Escaping epidemics through migration ? Quarantine measures under incomplete information about infection risk.

Alice Mesnard* Paul Seabright†

December 2008‡

Abstract

This paper explores the implications for public policy of the fact that individuals have incomplete information about their exposure to infectious disease and can take migration decisions before their health status is publicly observable. In a 2-period model we study conditions under which the presence of quarantine measures may lead to inefficient outcomes by strengthening individuals' interest in migration to escape centres of disease and thereby imposing negative externalities on other uninfected individuals. We show first that when the disease has an epicentre, the marginal migrant imposes a net negative externality. Secondly, quarantine policies may sometimes encourage migration instead of discouraging it. Thirdly, even when they succeed in discouraging migration, quarantine policies may lower social welfare, and even increase overall disease incidence, if they go too far, thereby discouraging those intra-marginal migrants for whom private benefits exceed private costs by more than the negative externality they impose on others.

JEL Classification: I18, O15, O19, R23.

Key words : migration, infectious diseases, public health, quarantine.

Corresponding author : alice_m@ifs.org.uk, Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE, UK. phone: 0044 2072914866. fax: 0044 8701334437.

*Institute for Fiscal Studies, CEPR and Toulouse School of Economics.

†Toulouse School of Economics (GREMAQ) and CEPR.

‡We are grateful to Jerome Adda, Emmanuelle Auriol, Jean-Paul Azam, David Bardey, Clive Bell, Pascaline Dupas, Michael Kremer, Javier Ortega, Patrick Rey, Marcos Vera-Hernandez, two anonymous referees and participants at the Cornell conference on social science and disease and the EUDN conference on development economics in Paris for valuable comments and advice. The usual disclaimer applies.

1 Introduction

This paper studies the effect of public policies to restrict migration by individuals suspected of carrying disease, when those individuals do not know for certain whether they have the disease but may have more information than the authorities about their probability of being carriers. It has long been known that migration affects the spread of disease, and this influence has for centuries been used to justify placing restrictions on the movement of individuals suspected of carrying infections[1]. For example, immigrants to the United States were screened for disease on Ellis Island and sometimes kept in quarantine until considered safe for the rest of the population (for a detailed description of how migrants were screened when leaving their origin country and at their arrival on Ellis Island, see Markel and Stern, 2002). Epidemiological studies have addressed how individual behaviour, among other factors, affects the spread of infections. However, the study of how individual behaviour in turn changes in response to the new incentives created by the occurrence of a disease is much less developed. The principal contribution of our paper is to bring the study of strategic behavior under uncertainty into the domain of epidemiology, and to analyze its impact, in interaction with public policies, on the overall impact of epidemic disease.

To our knowledge the work that has been done to date on strategic behavior in the context of disease considers the economic determinants of preventive behaviour such as vaccination or the adoption of safe sex. In particular, Philipson (2000) and Geoffard and Philipson (1996) show that, if demand for prevention treatments such as vaccines is prevalence elastic, then initially successful public health efforts typically run into diminishing returns, not simply for technical reasons but because the decline of a disease discourages prevention. Similar considerations apply to the factors determining the adoption of means of contraception such as barrier methods for Sexually Transmissible Diseases or the choice of partners (see for example, Kremer, 1996), and a growing literature now focuses on the microeconomic determinants of such individual decisions, in order to reach a better understanding of epidemiological patterns.

However, migration as a form of preventive behaviour has received very little attention, although evidence has accumulated that migration behaviour and epidemics are intrinsically linked. Migration behaviour can respond very rapidly to changes in the health environment, in particular when it suddenly deteriorates through epidemics. There are numerous historical instances of people fleeing plague or other infectious diseases by migrating to distant areas (see McNeill, 1997). During the Black Death, for instance, inhabitants from infected villages migrated to less infected villages in the neighbourhood. Much more recently, after the SARS outbreak in China, workers in urban areas were returned in large numbers to live with their families in safer rural areas[2].

Such behavioural responses are important for understanding the effectiveness of policy measures, in the context of an increasing demand for national and international regulation of disease control, especially

[1] The first international regulations on health were adopted by twelve European states during a conference held in Paris in 1851 following the failure of the early public health strategies using "quarantine and lazarettos" to prevent the importation of contagious diseases through the Mediterranean region. 137 regulations dealing with health issues for maritime traffic included articles that specifically mentioned plague, yellow fever and cholera, diseases for which international regulations and requirements continued until the late 20th century.

[2] see *Le Point*, 16 Mai 2003 pp50-51

given the very real fear that antibiotic resistance will soon lead to the recurrence of old diseases no longer curable by antibiotics and the emergence of new diseases that make regions such as Europe and North America more vulnerable to epidemics than they have been in the past.

Recently, the outbreak of SARS in China and its rapid transmission across the world (principally via air travel) demonstrated both the dramatic consequences of the lack of reporting by one state, and, subsequently, the effectiveness of strict health regulations applied quickly and simultaneously in different countries.

Even if it would be premature to draw conclusions regarding the long-term evolution of SARS, at least one question needs to be addressed urgently. How effective are quarantine measures and in what circumstances do they work best? Such measures have often been considered as the most efficient way to prevent individuals in an infected area from moving to a non infected area, and they continue to be used as a systematic response to epidemic outbreaks. For example, when the Black Plague arrived in Milan, three households were infected. The authorities immediately bricked up the doors and windows, leaving the human beings and the rats to their fate. "All perished and Milan was spared", as revealed by historical chronicles (for a review of these, see Benedictow, 2004).

But what happens if individuals have more accurate information on their risk probability than the state health authorities, and use this information asymmetry strategically? Can economists warn policy makers about unexpected effects of quarantine regulations, taking into account that individual behaviour changes in response to policy regulations? The aim of the paper is to study the possibility of such effects, which, to our knowledge, have not yet been considered in the debates on health and migration (see for example the review "Health and Migration" published by the International Organization for Migration). To our knowledge, only one related paper by Malani and Laxminaryan (2009) notes the possible perverse effects of imposing sanctions on trade partners as they will affect the incentives for surveillance and reporting of infectious diseases by health authorities in complex ways.

We develop a framework where migration responds to the prevalence of disease, to health regulations, and to the costs of migration. In the absence of quarantine regulations, the difference in disease prevalence between two areas provides the sole motive for migration and individuals choose whether or not to migrate by balancing the health benefits of doing so against the monetary (and possibly non-monetary) costs. Important assumptions that we adopt in our model are that individuals do not know whether or not they are infected but do know something about their prior exposure to the disease; they know more about the latter than do the health authorities. Incomplete information is the source of negative externalities as some sick individuals, who think erroneously that they are healthy, migrate and infect a larger number of healthy individuals at their destination. The health authorities may decide to apply quarantine measures to correct for excessive migration. However, asymmetric information about prior exposure has potentially important consequences for the effect of quarantine measures, which we will investigate. This assumption does not contradict the fact that health authorities have, or may have, considerable experience, as well as statistical information, about earlier outbreaks of a disease, including

its rate of propagation and incubation periods, which is captured in our model by the assumption that government knows the distribution of risk types in the population. We simply assume here that individuals know more about their idiosyncratic risk of being infected since they know more about their own characteristics (their past locations, personal habits, living conditions, networks etc...), whereas health authorities can only observe individuals' health status with a lag, once symptoms appear. It also means that the risk characteristics of individuals who migrate may differ in important ways from those of the general population, a fact that may have a significant impact on the subsequent evolution of the disease.

At this stage we do not study the effects of quarantine measures in general, many of which affect the mobility of goods (for a study of the effects of a ban on import of goods, see Anderson and James, 1998). But we study the effects of specific quarantine measures that affect the mobility of individuals and are taken by the health authorities of the country harbouring the focus of the disease, such as the ones taken by the Chinese authorities or, later on, by the Canadian authorities after the outbreak of SARS[3]. We show merely that behavioural responses to actual interventions may lead to perverse or undesirable effects. Specifically we show that these are more likely when the period during which individuals do not know whether they are infected, but have more information about their infection risk than the authorities, is long relative to the speed of propagation of the disease. Such effects are therefore more likely for diseases such as HIV/AIDS than for Ebola or SARS, with tuberculosis being an intermediate case.

Our model is very simple, and unrealistic in a number of ways that enable us to find analytical solutions. However, it illustrates three very useful and intuitive principles which apply much more generally than in this particular context, and which we believe will be helpful to analysts and policymakers in thinking about the impact of public policies towards disease. When disease incidence differs from one place to another, decisions to migrate involve private benefits and costs to the individual concerned which include changes in their risk of catching the disease. They also impose externalities on others, externalities that are negative and increasing in the number of uninfected individuals with which the migrant comes into contact, in the infectiousness of the disease, and in the probability that the migrant is a carrier. This much is uncontroversial. In our model we show that

- First, when the disease is concentrated in one place (the epicentre of an epidemic for instance), a decision to migrate away from the epicentre brings a potentially infected individual in contact with more uninfected individuals than she would have met had she remained where she was. Thus the typical migrant imposes a net negative externality as a result of her decision to migrate, and the marginal migrant (for whom, by definition, private benefits of migrating just equal the private costs of doing so) has a negative impact on social welfare. Laissez-faire will therefore lead to excessive migration. This provides a rationale for the frequent (and frequently justified) public policy response to epidemics, which is to attempt to restrict migration away from the epicentre by those who may be infected.

[3] We aknowledge that, in the past, most measures taken at the national level to stop the spread of a disease have been taken by the authorities of the countries which people are trying to reach. But given the increased need of controlling diseases at the international level, pressures on the authorities of the country harbouring the infectious diseases have increased.

- Secondly, and less obviously, not all policies that aim to restrict migration in fact do so. In particular, we distinguish two effects of quarantine policies. The first is that they raise migration costs, which lowers migration. For example, mandatory health certificates or test results may be required by health authorities to leave the epicentre of the disease. We call this a "type 1" quarantine measure. Another effect might be to impose a utility cost on individuals of remaining in the city where quarantine measures are effective, since they face a chance of being subjected to awkward and possibly dangerous restrictions on their movements. We call this a "type 2" quarantine measure. Such measures impose a welfare cost on those who suffer them, which tends to increase migration by those who are not currently subject to quarantine but fear they may become so if they remain where they are. Policies implemented without taking this effect into account may therefore have effects that are opposite from those intended.

- Thirdly, even policies that actually reduce migration may have an adverse impact on social welfare if they reduce migration "too much", and specifically if they discourage those intra-marginal migrants whose private benefits from migration substantially exceed their private costs of migration, by enough to outweigh the negative externality they impose on others. Overall disease prevalence may even increase if in the name of avoiding negative externalities the authorities discourage relatively low-risk individuals from escaping the epicentre of the disease, thereby increasing the probability that they will catch the disease there from infected individuals.

The design of quarantine policies needs therefore to be sensitive to the actual incentives for or against migration created by those policies, as well as to the need to consider the impact on intra-marginal as well as on marginal migrants. These messages, we believe, have an importance and relevance that extends far beyond the rather special circumstances of our model.

The paper proceeds as follows. In Section 2 we set out the assumptions of two-period model in which individuals face a choice of whether to migrate between two areas that differ in the prevalence of disease. In Section 3 we discuss the implications of different public policy measures under different structure of information and different types of diseases, before concluding in Section 4.

2 The Model

2.1 Cities and risk of infection

There are two cities, a and b, each containing a continuum of inhabitants. There are two time periods, $t = 0, 1$. The size of the population in period 0 is normalized to 1 in city a and N in city b, with $N > 1$. Apart from their size and their number of infected individuals, the two cities are identical. At the outbreak of the epidemic at period 0, r_a and r_b inhabitants are infected by the disease in city a and city b respectively; we assume that $r_a > r_b$, with minimal loss of generality since all it implies is that the initial population in which an epidemic is concentrated (the "epicentre") is small relative to the total potentially infected population. We also call r_a and r_b the initial "prevalence levels" of the disease in the

two cities. However, individuals may choose to migrate between the cities and this will affect prevalence levels in the second period. In equilibrium, and as a result of the various decisions affecting migration made by individuals and by the authorities, there will be N_a and N_b inhabitants infected by the disease in the two cities in period 1.

We write $s_{ijt} \in (h, s)$ for the state of health of individual i in city j in period t, where h is the healthy state and s is the sick state. An individual does not know her own state of health in period 0, learning this only in period 1. However, everyone (individuals plus the authorities) knows the prevalence levels of the disease r_a and r_b.

Nevertheless, individuals have some information about the extent of their prior exposure to the disease, which we interpret simply as the probability that they have already caught the disease in period 0. We will see that such incomplete information may generate inefficiencies if enough care is not taken in the design of policy measures. We write e_{ij} for the exposure of individual i in city j (it will often be convenient to drop the subscript j without risk of confusion); in the analysis below we make a variety of different assumptions about how e_{ij} is distributed across individuals.

We assume for the moment that the factors affecting previous exposure are purely random and have no bearing on the individual's future risk of catching the disease such that the risk for a previously uninfected individual of catching the disease in period 1 depends only on the infectiousness ϕ of the disease, and on the number of infected individuals N_j in the city where she decides to live[4]. We will refer to diseases of this kind as "type-insensitive" epidemics. Formally, we write this as follows:

$$pr(s_{ijt+1} = s | s_{ijt} = h) = \phi N_j \tag{1}$$

We will discuss later on the robustness of our results for other type of epidemics, for which the risk of future infection depends on the same factors that determined past exposure. We will refer to diseases of the latter kind as "type-sensitive". A good example would be a disease such as AIDS in which individual sexual behavior determines exposure, so that those whose behavior makes them more likely to have been infected in the past are also, if still uninfected, more likely to become infected in the future for given prevalence levels.

The fact that the risk of infection is increasing in the prevalence of the disease in the chosen city is important because it implies that individuals will choose whether or not to migrate between cities in response to differences in the prevalence levels[5]. However, not all individuals will make the same choices. This is not (in our simple framework) because of differences in their migration costs or in their ability to afford those costs, since we assume migration costs and incomes to be identical across individuals.

[4] The equations would be more complex and non-linear if we assumed that the risk depended not on the number but on the proportion of infected individuals in the city in question. However, we have no reason to think that this would affect the qualitative insights of the model, and the absolute numbers assumption has a natural interpretation: in a city centre which every citizen passes, the risk of infection is a function of the total number of dangerous pathogens in the environment, not the proportion of dangerous ones to innocuous ones.

[5] We deliberately ignore here the possibility that cities may differ in the quality of preventive or curative medical care available. It is possible that cities with high disease prevalence may be more, not less attractive to uninfected individuals, for instance if they also offer vaccination whose effects outweigh those of greater disease prevalence, or if prior residence is a condition of access to treatment. This qualification should be borne in mind when interpreting our results, though historically we do not believe such phenomena have been common.

Rather, their choices will differ because of their prior exposure to the disease. This happens in two ways. First, individuals are more likely to migrate, other things equal, if they have had lower prior exposure, since that increases the probability that they are healthy - only healthy individuals are at risk of catching the disease since infected individuals have caught it already. Secondly, under type-sensitive epidemics, individuals are more likely to migrate if their individual exposure is higher, since it makes them more sensitive to the environment in which they live. This means, other things equal, that migration is more likely given *higher* past exposure. The net impact of these two effects will be considered in more detail below.

2.2 Social welfare

Individual migration decisions affect social welfare and not just individual welfare. This is because they generate externalities - in this framework because an infected individual creates a risk for the uninfected individuals that surround her (the migration decision of an uninfected individual creates no externalities).

To see this, we first study the objective of public policy.

Denote m the migration costs, d the utility costs of being infected, n_s (n_h) the number of sick (healthy) individuals who migrate from city a to b. N_a (N_b) the number of infected individuals living in city a (b).

Social planners minimize I, the sum of the utility costs of infection plus the costs of migration, which we can write as:

$$I = d(N_a + N_b) + m(n_h + n_s) \tag{2}$$

As, after migration, $N_a = (1-r_a-n_h)(r_a - n_s)\phi - n_s + r_a$ and $N_b = (N-r_b+n_h)(r_b + n_s)\phi + n_s + r_b$, it is easy to derive

$$\frac{\partial I}{\partial n_h} = d\phi(r_b - r_a + 2n_s) + m \tag{3}$$

and show that, for n_s small, $\frac{\partial I}{\partial n_h} < 0$ as condition (7) holds. Moreover, $\frac{\partial I}{\partial n_s} > 0$ as:

$$\frac{\partial I}{\partial n_s} = d\phi(N - 1 - r_b + r_a + 2n_h) + m \tag{4}$$

Therefore public policy should minimize the number of sick migrants, n_s, which optimally should be equal to 0, and maximize the number of healthy migrants, n_h.

These optimal migration flows come from the specification of the infection function (equation 1), which implies that the total number of infected individuals is minimized when healthy individuals go to the city with fewer sick individuals and when sick individuals stay in the city with a high number of infected individuals.

2.3 The incentives for migration

To understand how individuals with different risk exposure take different decisions, assume for the moment that individuals migrate from one city to another on the basis of current prevalence levels in the two cities; in other words they have adaptive expectations about infection risk. This is a simplification because

it ignores the way in which individuals anticipate future evolution of prevalence levels in both cities by taking into account feedback effects of the decisions of individuals like themselves. However, it is useful to understand how incentives to migrate depend on the information individuals have on their prior exposure to the disease.

Denote y the per capita income generated in each city. Then the utility if an individual of type e_i chooses to migrate is written as[6]:

$$U_m = -m + e_i(y-d) + \phi r_b(1-e_i)(y-d) + (1-e_i)(1-\phi r_b)y$$

The utility if an individual of type e_i chooses not to migrate is:

$$U_n = e_i(y-d) + \phi r_a(1-e_i)(y-d) + (1-e_i)(1-\phi r_a)y$$

So individual i will wish to migrate as long as:

$$m < \phi(1-e_i)(r_a - r_b)d \tag{5}$$

Equation 5 shows that individ ual i will wish to migrate to the low prevalence city as long as the migration cost, m, is smaller than the migration benefit (on the right hand side of the inequality). This benefit is higher the lower is prior exposure e_i and the higher the difference between prevalence levels in the two cities. By re-arranging equation 5 we can derive a threshold exposure

$$e_A^* = 1 - \frac{m}{\phi d(r_a - r_b)} \tag{6}$$

such that all individuals having an exposure factor below e_A^* migrate to the less infected city. This simply reflects the fact that individuals who are more likely to be uninfected derive greater benefit from migrating and are more likely to do so. We will see in Section 3.3 that this conclusion is different in a model with type-sensitive infections, where both high and low risks may choose not to migrate.

It is evident that e_A^* is decreasing in m, so that higher values of m lower overall migration levels.

Moreover, to ensure that at least one individual will wish to migrate and that not all will do so we assume that

$$\phi(1-\overline{e})(r_a - r_b)d < m < \phi(1-\underline{e})(r_a - r_b)d \tag{7}$$

with $\underline{e}$ being the lowest and $\overline{e}$ being the highest value of e_i in city a.

In the remainder of the paper, we assume, instead, that migrants have rational expectations about prevalence levels in the two cities in period 1. That is, when they make their migration decisions in period 0 they do not assume that current prevalence levels will continue to hold in the future; they take into account that others are reasoning as they do. The rationale for this is not necessarily that individuals are perfectly far-sighted but rather that it determines a "migration equilibrium", such that, after migration

[6] Results would have been the same, if, instead of assuming risk neutrality, we had assumed that all individuals had the same degree of risk aversion. Moreover, we ignore the effect of heterogeneity in the degree of risk aversion in order to focus on one dimension of heterogeneity, that is in the degree of exposure to risk.

takes place, no individuals have any further interest in migrating, once they have taken into account the migration behaviour of other individuals.

Although the assumption of rational expectations seems significantly different from that of adaptive expectations about infection risk, and leads to quantitatively different behavioral responses, these responses and the equilibria to which they give rise are qualitatively no different from those that occur under adaptive expectations. All of the results we report below remain true under the hypothesis of adaptive expectations. The reason is not difficult to understand. Individuals migrate in response to perceived differences in infection risk between the two cities. If they take into account the fact that others will reason as they do, this will dampen but not completely offset their behavioral response, since migration will diminish the prevalence difference between cities that was the original spur to migration. We demonstrate this more formally in Lemma 2 below.

In the simple framework we have set out here, if individuals knew everything - and specifically, if they knew their own health status - there would be no public policy issue and no need for intervention, since infected individuals, who are the only ones whose migration decision creates externalities, would have no incentive to migrate. The problem arises because of incomplete information about health risk: migration occurs because individuals believe themselves to be healthy with some probability and therefore believe they will benefit from moving to a place where they are less likely to be exposed to infection. However, migration creates externalities because those same individuals have also a positive probability of being infected and therefore of infecting other healthy individuals with whom they come into contact.

In principle if the authorities know each individual's health risk they can determine which individuals should migrate and which should not. This seems unlikely because health risk will vary considerably among individuals for reasons the authorities are in no position to know. Alternatively, if the authorities know the distribution of health risk in the population, they may be able to calculate the externality imposed by the marginal migrant at a first-best optimum, and impose a tax on migration that just dissuades the marginal migrant from moving. We discuss below how this might be achieved. We show that this will work provided that the migrants are those whose risk of infection is below that of the marginal migrant, which will be true for certain types of disease but not of all types. As a result, we demonstrate that a real problem for public policy may arise if either the authorities do not know the distribution of health risk in the population, or if the disease type in question is one for which migration incentives are non-monotonic in risk type. We elaborate on these points below.

3 Effects of Policy Measures

To see the kind of perverse effects policy interventions may generate when individuals do not have perfect information on their infection status, we first study the case where individuals know everything, which we see will lead to a First-Best Private Optimum. Then, we introduce the case where individuals have heterogeneous risks of having been infected in the past, which will lead to them to respond differently to migration incentives. Since the authorities do not know the health status of each individual, they can

only apply policy measures (such as quarantine) that in principle affect both healthy and sick individuals. Furthermore, individuals will react to those measures on the basis of the knowledge of their own prior exposure, which affects their gains from migration.

We consider the case of type-insensitive infections where individuals have the same ex-ante risk of being infected in the future, conditional on being uninfected today. In this case, some asymptomatic sick people wish to migrate, because they erroneously believe they can avoid being infected. Although the externality imposed by migrating individuals on healthy individuals in their city of destination varies among individuals, provided the authorities know the distribution of this risk they can calculate the optimum level of migration and the externality imposed by the marginal migrant at this optimum: the risk threshold corresponding to the marginal migrant is unique, and those who migrate are those whose chance of infecting others is below that of the marginal migrant. This allows the authorities to set a (Pigouvian) tax on migration that implements the optimum.

Finally, we will turn to the case of type-sensitive infections where individuals have heterogeneous risks of being infected in the future, which are correlated to what they know about their past exposure. We show that in this case there are two "migration" thresholds. As health authorities do not know exactly about the risk type of individuals migrants, they cannot tax them for the negative externalities they generate, which gives rise to the absence of First-Best Policy.

3.1 First Best Laissez-Faire Optimum under Complete Information

We first assume that individuals know perfectly whether they are infected or not.

As the only migration gains are the differential in infected individuals between the two cities and there are positive migration costs, sick individuals prefer not to migrate, whereas healthy individuals migrate if migration gains are higher than migration costs, that is if :

$$m < d\phi(r_a - r_b) \tag{8}$$

So there is a "positive" selection of individuals through migration. In other words, there is a private incentive to "self-quarantine", which leads to first-best outcomes when people are perfectly informed about their own infection status : healthy individuals migrate into the city where the number of infected individuals is low and health authorities do not need to intervene to maximise social welfare.

Note that it makes no difference whether individuals have adaptive or rational expectations. Rational expectations would require individuals to take into account in their migration decision the fact that individuals like themselves may migrate and thereby affect the number of sick individuals in both cities at equilibrium. However, when individuals know their infection status, only healthy individuals have an interest in migrating, and the number of sick individuals in each city does not change. So the policy implications are exactly the same whether individuals have rational expectations or not.

In all cases, migration under laissez-faire leads to the First Best Private Optimum.

3.2 First Best Policy under Incomplete Information

3.2.1 Quarantine measures when healthy individuals face equal risks of future infection but have heterogeneous prior exposure

We now consider the case where individuals do not know perfectly whether they have been infected or not. If all individuals were unaware of their likelihood of being infected there would be no differences in resulting behavior and also no real public policy problem. Instead, however, we assume that individuals know their own prior exposure and therefore their probability of being infected, but that their risk of being infected in the future, conditional on being uninfected today, is unrelated to the factors that have made them more or less likely to be infected today. It is therefore those who are least likely to be already infected who are most likely to gain from migrating away from the epicentre of the infection. However, to the extent that some of them may be infected without knowing it, they impose a negative externality on uninfected individuals in their destination. We consider whether public policy can correct for this externality without knowing the exposure of particular individuals.

As individuals take into account the fact that other sick individuals may migrate to the low prevalence city, the utility if an individual of type e_i chooses to migrate is now written as:

$$U_m = -m + e_i(y-d) + \phi(1-e_i)(r_b+n_s)(y-d) + (1-e_i)(1-\phi(r_b+n_s))y$$

The utility if an individual of type e_i chooses not to migrate is:

$$U_n = e_i(y-d) + \phi(1-e_i)(r_a-n_s)(y-d) + (1-e_i)(1-\phi(r_a-n_s))y$$

So individual i will wish to migrate as long as net gains from migration are positive, i.e.:

$$0 \leq \phi d(1-e_i)(r_a - r_b - 2n_s) - m \tag{9}$$

Note that this differs from equation 5 by the inclusion of the term $2n_s$ - the migrant takes into account that others like herself will be migrating between periods 0 and 1.

We can study the impact of quarantine measures on migration incentives in general, without solving for the migration equilibrium level, n_s, by defining the exposure level e^* of the marginal migrant, defined as one for whom net gains from migration, defined by equation (9), equal zero. This marginal migrant is therefore:

$$e^* = 1 - \frac{m}{\phi d(r_a - r_b - 2n_s)} \tag{10}$$

It is useful to begin by defining $\mu_i^I \equiv \phi d(1-e_i)(r_a - r_b - 2n_s)$ as the gross gains from migration for individual of type e_i, and establishing

Lemma 1: μ_I is strictly decreasing in e_i.

Proof: For a given migration equilibrium level n_s, $\frac{\partial \mu_I}{\partial e_i} = -\phi d(r_a - r_b - 2n_s)$. By inequality 9, $(r_a - r_b - 2n_s) > 0$,so $\frac{\partial \mu_I}{\partial e_i} < 0$. QED

Lemma 1 shows that those who migrate will be all and only those whose exposure levels lie below the threshold; the fact that individuals have rational expectations has not changed matters in this respect. This also implies that n_s must be weakly increasing in e^*.

It is also straightforward to show that:

Lemma 2: e^* is strictly decreasing in m.

Proof: The proof is by contradiction. Suppose instead that e^* is weakly increasing in m. Since n_s is weakly increasing in e^*, and therefore in m, the denominator of the second term on the RHS of (10) is decreasing in m. Since the numerator of this second term is itself m this means that the second term is strictly increasing in m and therefore that e^* is strictly decreasing in m, contrary to our assumption. QED

Note that this establishes for rational expectations the equivalent monotonicity property that holds for e^*_A, the migration threshold under adaptive expectations, and underlines that rational expectations dampen but do not fully offset the behavioral responses under adaptive expectations.

This then allows us to state Proposition 1, proven in the Appendix:

Proposition 1: For type-insensitive epidemics, when exposure levels are distributed continuously over $[\underline{e}, \overline{e}]$:

a) For any m there exists a unique exposure threshold defined implicitly by $e = e^* = 1 - \frac{m}{\phi d(r_a - r_b - 2n_s)}$; all individuals with exposure levels below this threshold (and only these individuals) will migrate to the low prevalence city.

b) There exists a unique exposure threshold $e = \widetilde{e}$, which minimizes I, and a unique migration tax m^t which implements this optimum. Furthermore, $\widetilde{e} < e^*$, so that migration under laissez-faire is too high as compared to this optimum.

What does Proposition 1 tell us? Migration entails two types of change in welfare. First, migration imposes a negative externality as migrants will encounter a larger number of healthy individuals in the destination area than in their area of origin, thereby increasing the overall exposure of others to infection risk. Secondly, the marginal migrant has, by definition, no private gains from migration. Therefore the marginal migrant entails, in this case, net overall social welfare losses, as migration still entails negative externalities and private gains of the marginal migrant are zero. Proposition 1 simply says that, provided the distribution of migrants is continuous so that there is always a positive mass of migrants within an arbitrarily small distance of the exposure threshold, there will be too much migration under laissez-faire.

However, although the suboptimality of the laissez-faire equilibrium follows from Proposition 1 (provided e is continuously distributed), it does not follow that any given quarantine measure will improve

matters. If quarantine measures of type 1 are implemented, so that migration cost is now $m' > m$, the marginal migrant is now defined by $e^{**} = 1 - \frac{m'}{\phi d(r_a - r_b - 2n_s)}$ and we see that $e^{**} < e^*$ so that migration has decreased. However, e^{**} could be higher or lower than $\widetilde{e}$, so it does not follow that type 1 measures are necessarily an improvement on the status quo; they may "overshoot".

If quarantine measures of type 2 are implemented, so that there is a utility cost q to remaining in city a, the marginal migrant is now defined by $e^{**} = 1 - \frac{m-q}{\phi d(r_a - r_b - 2n_s)}$ and we see that $e^{**} > e^*$, so that migration has increased. Therefore implementing type 2 quarantine measures will actually be worse than doing nothing.

So even if health authorities do not observe individual types, they may apply first best policy measures based on the knowledge of the risk distribution in the population. But the details of the policy matter, and policies devised with good intentions but without attention to detail may end up being ineffective or even making things worse. We will see below that the type of policy measures that are desirable depends highly on the distribution of risk types.

3.2.2 A simple example: two exposure types

To illustrate this with a simple example, we will now consider the case where exposure types are not distributed continuously on $[\underline{e}, \overline{e}]$. Instead we assume that individuals can be of only two risk types, which we refer to as low risk and high risk. Thus $e_i \epsilon \{e_L, e_H\}$ with p the proportion of individuals with low exposure level.

In these circumstances Proposition 1 may no longer hold. Indeed, levels of migration under laissez-faire may be optimal. Also, pure-strategy equilibria may not exist since when all members of one risk type migrate this may reverse their migration incentives and may, in turn, make it impossible to implement the optimum with a migration tax.

We can understand intuitively that, with discrete risk types, there are some cases where policy measures may be welfare improving and other cases where policy measures may be welfare decreasing. This will depend on where the low and high types individuals are positioned compared to the migration threshold under laissez-faire, and specifically on whether either type, no type or both types may migrate in equilibrium; on how low is the exposure of low type individuals; and on how policy measures affect migration costs and the disutility of living in the epicentre of the disease.

More formally, we can show in Appendix Proposition 2:

Proposition 2: If exposure levels are discretely distributed between two types, so that the exposure factor $e_i \epsilon \{e_L, e_H\}$ with p the proportion of individuals with low exposure level, and if $e_L < 1 -$

$\frac{m}{\phi d(r_a - r_b - 2pe_L)} < e_H$, then the unique laissez-faire equilibrium has migration by all and only the low-exposure individuals, and there exists a threshold t such that if $e_L < t$ this outcome is second-best optimal[7], where t is the lower root of the quadratic equation $(A + Be_L + Ce_L^2) = 0$, where

$A \equiv mp - dp\phi(r_a - r_b)$;

$B \equiv d\phi(N - 1 + 2p + 2(r_a - r_b))$;

$C \equiv 2pd\phi$;

More straightforwardly we can show the results of two types of quarantine policy that can lead to sub-optimal outcomes:

Proposition 3: If exposure levels are discretely distributed between two types, so that the exposure factor $e_i \epsilon \{e_L, e_H\}$ with p the proportion of individuals with low exposure level, and if $e_L < 1 - \frac{m}{\phi d(r_a - r_b - 2pe_L)} < e_H$, then

a) if a quarantine intervention of type 1 raises migration cost to m' such that $e_L > 1 - \frac{m'}{\phi d(r_a - r_b)}$, the unique equilibrium has zero migration by both exposure types, which if $e_L < t$ is second-best sub-optimal since low-risk types are not separated from high-risk types;

b) if a quarantine intervention of type 2 imposes a disutility cost of q on sick individuals in city a such that $e_H < 1 - \frac{m-q}{\phi d(r_a - r_b - 2pe_L)}$, there exists no equilibrium in pure strategies, and the sole mixed-strategy equilibrium has a proportion $\pi < 1$ of high-risk types migrating, with π the solution to the equation $1 - \frac{m-q}{\phi d(r_a - r_b - 2pe_L - 2\pi(1-p)e_H)} = e_H$. If $e_L < t$ this is also suboptimal since a proportion of high-risk types are not separated from low-risk types.

Proof:

a) follows immediately from the definition of the migration threshold;

b) follows from noting that if all individuals in city a migrated to city b there would remain no infected individuals in city a, so migration would no longer be an equilibrium. QED

What do Propositions 2 and 3 tell us? As before, migration entails two types of change in welfare. First, migration imposes a negative externality as migrants will encounter a larger number of healthy individuals in the destination area than in their area of origin, thereby increasing the overall exposure of others to infection risk. Secondly, all migrants have a net positive "private" gain from migration as the distribution of exposure levels is discrete (so that migrants are all infra-marginal individuals, who gain a strictly positive private benefit from migrating). Proposition 2 simply says that, if low type individuals are low risk enough, the negative externality is more than compensated by the positive "private" gains, and migration of low type individuals increases overall social welfare. From this it follows naturally that,

[7] What we mean by "second-best" optimality is optimal "in the set of feasible policies where the government can control m and q".

in this case, discouraging low type individuals by quarantine measures of type 1 worsens social welfare (Proposition 3a) and that, encouraging high type individuals to migrate by quarantine measures of type 2 worsens social welfare (Proposition 3b), as high risk individuals who migrate entail excessive negative externalities.

Proposition 4 then immediately implies conditions under which welfare-improving interventions may exist:

Proposition 4: If exposure levels are discretely distributed between two types, so that the exposure factor $e_i \epsilon \{e_L, e_H\}$ with p the proportion of individuals with low exposure level, then:

a) if $1 - \frac{m}{\phi d(r_a - r_b - 2pe_L)} < e_L < e_H$, no individuals migrate, which is sub-optimal if $e_L < t$, but a second-best optimal migration can be induced by a quarantine intervention of type 2 such that $m - \phi d(r_a - r_b - 2pe_L)(1 - e_H) > q > m - \phi d(r_a - r_b - 2pe_L)(1 - e_L)$;

b) if $e_L < e_H < 1 - \frac{m}{\phi d(r_a - r_b - 2pe_L)}$, there exists no equilibrium in pure strategies, and the sole equilibrium has all low-risk types migrating and a proportion $\pi < 1$ of high-risk types migrating, with π the solution to the equation $1 - \frac{m}{\phi d(r_a - r_b - 2pe_L - 2\pi(1-p)e_H)} = e_H$.

If $e_L < t$, second-best optimal migration can be induced by a quarantine intervention of type 1 which raises m to m' such that $\phi d(r_a - r_b - 2pe_L)(1 - e_L) > m' > \phi d(r_a - r_b - 2pe_L)(1 - e_H)$;

c) if $e_L \geq t$ the second-best optimum has zero migration, which can be induced by a quarantine intervention of type 1 such that $m' > \phi d(r_a - r_b - 2pe_L)(1 - e_L)$.

Note that, once again, even where welfare-improving policy interventions exist they must be carefully designed to encourage only low type individuals to migrate. The quarantine measure should not "overshoot", moving from a situation in which neither type of individual migrates to one in which both types migrate, or vice versa - unless even the low-risk types are sufficiently high-risk that it is better for society that they do not migrate at all.

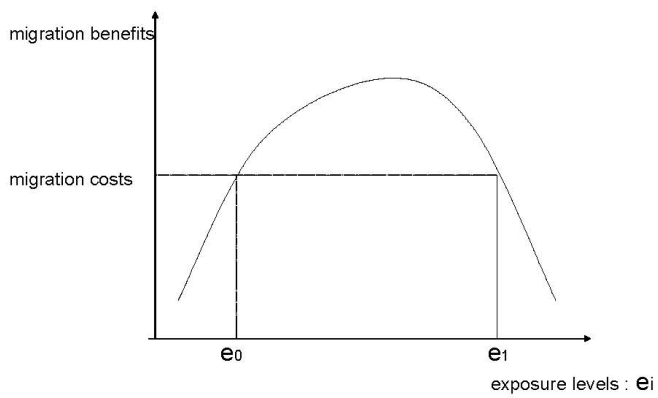
3.3 Absence of First Best Policy under Asymmetric Information

We now turn to the case of "type-sensitive epidemics", which are characterised by the fact that individuals with different exposure to the disease have also different risk of being infected in the future. For simplicity we will suppose that the risk of future infection of a healthy individual is simply proportional to past exposure (as well as to the disease- and city-specific factors just described):

$$pr(s_{ijt+1} = s | s_{ijt} = h) = e_{ij} \phi N_j \tag{11}$$

However, even without proportionality the qualitative insights of the model would remain unchanged so long as risk is increasing in past exposure.

For such "type-sensitive infections", migration incentives are more complicated as they are are non-monotonic in risk type, which determines two migration thresholds. So if the government does not know exactly the infectious statut of migrants, we will see that individuals cannot be taxed correctly.



We first solve for the new migration thresholds.

Using equation (11) instead of equation (1), the gross gains from migration to individual i become

$$\mu_S \equiv e_i(1-e_i)(r_a - r_b - 2n_s)d \tag{12}$$

where n_s is once again endogenous, and will depend on the distribution of exposure levels.

It is evident that, unlike μ_I, μ_S is not monotonic in e_i. Indeed, it follows immediately from (12) that

Proposition 5: When epidemics are type-sensitive, migrants have rational expectations and exposure levels are continuously distributed, for any given migration cost m and migration level n_s, there exist two thresholds e_0 and e_1, which are solutions to the quadratic equation $e_i - e_i^2 = m/d(r_a - r_b - 2n_s)$, such that any individual of risk type $e_i \epsilon [e_0, e_1]$ chooses to migrate.

An illustration of the migration thresholds is given in Figure 1.

Therefore, when types are continuously distributed, any change in migration cost will change the number of migrants in the neighborhood of both the low and the high thresholds. At laissez faire equilibrium, marginal migrants at both ends of the distribution will impose negative externalities through encountering a larger number of healthy individuals in the destination area than in their area of origin, thereby increasing the overall exposure of others to infection risk. At the same time, private gains of migration are zero at both migration thresholds. So, overall, there is always "too much" migration under laissez-faire when the distribution of exposure levels is continuous and the same policy implications follow as for a continuous distribution in the case of type insensitive epidemics.

It should be noted, however, that the magnitude of the externality imposed by the marginal migrants at the upper and lower thresholds respectively is not the same, since migrants at the high threshold increase the prevalence rate in the destination city by more than migrants at the low threshold. This means that any policy that imposes the same cost on all migrants (as is likely given the government's inability to distinguish risk types of apparently healthy individuals) will be inefficient compared to an optimal Pigouvian-type tax that taxed each individual according to the externality she imposed on others. Such a Pigouvian policy is infeasible here because of asymmetric information, which is one of the difficulties faced by public intervention in this context.

Note that the intuition of the welfare analysis for type insensitive epidemics in the simple examples studied above when exposure levels were discretely distributed continues to hold in this case : there are some cases where policy measures may be welfare improving and other cases where policy measures may be welfare decreasing. Once again this will depend on where the low and high types individuals are positioned compared to the two migration thresholds under laissez-faire such that either type, no type or both types may migrate in equilibrium; on how low low type individuals are; and on how policy measures are designed to affect migration costs or/and the disutility of living in the epicentre of the disease.

The point remains that policy needs to be designed with a careful attention to detail - not only can quarantine measures encourage instead of discourage migration, but they may affect individuals with quite different categories of risk exposure, with consequently quite different impacts on the welfare of others.

4 Discussions and conclusion

Our model has, first, highlighted the fact that the effectiveness of quarantine measures in the presence of epidemics are highly sensitive both to the type of the disease and to the information individuals have about their risk of being infected.

Since migration is motivated by the difference in prevalence levels between the two cities, individuals who are less likely to be infected derive greater benefit from migrating away from the epicentre of the disease. This leads to a private incentive to "self-quarantine", which, when people are perfectly informed about their own infection status, leads to first-best outcomes.

The above incentive is imperfect when people have imperfect information about their own infection status. In this case, migration imposes net negative externalities by increasing the rate of exposure faced by the uninfected outside the epicentre of the epidemic. In and of itself, this problem can be solved through traditional Pigovian taxes to implement the optimum migration level, as the authorities can calculate the externality imposed by the marginal migrant and correct for it.

However, traditional policy correctives may fail when epidemics are type-sensitive, since individuals vary in their ex ante risk of acquiring the disease, and this variation is likely to be unobservable to policymakers. In this case, there are two migration thresholds and the externality due to migration varies across individuals but policymakers are not be able to construct a schedule of taxes that varies

correspondingly. If they impose a uniform tax on migration, the costs imposed on the low-risk marginal migrant will likely be too high, and that on the high-risk individuals likely too low. In these conditions, the loss imposed on the low-risk (high risk) may be larger (smaller) than the gains enjoyed by the society that averts migration by some infected individuals, and it is not obvious whether policies restricting migration may or may not end up improving welfare on balance.

The points above imply that when modes of transmission are unknown and ex ante infection risk is more likely to be continuous, quarantine measures or other disincentives to migration may be beneficial provided they are designed with caution so that they do not overshoot. On the other hand, the probability that policy intervention produces perverse outcomes increases when exposure risks are known to individuals but not perfectly so, in a way that is unobservable to policy makers (e.g., the case of a sexually transmitted disease).

Second, our paper has highlighted the fact that although quarantine of individuals who have been identified as sick reduces (obviously) the propensity *of these individuals* to migrate and spread the disease, the threat of quarantine increases the propensity to migrate of other individuals who have not yet been fallen sick but who know themselves to be at risk.[8] This is surely one of the reasons why the health authorities in Toronto encouraged self-applied quarantine measures after the SARS outbreak. The idea was to encourage individuals voluntarily to adopt quarantine measures and go to the hospitals to get preventive treatment in case they were at risk of having been infected.

It is worth noting that, to the extent that individuals in fact migrate for reasons other than their infection risk, individuals who already know themselves to be infected may nevertheless wish to migrate following ill-considered quarantine measures and may thus pose a public policy problem. The points we make here about the risk of increasing rather than diminishing migration through ill-considered quarantine regulations would still be relevant.

How important these perverse effects are in practice will clearly depend on the degree of **i**ncomplete information and of asymmetry between individuals and the authorities, which will depend on the nature of the disease. This will also depend on how long is the incubation period of the disease during which such imperfections can be expected to last. That diseases differ greatly in the extent of imperfect information and asymmetries explains, among other reasons, why quarantine may be comparatively effective at halting the spread of a disease such as SARS, while it would be ineffective or even counter-productive at halting the spread of HIV/AIDS.

Similar considerations apply to many of the other "new" pathogens that have emerged in the last 25 years, including not only the Ebola virus, HIV and Hepatitis C but also lesser known pathogens such as Legionella pneumophilia, E coli 0157:H7, Borrelia burgdorferi, Helicobacter pylori, Hantavirus, Cryptosporidia, Ehrlichiosis, H5N1 (or Avian flu), and Nipah. Most of these pathogens have incubation

[8] For an account of the reactions of individuals to this type of quarantine restrictions during the SARS epidemic in China, including ways in which restrictions were evaded, see "In Liaoning" by Jon Cannon, *London Review of Books*, 25(11), 5 June 2003.

periods greater than standard travel times. For example, after the SARS virus enters the body, it requires 3-10 days incubation period before the disease appears, which is much longer than the duration of any trip by air. Hence an adequate response to the outbreak is indeed to encourage individuals to go to health centers when they suspect they may have been infected with SARS, instead of applying traditional non-voluntary quarantine measures to restrict migration away from the epicentre by those who may be infected[9].

We may also want to consider the case of re-emerging infectious diseases such as malaria, tuberculosis, dengue, yellow fever or cholera. However, for these diseases, it may be important to consider disease prevalence as a more long-term characteristic of a given area and, hence, a determinant of both the long-term location of individuals and the development process, as we study in a companion paper on migration and endemic disease, Mesnard and Seabright (2008).

We cannot claim to have done more than indicate in a simple and stylized context the complexity of the considerations that public health policies must take into account, but we hope to have shown that when individuals have incomplete private information about their exposure to disease, their strategic behaviour may have important effects on public health outcomes that policymakers cannot afford to ignore. Furthermore, some much simpler and more general messages can be derived than just the conclusion that everything is complicated and the authorities need to take the specific circumstances of the disease into account. As we stated in the introduction, three principles in particular apply much more widely than in the specific circumstances of our model. First, when the disease has an epicentre, the marginal migrant imposes a net negative externality. Secondly, quarantine policies may encourage migration instead of discouraging it, specifically on the part of individuals who are not currently sick but fear they may become so in the future. Thirdly, even when they succeed in discouraging migration, quarantine policies may lower social welfare, and even increase overall disease incidence, if they go too far, thereby discouraging those intra-marginal migrants for whom private benefits substantially exceed private costs, by enough to outweigh the negative externality they impose on others. These principle are ones that could help in designing better quarantine policies even in circumstances to which the assumptions of our model do not literally apply.

5 Bibliography

Anderson, Kym and James Sarah, 1998. Externalities, optimal policy, quarantine, risk and uncertainty, SPS measures. CEPR Working Paper DP1934.

Benedictow, Ole, 2004. The Black Death 1346-1353 : the Complete History. Boydell Press.

Geoffard, Pierre -Yves and ,Thomas Philipson, 1996. Rational epidemics and their public control. International Economic Review, 37 (3), 603-624.

[9]Moreover, according to current data, infected people do not pass on the SARS virus to others during the incubation period, but become infectious only when the first symptoms appear-like cough or sneezing which spread droplets containing virus particules.

Gersovitz, Mark and Jeffrey S Hammer, 2004. The economical control of infectious diseases. Economic Journal, vol 114, No492, 1-27.

Kremer, Michael, 1996. Integrating behavioral choice into epidemiological models of AIDS. Quaterly Journal of Economics, 549-573.

Malani, Anup and Ramanan Laxminaryan, 2007: Incentives for Surveillance and Reporting of Infectious Diseases. University of Chicago Law School working paper.

McNeill, William H, 1977. Plagues and People. New York, Doubleday.

Markel, Howard and Alexandra M Stern, 2002. The foreignness of Germs: The persistent association of immigrants and disease in American society. The Milbank Quarterly, Vol 80, No.4.

Mesnard Alice and Paul Seabright, 2008. Migration and the equilibrium prevalence of infectious diseases. CEPR working paper DP6651.

Philipson,Thomas, 2000. Handbook of Health Economics. Culyer and Newhouse (Eds.), North-Holland.

6 Appendix

Proof of Proposition 1:

a) Assumption (7) generalised to the case with rational expectations (i.e. with n_s endogenous) implies that $\overline{e} > 1 - \frac{m}{\phi d(r_a - r_b - 2n_s)} > \underline{e}$. Since e is continuous on $[\underline{e}, \overline{e}]$, by the intermediate value theorem there exists $e^* = 1 - \frac{m}{\phi d(r_a - r_b - 2n_s)}$. Lemma 1 implies that individuals with $e_i > e^*$ have gross gains from migration less than m, and individuals with $e_i < e^*$ have gross gains from migration greater than m.

b) We can define $\widetilde{e}$ as the level of e at which $\frac{dI}{de} = 0$, provided the second order conditions for a minimum hold. We define $\widetilde{n}_s$ and $\widetilde{n}_h$ as the corresponding values of n_s and n_h. Note that

$$\frac{dI}{de} = f(e)e\frac{\partial I}{\partial n_s} + f(e)\,(1-e)\,\frac{\partial I}{\partial n_h} \tag{13}$$

with f(e) the density of the distribution at e.

Substituting equations 3 and 4 and collecting terms in m yields

$$\frac{dI}{de} = f(e)\left\{m + (1-e)\left[d\phi\left(r_b - r_a + 2n_s\right)\right] + e\left[d\phi\left(N - 1 - r_b + r_a + 2n_h\right)\right]\right\} \tag{14}$$

Equation 10 implies that $m + (1-e^*)\left[d\phi\left(r_b - r_a + 2n_s\right)\right] = 0$, so when $\frac{dI}{de}$ is evaluated at $e = e^*$, the first two terms on the RHS of equation (14) are zero. Thus we are left with

$$\frac{dI}{de^*} = f(e^*)e^*\left[d\phi\left(N - 1 - r_b + r_a + 2n_h\right)\right] \tag{15}$$

which is strictly positive given that $f(e^*)$ is strictly positive. This implies that at $e = e^*$migration is higher than the optimum : a reduction in e would strictly reduce costs.

The migration tax m^t that implements the optimum can be calculated by evaluating equation (14) at $e = \widetilde{e}$ and noting that

$$m + m^t + (1 - \widetilde{e}) \left[d\phi \left(r_b - r_a + 2\widetilde{n}_s \right) \right] = 0 \tag{16}$$

which implies that

$$m^t = \widetilde{e} \left[d\phi \left(N - 1 - r_b + r_a + 2\widetilde{n}_h \right) \right] \tag{17}$$

The uniqueness of m^t follows from the fact that it is strictly increasing in $\widetilde{e}$, since $\widetilde{n}_h$ is itself strictly increasing in $\widetilde{e}$. If they know the distribution of e_i (and therefore the values of n_h and n_s for any threshold e), the authorities can calculate $\widetilde{e}$ using (14) and thereby calculate m^t.

To verify that the second order conditions for a minimum hold, write the derivative of equation (14) as:

$$\frac{d^2 I}{de^2} = d\phi \left(N - 1 + 2n_h - 2n_s + 2e \right) + 2d\phi(r_a - r_b) \tag{18}$$

which is strictly positive since $N > 1, (r_a - r_b) > 0$ and $e > n_s$. QED

Proof of Proposition 2:

a) From the conditions on e_L and e_H it follows immediately that all and only low-exposure individuals migrate. Thus $n_s = pe_L$ and $n_h = p\left(1 - e_L\right)$ and we can define I^*, the "intermediate migration" social disutility level, by

$$\begin{aligned} I^* \quad = \quad & mp + d(1 - r_a - p\left(1 - e_L\right)) \left(r_a - pe_L\right) \phi + r_a - pe_L + \\ & \left(N - r_b + p\left(1 - e_L\right)\right) \left(r_b + pe_L\right) \phi + r_b + pe_L \end{aligned}$$

We first show that this is lower than the no-migration disutility level, defined by :

$$I^0 = d\left[\left(1 - r_a\right)\left(r_a\right)\phi + r_a + \left(N - r_b\right)\left(r_b\right)\phi + r_b\right]$$

Subtracting yields:

$I^* - I^0 = p(A + Be_L + Ce_L^2)$, where

$A \equiv m - d\phi(r_a - r_b)$;

$B \equiv d\phi(N - 1 + 2p + 2(r_a - r_b))$;

$C \equiv -2pd\phi$.

Noting that equation 9 implies that $A < 0$, that $B > 0$ because $N > 1$ and $r_a > r_b$, and that $C < 0$, it follows that $A + Be_L + Ce_L^2$ is a quadratic function with a negative intercept, which admits two positive real roots. Substituting $e_L = 1$ and noting that $A + B + C > 0$, we can show that, for $e_L = 1$, $I^* - I^0$ is positive. So only one of the two roots lies between 0 and 1 . We denote this root t.

Therefore $I^* - I^0$ is negative if $e_L < t$ where t is the lower root of the quadratic equation $A + Be_L + Ce_L^2 = 0$.

To see that I^* is also lower than the all-migration level I^a in which all individuals leave city a for city 1, note that as $n_h + n_s = 1$ and, after migration, $N_b = (N - r_b + (1 - r_a))(r_b + r_a)\phi + r_a + r_b$, we can write

$$I^a = dN_b + m = d[DE\phi + E] + m$$

where $D \equiv (N - r_b + (1 - r_a))$ and $E \equiv r_a + r_b$

We can write

$$I^* = d[FG\phi + HI\phi + E] + mp$$

where $F \equiv (1 - r_a - p(1 - e_L)$,

$G \equiv (r_a - pe_L)$,

$H \equiv (N - r_b + p(1 - e_L))$,

$I \equiv (r_b + pe_L)$.

Therefore we obtain easily: $I^a - I^* = m(1 - p) + d\phi(DE - FG - HI)$.

Noting that $D = H + F$ and $E = G + I$, we can write:

$I^a - I^* = m(1 - p) + d\phi(HG + FI)$, which is necessarily strictly positive. QED