Patient Mobility, Health Care Quality and Welfare

Kurt R. Brekke (Norwegian School of Economics)

Rosella Levaggi (U Brescia) Luigi Siciliani (U York)

Odd Rune Straume (U Minho - NIPE)

1 Introduction

1.1 Research questions

What are the effects of interjurisdictional patient mobility on

- health care quality?
- regional welfare?
- global welfare?

1.2 Applications

- Cross-border patient mobility in the EU:
  - still very low (1% of public health care spending in 2006);
  - new EU law passed in March 2011 that gives EU citizens the right to choose among health care providers in all EU member states;
  - the principles of cross-border reimbursement are still not very specific.

- Within-country patient mobility between separate jurisdictions:
  - Sweden (Counties); Italy (Regions); Canada (Provinces).

1.3 Theoretical analysis

1.3.1 Modelling framework

- Hotelling model with two regions: high-skill and low-skill.

- Health care financed by income taxation.

- Health care quality and income tax rate decided by regional policy makers.

1.3.2 Analysis

- Decentralisation without mobility versus centralisation.

- Mobility versus no mobility in a decentralised system.

- Comparison of different cross-border reimbursement schemes:
  - zero transfer payment;
  - transfer payment equal to marginal cost;
  - transfer payment that maximises global welfare;
  - a transfer payment system with two prices.

1.4 Related literature

- Quality competition between health care providers with fixed prices.

- Fiscal federalism (cross-border shopping).

1.5 Outline of the paper

1. Introduction

2. Model

3. The first-best solution

4. The centralised solution

5. Decentralised health care provision

6. Concluding remarks

2 Model

- Hotelling with unit demand and unit patient mass.

- Patients uniformly distributed on $L = [0, 1]$.

- Two regions:

  – Region 1: $L_1 = \left[0, \frac{1}{2}\right]$.

  – Region 2: $L_2 = \left[\frac{1}{2}, 1\right]$.

- Two health care providers:
  - The provider in Region 1 is located at 0.
  - The provider in Region 2 is located at 1.

- Health care provision is publicly funded through general income taxation and is free at the point of consumption.

- The utility of a patient who is located at $x_i \in L_i$ and treated by the provider in Region j, located at z_j, is given by

$$U\left(x_i, z_j\right) = \begin{cases} y\left(1-\tau\right)+v+\beta q_j - t\left|x_i - z_j\right| & if \quad i=j \\ y\left(1-\tau\right)+v+\beta q_j - t\left|x_i - z_j\right| - f & if \quad i\neq j \end{cases}.$$

- A fraction $1-\lambda$ of the patients have prohibitively high values of f.

- λ is constant at each point in L.

- The cost for Region i of providing D_i treatments with quality q_i is given by

$$C_i = cD_i + G(\theta_i, q_i),$$

where $G(\theta_i, q) > (<)\, G\left(\theta_j, q\right)$ and $G_q(\theta_i, q) > (<)\, G_q\left(\theta_j, q\right)$ for all $q \geq 0$, if $\theta_i > (<)\, \theta_j$.

- Region 1 is the high-skill region, with superior technology for providing health care quality:

$$\theta_1 < \theta_2.$$

3 The first-best solution

- A utilitarian supraregional policy maker chooses provider quality and patient allocation:

$$\max_{x,q_1,q_2} W = y(1-\tau) + \lambda \left(\begin{array}{c} \int_0^x (v + \beta q_1 - ts)\, ds \\ + \int_x^1 (v + \beta q_2 - t(1-s))\, ds \end{array} \right) + (1-\lambda) \left(\begin{array}{c} \int_0^{\frac{1}{2}} (v + \beta q_1 - ts)\, ds \\ + \int_{\frac{1}{2}}^1 (v + \beta q_2 - t(1-s))\, ds \end{array} \right)$$

subject to the budget constraint

$$\tau y = c + G(\theta_1, q_1) + G(\theta_2, q_2).$$

- First-best qualities characterised by

$$\left(q_1^{fb}\right) : \frac{\beta}{2}\left[1 + \frac{\lambda\beta}{t}\left(q_1^{fb} - q_2^{fb}\right)\right] = G_{q_1}\left(\theta_1, q_1^{fb}\right),$$

$$\left(q_2^{fb}\right) : \frac{\beta}{2}\left[1 - \frac{\lambda\beta}{t}\left(q_1^{fb} - q_2^{fb}\right)\right] = G_{q_2}\left(\theta_2, q_2^{fb}\right),$$

which imply $q_1^{fb} > q_2^{fb}$ and $x^{fb} = \frac{1}{2}\left(1 + \frac{\beta}{t}\left(q_1^{fb} - q_2^{fb}\right)\right) > \frac{1}{2}$.

4 The centralised solution

- Provider qualities are chosen by a utilitarian central policy maker but patients are free to choose their preferred provider:

$$\max_{q_1,q_2} W = y(1-\tau) + \lambda \begin{pmatrix} \int_0^{\widehat{x}} (v + \beta q_1 - ts)\, ds \\ + \int_{\widehat{x}}^1 (v + q_2 - t(1-s))\, ds \end{pmatrix}$$
$$+ (1-\lambda) \begin{pmatrix} \int_0^{\frac{1}{2}} (v + \beta q_1 - ts)\, ds \\ + \int_{\frac{1}{2}}^1 (v + \beta q_2 - t(1-s))\, ds \end{pmatrix},$$

subject to

$$\tau y = c + G(\theta_1, q_1) + G(\theta_2, q_2),$$

where

$$\widehat{x} = \frac{1}{2}\left(1 + \frac{\beta}{t}(q_1 - q_2)\right).$$

- Straightforward to show that $q_i^c = q_i^{fb}$ and $\widehat{x}\left(q_1^c, q_2^c\right) = x^{fb}$.

5 Decentralised health care provision

- In each region, the optimal health care quality is chosen to maximise the utility of the patients living in that region.

- The cost of health care in Region i is financed by a proportional income tax τ_i.

5.1 No patient mobility

- The optimisation problem of the policy maker in Region i :

$$\max_{q_i} W_i = \frac{y}{2}(1 - \tau_i) + \int_0^{\frac{1}{2}} (v + \beta q_i - ts)\, ds,$$

subject to

$$\frac{\tau_i y}{2} = \frac{c}{2} + G(\theta_i, q_i).$$

- Optimal health care quality characterised by

$$\frac{\beta}{2} = G_{q_i}(\theta_i, q_i^n).$$

- Easy to verity that $q_1^n < q_1^{fb}$ and $q_2^n > q_2^{fb}$.

- Would both regions benefit from centralised policy making?
  - The high-skill region would be better off, while the low-skill region would be worse off.

- Welfare effects of centralisation for the low-skill region:
  - some of the potentially mobile patients get access to higher-quality health care in the high-skill region;
  - the remaining patients experience a drop in health care quality;
  - the tax burden is different (with ambiguous sign).

5.2 Interjurisdictional patient mobility

- The health care provider in the high-skill region cannot turn down patients who travel from the low-skill region to obtain treatment.

- The low-skill region pays a per-patient transfer price p to the high-skill region.

- The two policy makers choose qualities simultaneously and non-cooperatively.

- Region 1:

$$\max_{q_1} W_1 = \frac{y}{2}(1-\tau_1) + \int_0^{\frac{1}{2}} (v + \beta q_1 - ts)\, ds,$$

subject to

$$\frac{\tau_1 y}{2} = \frac{c}{2} - (p-c)\,\lambda\left(\widehat{x} - \frac{1}{2}\right) + G(\theta_1, q_1).$$

- Region 2:

$$\begin{aligned}\max_{q_2} W_2 &= \frac{y}{2}(1-\tau_2) + \lambda \int_{\frac{1}{2}}^{\widehat{x}} (v + \beta q_1 - ts)\, ds \\ &\quad + (1-\lambda) \int_{\frac{1}{2}}^{\widehat{x}} (v + \beta q_2 - t(1-s))\, ds \\ &\quad + \int_{\widehat{x}}^{1} (v + \beta q_2 - t(1-s))\, ds,\end{aligned}$$

subject to

$$\frac{\tau_2 y}{2} = \frac{c}{2} + (p-c)\lambda\left(\widehat{x} - \frac{1}{2}\right) + G(\theta_2, q_2).$$

- Nash equilibrium qualities characterised by

$$(q_1^m) : \frac{\beta}{2}\left(1 + \frac{\lambda}{t}(p - c)\right) = G_{q_1}(\theta_1, q_1^m),$$

$$(q_2^m) : \frac{\beta}{2}\left(1 + \frac{\lambda}{t}(p - c - \beta(q_1^m - q_2^m))\right) = G_{q_2}(\theta_2, q_2^m).$$

- In equilibrium: $q_1^m > q_2^m$, and

$$\frac{\partial q_1^m}{\partial p} > 0, \quad \frac{\partial q_2^m}{\partial p} > 0.$$

- A higher transfer price intensifies quality competition between the two regions.

5.2.1 No transfer payment

- Suppose that $p = 0$.

- The Nash equilibrium is characterised by

$$(q_1^m) : \frac{\beta}{2}\left(1 - \frac{\lambda c}{t}\right) = G_{q_1}\left(\theta_1, q_1^m\right),$$

$$(q_2^m) : \frac{\beta}{2}\left(1 - \frac{\lambda}{t}\left(c + \beta\left(q_1^m - q_2^m\right)\right)\right) = G_{q_2}\left(\theta_2, q_2^m\right).$$

- Compared with no mobility:
  - $q_1^m < q_1^n < q_1^{fb}$ and $q_2^{fb} \gtrless q_2^m < q_2^n$;
  - allowing for interjurisdictional patient mobility without transfer payments leads to a "race-to-the-bottom" in terms of health care quality;
  - welfare goes down in both regions.

5.2.2 Transfer payment equal to marginal cost

- Suppose that $p = c$.

- The Nash equilibrium is characterised by

$$(q_1^m) : \frac{\beta}{2} = G_{q_1}\left(\theta_1, q_1^m\right),$$

$$(q_2^m) : \frac{\beta}{2}\left(1 - \frac{\lambda\beta}{t}\left(q_1^m - q_2^m\right)\right) = G_{q_2}\left(\theta_2, q_2^m\right).$$

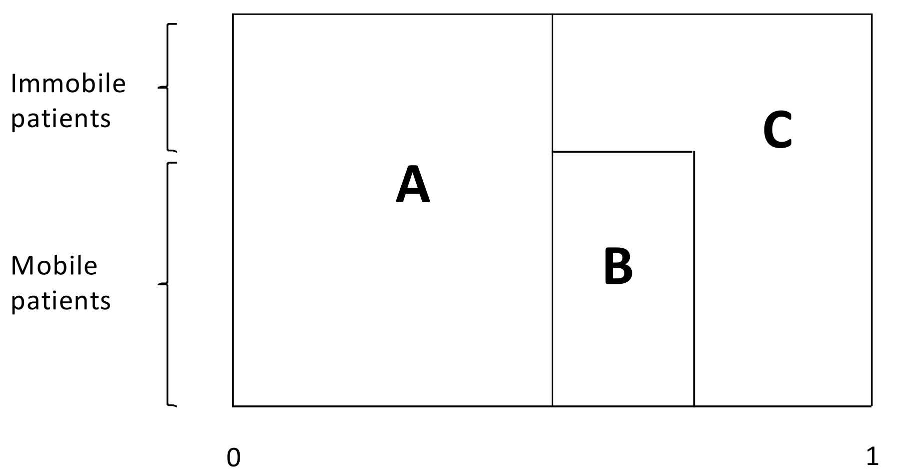
- Compared with no mobility:
  - $q_1^m = q_1^n < q_1^{fb}$ and $q_2^{fb} < q_2^m < q_2^n$;
  - mobility has no effect on welfare in the high-skill region;
  - welfare goes up in the low-skill region.

- Mobility with $p = c$ has three different welfare effects in the low-skill region:
  - some patients get access to higher-quality health care in the other region;
  - the remaining patients experience a drop in health care quality;
  - the tax burden goes down.

- At regional level, allowing for mobility with $p = c$ is a Pareto improvement.

- However, there are winners and losers within the low-skill region:

Figure 1. Effect of patients' mobility when transfer payment is equal to marginal cost



Patients in area A are indifferent
Patients in area B gain
Patients in area C lose

- The losers will outnumber the winners if
  - the number of immobile patients is sufficiently low;
  - travelling costs are sufficiently high;
  - the marginal utility of quality is sufficiently low;
  - the skill-difference is sufficiently low.

5.2.3 Optimal transfer payment

- Suppose p is chosen to maximise global welfare: $p = p^*$.

- Even if $p = p^*$, the first-best outcome cannot be implemented with decentralised policy making.

- The optimal transfer payment implies underprovision (overprovision) of quality in the high-skill (low-skill) region.

- With quadratic quality costs, the optimal transfer price is given by

$$p^* = c + \frac{t\beta^2 (\theta_2 - \theta_1)}{(\theta_1 + \theta_2)\left(2t\theta_1 - \lambda\beta^2\right)}.$$

- Notice that:
  - the optimal price is higher than marginal costs;
  - increased patient mobility (λ) increases the optimal price.

- Compared with no mobility:
  - $q_1^n < q_1^m < q_1^{fb}$ and $q_2^{fb} < q_2^m < q_2^n$;
  - welfare goes up in the high-skill region;
  - welfare goes down in the low-skill region.

5.2.4 Transfer payments with two prices

- Suppose Region 2 pays p_2 while Region 1 receives p_1, for each patient from Region 2 seeking treatment in Region 1.

- If $p_1 \neq p_2$, there is an extra tax bill of $(p_1 - p_2)\,\lambda\left(\widehat{x} - \frac{1}{2}\right)$.

- Suppose that tax payers in Region 1 pay a share α of the extra tax bill, while tax payers in Region 2 pay the remaining share $1 - \alpha$.

- The two budget constraints are now given by

$$\begin{aligned}\frac{\tau_1 y}{2} &= c\left(\frac{1}{2}+\left(\widehat{x}-\frac{1}{2}\right)\right)+G\left(q_1,\theta_1\right)-p_1\lambda\left(\widehat{x}-\frac{1}{2}\right)\\ &\quad+\alpha\left(p_1-p_2\right)\lambda\left(\widehat{x}-\frac{1}{2}\right)\end{aligned}$$

and

$$\begin{aligned}\frac{\tau_2 y}{2} &= \left((1-\lambda)\left(\widehat{x}-\frac{1}{2}\right)+1-\widehat{x}\right)c+G\left(q_2,\theta_2\right)+p_2\lambda\left(\widehat{x}-\frac{1}{2}\right)\\ &\quad+(1-\alpha)\left(p_1-p_2\right)\lambda\left(\widehat{x}-\frac{1}{2}\right).\end{aligned}$$

- The Nash equilibrium is characterised by

$$(q_1^m) : \frac{\beta}{2}\left[1 + \frac{\lambda}{t}\left[(1-\alpha)\,p_1 + \alpha p_2 - c\right]\right] = G_{q_1}\left(\theta_1, q_1^m\right),$$

$$(q_2^m) : \frac{\beta}{2}\left[1 + \frac{\lambda}{t}\left[(1-\alpha)\,p_1 + \alpha p_2 - c - \beta\left(q_1^m - q_2^m\right)\right]\right] = G_{q_2}\left(\theta_2, q_2^m\right).$$

- For any pair of prices, (p_1, p_2), a uniform price

$$p = (1-\alpha)\,p_1 + \alpha p_2$$

yields exactly the same outcome.

- Thus, a more sophisticated transfer payment scheme with two prices does not improve on the solution with $p = p^*$.

6 Summary

- Patient mobility is potentially beneficial for global welfare when regions differ in their ability to provide health care quality.

- However, the welfare effects of imposing patient mobility crucially depends on the design of the transfer payment scheme.

- The transfer price needs to be sufficiently high in order to avoid a "race-to-the-bottom" effect.

- Allowing for patient mobility with a transfer price equal to marginal cost is a Pareto improvement at regional (but not individual) level.

- The transfer price that maximises global welfare does not constitute a Pareto improvement.

- A more complex transfer pricing system with different prices for the "exporting" and "importing" regions does not bring any additional advantages.