned}
\frac{\partial\tau_M}{\partial F} &= -\frac{3\left(c_H-\pi-c_M\right)}{2t\overline{y}_M\left(1+\psi_M\right)}\leq 0, & (32)\\
\frac{\partial\tau_L}{\partial F} &= -\frac{3\left(c_H+c_M-2\pi-2c_L\right)}{2t\overline{y}_L\left(1+\psi_L\right)}\leq 0, & (33)
\end{aligned}
$$

and where ρ_M and ρ_L are defined by (A8) in Appendix A.

The quality of the *high-income* region is not affected by lower non-monetary mobility costs, since patient mobility is budget neutral and therefore has no implication for the tax rate in Region H.

Health-care quality in *middle-income* region goes down when non-monetary mobility costs reduce. We identify three different effects. For given quality levels, a lower F leads to less demand for health care in the middle-income region and therefore lower marginal benefit of quality in this region. This effect tends to reduce quality. More patient mobility also increases the number of patients who have to pay a copayment, which reduces the marginal utility of income and tightens the budget constraint. This effect also tends to reduce quality. Finally, as long as $\pi < c_H - c_M$, migration of patients to the high-income region is costly for the government of the middle-income region, implying that higher mobility increases the income tax rate ($\partial\tau_M/\partial F < 0$), which also contributes to lower quality provision in Region M. Thus, the three direct effects lead to lower quality in the middle-income region, which *in equilibrium* reinforces the positive effect on patient flows to the high-income region induced by a lower F, i.e.,

$$\frac{d\Phi_{MH}^*}{dF} = \frac{\partial \Phi_{MH}}{\partial F} + \frac{\partial \Phi_{MH}}{\partial q_M}\frac{\partial q_M^*}{\partial F} + \frac{\partial \Phi_{MH}}{\partial q_H}\frac{\partial q_H^*}{\partial F} = -\frac{1}{2t}\left(1 + b\frac{\partial q_M^*}{\partial F}\right) < 0. \tag{34}$$

For the *low-income* region, there are direct effects of a reduction in F, and an indirect effect through the strategic responses to quality changes in other regions. The direct effects tend to reduce quality and are equivalent to the ones described above for the middle-income region. However, the indirect effect tends to increase quality. As discussed in Section 3, qualities in the low- and middle-income regions are strategic substitutes. Since the quality of the middle-income region decreases in response to higher mobility, the indirect effect induces the low-income region to increase quality. The net effect on equilibrium patient mobility in the low-income region is therefore generally ambiguous, i.e.,

$$\begin{aligned} \frac{d\left(\Phi_{LM}^* + \Phi_{LH}^*\right)}{dF} &= \frac{\partial\left(\Phi_{LM} + \Phi_{LH}\right)}{\partial F} + \frac{\partial\left(\Phi_{LM} + \Phi_{LH}\right)}{\partial q_L}\frac{\partial q_L^*}{\partial F} + \frac{\partial \Phi_{LM}}{\partial q_M}\frac{\partial q_M^*}{\partial F} + \frac{\partial \Phi_{LH}}{\partial q_H}\frac{\partial q_H^*}{\partial F} \\ &= -\frac{1}{t}\left(1 + b\frac{\partial q_L^*}{\partial F}\right) + \frac{b}{2t}\frac{\partial q_M^*}{\partial F} \lesseqgtr 0. \end{aligned} \tag{35}$$

Thus, the negative quality response in the middle-income region and the ambiguous quality response in the low-income region makes the overall effect on patient mobility out of the low-income region indeterminate.

We summarise the above-described effects as follows:

Proposition 1 *A reduction in the non-monetary cost of mobility has no effect on quality in the high-income region and reduces quality in the middle-income region. The effect on quality in the low-income region is a priori indeterminate and depends on the sum of three different effects: for given quality levels, a reduction in F leads to (i) higher demand for cross-border health care and (ii) higher total cost of health care, which is paid partly by migrating patients and partly by the remaining tax payers in Region L. These two direct effects, which contribute to lower quality provision in Region L, are mitigated by the indirect effect of (iii) lower quality in Region M, which instigates a positive quality response from the low-income region.*

5.2 Patient copayments

As an alternative to reducing non-monetary costs of mobility, suppose that policymakers stimulate cross-border patient mobility by reducing patient copayments (or other monetary costs of mobility). The effects of a change in π on qualities are given by[20]

$$
\begin{aligned}
\frac{dq_H^*}{d\pi} &= 0, && (36)\\
\frac{dq_M^*}{d\pi} &= \left(-\frac{d^2W_M}{dq_M^2}\right)^{-1}\frac{d^2W_M}{dq_M d\pi}, && (37)\\
\frac{dq_L^*}{d\pi} &= \left(-\frac{d^2W_L}{dq_L^2}\right)^{-1}\left(\frac{d^2W_L}{dq_L d\pi}+\frac{dq_M^*}{d\pi}\frac{d^2W_L}{dq_L dq_M}\right). && (38)
\end{aligned}
$$

where

$$
\begin{aligned}
\frac{d^2W_M}{dq_M d\pi} &= \frac{b}{2t}\left(u_{\pi_M}-\rho_M\frac{\partial\tau_M}{\partial\pi}\right)+\widetilde{K}_M\left[\frac{3\rho_M}{2t}-\overline{u}_{\tau_M\tau_M}\left(\frac{3\left(\Delta u_Y\right)^2}{2tu''(\cdot)}-1\right)\frac{\partial\tau_M}{\partial\pi}\right]\\
&\quad-\frac{\left[\overline{u}_{\tau_M}+3\Delta u'\left(\lambda_M\phi_{MH}^R y^R+(1-\lambda_M)\phi_{MH}^P y^P\right)\right]}{(1+\psi_M)}\left(\frac{b}{2t\overline{y}_M}-3u''(\cdot)\widetilde{K}_M\right), \qquad (39)
\end{aligned}
$$

$$
\begin{aligned}
\frac{d^2W_L}{dq_L d\pi} &= \frac{b}{t}\left(u_{\pi_L}-\rho_L\frac{\partial\tau_L}{\partial\pi}\right)+\widetilde{K}_L\left[\frac{3\rho_L}{t}-\overline{u}_{\tau_L\tau_L}\left(\frac{3\left(\Delta u_Y\right)^2}{tu''(\cdot)}-1\right)\frac{\partial\tau_L}{\partial\pi}\right]\\
&\quad-\frac{1}{(1+\psi_L)}\left(\frac{b}{2t\overline{y}_L}-3u''(\cdot)\widetilde{K}_L\right)\begin{array}{c}\left\{\overline{u}_{\tau_L}+3\Delta u_Y\times\left[\lambda_L\left(\phi_{LH}^R+\phi_{LM}^R\right)y^R\right.\right.\\ \left.\left.+(1-\lambda_L)\left(\phi_{LH}^P+\phi_{LM}^P\right)y^P\right]\right\}\end{array}, \qquad (40)
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial\tau_M}{\partial\pi} &= -\frac{2t\Phi_{MH}+3\left(c_H-\pi-c_M\right)u_{\pi_M}}{2t\overline{y}_M\left(1+\psi_M\right)}<0, && (41)\\
\frac{\partial\tau_L}{\partial\pi} &= -\frac{2t\left(\Phi_{LH}+\Phi_{LM}\right)+3\left(c_H+c_M-2\pi-2c_L\right)u_{\pi_L}}{2t\overline{y}_L\left(1+\psi_L\right)}<0. && (42)
\end{aligned}
$$

For given quality levels, a reduction in π lowers the cost of seeking care in a different region and therefore increases the demand for cross-border health care. However, differently from a change in non-monetary costs, a reduction in copayments also tightens the budget constraints of the middle-

[20] The results in this section would be qualitatively similar if we instead assume a cost-sharing system where a patient from Region j seeking treatment in Region i pays $\pi^{ji}=\vartheta\left(p_i-p_j\right)=\vartheta\left(c_i-c_j\right)$, where $\vartheta\in(0,1)$. In such a system, the special case $\vartheta=1$ (which corresponds to $\pi=c_i-c_j$ in the current model set-up) would imply that cross-border patient mobility is budget neutral (i.e., $\partial\tau_i/\partial F=0$).

and low-income regions: a smaller copayment increases the tax rate necessary to finance patient mobility.

As before, the quality of the *high-income* region is not affected by higher demand due to lower π, since patient mobility is budget neutral for Region H. Moreover, since no patients from the high-income region moves to a different region, there are no copayments paid.

The effect of a lower copayment in the *middle-income* region is qualitatively similar to a reduction in non-monetary costs F, and therefore tends to reduce quality. However, it contains one additional effect, given by the last term in (39). Since a lower copayment makes mobility more expensive for the government in the middle-income region, this region has stronger incentives to raise quality in order to limit migration to other regions: less expenses are now paid directly by the patient and mobility has a stronger negative effect on the government's finances. This effect could potentially reverse the effect on the quality provision in the middle-income region: quality may *increase* rather than decrease. The effect on equilibrium patient flows from the middle-income to the high-income region is therefore also generally ambiguous, i.e.,

$$\frac{d\Phi_{MH}^*}{d\pi} = \frac{\partial \Phi_{MH}}{\partial \pi} + \frac{\partial \Phi_{MH}}{\partial q_M}\frac{\partial q_M^*}{\partial \pi} + \frac{\partial \Phi_{MH}}{\partial q_H}\frac{\partial q_H^*}{\partial \pi} = -\frac{1}{2t}\left(u_{\pi_M} + b\frac{\partial q_M^*}{\partial \pi}\right) \lesseqgtr 0 \tag{43}$$

A similar direct effect of a copayment reduction, given by the last term in (40), can be identified for the *low-income* region. However, the overall effect on quality in the low-income region also depends on an indirect effect, namely the strategic response to quality changes in the middle-income region. Suppose that the sum of the direct effects are such that a copayment reduction tends to reduce (increase) quality in the low- and middle-income region. Because of strategic substitutability, the reduction (increase) in quality in the low-income region is attenuated by the strategic response to the reduction (increase) of quality in the middle-income region, which – in isolation – tends to increase (reduce) quality in the low-income region. The effect on equilibrium patient flows from the low-income region to the high-income and middle-income regions is therefore also generally ambiguous:

$$\begin{aligned} \frac{d\left(\Phi_{LM}^* + \Phi_{LH}^*\right)}{d\pi} &= \frac{\partial\left(\Phi_{LM} + \Phi_{LH}\right)}{\partial \pi} + \frac{\partial\left(\Phi_{LM} + \Phi_{LH}\right)}{\partial q_L}\frac{\partial q_L^*}{\partial \pi} + \frac{\partial \Phi_{LM}}{\partial q_M}\frac{\partial q_M^*}{\partial \pi} + \frac{\partial \Phi_{LH}}{\partial q_H}\frac{\partial q_H^*}{\partial \pi} \\ &= -\frac{1}{t}\left(u_{\pi_L} + b\frac{\partial q_L^*}{\partial \pi}\right) + \frac{b}{2t}\frac{\partial q_M^*}{\partial \pi} \lesseqgtr 0. \end{aligned} \tag{44}$$

The effects of lower monetary mobility costs on regional quality provision are summarised as follows:

Proposition 2 *A reduction in patient copayments has no effect on quality in the high-income region. The effect on quality in the low- and middle-income regions is a priori indeterminate and depends on the following three direct effects: for given quality levels, a reduction in π leads to (i) higher demand for cross-border health care and (ii) higher total cost of health care, which is paid partly by migrating patients and partly by the remaining tax payers in the exporting regions. These two effects, which contribute to lower quality provision in Regions L and M, are mitigated by the fact that (iii) the cost of patient emigration increases, since domestic tax payers have to pay a larger fraction of the higher treatment costs, which instigates a positive quality response from the exporting regions. In addition, there is an indirect effect on quality provision in Region L because of the strategic response to quality changes in Region M.*

A comparison of Propositions 1 and 2 reveals that a reduction of mobility costs might have qualitatively different effects on regional quality provision depending on whether these mobility costs are monetary or non-monetary. More specifically, whereas a reduction in non-monetary mobility costs unambiguously leads to a quality reduction in the middle-income region, this might not be the case if, instead, monetary mobility costs are reduced. In order to shed some more light on this, and derive some explicit conditions under which lower monetary mobility costs will lead to a quality increase in Region M, let us consider the following specific assumptions:

A1 The copayment π is initially *small.*

A2 $\lambda_M = \frac{1}{2}$.

A3 $u(Y) = \alpha Y - \frac{\beta}{2}Y^2$, with $\alpha > \beta y^R$ and $\beta > 0$.

Under these assumptions, the following result applies:[21]

Proposition 3 *Suppose that A1-A3 hold. The following three conditions are then sufficient for a copayment reduction to increase equilibrium quality provision in the middle-income region: (i) income inequality between rich and poor is sufficiently small, (ii) the difference in treatment costs between the high- and middle-income regions is sufficiently small, and (iii) the non-monetary mobility cost F is sufficiently high.*

[21] See Appendix D for a formal proof.

The above proposition confirms that a parameter set for which reduced monetary mobility costs leads to higher quality provision in the middle-income region exists. If the three conditions in Proposition 3 are satisfied, the first two effects detailed in Proposition 2 are sufficiently small to be outweighed by the third effect.

If λ_L is sufficiently close to 1/2, similar conditions guarantee that the direct effects of a lower copayment are negative with respect to quality provision also in the low-income region. Thus, in this case, if the indirect effect is sufficiently low (i.e., if the degree of strategic substitutability in quality provision for the low- and middle-income regions is sufficiently weak), a marginal reduction in monetary (non-monetary) mobility costs leads to higher (lower) quality provision in both the middle- and the low-income region.

5.3 Regional welfare

How does an increase in the demand for cross-border health care – because of lower monetary or non-monetary mobility costs – affect social welfare in each of the three regions? Before proceeding to answer this question, it is instructive to consider some general characteristics of the welfare comparative statics. The welfare in each region is given by $W_i(q_H^*(x), q_M^*(x), q_L^*(x), x)$, where x is the parameter of interest. Notice first that, due to the envelope theorem, $\partial W_i^*/\partial q_i^* = 0$. Moreover, because of the assumption $p_i = c_i$, indirect welfare effects in a particular region only come from quality changes in neighbouring regions to which the region in question is *exporting* patients. Assuming that the patient flow is from Region j to Region i, indirect welfare effects are given by

$$\frac{\partial W_j^*}{\partial q_i^*} = b\Phi_{ji} + \frac{\overline{u}_{\tau j}}{3}\frac{\partial \tau_j}{\partial q_i} \tag{45}$$

and are generally ambiguous. On the one hand, higher quality in Region i increases the utility of the patients who travel from Region j (first term in (45)). On the other hand, higher quality in Region i increases patient export to this region, which – if $\pi < c_i - c_j$ – implies a higher tax burden for the residents of Region j (second term in (45)).

5.3.1 Administrative mobility costs

Consider a reduction in F. For the *high-income* region the effect is zero: there are no patients moving from the high-income region to other regions, and mobility is budget neutral.

The effect of lower non-monetary mobility costs on welfare in the *low-income* region is

$$\frac{dW_L^*}{dF} = \frac{\partial W_L^*}{\partial F} + \frac{\partial W_L^*}{\partial q_M^*}\frac{\partial q_M^*}{\partial F}, \tag{46}$$

where

$$\frac{\partial W_L^*}{\partial F} = -\left(\Phi_{LH} + \Phi_{LM}\right). \tag{47}$$

A reduction in F has three different effects – two positive and one negative – on welfare in the low-income region: (i) it directly reduces the cost for all patients who move, which has a positive welfare effect; it also leads to lower quality in Region M, which (ii) reduces utility for those patients who travel to Region M, but also (iii) implies a tax reduction because of less (costly) patient emigration to the middle-income region.

Finally, the effect of lower non-monetary mobility costs in the *middle-income* region is simply given by

$$\frac{dW_M^*}{dF} = \frac{\partial W_M^*}{\partial F} = -\Phi_{MH} < 0. \tag{48}$$

Thus, lower costs of mobility has an unambiguously positive effect on welfare in Region M, because of the cost reduction for those patients who travel to Region H for treatment.

Proposition 4 *A reduction in the non-monetary cost of mobility has no effect on welfare in the high-income region and a positive welfare effect in the middle-income region. The welfare effect in the low-income region is a priori indeterminate and depends on three (one direct and two indirect) effects. The direct effect is positive: (i) lower costs for migrating patients. The indirect effects are caused by the quality reduction in Region M, which leads to (ii) lower taxes in Region L because of the reduction in patient migration to Region M, and (iii) lower quality of health care for patients who still migrate to Region M. The overall welfare effect is positive if the first two effects outweigh the third effect, and vice versa.*

These welfare results suggest that there might be both winners and losers from a policy of facilitating cross-border patient mobility by reducing administrative costs, and that the potential losers are patients and tax payers in the low-income region.

5.3.2 Patient copayments

As for the case of non-monetary mobility costs, a reduction in the monetary cost π has no effect on welfare in the *high-income* region: there are no patients moving from the high-income region to other regions, and mobility is budget neutral.

However, in contrast to the effect of non-monetary mobility costs, the direct effects of a copayment reduction on welfare in the *low-income* and *middle-income* regions are ambiguous, and given by

$$\frac{\partial W_L^*}{\partial \pi} = -\left(\Phi_{LH} + \Phi_{LM}\right) u_{\pi_L} - \frac{\overline{u}_{\tau_L}}{3} \frac{\partial \tau_L}{\partial \pi} \tag{49}$$

and

$$\frac{\partial W_M^*}{\partial \pi} = -\Phi_{MH} u_{\pi_M} - \frac{\overline{u}_{\tau_M}}{3} \frac{\partial \tau_M}{\partial \pi}. \tag{50}$$

On the one hand, and similarly to a reduction in non-monetary mobility costs, a lower copayment increases utility for those who move to a different region to obtain health care. On the other hand, it increases the tax rate necessary to finance patient emigration.

In the middle-income region, the welfare effect is given only by the direct effect in (50), whereas the overall welfare effect in the low-income region is

$$\frac{dW_L^*}{d\pi} = \frac{\partial W_L^*}{\partial \pi} + \frac{\partial W_L^*}{\partial q_M^*} \frac{\partial q_M^*}{\partial \pi}, \tag{51}$$

where the sign of the indirect effect (because of strategic substitutability) is generally ambiguous.

Proposition 5 *A reduction in patient copayments has no effect on welfare in the high-income region. In the low- and middle-income regions, the welfare effect is a priori indeterminate and depends on two counteracting direct effects: a reduction in π leads to (i) lower costs for migrating patients, but (ii) higher taxes necessary to finance patient export. In addition, the overall welfare effect in the low-income region is also determined by indirect effects caused by quality changes in Region M, as detailed in Proposition 4.*

Similarly to the effects on regional quality provision, Proposition 5 reveals that the effect of lower mobility costs on welfare in the middle-income region might crucially depend on whether these costs are monetary or non-monetary. As in Section 5.2, we can further characterise the effect of lower monetary mobility costs by considering a specific parametric example:[22]

[22] See Appendix D for a formal proof.

Proposition 6 *Suppose that A1-A3 hold. For $c_H > c_M$, the following two conditions are then sufficient for a reduction in patient copayment to reduce welfare in the middle-income region: (i) income inequality between rich and poor is sufficiently small, (ii) the non-monetary mobility cost F is sufficiently large.*

This proposition confirms the existence of a parameter set for which the welfare effect of lower mobility costs in the middle-income region qualitatively depends on whether these costs are monetary or non-monetary. For this parameter set, the total utility gain of lower copayments for migrating patients is outweighed by the corresponding increase in taxes for the remaining population in Region M. If λ_L is sufficiently close to 1/2, a similar condition ensures that the direct welfare effects of lower patient copayments (cf. Proposition 5) are negative also in Region L.

Generally, Propositions 4 and 5 suggest that a policy of stimulating cross-border patient mobility might have adverse welfare effects at regional level. When seen in conjunction, these two propositions also suggest that such adverse effects might be less likely if the policy implies a reduction in non-monetary, rather than monetary, mobility costs. For example, under the conditions given by Proposition 6, and if the indirect effect of quality in Region M on welfare in Region L is sufficiently small, a reduction in monetary (non-monetary) mobility costs will reduce (increase) welfare in both the low- and the middle-income region.

6 Income inequality

In this section we exploit the structural richness of our model to analyse how regional quality provision depends on the degree of income inequality – both across and within regions – when patients have the option to seek treatment outside their own region.

6.1 Inter-regional income inequality

In order to study the effects of inter-regional income inequality on regional quality provision, we assume that $\lambda_H = \lambda_M + \delta$ and $\lambda_L = \lambda_M - \delta$, where δ measures the degree of income dispersion across regions. An increase in δ has no effect on the income distribution in the middle-income region, increases the proportion of rich individuals in the high-income region and reduces it in the low-income region.

How does an increase in inter-regional income dispersion affect quality provision? The effects are given by

$$\frac{dq_H^*}{d\delta} = \left(-\frac{d^2W_H}{dq_H^2}\right)^{-1} \frac{d^2W_H}{dq_H d\delta}, \tag{52}$$

$$\frac{dq_M^*}{d\delta} = \left(-\frac{d^2W_M}{dq_M^2}\right)^{-1} \frac{dq_H^*}{d\delta}\frac{d^2W_M}{dq_M dq_H}, \tag{53}$$

$$\frac{dq_L^*}{d\delta} = \left(-\frac{d^2W_L}{dq_L^2}\right)^{-1} \left[\frac{d^2W_L}{dq_L d\delta} + \frac{dq_M^*}{d\delta}\frac{d^2W_L}{dq_L dq_M} + \frac{dq_H^*}{d\delta}\frac{d^2W_L}{dq_L dq_H}\right], \tag{54}$$

where $d^2W_H/dq_H d\delta > 0$, while the sign of $d^2W_L/dq_L d\delta$ is indeterminate.[23]

As expected, increased inter-regional income inequality leads to higher quality provision in the *high-income* region. The intuition is relatively simple and consists of two effects. A higher income implies that, for a given tax rate, the average expected marginal utility of income is lower, and the tax rate necessary to finance health care is also lower. Both effects reduce the marginal cost of quality provision. There are no indirect effects since the quality choice of the high-income region is independent of qualities in other regions.

Since the *middle-income* region maintains the same average income, there are no direct effects on quality. However, since qualities are strategic substitutes, the increase in quality by the high-income region triggers a reduction in quality for the middle-income region.

In the *low-income* region, there are several direct and indirect effects. First, an increase in dispersion reduces the average income in the low-income region, which tends to reduce quality due to the higher tax rate and the higher marginal utility of income. Second, an increase in the share of poor patients reduces overall mobility, which increases incentives to provide quality for two reasons: (i) more patients benefit from the quality investment and (ii) lower mobility reduces the tax rate and therefore the marginal cost of quality provision. Finally, there are two indirect effects going in opposite direction: the increase (reduction) in quality in the high- (middle-) income region triggers lower (higher) incentives for quality provision in the low-income region.

Proposition 7 *An increase in inter-regional income inequality leads to higher quality in the high-income region and lower quality in the middle-income region. The quality effect in the low-income region is a priori indeterminate because of two counteracting direct effects: (i) lower average income, which also leads to (ii) less demand for cross-border health care. The first (second) effect discourages (encourages) quality provision. In addition, quality provision in the low-income region is also (iii)*

[23] See Appendix C for explicit expressions.

encouraged by the quality reduction in Region M and (iv) discouraged by the quality increase in Region H.

6.2 Intra-regional income inequality

Let us finally consider how increased income inequality in a particular region affects quality provision in the same and (potentially) other regions. We model income dispersion within Region i as a mean-preserving spread θ such that $\widetilde{y}_R^i := y_R + \frac{\theta}{\lambda_i}$ and $\widetilde{y}_P^i := y_P - \frac{\theta}{1-\lambda_i}$. This definition implies collecting $\frac{\theta}{1-\lambda_i}$ euros from each of the poor and distributing this amount by giving $\frac{\theta}{\lambda_i}$ to each of the rich. Income inequality is increased without affecting average income.

The effect of higher income dispersion in the *high-income* region on the same region's optimal quality provision is positive, since

$$\frac{dq_H^*}{d\theta} = -\frac{d^2 W_H}{dq_H d\theta} / \frac{d^2 W_H}{dq_H^2}. \tag{55}$$

with

$$\frac{dW_H}{dq_H d\theta} = -\left[u'\left(Y_H^R\right) - u'\left(Y_H^P\right) + u''\left(\cdot\right)\left(Y_H^R - Y_H^P\right)\right] \frac{K'\left(q_H\right)}{\overline{y}_H} > 0. \tag{56}$$

Higher income dispersion implies that the rich bear a larger share of the total tax burden. A tax reduction will therefore benefit the rich to a higher degree, which implies that the average utility gain of a marginal tax reduction is lower. This, in turn, implies that the optimal tax rate, and therefore the optimal quality provision, is higher. Given strategic substitutability between regions, a higher income dispersion in the high-income region will then ultimately lead to lower quality in the other two regions.

The effect of higher income dispersion in the *middle-income* region on that region's quality provision is given by

$$\frac{dq_M^*}{d\theta} = -\frac{d^2 W_M}{dq_M d\theta} / \frac{d^2 W_M}{dq_M^2}, \tag{57}$$

where[24]

$$\frac{d^2 W_M}{dq_M d\theta} = \frac{\partial^2 W_M}{\partial \tau_M \partial \theta}\frac{\partial \tau_M}{\partial q_M} = \frac{\partial \tau_M}{\partial q_M} u'\left(Y_M^P\right) - u'\left(Y_M^R\right) - u''\left(\cdot\right)\left(Y_M^R - Y_M^P\right) \tag{58}$$
$$-3\Delta u_Y \frac{\partial \tau_M}{\partial q_M}\left(\frac{1}{2t}\left(\Delta u_Y\left(Y_M^R - Y_M^P\right)\right) + \phi_{MH}^R - \phi_{MH}^P\right)$$

The sign of this expression is a priori ambiguous and depends on the sign of the expression in the square brackets. The first line in (58) is positive and is similar to the one given by (56) for the high-income region. The second line is negative and therefore pulls in the opposite direction. Although higher income dispersion does not affect total patient export (for given quality levels), it affects the *composition* of the patients who choose to travel out of the region, with an increase in the share of rich patients. Thus, higher income dispersion implies that a larger share of rich patients have to pay a copayment π for health care abroad, which, all else equal, increases the marginal utility of income for the rich (on average) and therefore counteracts the effect of higher income dispersion.

Finally, the own-region effects of higher income dispersion in the *low-income* region are equivalent to the ones of the middle-income region described above and therefore not explicitly presented. The only qualitative difference is that a change in quality provision in Region L has no spillover effects to other regions.

Proposition 8 *Higher income inequality in Region H leads to higher quality in the high-income region and lower quality in the middle- and low-income regions, whereas higher income inequality in either Region M or Region L has no effect on quality provision in the high-income region and indeterminate effects on quality in the low- and middle-income regions. This indeterminacy is caused by two counteracting effects: (i) rich people take a larger share of the total tax burden, which stimulates quality provision, but (ii) rich people constitute a larger share of the migrating patients, which counteracts the first effect. In addition, a quality increase (decrease) in Region M will contribute to a quality decrease (increase) in Region L.*

[24] Notice that Δu_Y and Φ_{MH} do not depend on θ (i.e., higher income inequality increases the number of rich patients and reduces the number of poor patients who travel out of the region to be treated, but the net effect is zero), implying that $\partial^2 W_M/\partial q_M \partial\theta = \partial \tau_M/\partial\theta = \partial^2 \tau_M/\partial q_M \partial\theta = 0$.

7 Conclusions

Cross-border patient mobility is an important issue across countries – as exemplified by the new regulation in the EU – and across regions within countries with regional health-care provision, such as Canada, Italy and Sweden. In this paper, we study the consequences of cross-border patient mobility on the quality of health care and the corresponding regional welfare effects. We develop a Salop model with three regions; a high-income, a middle-income, and a low-income region. In each region, health-care quality is set by a policy maker maximising regional welfare subject to health-care costs being financed by taxation. Since the marginal cost of taxation is decreasing in income (due to decreasing marginal utility of income), health-care quality is increasing in the regions' income level. Thus, patient mobility occurs from lower-income regions with poorer health-care quality to higher-income regions with better health-care quality.

We focus on the (interior) equilibrium where (i) the high-income region attracts patients from both the low- and middle-income regions and (ii) the middle-income region attracts patients from the low-income region. Profitability of cross-border patient mobility depends on the transfer payment scheme and we assume DRG-pricing, where the importing region receives a price equal to marginal treatment cost for migrating patients.

While our analysis produces a rich set of results regarding regional effects of cross-border patient mobility on quality provision and welfare, we would like to highlight three different results: First, an increase in patient mobility driven by a reduction in non-monetary mobility costs has no effect on quality in the high-income region, but reduces quality in the middle-income and, if indirect effects are small, also reduces quality in the low-income region. Thus, and perhaps counter-intuitively, patient mobility can have adverse effects on the quality of care in lower-income regions exporting patients to higher-income regions, and can therefore increase dispersion in health care quality between high- and low-income regions. This result may explain the delay in the application of the EU Directive in several member countries.[25]

Second, lower patient copayment for cross-border health care has no effect on quality in the high-income region, and has an indeterminate effect on the quality in the low- and middle-income regions. This result shows that whether increased cross-border patient mobility amplifies or dampens dispersion in health care quality across different countries might crucially depend on the exact mechanism that stimulates mobility.

[25] Evaluative study on the crossborder Healthcare Directive (2011/24/EU) Final report 21 March 2015.

Third, an increase in inter-regional income dispersion increases quality in the high-income region, reduces quality in the middle-income region, whereas the effect on quality provision in the low-income region is indeterminate. This result might assist in predicting the likely effects of austerity and the economic crisis, which has affected EU Member States in a differential way, and has been highlighted in the recent change of wind in the decisions by the European Court of Justice that has ruled against patients asking reimbursement for treatment abroad (Elchinov, Luca and Petru)[26], where patients were coming from countries with relatively lower income (i.e., Romania and Bulgaria). The concern was that, as a result of mobility, quality may decrease for those patients who do not seek care abroad. In summary, the consequences and implications of cross-border patient mobility are far from straightforward.

By way of concluding, we would like to highlight some limitations of our study. First, our results are derived assuming DRG-pricing. While this is a widely used pricing scheme for hospital care in most Western countries, different regions may bilaterally agree on a different way of pricing cross-border care. However, there is still an underlying problem that a patient that is profitable to treat for the importing region might be unprofitable for the exporting region to send. Clearly, designing an optimal payment scheme for cross-border patients is a key challenge. For the derivation of some results we also assume that the DRG price is equal to the treatment cost since this is what we observe in most countries. Other countries set the DRG price above the treatment cost to allow for fixed costs and capital expenses.[27] Under this scenario, even the high-income region would respond to mobility by increasing quality to attract additional patients with positive profit margin. In turn, this would generate additional effects to the middle- and low-income regions on top of those already identified in the current analysis, which depends on the degree of substitution in quality across regions.[28]

Second, our analysis has focused on cross-border patient mobility induced by quality differences across regions. Another important reason for cross-border mobility is impaired access to health care at home due to capacity constraints and long waiting lists. We could interpret waiting time

[26]ECJ judgment in Elchinov, C-173/09, EU:C:2010:581 (from the Bulgaria to Germany); ECJ order in Luca, C-430/12, EU:C:2013:467 (from the Romania to Austria); ECJ judgment in Petru, C-268/13, EU:C:2014:2271 (from Romania to Germany).

[27]An alternative justification for the price mark up to be positive is that the hospital engages in upcoding and receives a price designed for example for patients with additional comorbidities while the patient has none or limited ones. This would involve the explicit modelling of patients with different severity which is outside of the scope of the current analysis.

[28]Such additional effects are often indeterminate and bring limited additional insights into the behaviour of the low- and middle-income country.

as negative quality in our framework, implying that poorer regions have longer waiting times than richer regions inducing patients to migrate to obtain quicker access to care. However, this interpretation ignores capacity constraints. If the richer regions that import patients have excess capacity, then our results would hold. However, if the richer regions have also capacity constrains, then cross-border mobility has a direct adverse effect on the quality (waiting time) of the patients of the richer region, which is not captured by our model. One possible way to account for capacity constraints is to allow for increasing marginal treatment costs. Under this assumption, most of our results will hold as long as the regulated price (DRG-price) is set at the marginal treatment cost in each region. If not, the analysis would be much more involved. We leave this for future research.

Third, in our positive analysis we treat the copayment as exogenous, and restrict its possible values to be in line with the new EU directive which entitles migrating patients the right to receive cost reimbursement only up to what they would have in their home country. Endogenising the copayment would require a model that allows for uncertainty in the health status of the patient. We assume that the patient has already fallen sick and the copayment has already been determined in a previous stage. Adding a proper moral hazard set-up to our model is likely to be intractable. We conjecture that, as suggested by standard insurance theory, the copayment will be determined such that it trades off the benefits from insurance against the costs of overconsumption. We restrict copayments to be between zero and the difference in treatment costs. Zero copayments can be justified from an equity perspective, and many countries of the NHS type set zero copayments for elective care. Positive but low copayments are also consistent with low elasticity of demand for elective care. The exploration of optimal copayments when patients can move across regions seems an interesting venue for future research.

Finally, our analysis is mainly positive and we have not explicitly defined a first best scenario. Suppose for example that a super-regional entity maximises the welfare across the three regions and sets quality in each region taking into account the marginal cost in each region but internalises patients' benefits from other regions. The quality in the high-income region would then be higher then the equilibrium quality level derived in our analysis, since, in our model, the policy maker in the high-income region does not internalise the benefits of patients from other regions. Similarly, the quality in the other regions will generally differ from those derived in our model (depending on the price for importing and exporting patients). A super-regional entity could then design an optimal system of transfers which induces regions to internalise externalities from other regions.

In practice, however, a super-regional entity like this is unlikely to have the authority to impose such system of transfers.[29] An alternative approach would be to calculate a cooperative solution to our game and provide a Pareto-efficient solution which is constrained by the available instruments, which we leave for future research.

References

[1] Aiura, H., Sanjo, Y., 2010. Privatization of local public hospitals: effect on budget, medical service quality, and social welfare. International Journal of Health Care Finance & Economics, 10, 275–299.

[2] Andritsos, D.A., Tang, C.S., 2013. The impact of cross-border patient movement on the delivery of healthcare services. International Journal of Production Economics, 145, 702-712.

[3] Andritsos, D.A., Tang, C.S., 2014. Introducing competition in healthcare services: The role of private care and increased patient mobility. European Journal of Operational Research, 234, 898–909.

[4] Barros, P.P., Siciliani, L., 2011. Public-private interface, chapter 15, p. 927-1002, in Pauly, M., McGuire, T., Barros, P.P. (eds) Elsevier Handbook in Health Economics, Volume 2.

[5] Beckert, W., Christensen, M., Collyer, K. 2012. Choice of NHS-funded hospital services in England. Economic Journal, 122, 400–417.

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

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

[8] Brekke, K.R., Gravelle, H., Siciliani, L., Straume, O.R., 2014a. Patients' choice, mobility and competition among health care providers Health Care Provision and Patient Mobility – Health Integration in the European Union (eds., Levaggi R., Montefiori, M.), Developments in Health Economics and Public Policy, 12, 1-26, Springer.

[29] The first-best solution that maximises the sum of regional welfare would imply a full redistribution of income across regions and equal quality of care in all regions. This is clearly an unrealistic outcome.

[9] Brekke, K.R., Levaggi, R., Siciliani, L., Straume, O.R., 2014b. Patient mobility, health care quality and welfare. Journal of Economic Behavior and Organization, 105, 140–157.

[10] Busse, R., Geissler, A., Quentin, W., Wiley, M., 2011. Diagnosis-Related Groups in Europe: Moving towards transparency, efficiency and quality in hospitals. European Observatory on Health Systems and Policies Series. McGraw-Hill Open University Press, New York.

[11] Cohen, I.G., 2015. Patients with passports: Medical Tourism, Law and Ethics, Oxford University Press.

[12] Commission of the European Communities, 2006. Communication from the commission. Consultation regarding community action on health services. Brussels, September 26, 2006, SEC(2006) 1195/4.

[13] Directive 2011/24/EU of the European Parliament and of the Council of 9 March 2011 on the application of patients' rights in cross-border healthcare

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

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

[16] Gravelle, H., Sivey, P., 2010. Imperfect information in a quality-competitive hospital market. Journal of Health Economics, 29, 524–535.

[17] Kanbur, R., Keen, M., 1993. Jeux sans frontières: tax competition and tax coordination when countries differ in size. American Economic Review, 83, 877–892.

[18] Levaggi, R., Moretto, M., Pertile, P., 2012. Static and dynamic efficiency of irreversible health care investments under alternative payment rules. Journal of Health Economics, 31, 169–179.

[19] Levaggi, R., Moretto, M., Pertile, P., 2014. Two-part payments for the reimbursement of investments in health technologies. Health Policy, 115, 230–236.

[20] Nielsen, S.B., 2001. A simple model of commodity taxation and cross-border shopping. Scandinavian Journal of Economics, 103, 599–624.

[21] OECD, 2014. Health at Glance: Europe 2014, OECD Publishing, Paris.

[22] Petretto, A., 2000. On the cost-benefit of the regionalisation of the National Health Service. Economics of Governance, 1, 213–232.

[23] Salop, S. C., 1979. Monopolistic competition with outside goods. Bell Journal of Economics, 10, 141–156.

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

[25] Trandel, G.A., 1994. Interstate commodity tax differentials and the distribution of residents. Journal of Public Economics, 53, 435–457.

[26] Wang, Y.-Q., 1999. Commodity taxes under fiscal competition: Stackelberg equilibrium and optimality. American Economic Review, 89, 974–981.

Appendix A: Further details of the welfare maximisation problems

Welfare expressions

Aggregate utility of residents in the high-, middle- and low-income regions, respectively, is given by

$$W_H := 2\int_0^{\frac{1}{6}} (v + bq_H - tx)\,dx + \frac{\lambda_H}{3} u\left(Y_H^R\right) + \frac{1-\lambda_H}{3} u\left(Y_H^P\right), \tag{A1}$$

$$\begin{aligned} W_M \ : \ &= \lambda_M \int_0^{\frac{1}{6}-\phi_{MH}^R} \left(v + bq_M + u\left(Y_M^R\right) - tx\right) dx \\ &+\lambda_M \int_0^{\phi_{MH}^R} \left(v + bq_H + u\left(\widehat{Y}_M^R\right) - t\left(\frac{1}{6}+x\right) - F\right) dx \\ &+ (1-\lambda_M) \int_0^{\frac{1}{6}-\phi_{MH}^P} \left(v + bq_M + u\left(Y_M^P\right) - tx\right) dx \\ &+ (1-\lambda_M) \int_0^{\phi_{MH}^P} \left(v + bq_H + u\left(\widehat{Y}_M^P\right) - t\left(\frac{1}{6}+x\right) - F\right) dx \\ &+ \int_0^{\frac{1}{6}} \left(v + bq_M + \lambda_M u\left(Y_M^R\right) + (1-\lambda_M) u\left(Y_M^P\right) - tx\right) dx \end{aligned} \tag{A2}$$

and

$$\begin{aligned} W_L \ : \ &= \lambda_L \int_0^{\frac{1}{6}-\phi_{LH}^R} \left(v + bq_L + u\left(Y_L^R\right) - tx\right) dx \\ &+\lambda_L \int_0^{\frac{1}{6}-\phi_{LM}^R} \left(v + bq_L + u\left(Y_L^R\right) - tx\right) dx \\ &+\lambda_L \int_0^{\phi_{LH}^R} \left(v + bq_H + u\left(\widehat{Y}_L^R\right) - t\left(\frac{1}{6}+x\right) - F\right) dx \\ &+\lambda_L \int_0^{\phi_{LM}^R} \left(v + bq_M + u\left(\widehat{Y}_L^R\right) - t\left(\frac{1}{6}+x\right) - F\right) dx \\ &+ (1-\lambda_L) \int_0^{\frac{1}{6}-\phi_{LH}^P} \left(v + bq_L + u\left(Y_L^P\right) - tx\right) dx \\ &+ (1-\lambda_L) \int_0^{\frac{1}{6}-\phi_{LM}^P} \left(v + bq_L + u\left(Y_L^P\right) - tx\right) dx \\ &+ (1-\lambda_L) \int_0^{\phi_{LH}^P} \left(v + bq_H + u\left(\widehat{Y}_L^P\right) - t\left(\frac{1}{6}+x\right) - F\right) dx \\ &+ (1-\lambda_L) \int_0^{\phi_{LM}^P} \left(v + bq_M + u\left(\widehat{Y}_L^P\right) - t\left(\frac{1}{6}+x\right) - F\right) dx. \end{aligned} \tag{A3}$$

Budgetary effects of quality provision

Using (14)-(16) in Section 3, the effects of a unilateral quality increase in Region H on the income tax rate in each of the three regions are given by

$$\frac{\partial \tau_H}{\partial q_H} = \frac{3}{\overline{y}_H} \left(K'\left(q_H\right) - \left(p_H - c_H\right) \frac{b}{t} \right), \tag{A4}$$

$$\frac{\partial \tau_i}{\partial q_H} = \frac{3b\left(p_H - \pi - c_i\right)}{2t\overline{y}_i\left(1 + \psi_i\right)}, \quad i = L, M, \tag{A5}$$

where

$$\psi_M := \frac{p_H - \pi - c_M}{2t\overline{y}_M} \rho_M, \tag{A6}$$

$$\psi_L := \frac{p_H + p_M - 2\pi - 2c_L}{2t\overline{y}_L} \rho_L \tag{A7}$$

and

$$\rho_i := \lambda_i y_R \left(u'\left(\widehat{Y}_i^R\right) - u'\left(Y_i^R\right) \right) + \left(1 - \lambda_i\right) y_P \left(u'\left(\widehat{Y}_i^P\right) - u'\left(Y_i^P\right) \right) > 0, \quad i = L, M. \tag{A8}$$

Similarly, the effects of a unilateral quality increase in Region M are given by

$$\frac{\partial \tau_H}{\partial q_M} = \frac{3\left(p_H - c_H\right) b}{2t\overline{y}_H}, \tag{A9}$$

$$\frac{\partial \tau_M}{\partial q_M} = \frac{3\left(K'\left(q_M\right) - \left(p_H - \pi + p_M - 2c_M\right) \frac{b}{2t}\right)}{\overline{y}_M\left(1 + \psi_M\right)}, \tag{A10}$$

$$\frac{\partial \tau_L}{\partial q_M} = \frac{3b\left(p_M - \pi - c_L\right)}{2t\overline{y}_L\left(1 + \psi_L\right)}. \tag{A11}$$

Finally, the effects of a unilateral quality increase in Region L are given by

$$\frac{\partial \tau_H}{\partial q_L} = \frac{3\left(p_H - c_H\right) b}{2t\overline{y}_H}, \tag{A12}$$

$$\frac{\partial \tau_M}{\partial q_L} = \frac{3b\left(p_M - c_M\right)}{2t\overline{y}_M\left(1 + \psi_M\right)}, \tag{A13}$$

$$\frac{\partial \tau_L}{\partial q_L} = \frac{3\left(K'\left(q_L\right) - \left(p_H + p_M - 2\pi - 2c_L\right) \frac{b}{2t}\right)}{\overline{y}_L\left(1 + \psi_L\right)}. \tag{A14}$$

Second-order conditions

The second-order conditions for each of the three welfare maximisation problems are

$$\frac{d^2W_H}{dq_H^2} = -\frac{\overline{u}_{\tau_H}}{\overline{y}_H}K''(q_H)+\frac{3}{\overline{y}_H^2}\left(K'(q_H)-(p_H-c_H)\frac{b}{t}\right)^2\left[\lambda_H u''\left(Y_H^R\right)\left(y^R\right)^2+(1-\lambda_H)u''\left(Y_H^P\right)\left(y^P\right)^2\right]<0, \tag{A15}$$

$$\begin{aligned}\frac{d^2W_M}{dq_M^2} = & \frac{b^2}{2t}\\ & +\frac{b}{2t}\left\{\lambda_M y^R\left[u'\left(\widehat{Y}_M^R\right)-u'\left(Y_M^R\right)\right]+(1-\lambda_M)y^P\left[u'\left(\widehat{Y}_M^P\right)-u'\left(Y_M^P\right)\right]\right\}\frac{\partial\tau_M}{\partial q_M}\\ & +\frac{1}{3}\left(\frac{\partial\tau_M}{\partial q_M}\right)^2\left[\lambda_M u''\left(Y_M^R\right)\left(y^R\right)^2+(1-\lambda_M)u''\left(Y_M^P\right)\left(y^P\right)^2\right]\\ & +\frac{1}{3}\left(\frac{\partial\tau_M}{\partial q_M}\right)^2\lambda_M\phi_{MH}^R\left(y^R\right)^2\left[u''\left(\widehat{Y}_M^R\right)-u''\left(Y_M^R\right)\right]\\ & +\frac{1}{3}\left(\frac{\partial\tau_M}{\partial q_M}\right)^2(1-\lambda_M)\phi_{MH}^P\left(y^P\right)^2\left[u''\left(\widehat{Y}_M^P\right)-u''\left(Y_M^P\right)\right]\\ & -\left[\overline{u}_{\tau_M}+3\lambda_M\phi_{MH}^R y^R\left[u'\left(\widehat{Y}_M^R\right)-u'\left(Y_M^R\right)\right]+3(1-\lambda_M)\phi_{MH}^P y^P\left[u'\left(\widehat{Y}_M^P\right)-u'\left(Y_M^P\right)\right]\right]\\ & \times\left\{\frac{K''(q_M)}{\overline{y}_M(1+\psi_M)}+\left(\frac{\partial\tau_M}{\partial q_M}\right)^2\frac{(p_H-c_M)}{6t\overline{y}_M(1+\psi_M)}\times\Theta\right\}<0,\end{aligned} \tag{A16}$$

where $\Theta := \lambda_M\left(y^R\right)^2\left[u''\left(\widehat{Y}_M^R\right)-u''\left(Y_M^R\right)\right]+(1-\lambda_M)\left(y^P\right)^2\left[u''\left(\widehat{Y}_M^P\right)-u''\left(Y_P^M\right)\right]$ and

$$\begin{aligned}\frac{d^2W_L}{dq_L^2} = & \frac{b^2}{t}+\frac{b}{t}\frac{\partial\tau_L}{\partial q_L}\lambda_L\left(\phi_{LH}^R+\phi_{LM}^R\right)y^R\left(u'\left(\widehat{Y}_L^R\right)-u'\left(Y_L^R\right)\right)\\ & +\frac{b}{t}\frac{\partial\tau_L}{\partial q_L}(1-\lambda_L)\left(\phi_{LH}^P+\phi_{LM}^P\right)y^P\left(u'\left(\widehat{Y}_L^P\right)-u'\left(Y_L^P\right)\right)\\ & +\frac{b}{t}\rho_L\frac{\partial\tau_L}{\partial q_L}\\ & +\left(\frac{\partial\tau_L}{\partial q_L}\right)^2\frac{1}{3}\left(\lambda_L u''\left(Y_L^R\right)\left(y^R\right)^2+(1-\lambda_L)u''\left(Y_L^P\right)\left(y^P\right)^2\right)\\ & +\left(\frac{\partial\tau_L}{\partial q_L}\right)^2\lambda_L\left(y^R\right)^2\left(u'\left(\widehat{Y}_L^R\right)-u'\left(Y_L^R\right)\right)^2\\ & +\left(\frac{\partial\tau_L}{\partial q_L}\right)^2(1-\lambda_L)\left(y^P\right)^2\left(u'\left(\widehat{Y}_L^P\right)-u'\left(Y_L^P\right)\right)^2\\ & -\left[3\overline{u}_{\tau_L}+3\lambda_L\left(\phi_{LH}^R+\phi_{LM}^R\right)y^R\left(u'\left(\widehat{Y}_L^R\right)-u'\left(Y_L^R\right)\right)\right]\frac{K''(q_L)}{\overline{y}_L(1+\psi_L)}\\ & -\left[3(1-\lambda_L)\left(\phi_{LH}^P+\phi_{LM}^P\right)y^P\left(u'\left(\widehat{Y}_L^P\right)-u'\left(Y_L^P\right)\right)\right]\frac{K''(q_L)}{\overline{y}_L(1+\psi_L)}<0.\end{aligned} \tag{A17}$$

From (A15) we see that the second-order condition for optimal quality provision in Region H always holds (as long as quality costs are strictly convex or the utility function is strictly concave). The two other conditions hold if the quality cost function is sufficiently convex or if the utility function is sufficiently concave.

Appendix B: Equilibrium existence

We are considering a Nash equilibrium where $q_H^* > q_M^* > q_L^*$ and which is an interior solution where some, but not all, of both rich and poor patients in each region travel to a neighbouring region for health care treatment. Formally, such an interior solution requires $0 < \phi_{MH}^k < \frac{1}{3}$, $\phi_{LH}^k > 0$, $\phi_{LM}^k > 0$ and $\phi_{LH}^k + \phi_{LM}^k < \frac{1}{3}$, where $k = R, P$. To derive the conditions for such an equilibrium to exist, consider first the special case where $c_H = c_M = c_L$, which also implies $\pi = 0$. With the assumption of marginal cost pricing, this further implies that patient migration is budget neutral for all regions. From (18)-(20), the first-order conditions for the candidate equilibrium are then given by

$$\frac{\partial W_H}{\partial q_H} = \frac{b}{3} - \frac{\overline{u}_{\tau_H}}{\overline{y}_H} K'(q_H) = 0, \tag{B1}$$

$$\frac{\partial W_M}{\partial q_M} = b\left(\frac{1}{3} - \Phi_{MH}\right) - \frac{\overline{u}_{\tau_M}}{\overline{y}_M} K'(q_M) = 0, \tag{B2}$$

$$\frac{\partial W_L}{\partial q_L} = b\left(\frac{1}{3} - (\Phi_{LH} + \Phi_{LM})\right) - \frac{\overline{u}_{\tau_L}}{\overline{y}_L} K'(q_L) = 0. \tag{B3}$$

Notice that

$$\frac{\partial\left(\frac{\overline{u}_{\tau_i}}{\overline{y}_i}\right)}{\partial \lambda_i} = -\frac{\left[u'\left(Y_i^P\right) - u'\left(Y_i^R\right)\right] y^R y^P}{\overline{y}_i^2} < 0, \quad i = H, L, M, \tag{B4}$$

which implies

$$\frac{\overline{u}_{\tau_H}}{\overline{y}_H} < \frac{\overline{u}_{\tau_M}}{\overline{y}_M} < \frac{\overline{u}_{\tau_L}}{\overline{y}_L}.$$

Thus, for $q_H = q_M = q_L$, which implies $\Phi_{LH} = \Phi_{LM} = \Phi_{MH} = 0$, it follows that

$$\frac{\partial W_H}{\partial q_H} > \frac{\partial W_M}{\partial q_M} > \frac{\partial W_L}{\partial q_L},$$

which, due to the concavity of W_i, implies $q_H^* > q_M^* > q_L^*$. Since $\Phi_{LH} > \Phi_{LM}$ for $q_H^* > q_M^*$, it follows that cross-border mobility will reinforce the quality dispersion across the three regions. Using the definition of ϕ_{ij}^k, given by (7) in Section 2, an interior solution requires that the following conditions hold:

$$\frac{F}{b} < q_H^* - q_M^* < \frac{F}{b} + \frac{2t}{3b}, \tag{B5}$$

$$q_M^* - q_L^* > \frac{F}{b}, \tag{B6}$$

$$q_H^* + q_M^* - 2q_L^* < \frac{2t}{3b} + \frac{2F}{b}. \tag{B7}$$

It is straightforward to see that all these conditions hold if F is sufficiently small and t sufficiently high. All the above conditions are derived for the case of $c_H = c_M = c_L$, which implies $\pi = 0$. Differences in marginal treatment costs across the regions (and a positive patient copayment) will affect the marginal cost of quality provision in the middle- and low-income regions in an indeterminate way. For a given quality difference, a positive copayment will also reduce the threshold value of F below which an interior solution exists. However, by continuity, equilibrium existence is guaranteed also for sufficiently small treatment cost differences across the regions.

Summing up, a *sufficient* condition for the existence of a Nash equilibrium with $q_H^* > q_M^* > q_L^*$ is that marginal treatment cost differences across regions are relatively small. This equilibrium is an interior solution if F is sufficiently low and t is sufficiently large.

Appendix C: Comparative statics

Notice first that our assumptions of $p_i = c_i$ and $u'''(\cdot) = 0$ implies $d^2W_H/dq_Hdq_M = d^2W_H/dq_Hdq_L = d^2W_M/dq_Mdq_L = 0$. Totally differentiating the FOCs with respect to qualities and a parameter x, we obtain

$$\begin{vmatrix} \frac{d^2W_H}{dq_H^2} & 0 & 0 \\ \frac{d^2W_M}{dq_Mdq_H} & \frac{d^2W_M}{dq_M^2} & 0 \\ \frac{d^2W_L}{dq_Ldq_H} & \frac{d^2W_L}{dq_Ldq_M} & \frac{d^2W_L}{dq_L^2} \end{vmatrix} \begin{vmatrix} dq_H \\ dq_M \\ dq_L \end{vmatrix} + \begin{vmatrix} \frac{d^2W_H}{dq_Hdx} \\ \frac{d^2W_M}{dq_Mdx} \\ \frac{d^2W_L}{dq_Ldx} \end{vmatrix} dx = 0, \tag{C1}$$

where $d^2W_M/dq_Mdq_H < 0$, $d^2W_L/dq_Ldq_H < 0$ and $d^2W_L/dq_Ldq_M < 0$ (see eq. (26) in Section 4 for the full expressions). The determinant of the above matrix is

$$\Delta = \frac{d^2W_H}{dq_H^2}\frac{d^2W_M}{dq_M^2}\frac{d^2W_L}{dq_L^2} < 0. \tag{C2}$$

The effect of a marginal change in x on quality provision in Region H is given by

$$\frac{dq_H^*}{dx} = -\frac{\frac{d^2W_M}{dq_M^2}\frac{d^2W_L}{dq_L^2}\frac{d^2W_H}{dq_Hdx}}{\Delta} = \left(-\frac{d^2W_H}{dq_H^2}\right)^{-1}\frac{d^2W_H}{dq_Hdx}. \tag{C3}$$

Since quality in the high-income region is independent of other qualities, there is only one direct effect and no indirect effects.

The effect of a marginal change in x on quality provision in Region M is given by

$$\begin{aligned}\frac{dq_M^*}{dx} &= -\frac{1}{\Delta}\left[\frac{d^2W_M}{dq_Mdx}\frac{d^2W_H}{dq_H^2}\frac{d^2W_L}{dq_L^2} - \frac{d^2W_H}{dq_Hdx}\frac{d^2W_L}{dq_L^2}\frac{d^2W_M}{dq_Mdq_H}\right] \\ &= \left(-\frac{d^2W_M}{dq_M^2}\right)^{-1}\left[\frac{d^2W_M}{dq_Mdx} + \frac{dq_H^*}{dx}\frac{d^2W_M}{dq_Mdq_H}\right].\end{aligned} \tag{C4}$$

The first term is the direct effect of x on q_M, while the second term the indirect effect through a quality change in the high-income region.

Finally, the effect of a marginal change in x on quality provision in Region L is given by

$$\begin{aligned}\frac{dq_L^*}{dx} &= -\frac{1}{\Delta}\left[\frac{d^2W_L}{dq_Ldx}\left(\frac{d^2W_H}{dq_H^2}\frac{d^2W_M}{dq_M^2}\right) - \frac{d^2W_M}{dq_Mdx}\frac{d^2W_H}{dq_H^2}\frac{d^2W_L}{dq_Ldq_M}\right] \\ &\quad +\frac{1}{\Delta}\frac{d^2W_H}{dq_Hdx}\left(\frac{d^2W_M}{dq_M^2}\frac{d^2W_L}{dq_Ldq_H} - \frac{d^2W_M}{dq_Mdq_H}\frac{d^2W_L}{dq_Ldq_M}\right) \\ &= \left(-\frac{d^2W_L}{dq_L^2}\right)^{-1}\left[\frac{d^2W_L}{dq_Ldx} + \frac{dq_M^*}{dx}\frac{d^2W_L}{dq_Ldq_M} + \frac{dq_H^*}{dx}\frac{d^2W_L}{dq_Ldq_H}\right].\end{aligned} \tag{C5}$$

The expressions for $d^2W_H/dq_Hd\delta$ and $d^2W_L/dq_Ld\delta$ from section 6.1 are given by

$$\frac{d^2W_H}{dq_Hd\delta} = K'(q_H)\left(u'\left(Y_H^P\right) - u'\left(Y_H^R\right)\right)\frac{y^Ry^P}{\overline{y}_H^2} - \overline{u}_{\tau_{HH}}\frac{\tau_H\left(y^R - y^P\right)}{\overline{y}_H^2} > 0 \tag{C6}$$

and

$$\begin{aligned}\frac{d^2W_L}{dq_Ld\delta} &= b\left(\phi_{LH}^P - \phi_{LH}^R + \phi_{MH}^P - \phi_{MH}^R\right) - \frac{b}{t}\rho_L\frac{\partial\tau_L}{\partial\delta} \\ &\quad +\widetilde{K}_L\frac{y^Py^R}{\overline{y}_L}\left(u'\left(Y_H^R\right) - u'\left(Y_H^P\right) + \Delta u_Y\left(\phi_{LH}^R + \phi_{LM}^R - \left(\phi_{LH}^P + \phi_{LM}^P\right)\right)\right) \\ &\quad +\widetilde{K}_L\frac{\partial\tau_L}{\partial\delta}\left[\frac{3}{t}\left(\Delta u_Y\right)^2\left(\lambda_L\left(y^R\right)^2 + \left(1-\lambda_L\right)\left(y^P\right)^2\right) - \overline{u}_{\tau_L\tau_L}\right] \\ &\quad +\frac{\widetilde{K}_L}{\left(1+\psi_L\right)\overline{y}_L}\times\left[\Delta u_Y\left(y^P - y^R\right)\frac{\left(c_H + c_M - 2\pi - 2c_L\right)}{2t}\overline{y}_L\right] \\ &\quad \times\left[\overline{u}_{\tau_L} + 3\Delta u_Y\left(\lambda_L\left(\phi_{LH}^R + \phi_{LM}^R\right)y^R + \left(1-\lambda_L\right)\left(\phi_{LH}^P + \phi_{LM}^P\right)y^P\right)\right]\end{aligned} \tag{C7}$$

where

$$\frac{\partial \tau_L}{\partial \delta} = \frac{\tau_L \left(y^R - y^P\right) - \left(\left(c_H - \pi - c_L\right)\left(\phi_{LH}^R - \phi_{LH}^P\right) + \left(c_M - \pi - c_L\right)\left(\phi_{LM}^R - \phi_{LM}^P\right)\right)}{\overline{y}_L}. \tag{C8}$$

Appendix D: Proofs

Proof of Proposition 3

A1 is interpreted as π being so small that we can, by continuity, evaluate the effect of a marginal change in π at $\pi = 0$. If we set $\pi = 0$, then $\widehat{Y}_R^M = Y_R^M$, $\widehat{Y}_P^M = Y_P^M$, and $\rho_M = \psi_M = \Delta u_Y = 0$, so that

$$\left.\frac{d^2 W_M}{dq_M d\pi}\right|_{\pi=0} = \frac{b}{2t}\left(u_{\pi_M} - \frac{\overline{u}_{\tau_M}}{\overline{y}_M}\right) + \widetilde{K}_M\left(\overline{u}_{\tau_M \tau_M}\frac{\partial \tau_M}{\partial \pi} + 3\overline{u}_{\tau_M} u''(\cdot)\right). \tag{D1}$$

A2 implies that $y_R = \overline{y}_M + \mu$ and $y_P = \overline{y}_M - \mu$, where is μ some positive number. Using this definition of μ, substituting for $\frac{\partial \tau_M}{\partial \pi} = -\frac{2t\Phi_{MH} + 3(c_H - c_M)u_{\pi_M}}{2t\overline{y}_M}$, and also applying A3, we obtain

$$\left.\frac{d^2 W_M}{dq_M d\pi}\right|_{\pi=0} = \frac{b\theta\left(1-\tau_M\right)\mu^2}{2t\overline{y}_M} + \theta\widetilde{K}_M\left[\frac{\overline{y}_M^2 + \mu^2}{\overline{y}_M}\left(\Phi_{MH} + \frac{3\left(c_H - c_M\right)u_{\pi_M}}{2t}\right) - 3\overline{u}_{\tau_M}\right]. \tag{D2}$$

A marginal increase of π reduces q_M^* if the expression in (D2) is negative. The first term is positive, while the two terms in the square brackets have opposite signs. By visual inspection, the two positive terms are dominated by the negative term if three conditions are satisfied: (i) the income inequality between rich and poor (measured by μ) is sufficiently small, (ii) the non-monetary mobility cost F is sufficiently large (implying that Φ_{MH} is sufficiently small) and (iii) the treatment cost difference $c_H - c_M$ is sufficiently small. *Q.E.D.*

Proof of Proposition 6

A1 is interpreted as π being so small that we can, by continuity, evaluate the effect of a marginal change in π at $\pi = 0$. Applying A1-A3, using the definition of μ given in the proof of Proposition 3, and substituting for $\frac{\partial \tau_M}{\partial \pi} = -\frac{2t\Phi_{MH} + 3(c_H - c_M)u_{\pi_M}}{2t\overline{y}_M}$, we obtain

$$\left.\frac{\partial W_M^*}{\partial \pi}\right|_{\pi=0} = u_{\pi_M}\left\{-\Phi_{MH} + \frac{1}{3}\left(\Phi_{MH} + \frac{3\left(c_H - c_M\right)u_{\pi_M}}{2t}\right)\left(1 + \mu\frac{u'\left(Y_M^R\right) - u'\left(Y_M^P\right)}{2u_{\pi_M}\overline{y}_M}\right)\right\}. \tag{D3}$$

Substituting for $u'\left(Y_M^R\right) - u'\left(Y_M^P\right) = -2\beta\left(1-\tau_M\right)\mu$, we can re-write (D3) as

$$\begin{aligned} \left.\frac{\partial W_M^*}{\partial \pi}\right|_{\pi=0} = & -\frac{u_{\pi_M}\Phi_{MH}}{3}\left\{2+\frac{\beta\left(1-\tau_M\right)\mu^2}{u_{\pi_M}\overline{y}_M}\right\} \\ & +u_{\pi_M}\left\{\frac{\left(c_H-c_M\right)u_{\pi_M}}{2t}\left(1-\frac{\beta\left(1-\tau_M\right)\mu^2}{u_{\pi_M}\overline{y}_M}\right)\right\}. \end{aligned} \tag{D4}$$

A marginal increase in π will increase welfare in Region M if the expression in (D4) is positive. The first term is negative while the second term is positive if μ is sufficiently small. Thus, for a given value of $c_H - c_M > 0$, the expression is positive if two conditions are met: (i) F is sufficiently large (such that Φ_{MH} is sufficiently small) and (ii) μ is sufficiently small. *Q.E.D.*

Figure 1. Patients' mobility. Rich patients

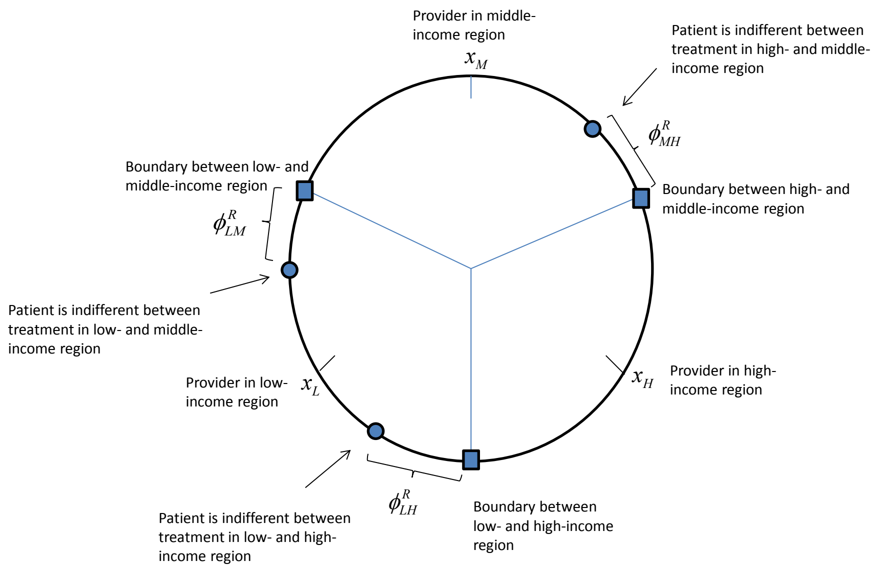

Figure 2. Patients' mobility. Poor patients