Changing Research Perspectives on the Global Health Workforce

Till Bärnighausen, David E. Bloom

July 2009

PGDA Working Paper No. 46
http://www.hsph.harvard.edu/pgda/working.htm

The views expressed in this paper are those of the author(s) and not necessarily those of the Harvard Initiative for Global Health. The Program on the Global Demography of Aging receives funding from the National Institute on Aging, Grant No. 1 P30 AG024409-06.
We thank Larry Rosenberg, Sherry A. Glied, and Peter C. Smith for valuable comments and Gaargi Ramakrishnan for research support. A shorter version of this paper will appear as a chapter in P. Smith & S. Glied (Eds.), Oxford Handbook of Health Economics. New York: Oxford University Press (Bärnighausen & Bloom, 2010 (forthcoming)).
ABSTRACT

Past research on the health workforce can be structured into three perspectives—“health workforce planning” (1960 through 1970s); “the health worker as economic actor” (1980s through 1990s); and “the health worker as necessary resource” (1990s through 2000s). During the first phase, shortages of health workers in developed countries triggered the development of four approaches to project future health worker requirements. We discuss each approach and show that modified versions are experiencing a resurgence in current studies estimating health worker requirements to meet population health goals, such as the United Nations’ health-related Millennium Development Goals. A perceived “cost explosion” in many health systems shifted the focus to the study of the effect of health workers’ behavior on health system efficiency during the second phase. We review the literature on one example topic: health worker licensure. In the last phase, regional health worker shortages in developing countries and local shortages in developed countries led to research on international health worker migration and programs to increase the supply of health workers in underserved areas. Based on our review of existing studies, we suggest areas for future research on the health workforce, including the transfer of existing approaches from developed to developing countries.

Till Bärnighausen
Harvard School of Public Health
Department of Global Health and Population
665 Huntington Ave.
Boston, MA 02115
and Africa Centre for Health and Population Studies
University of KwaZulu-Natal
tbarnighausen@africacentre.ac.za

David E. Bloom
Harvard School of Public Health
Department of Global Health and Population
665 Huntington Ave.
Boston, MA 02115
dbloom@hsph.harvard.edu
I. Introduction

“People deliver health” (Joint Learning Initiative, 2004). The health workforce, i.e., the people who are “primarily engaged in action with the primary intent of enhancing health” (WHO, 2006b), diagnose illnesses, heal, care for people, monitor health outcomes, support treatment adherence, provide medical information and prevent diseases. The importance of health workers as decision makers and service providers in health systems is obvious. It is matched by their impact on health spending. WHO estimates that across countries worldwide about 50% of total public and private health expenditure (including capital costs) is spent on health worker wages, salaries and allowances (Hernandez, Dräger, Evans, Tan-Torres Edejer, & Dal Poz, 2006). In addition, health workers are indirectly responsible for a large proportion of recurrent spending on health care through drug prescription, hospital admissions and other resource use in providing diagnostic, therapeutic and palliative services.

In the following, we review the health economics and health systems literature on the health workforce. We organize the review according to three perspectives on health workers, which correspond roughly to chronological phases of academic publication: health workforce planning (1960s and 1970s), the health worker as economic actor (1980s and 1990s), and the health worker as necessary resource (1990s and 2000s).

A major research focus of studies on the health care workforce in the 1960s and 1970s was on models to predict future health staffing needs. This research was triggered by shortages of specific types of health workers in developed countries and by reports from socialist countries that health manpower planning can aid health policy makers in ensuring an adequate supply of health workers. Planning models developed in this phase continue to be applied today both in developed and developing countries.

In the 1980s and 1990s, the research focus shifted to the study of health workers’ effect on allocative and technical efficiency in health systems. Research in this phase was motivated by the perception of rising health expenditures in developed countries and the belief that health workers do not always act in their patients’ best interest. Many of the important

---

1 Hernandez and colleagues identified data from 29 countries in Europe, 7 countries in the Americas, 4 countries in the Western Pacific region, 2 countries in Africa, and 1 country in the Eastern Mediterranean. It is thus possible that the true average across all countries worldwide substantially differs from this estimate. Saltman and Von Otter estimated that 65% to 80% of recurrent health expenditure is spent on health worker remuneration (Saltman & Von Otter, 1995); for Africa, Huddart and Picazo estimated this proportion to lie between 50% and 70% (Huddart & Picazo, 2003).
research topics in this phase are covered in other chapters of the handbook, including agency (chapter 26), provider payment and incentives (chapter 27), competition among providers (chapter 29), performance and productivity (chapters 30 and 31), and medical decision-making (chapter 33). In this chapter, we review the literature on one research theme in this phase that is not discussed in detail elsewhere in this book (health worker licensure).

Starting in the 1990s and continuing in this decade, health workers became increasingly viewed as a resource that is necessary to achieve population health goals. In many geographical areas, health policies could not be implemented because health workers were lacking. In developed countries, essential health care could not be delivered in rural and remote areas because the human resources for such delivery were not available. In developing countries, it became apparent that population health goals – such as the Millennium Development Goals (MDGs) to “reduce child mortality”, “improve maternal health”, and “combat HIV/AIDS, malaria and other diseases” (United Nations, 2009) could not be attained unless the sizes of many national health workforces were dramatically increased. As major themes in this phase, we review the research on programs to increase the supply of health workers to underserved areas in developed countries and studies measuring and analyzing international health worker migration.

Table 1 provides an overview of the three perspectives on health workers. In separate subchapters below, we describe the health policy backgrounds that gave rise to each of the perspectives and then review relevant literature. The three perspectives are of course highly stylized; the backgrounds are reductionist descriptions of much richer policy contexts; and the periods overlap. However, the perspectives serve useful purposes in this subchapter in framing past research, structuring our exposition, and laying out a research agenda on the health workforce for coming years.
Table 1: Three perspectives on the health workforce

<table>
<thead>
<tr>
<th>Period</th>
<th>First perspective: Health workforce planning</th>
<th>Second perspective: The health worker as economic actor</th>
<th>Third perspective: The health worker as necessary resource</th>
</tr>
</thead>
<tbody>
<tr>
<td>Period</td>
<td>1960s and 1970s</td>
<td>1980s and 1990s</td>
<td>1990s and 2000s</td>
</tr>
<tr>
<td>Background</td>
<td>• General health worker shortages</td>
<td>• Rising health expenditures</td>
<td>• Local health worker shortages</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Cost containment policies</td>
<td>• Focus of international development aid on population health</td>
</tr>
<tr>
<td>Geographical focus</td>
<td>Developed countries</td>
<td>Developed countries</td>
<td>Developing countries and developed countries</td>
</tr>
<tr>
<td>Research themes</td>
<td>Health workforce planning models</td>
<td>Application of microeconomic theory to the study of health worker behavior</td>
<td>• Interventions to increase the supply of health workers to medically underserved areas</td>
</tr>
<tr>
<td></td>
<td>• Need</td>
<td>• Agency</td>
<td>• Health worker migration from developing to developed countries</td>
</tr>
<tr>
<td></td>
<td>• Demand</td>
<td>• Provider payment and incentives</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Service targets</td>
<td>• Competition among providers</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Population ratio</td>
<td>• Performance and productivity</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Medical decision-making</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Health worker licensure</td>
<td></td>
</tr>
</tbody>
</table>
II. First phase: Health workforce planning

II.1 Background


Second, reports from socialist countries, such as the USSR, asserted that health manpower planning could produce valid projections of future health workforce requirements, aiding policy makers in designing workforce policies to ensure sufficient supplies of health workers (Daniels, 1974; Popov, 1971). Third, WHO identified a number of technical difficulties that policy makers would inevitably face in ensuring adequate future supplies of health workers both in developing and developed countries (Hall & Mejia, 1978).
“Manpower requires the longest preparatory period of all the health resources and cannot be improvised. It is also subject to a certain inherent inertia, in that the rigidity of the health and education systems and the attitudes of health workers do not make for easy mobility or conduce to improving geographical and occupational distribution. Nor can manpower be stored or discarded. If it is to be available at the proper time, it has to be planned for in advance in the right amount and type – no more and no less than is needed. Because they are subject to obsolescence, manpower abilities and skills also need to be maintained by means of permanent supervision and continuing education.”

WHO advocated health workforce planning as a method of ensuring a sufficient supply of health workers despite these difficulties. Initially, WHO had defined health workforce planning as “the process of estimating the number of persons and the kind of knowledge, skills, and attitudes they need to achieve predetermined health targets and ultimately health status objectives” (Hornby, Mejia, Ray, & Simeonov, 1976). The organization envisioned that health workforce planners would calculate future health worker requirements through mathematical modeling, using detailed data on population projections, disease burdens, health services, and capacities of health care facilities and education institutions. The planners would prepare several health workforce scenarios and present them to policy makers, who, in turn, would select the best option for implementation. However, WHO soon came to realize that health policy-making did not follow this rational model.

“What are the lessons to be learned from experience with manpower planning? Such planning is now more concerned with the actual decision-making process and implementation of decisions than with manipulation of numbers in forecasting demand and supply … Planning efforts are unlikely to be effective if due account is not take of the social, economic, and – most important – the political milieu in which they take place” (Hornby et al., 1976).

“[M]anpower studies or reports of commissions, however, sophisticated, do not necessarily lead to the development, much less to the implementation of a plan, or to an integrated process of health manpower development unless the necessary social, economic, and political conditions and a definite national political will are present” (Hall & Mejia, 1978).

“In a word, the health manpower planning process came to be recognized in WHO, as it was in countries, as being more political and social than mathematical” (Fülöp & Roemer, 1982).
The 1978 WHO report *Health Manpower Planning: Principles, Methods, Issues* (Hall & Mejia, 1978) thus broadened the objectives of planning to include implementation plans and communication strategies and emphasized the importance of the “political dimension”, “leadership readiness for and commitment to change”, “enabling legislation for planning and subsequent plan implementation”, and “administrative capacity and willingness to implement the plan” (Hall & Mejia, 1978).

II.2 Literature

While the focus of health workforce planning thus moved away from mathematical modeling towards management and policy-making, the initial emphasis on estimation led to the development of the four main approaches – still used today – to plan national or regional health workforce requirements: the need, demand, service targets, and population ratio approach (see Table 2 for an overview). The approaches differ in their scientific stance, which, in turn, determines the type of information that is taken into account in the planning effort. The need approach takes a normative stance. Experts use epidemiological information to estimate the future occurrence of disease cases in a population. The number and types of health workers necessary to provide the services to adequately treat all disease cases are then calculated, using information on the amount of health worker time required to provide certain services. In contrast, the demand approach takes a positive stance. Future demand for health services is predicted from current demand by assuming that the relationships between demand and its determinants (such as population size and income) will remain unchanged while the level of the determinants changes. Predicted health service demand is then translated into human resource requirements.

The service targets approach usually takes a normative stance but, unlike the need approach, allows for constraints to the provision of health care in calculating future service requirements. Such constraints include health care capacity and technology on the supply side and ability and willingness to pay for health care on the demand side. Finally, the population ratio approach can either take a normative or a positive stance. Future health worker requirements are predicted using population growth predictions and health worker-to-population ratios derived from studies or deliberations among policy makers.

---

2 These approaches are defined well in the 1978 WHO book on health manpower planning by Hall and Mejia (Hall & Mejia, 1978). More recent reviews used similar categorizations of planning approaches (Dreesch, Dolea, Dal Poz, Goubarev, Adams, Aregawi et al., 2005; Markham & Birch, 1997; Murphy, 2002; O'Brien-Pallas, Baumann, Donner, Murphy, Lochhaas-Gerlach, & Luba, 2001).
The need, demand, and service targets approach require the calculation of health services requirements as a step preceding the estimation of health worker requirements. Since the need and the service targets approach are normative, they are likely to stimulate the study of health worker productivity, such as analyses of the types and combinations of health workers that can produce desired numbers of health services most efficiently. In contrast, the demand approach does not motivate the study of health worker productivity, because it takes current modes of health services delivery as given, and the population ratio approach does not require a translation of health services into health human resources requirements and thus adds little information to the study of health worker productivity.

The difference between projected health worker requirements and the current supply of health workers is likely to be largest if the need approach is used for projection because – unlike the other approaches – the need approach neither takes account of a country’s current level of health care provision nor considers a country’s capacity to deliver health care. As a result, in many developing countries the needed number of health workers substantially exceeds the current supply of health workers (Joint Learning Initiative, 2004; WHO, 2006b). In contrast, in developed countries the needed health worker numbers may be significantly lower than the number of currently available workers, either because health service demand exceeds need or a country has produced more health workers than necessary to satisfy either need or demand (R. A. Cooper, 2004; R. A. Cooper, Getzen, & Laud, 2003).³

Both the need and the service targets approach require that health policy makers can effectively control the health care sector in order to ensure that the health workers in a country do indeed provide the services that have been determined to be needed and do not provide unneeded services demanded by consumers. In contrast, as Hall and Mejia write, studies following the demand approach “have the greatest applicability in countries … in which the government plays a relatively passive role in relation to health sector development and hence is more concerned with anticipating than shaping the future” (Hall & Mejia, 1978).

³ Two general observations explain why health worker need and status quo are likely to differ in opposite directions in developing and developed countries. First, health care need decreases with income (Deaton, 2003; Riley, 2001). Second, the supply and utilization of health workers increases with income (R. A. Cooper et al., 2003).
Table 2: Four approaches to health workforce planning

<table>
<thead>
<tr>
<th>Approach</th>
<th>Need</th>
<th>Demand</th>
<th>Service targets</th>
<th>Population ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scientific stance</td>
<td>Normative</td>
<td>Predictive</td>
<td>Normative</td>
<td>Normative or predictive</td>
</tr>
<tr>
<td>Origin</td>
<td>Epidemiology, Medicine</td>
<td>Economics</td>
<td>Health services research</td>
<td>Geography</td>
</tr>
<tr>
<td>Assumption⁴</td>
<td>(All) health needs should and can be met</td>
<td>Current level and distribution of health services adequate</td>
<td>Planned coverage of specific priority health services can be achieved within the time period of projection</td>
<td>An adequate health worker-to-population ratio can be identified</td>
</tr>
<tr>
<td>Data requirements</td>
<td>High</td>
<td>High</td>
<td>Low to high</td>
<td>Low</td>
</tr>
<tr>
<td>Health worker requirements derived from health services projections</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Stimulates study of health worker productivity</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Expected difference between projections and status quo*</td>
<td>Large</td>
<td>Small</td>
<td>Intermediate</td>
<td>Small or large</td>
</tr>
<tr>
<td>Level of sectoral control necessary for implementation</td>
<td>High</td>
<td>Low</td>
<td>High</td>
<td>Low or high</td>
</tr>
</tbody>
</table>

*First approach to estimating health workforce requirements: need*

A good example of an analysis following the need approach to estimating health workforce requirements is a 1972 study by Schönfeld, Heston and Falk calculating the numbers of physicians required for primary medical care in the United States (US) (Schönfeld, Heston, & Falk, 1972). Schönfeld and colleagues used data from the National Center for Health Statistics to predict the annual numbers of cases of different diseases requiring attention by two categories of primary care physicians (pediatricians and internists). They then calculated

⁴ Sources: (Dreesch et al., 2005; Hall & Mejia, 1978; Murphy, 2002).
the total number of pediatrician and internist hours needed to adequately treat all disease cases using treatment norms based on health worker opinion.

“For this specific study, the basic clinical data for number of services and time per unit of service have been accumulated from 617 interviews with physicians, expressing the judgments of 24 pediatricians and 55 internists regarding what should be done to provide good care in each of 80 diseases occurring in children and 170 diseases of adults” (Schönfeld et al., 1972).

Finally, in order to estimate health worker need, the authors divided the total pediatrician hours required in one year by the average number of hours a pediatrician works per year (2,227 in the period 1965 to 1967) and the total internist hours by the average number of hours an internist works per year (2,198). The study concluded that 133 primary care physicians were needed per 100,000 persons, while only about half that number were available at the time (Schönfeld et al., 1972).

Other applications of the need to approach to estimating health worker requirements in the US include the Lee-Jones report of 1933 (Lee & Jones, 1933) and the “adjusted needs-based model” developed by the Graduate Medical Education National Advisory Committee (GMENAC) of 1980 (Graduate Medical Education National Advisory Committee (GMENAC), 1980). In the period 1960-1980, the need approach to estimating health worker requirements was further applied in the USSR (Ministry of Health, 1967), in Latin America (Center for Development Studies (CENDES), 1965), and in Sri Lanka (Hall & Mejia, 1978). A recent example of a study following the need approach is an analysis by Birch and colleagues who estimated the need for registered nurses in the Atlantic Region of Canada (Birch, Kephart, Tomblin-Murphy, O'Brien-Pallas, Alder, & MacKenzie, 2007).

The need approach is appealing to health workers because it is clearly rooted in the sciences of epidemiology and medicine and its underlying ethic of health care is often embraced by the health care professions (Bodenheimer & Grumbach, 2002; Cookson & Dolan, 2000) – health care should be provided according to need and independently of other characteristics of individual patients. However, the approach has been criticized on conceptual and practical grounds. Conceptually, as Klarman (1969) points out, “[w]hether need is a desirable standard hinges on society’s willingness to accord an absolute priority to health services, regardless of cost” (Klarman, 1969). Some modifications of the need approach take this criticism into account. In some studies, only those diseases whose treatments met certain
cost-benefit criteria were taken into account in planning (Hall & Mejia, 1978). Practically, the morbidity data for need projection may not be available in many countries or may be measured with considerable uncertainty. In addition, the need approach may not be robust across geographical areas, practice settings, or time periods. For instance, Schönfeld, Heston, and Falk cautioned against generalization of their study findings because

“the interviewed physicians of this study are mainly in solo or limited partnership practice in one geographical area (New Haven, Connecticut), and their opinions apply to only one point in time – the present. It may therefore be necessary to have these current views and the derivative standards of what constitutes good care reviewed by physicians in other parts of the country and in other types of practices” (Schönfeld et al., 1972).

The results from the GMENAC study could not be replicated in a later study when treatment norms were developed through analysis of visit records in three large health maintenance organizations (HMOs) or by a Delphi panel of pediatricians belonging to the three HMOS instead of by the national panel of pediatricians used in the original study (Weiner, Steinwachs, Shapiro, Coltin, Ershoff, & Connor, 1987). The analyses of visit records obtained from the HMOs Maxicare, Med Centers Health Plan, and Harvard Community Health Plan yielded estimated numbers of pediatricians that were 59%, 84%, and 92% of the GMENAC estimate, respectively. The Delphi panel of pediatricians from these three HMOs yielded estimated numbers that were 39%, 48%, and 33% of the GMENAC estimate, respectively.

Moreover, even if data and methods allowed valid estimation of health worker requirements according to health need and a country had sufficient resources and the political will to provide all the required health workers, it would be unlikely that health care would be provided exactly according to policy plans. Patients may not demand needed health services and may demand unneeded ones; health workers may not supply needed services and may supply unneeded ones (Murphy, 2002). Finally, since the treatment norms used in the need approach are based on current medical practice, the approach may inhibit innovation in the delivery of health care. In many cases, current medical practice may be replaced by alternative means of delivery, e.g., by substituting one type of health worker with another or by increasing the use of medical information technology.

In sum, the need approach may only be feasible in situations where detailed data on disease prevalence and incidence are available, the health sector can draw on adequate resources to
fund needed health workers, and the strength of sectoral control is such that it can be assured that most utilized health care is needed. In such settings, health policy makers should undertake careful sensitivity analyses of health worker requirements in order to assess how much estimates will change if expert opinion or health care technology changes.

Second approach to estimating health workforce requirements: demand

An example of an analysis following the demand approach is a 1975 study by Hall, Reinke and Lawrence on health workforce requirements in Chile (Hall, Reinke, & Lawrence, 1975). The researchers first estimated in regression analyses the effects of geographical location, income, sex, age, insurance coverage, and education on three measures of health care demand (doctor visits, dentist visits, hospital nights). They then used predictions of the independent variables to estimate future demand for health services (10 years and 20 years after 1968, the year for which the equations were estimated). Finally, they converted predicted demand into health worker requirements, using estimates of productivity and staffing patterns of hospitals and facilities providing ambulatory and dental care, which were based on field observation and expert opinion (Hall, 1971).

Other health manpower studies employing a demand approach include Baker and Pearlman’s 1967 study on Taiwan (Baker & Perlman, 1967), Fein’s 1967 study on the US (Fein, 1967), and Hall’s 1969 study on Peru (Hall, 1969). Recent examples are a study predicting the demand for accident and emergency services in the UK (Chambers & Johnson, 1986) and a 2008 WHO study projecting the global demand for physicians (Scheffler, Liu, Kinfu, & Dal Poz, 2008).

Whether the predictive nature of the demand approach is an advantage or a disadvantage will depend on the intended use of the estimated health worker requirements. In situations in which policy makers intend to merely adjust the status quo of health services delivery to changes in a few factors that are outside their control (such as population size and composition, economic growth, or migration) demand studies may provide sufficient information about future health worker requirements. Demand studies may also provide baseline estimates for further normative analyses in countries where the government intends to significantly change the pattern of health care utilization.

While demand studies could thus inform health policy in many circumstances, they are often not feasible because the data necessary to estimate demand functions are lacking (Dreesch et
al., 2005). In addition, demand-based studies may not yield reliable estimates of future health worker requirements. First, they commonly do not take into account (as, for example, the studies mentioned above) that demand and supply are simultaneously determined. Estimates of health worker requirements that do not account for the interaction between demand and supply in the health care market are thus likely to be biased. Research on health worker requirements following the demand approach may benefit from increased use economic models of health care demand (Benham, 1971; M. S. Feldstein, 1967, 1971). Second, demand-based studies usually assume that the relationships between demand and its determinants will remain unchanged in the future, an assumption that is likely to be violated in many situations. Third, they assume that projections of the development of the determinants of health care demand are reasonably accurate. This assumption may not hold true; for instance, per-capita income may not rise to predicted levels because of an unexpected economic crisis. Fourth, they may wrongly assume generalizability of findings from the particular setting of one study to other settings. For instance, one of the determinants commonly taken into account in demand-based studies is income. However, estimates of the income elasticity of health care demand in microeconomic studies (which are often used for demand-based forecasting (Hall & Mejia, 1978; Klarman, 1969, 1973)) vary widely, even if they cover similar geographic areas and time periods, suggesting that these estimates are quite uncertain (see, for instance, (R. Anderson & Benham, 1970; Fein, 1967; P. J. Feldstein, 1964; Gorham, 1967; Newhouse & Phelps, 1976; Rosett & Huang, 1973; Silver, 1970; Stigler, 1956)). In sum, studies of health worker requirements following the demand approach can be useful in certain situations. In many other situations, however, they may not provide the information required to plan and implement future health policies, may provide the required information only with considerable uncertainty, or may not be feasible to conduct at all.

**Third approach to estimating health workforce requirements: service targets**

Hall and Mejia (Hall & Mejia, 1978) describe an example of a study deriving health worker requirements from service targets. Colombian health planners first identified “priority services” for a population of nine million people lacking access to basic health care, based on “morbidity surveys carried out earlier, the accumulated experience of pilot simplified medicine programmes, statistics on service utilization, referral rates, and international experience”. The planners then calculated the numbers of different types of health workers
(health promoters and nurse auxiliaries working in communities and physicians and support staff working at health posts and in hospitals), using “normative techniques”. Many of the more recent examples of estimations of health worker requirements follow a service target approach. For instance, Kurowski, Wyss, Abdulla and Mills (Kurowski & Mills, 2006; Kurowski, Wyss, Abdulla, & Mills, 2007) estimated the numbers of health workers required to deliver 33 “priority interventions” identified by the Commission on Macroeconomics and Health as important to achieving the MDGs (Commission on Macroeconomics and Health, 2001; Working Group 5 of the Commission on Macroeconomics and Health, 2002). Based on WHO guidelines, the researchers broke the interventions down into tasks requiring different levels of health worker skill. Next, they measured the time needed to complete the different tasks through structured interviews with randomly selected health workers in Tanzania. Then, they estimated through a time and motion study (Robbins, Bergman, Stagg, & Coulter, 2003) the proportion of health worker time spent providing health care (staff productivity) and the proportion of productive staff time spent providing the selected priority interventions (service productivity). Finally, the researchers estimated the numbers of nurses, midwives, physicians, and technicians necessary to provide the “priority interventions” in Tanzania under three productivity scenarios (assuming that staff and service productivity remained unchanged, deteriorated, or improved). Dreesch and colleagues (Dreesch et al., 2005) proposed a similar method to estimate the health workforce required to achieve the MDGs. First, they selected priority interventions, taking into account “incidence and prevalence of health problems, demographic characteristics of the population, and (internationally) the targets set forth by the MDGs, or (at country level) targets identified in health sector strategic plans”. Next, they identified in functional job analysis (F. I. Moore, 1999) the tasks required to deliver the interventions. Then, they estimated the time required to complete the tasks based on observation (Hurst, 2002) or expert opinion. Finally, they considered whether required times per task could be reduced by improving productivity through task reorganization.

As these examples demonstrate, the main differences between the two normative planning approaches (the service targets and the need approach) are in the scope of health care considered and in the constraints taken into account. The service targets approach decides on priority interventions, taking into account health care need but also (explicitly or implicitly) existing constraints to meeting that need; in contrast, the need approach starts by identifying health care need and assumes that all need can be met. The former approach will thus yield
more realistic human resources requirements than the latter. The service targets approach has the further advantages that it explicitly disaggregates need into different components, is easy to communicate, can be easily combined with other planning approaches, and facilitates the study of health worker productivity (Hall & Mejia, 1978). Moreover, unlike the need and the demand approach, the service targets approach does not depend on detailed local data to yield useful estimates of health workforce requirements.

On the other hand, just like the needs approach, the service targets approach has the disadvantage that it assumes that consumers demand and health workers provide precisely those services that the planners have used to derive health worker requirements. In most settings, however, the overlap between the health services consumed and those planned will be imperfect. The health services approach will be most useful in countries where the government can exert substantial control over health services and is willing to use this power to maximize delivery of planned services and minimize delivery of unplanned ones.

**Fourth approach to estimating health workforce requirements: population ratio**

Health worker-to-population ratios have been frequently used in planning manpower requirements. For instance, following a WHO recommendation Thailand aimed at achieving a physician-to-population ratio of 1:5000 in the period 1972-1976 (Chunharas, 1998). Bahrain tried to attain a physician-to-population ratio of 1:650 in the period 1998-2005 (Ahmed, Fateha, & Benjamin, 2000). In the 1993 World Development Report, the World Bank stated that “[t]he public health and minimum essential clinical interventions require about 0.1 physician per 1,000 population and between 2 and 4 graduate nurses per physician” (World Bank, 1993); and in the 2006 World Health Report, WHO identified a “needs-based sufficiency” threshold of 2.5 health workers per 1,000 population (WHO, 2006b):

“To achieve a global assessment of shortfall, the Joint Learning Initiative (JLI), a network of global health leaders, launched by the Rockefeller Foundation, suggested that, on average, countries with fewer than 2.5 health care professionals (counting only doctors, nurses and midwives) per 1,000 population failed to achieve an 80% coverage rate for deliveries by skilled birth attendants or for measles immunization.”

“This method of defining a shortage, whether global or by country, is driven partly by the decision to set the minimum desired level of coverage at 80% and partly by the empirical identification of health worker density associated with that level of coverage. Using a similar
“threshold” method and updated information on the size of the health workforce obtained for this report, the JLI analysis has been repeated for skilled birth attendants . . . A remarkably similar threshold is found at 2.28 health care professionals per 1000 population, ranging from 2.02 to 2.54 allowing for uncertainty. The 57 countries that fall below this threshold and which fail to attain the 80% coverage level are defined as having a critical shortage.”

The popularity of the population ratio approach to health workforce planning stems from the ease and flexibility of its use. The data to estimate current health worker-to-population ratios are available in most countries worldwide for some time periods, and ratio objectives are easy to communicate to policy makers and planners. On the other hand, the approach suffers from many limitations. First, population ratios are necessarily averages across geographical regions, subregions of which may have ratios that differ widely from the averages. For instance, Bennet, Hall, Lutwana and Rado in a study in 1965 found an overall physician-to-population ratio of 1:18,500 in Kenya, Uganda, and the then Tanganyika. This average, however, could be decomposed into an average ratio of 1:4,400 for urban areas and an average ratio of 1:31,000 for rural areas where more than 90% of the population lived (Bennett, Hall, Lutwana, & Rado, 1965). Figure 1 shows the quotients of urban nurse-to-population ratios divided by the rural nurse-to-population ratios for country-years for which these data were available in the WHO Global Atlas of the Health Workforce (WHO, 2009); Figure 2 shows the urban-to-rural quotients for physicians based on the same data source. The two figures show large differences between health worker population ratios in urban and rural areas in most countries for which data were available, suggesting that for many planning purposes pooled rural-urban averages will not be sufficiently disaggregated.

Second, unlike the other three health worker planning approaches, the population ratio approach does not require an intermediary step of estimating health services in order to calculate the health worker numbers. It is thus the approach that will be least likely to lead to a focus on health worker productivity in planning future worker requirements. Third, the

---

5 The graphs show the ratios for all country-years for which data were available in the WHO Global Atlas of the Health Workforce with one exception. We do not show the urban-to-rural ratio for physicians for Bhutan 2004. Because the ratio is quite large (599), the differences between the other ratios in the graphs would have been difficult to discern, had we included it in the display.

6 Note that while the WHO Global Atlas is the most reliable dataset on health worker population ratios available (Anand & Bärnighausen, 2004), it cannot be ruled out that the differences health worker population ratios between urban and rural areas shown in the two figures can be partially explained by differences in data sources and data definitions (such as definitions of physicians and nurses or rural and urban areas) (WHO, 2009). However, it is unlikely that improved adjustment for these differences would alter the conclusion that rural-to-urban health worker population ratios vary widely across countries.
approach does not explain which factors drive health worker requirements, except for changes in crude population size. Thus, the population ratio approach to health workforce planning is mainly useful for international comparison and as a first indicator of a country’s overall health human resources situation (Bossert, Bärnighausen, Bowser, Mitchell, & Gedik, 2007), rather than as the main planning tool to estimate health worker requirements.

**Figure 1: Quotients of urban nurse-to-population ratios divided by rural nurse-to-population ratios**
Figure 2: Quotients of urban physician-to-population ratios divided by rural physician-to-population ratios.

<table>
<thead>
<tr>
<th>Country</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gabon 2004</td>
<td>0.9</td>
</tr>
<tr>
<td>Sao Tome and Principe 2004</td>
<td>11.0</td>
</tr>
<tr>
<td>Uruguay 2004</td>
<td>11.6</td>
</tr>
<tr>
<td>Guinea 2004</td>
<td>6.4</td>
</tr>
<tr>
<td>Guinea-Bissau 2004</td>
<td>5.0</td>
</tr>
<tr>
<td>DRG Congo 2004</td>
<td>4.5</td>
</tr>
<tr>
<td>Pakistan 2004</td>
<td>4.2</td>
</tr>
<tr>
<td>Liberia 2003</td>
<td>3.9</td>
</tr>
<tr>
<td>Gambia 2003</td>
<td>3.3</td>
</tr>
<tr>
<td>Ghana 2004</td>
<td>3.1</td>
</tr>
<tr>
<td>Djibouti 2004</td>
<td>2.9</td>
</tr>
<tr>
<td>Maldives 2004</td>
<td>2.4</td>
</tr>
<tr>
<td>Cameroon 2004</td>
<td>2.1</td>
</tr>
<tr>
<td>Iraq 2004</td>
<td>2.0</td>
</tr>
<tr>
<td>Benin 2004</td>
<td>2.0</td>
</tr>
<tr>
<td>Mauritius 2004</td>
<td>1.7</td>
</tr>
<tr>
<td>Algeria 2002</td>
<td>1.7</td>
</tr>
<tr>
<td>Tanzania 2002</td>
<td>1.5</td>
</tr>
<tr>
<td>Mauritania 2004</td>
<td>1.0</td>
</tr>
<tr>
<td>Gabon 2004</td>
<td>0.9</td>
</tr>
</tbody>
</table>
III.  Second phase: The health worker as economic actor

III.1  Background

Starting in the 1970s and continuing in the 1980s and 1990s, policy makers in developed countries became increasingly worried about rising health expenditures (Altman & Levitt, 2002; Boccuti & Moon, 2003; R.G. Evans, Barer, & Hertzman, 1991; Gray, 1998). The average annual growth rate of health care expenditure in real terms for 18 countries belonging to the OECD was 3.0% in the period 1980-1990 and 3.3% in the period 1990-2001 (Huber & Orosz, 2003). Health care expenditures throughout the developed world rose in terms of both per-capita spending on health (Huber, 1999; Huber & Orosz, 2003; Martin, Whittle, & Levit, 2001; Schieber, Poullier, & Greenwald, 1993, 1994) and the proportion of per-capita gross domestic product (Huber, 1999; Huber & Orosz, 2003; Schieber et al., 1993, 1994).

At the same time, the apparent health worker shortage of the past period gave way to a perceived oversupply (O. Adams, 1989; Budetti, 1981; Bui, 1986; Charatan, 1996; Ellwood & Ellwein, 1981; Ginzberg, 1983; Grayson, 1978; "HEW advisers eye physician 'oversupply','" 1980; Hosokawa & Roberts, 1982; Macleish, 1985; Menken, 1981; Menken & Sheps, 1984; "Most medical staffs have an oversupply of specialists, but don't plan to take action to correct this," 1995; Placone & Wallace, 1982; Riska, 1995; Sabiston, 1984; Salmon & Culbertson, 1985; Schroeder, 1984; Selby, 1981; Tierney, Waters, & Williams, 1980; Whitcomb, 1995; M. S. White & Culbertson, 1981; R. I. White, 1992; Yager, 1987). The focus of research on health workers thus shifted from planning for future supplies to investigating the extent to which health workers contribute to increases in health expenditures.

---

7 In this study, health care spending was adjusted for inflation using the gross domestic product deflator (Huber & Orosz, 2003).

8 While most macroeconomic studies of the income elasticity of health care expenditure estimated values greater than unity (Gertham & Jonsson, 1991; Gertham, Sogaard, Anderson, & Jonsson, 1992; Hitiris, 1997; Hitiris & Posnett, 1992; Milne & Molana, 1991; Newhouse, 1977), the estimates vary substantially and some studies find income elasticities smaller than unity (Parkin, McGuire, & Yule, 1987; Selvanathan & Selvanathan, 1993). Reasons for the divergent results include difference in datasets, econometric models, and approaches to adjusting for purchasing power in studies with cross-country data (Okunade & Suraratdecha, 2000).

9 Claims of health worker shortages or oversupplies imply that quantity standards for health workers have been set. Such quantity standards are difficult to determine and usually quite uncertain (Ginzberg, 1983), as described in the preceding subchapter on health worker planning, making it difficult to validate claims of shortages or oversupplies. For the argument in this background section, however, it is unimportant whether such claims are valid. We merely need to show that policy makers and researchers perceived an oversupply of health workers where previously they had perceived a shortage.
and whether changes in the administration of health workers can help contain costs (Carrin & Hanvoravongchai, 2003; Mehrotra, Dudley, & Luft, 2003; Stone, 1997). Health workers came to be seen as profit-maximizing economic actors individually and as rent-seeking professional groups collectively. Much of the research in this phase was motivated by the belief that health workers do not always act in their patients’ best interest, exploiting market failures in health care for their own gain and reducing social welfare in the process. In health economics research, this belief was exemplified by the intensive study of supplier-induced demand, defined by McGuire in the *Handbook of Health Economics* (McGuire, 2000) as follows:

“Supplier induced demand exists when the physician influences a patient’s demand for care against the physician’s interpretation of the best interest of the patient.”

In commentaries in health policy journals, the belief was expressed more blatantly. “Is the health workforce a barrier to cost containment?” asks Montoya in *The American Journal of Managed Care* (Montoya, 2000) and as Aaron writes in *Health Affairs* (Aaron, 2002)

“Hospital wings echoed with unoccupied beds. Doctors’ garages were filled with Mercedes Benzes. And employers were reeling under rising health care premium costs.”

While the view of health workers as profit-maximizing economic actors was narrow (neglecting, for instance, the influence of such factors as altruism and professional ethics in guiding the actions of nurses and doctors (Le Grand, 1997)), the rigorous application of microeconomic theory to health worker behavior was fruitful, leading to important insights on a range of topics, such as competition among providers, health worker licensure, information asymmetry and health worker agency (Blomqvist, 1991), health worker performance and productivity (Cutler & Berndt, 2001), provider payment and incentives (Sloan & Kasper, 2008), and medical decision making. As described in the introduction section, other chapters in this handbook are devoted to many of these topics (chapters 26, 27, 29, 30, 31, and 33). In the literature review in this subchapter, we thus cover only one topic, health worker licensure, which is not discussed in detail elsewhere in this book.

---

10 Many recent economic models of health worker behavior take a more differentiated view of factors explaining health worker behavior (Encinosa, Gaynor, & Rebitzer, 2007; R. G. Frank & Zeckhauser, 2007; Olbrich, 2008; Schneider & Ulrich, 2008).
III.2 Health worker licensure

Health worker licensure, i.e., the authorization of health workers to practice their profession, has existed since the medieval period (Garcia-Ballester, McVaugh, & Rubio-Vela, 1989; Loewy, 1989). Over time it has become increasingly structured and standardized (Rowe & Garcia-Barbero, 2005). Today, licensure of health workers has been introduced into national or subnational legislation in many countries worldwide (see, for instance, (Gaumer, 1984; Graddy, 1991a; Muzondo & Pazderka, 1980; Rowe & Garcia-Barbero, 2005; The Licentiate Committee of the Medical Council of Hong Kong, 2008; WHO, 2006a)). In order to obtain a health worker license, candidates must have completed formal training in a recognized health care education institution and must usually meet further requirements such as completion of practical training (e.g., an internship), passage of licensure examinations, proof of absence of a criminal record, and swearing of a professional oath (such as the Hippocratic oath or the oath of Geneva). Licensing may be exercised by the Ministry of Health or by independent professional bodies, such as a chamber, order, college or council (Rowe & Garcia-Barbero, 2005). Clearly specified conditions (for instance, malpractice or substance use) can lead to revocation of a health worker license.

Theories of licensure

There are two families of theories to explain the existence and specific patterns of legislation, including health worker licensure (T. G. Moore, 1961; Noether, 1986; Paul, 1984; Posner, 1974). According to “public interest” theories policy makers supply regulation in response to the demand of the public for the correction of a market failure. According to “interest group” theories, on the other hand, policy makers supply regulation in response to the demands of interest groups trying to increase the incomes of their members (Peltzman, 1976; Stigler, 1971). The two families of theories offer different explanations for the existence of health worker licensure and lead to different predictions regarding licensure effects.

---

11 Licensure has to be distinguished from two other types of state regulation of professions, certification and registration (Gaumer, 1984). Certification restricts the use of a particular occupational title to individuals meeting specified requirements, but does not bar individuals who do not meet the requirements from practice in the occupation. Registration simply requires that certain information on a health worker (such as name, address, and place of occupational training) is recorded by a state agency.
a. “Public interest” theories

A common version of the “public interest” explanation for the existence of licensure is “the belief that the public interest will be best served if the poorly trained, incompetent, and unethical people are kept from practicing”, because such exclusions will increase the average quality of health services (Gaumer, 1984). This version is not motivated by an economic model of behavior. Arrow (Arrow, 1963) developed a “public interest” theory of licensure, which rests on the assumption that health workers know their own abilities – and the more capable health workers have higher-paying alternative opportunities for employment. Because patients cannot easily judge the quality of individual health workers, the price of a health service does not differ by health worker and reflects the average quality of the service in the market – i.e., a setup analogous to the one in Akerlof’s classic analysis of “The market for ‘lemons’” (Akerlof, 1970).

Health workers with above-average ability withdraw from this market because their alternative employment opportunities provide them with higher incomes. The withdrawal lowers the average quality of health workers and the prices of health services fall, causing those among the remaining workers whose abilities exceed the new average ability to exit the market. This process repeats itself until only the least capable health workers are left to supply health services. In this analysis, the market fails because of informational asymmetry between health workers and consumers of health services. Searching out information to overcome this asymmetry may be more expensive for the individual consumer than the expected gain in welfare (Leland, 1979). Consumers may find it costly to obtain information on health worker quality because most people consume the services of a particular health worker only rarely and the outcome of health care may not be observable for a long time after service delivery and may depend on factors other than health worker quality (Henderson, McQuire, & Mooney, 1987). In this situation, licensure may serve as a less expensive means of reducing the informational asymmetry between health worker and consumer by screening out those workers who do not meet defined minimum quality standards – in Arrow’s words “[t]he general uncertainty about the prospects of medical treatment is socially handled by rigid entry requirements” (Arrow, 1963).

However, if the only purpose of licensure is to reduce information asymmetry between health workers and consumers, certification – which, unlike licensure, does not exclude anyone from practicing as a health worker – should serve that purpose at least equally well without
reducing consumer choice. While licensure excludes low-quality health workers from professional practice, certification informs consumers about the relative quality of health workers. Several authors have thus argued that the selection of health worker licensure over certification is evidence that health worker interests dominate the public interest in the regulation of health care (Friedman, 1962; Leffler, 1978; Leland, 1979; T. G. Moore, 1961). Against this argument, Svorny (Svorny, 1987) holds that licensure is more effective than certification in ensuring that health workers provide high-quality care because, unlike loss of licensure, loss of certification does not prevent health workers from practicing altogether. A health worker who loses certification (which she may regain) may have to reduce the prices of her services, but will still be able to earn an income providing health care. In contrast, a health worker who loses her license permanently forfeits the entire premium stream derived from the investment in career-specific training. Compared to loss of certification, revocation of licensure should lead to a larger loss of expected net present value (of the future returns on health care education) and thus serve as a stronger incentive against providing low-quality health care than certification.

**b. “Interest group” theories**

“Interest group” explanations for the existence of health worker licensure originate in the belief that licensure has been introduced in response to the demands of interest groups aiming to ensure through licensure that incumbent health workers earn economic rents (Friedman, 1962; Kessel, 1958, 1970; Rayack, 1967). Peltzman (Peltzman, 1976) developed a formal model of the influence of interest groups on regulation (such as licensure legislation), based on a study by Stigler (Stigler, 1971). According to the model, self-interested legislators maximize the expected number of votes in their favor

\[
M = n \cdot f - (N - n) \cdot h,
\]

where \(n\) is the number of potential voters belonging to the interest group that is set to benefit from the regulation, \(f\) is the average probability that a member of the interest group will vote for the legislator, \(N\) is the total number of potential voters, and \(h\) is the average probability that a member of the group which incurs a loss due to the legislation (every non-\(n\)) votes against the legislator (Peltzman, 1976). In this model, the probability of a supportive vote

---

12 Investment in career-specific training includes tuition, living expenses during schooling, and the opportunity cost of time attaining a health care education.
from the beneficiary group \((f)\) is a function of the net benefit per interest group member \(g\) which, in turn, is a function the total dollar amount transferred to the beneficiary group \((T)\), the total amount of dollars spent by the group to mitigate opposition \((K)\), and the cost of organizing the beneficiary group to support the legislator and to mitigate opposition \((C(n))\)

\[
g = \frac{T - K - C(n)}{n}.
\]

The legislator chooses the size of the group she will benefit \((n)\), and offers the group \(T\) in return for \(K\). The probability of opposition \(h\), on the other hand, increases with the rate at which the income of a member of \(N\) is taxed in order to pay for \(T\) and decreases with

\[
z = \frac{K}{(N - n)}.
\]

Note that, assuming diminishing returns to the per-capita net benefit, the legislator who maximizes \(M\) will benefit both \(n\) and \((N - n)\) to some extent. Peltzman’s model further predicts that the cost of interest group organization \(C(n)\) and “imperfect information about both the gains and losses of regulatory decisions” restrict the size of the group benefiting from the regulation. Furthermore, “the costs of using the political process” limit the size of the gains of the beneficiary group (Peltzman, 1976). Based on the “interest group” theories of regulation, we would expect health worker groups to be successful in increasing their rents through regulation, because – in comparison to many other occupational groups – they are small, well-informed, and well-organized in professional organizations (Friedman, 1962; Kessel, 1970; Peltzman, 1989; Stigler, 1971).

**Literature on health worker licensure**

Three categories of health worker licensure studies can be distinguished. The first category examined whether health care markets are competitive or monopolistic. In this literature, monopolistic market structures are taken as evidence that licensure has benefited health

\[13\] Using a setup similar to Peltzman’s, Becker (Becker, 1983) developed another version of an “interest group” theory of regulation in order to investigate the effect of deadweight loss on the size of the gain that the beneficiary group can achieve through the regulatory process. The deadweight loss is the difference between the gains of the beneficiary group and the losses of the public due to regulation. Becker’s model predicted that deadweight loss is a constraint on the gains of the beneficiary group. Because deadweight loss increased at an increasing rate as regulation moves output away from the efficient level, the beneficiary group would need to exert increasing marginal pressure in order to overcome opposition to the regulation. The model further predicted that deadweight loss motivates a search for greater efficiency in regulatory redistribution because neither the beneficiary nor the loser group opposes changes that eliminate some deadweight loss.
workers and hurt consumers. The second category of licensure studies analyzed the effects of changes in licensure regulation on health worker incomes or on the quantity, price, or quality of health services. Studies in this category infer from relationships between these outcomes and the stringency of licensure requirements whether “interest group” or “public interest” explanations of licensure hold true. The third category of studies investigated whether the political power of health worker interest groups is a determinant of licensure legislation. A positive relationship between interest group power and licensure legislation is taken as evidence that the interests of health workers have dominated those of consumers in influencing licensure legislation. Below, we describe the analytical approaches that have been employed in each of the three categories. For each approach, we review the findings of studies and critically discuss the evidence on reasons for the existence of health worker licensure.

a. First category of licensure studies: market structure

One approach to determining the level of competitiveness in the health care market has been to estimate the return to health worker education and to compare it to the returns to education in other occupations. Studies employing this approach generally found significant positive returns to medical education. However, depending on the time period of observation, the particular physician specialties, the comparison occupations, and the estimation technique, the estimated returns to education varied widely (Burstein & Cromwell, 1985; Friedman & Kuznet, 1945; Leffler, 1978; Leffler & Lindsay, 1980; Lindsay, 1973; Marder & Willke, 1991; Maurizi, 1975; Weeks, Wallace, Wallace, & Welch, 1994; Wilkinson, 1966). Moreover, a number of estimation problems were identified in these studies. First, the results were sensitive to the choice of discount rate and there is no consensus as to which discount rate to use to estimate returns to medical education (Leffler, 1978). Second, the studies commonly failed to take into account progressive taxation (Leffler, 1978) and the cost of malpractice insurance. Third, many of the earlier studies of return to medical education did not control for differences in hours worked across occupations (Lindsay, 1973). Where physicians earn more but work longer hours than comparison occupations, adjustment for hours worked reduces cross-occupation differences in rates of return. Newer studies took this criticism into account. Burstein and Cromwell (1985) found hours-adjusted internal rates of return (IRR) to education for all physicians of about 12 percent in 1980 (Burstein & Cromwell, 1985); Marder and Willke estimate hours adjusted IRRs to vary across specialty
between about 2 percent for Pediatrics to 58 percent for Pathology (Marder & Willke, 1991); and Weeks and colleagues calculated hours adjusted IRRs of about 16 percent for primary care physicians and 21 percent for specialist physicians, while the comparison occupations received similar or higher returns on their education (dentists: 21 percent; attorneys: 25 percent; business people: 29 percent) (Weeks et al., 1994). Thus, studies since 1985 that calculated hours-adjusted IRR suggest that for many medical specialties the stringent licensure requirements did not led to excessive returns.

Another approach to assess whether health care markets are competitive or monopolistic has been to investigate the relationship between physician density per population and the price of physician services. A positive association between density and price was interpreted as evidence of a monopolistic market (Newhouse, 1970) or as evidence that physicians induce demand for their services (Cromwell & Mitchell, 1986; R. G. Evans, Parish, & Sully, 1973; Fuchs, 1978; Tussing & Wojtowycz, 1986). However, as Sloan and Feldman (Sloan & Feldman, 1978) and Newhouse (Newhouse, 1978) pointed out, the inability to control for differences in quality of physician services in these studies rendered their results of little explanatory value.

“Recall that analysis of a competitive market assumes a fixed and homogeneous product. But physicians in areas with greater concentration of physicians spend more time with their patients and patients in such areas spend less time waiting for their physicians. More time with the physicians and less waiting time are both desirable characteristics that lead to more physician time per visit. In a competitive market, visits using more physician time would be more expensive. As a result, even if the market were competitive, there could be a positive relationship between the physician/population ratio and the price of a (non-homogeneous) physician visit” (Newhouse, 1978).

A third approach has been to estimate an index of market power based on comparison of marginal revenue to marginal cost in the health services industry (Bresnahan, 1982; Panzar & Rosse, 1987). Wong (Wong, 1996) did not find evidence that the marginal revenue exceeds marginal costs in the market for physician services in the US, while Seldon, Jung and Cavazos (Seldon, Jung, & Cavazos, 1998) observed that the price for physician services was maintained above marginal costs. The results regarding physician market power from studies following this approach are thus inconclusive.
In sum, the studies in this category do not provide decisive evidence for monopolistic structures in health care markets. Moreover, none of the studies took consumer benefits into account. It is possible that consumers are better off in a less competitive market than in a more competitive one.\(^{14}\) For instance, licensure legislation may restrict entry of health workers into the market and at the same time improve the quality of health care. If the outward shift in demand caused by the latter effect exceeds the inward shift in supply caused by the former, consumer welfare will increase. Even if the studies in this category had provided strong evidence for monopolistic structures in health care markets, this evidence would have been insufficient to rule out the possibility that such structures were serving the public interest.

*Second category of licensure studies: licensure effects*

A second category of licensure studies has investigated the effects of changes in licensure legislation on outcomes in order to establish whether “public interest” or “interest group” explanations of licensure hold true. Outcomes considered in this category of studies included health worker incomes and the price, quality and quantity of health services. A number of studies found that health workers’ incomes (G. M. Anderson, Halcoussis, & Lowenberg, 2000; Friedman & Kuznet, 1945; Leffler, 1978; Muzondo & Pazderka, 1980; W. D. White, 1978) or the price of health services and products (Benham & Benham, 1975; Shepard, 1978) increased when licensure regulations became stricter. The studies inferred from the results that licensure benefited health workers and not consumers. Yet, these studies suffer from the same fundamental weakness that studies investigating the level of competitiveness in the health services market suffer from: Their results do not rule out that the quality-improving effect of licensure legislation has shifted health services demand to such an extent that aggregate consumption and consumer welfare increase despite the supply-restricting effect of the legislation (Svorny, 1987). Hence the results cannot be used to determine whether licensure legislation benefited the public. Two other approaches in this category of studies are better suited for this purpose.

\(^{14}\) Conversely, lack of evidence of monopolistic structures does not imply that licensure was established to protect the public. As Svorny (1987) points out, “there must have been a one-time gain to practicing physicians when restrictions on practice of medicine were imposed. As in other protected markets, entry may dissipate any rents over time.”
First, it is theoretically possible to exclude the possibility that health worker licensure benefited consumers through analysis the quality of health services. If health services quality did not improve in response to licensure regulation, it could be ruled out that the legislation benefited the consumers. (In contrast, if health services quality did improve in response to licensure legislation, additional information would be necessary in order to determine whether the legislation increased consumer welfare, because the quality improvement may or may not have been sufficient in size to compensate for a reduction in consumer welfare due to supply-restricting effects of licensure.)

Three studies investigated the effect of licensure on health service quality. Carroll and Gaston (Carroll & Gaston, 1981) found that some licensure requirements for dentists in the US led to a reduction in the quality of dental services; Feldman and Begun (Feldman & Begun, 1985) found that licensure legislation in the US increased the quality of optometric services; and Haas-Wilson (Haas-Wilson, 1986) did not find any significant effect on optometric service quality of four state restrictions on optometric practice which, if violated, could have led to revocation of optometric license. The three studies have thus come to very different conclusions regarding the effect of licensure on health care quality. More important, it is highly unlikely that the quality measures used in the three studies captured all, or even the most important, dimensions of health services quality. Carroll and Gaston (Carroll & Gaston, 1981) measured quality as waiting time for dental appointments, while Feldman and Begun (Feldman & Begun, 1985) measured it as length and thoroughness of eye exams, and Haas-Wilson (Haas-Wilson, 1986) measured it as the thoroughness of eye exams. The studies thus analyzed the effect of licensure legislation on some specific indicators of process quality, while ignoring other quality dimensions (such as outcome quality). However, in order to rule out the possibility that health worker licensure legislation benefited consumers, information on the effects of licensure on all quality dimensions is necessary.

Second, analysis of the effect of licensure legislation on equilibrium consumption can be used to determine whether consumers benefited from such legislation. If licensure legislation caused an outward shift in the demand for health services that is greater than the inward shift in supply, equilibrium consumption would have increase and consumers would have benefited from the legislation; if, on the other hand, the outward shift in demand were smaller

---

15 We use the word ‘quality’ here to mean all possible dimensions of health service quality as well as increases in certainty about the level of quality in the market for health services.
than the inward shift in supply, equilibrium consumption would have decreases and consumer welfare would have been reduced.16

Svorny (Svorny, 1987) investigated the effect of two particular medical licensure regulations (basic science certification and citizenship) on the consumption of physician services. She found that, controlling for a range of socioeconomic, population, and health system characteristics, more stringent licensure requirements in a US state were associated with lower numbers of physicians per population in the state, and concluded that “the interests of organized medicine dominated those of consumers in influencing the medical regulatory supply process” (Svorny, 1987). Adams, Ekelund and Jackson examined the effect of a range of variables capturing the stringency of midwife licensure and other midwifery regulations on the consumption of the services of certified nurse-midwives (CNM) (Adams 2003). They found that more stringent midwifery regulations in a US state led to lower consumption of midwifery services in the state and concluded that “supply-restricting effects dominated quality assurance in the U.S. market for CNM services” (A. F. Adams, Ekelund, & Jackson, 2003).

Both studies have technical limitations. Svorny (Svorny, 1987) assumed that her outcome variable, physician population density by US state, was a measure of the quantity of physician services consumed in the states. However, it is plausible that physicians in states with lower physician population densities provided on average more services per year than physicians in states with higher densities. The conclusion that licensure decreases equilibrium consumption of health services may thus be wrong. Adams, Ekelund and Jackson (A. F. Adams et al., 2003) controlled for endogeneity between their outcome variable (the quantity of midwife services) and the licensure and regulation variables using a number of instrumental variables. However, the instrumental variables (hospital charges for an uncomplicated vaginal delivery, number of CNM per-capita, physician deliveries as a percentage of total deliveries, total population in the state) may be invalid because they are likely to have affected the outcome variable through many pathways other than the licensure and regulation variables. The coefficients in the instrumental variable regressions may thus be biased.

---

16 This identification of the net effect of licensure legislation on consumer benefits is possible because we assume that such legislation never leads to an outward shift in supply. If this assumption were not met, consumer welfare could decrease when equilibrium consumption increases.
In sum, analyses of the licensure effect on health worker income or the price of health services cannot determine whether licensure served the public interest. Analyses of the licensure effect on health service quality can theoretically rule out that licensure benefited the public, but past studies were not able to demonstrate that licensure was in the public interest (and it seems unlikely that future studies will) because of practical difficulties in measuring comprehensively all dimensions of health service quality that can be affected by licensure legislation. The most promising approach within the second category of studies is to evaluate the licensure effect on the quantity of health services. The results of the two studies that used this approach suggest that licensure reduced social welfare, but technical issues diminish the strength of this evidence.

b. Third category of licensure studies: political power

A third category of licensure studies has analyzed the relationship between the political power of health worker interest groups and the passage of licensure legislation and other occupational regulation. Variables used to proxy interest group power were found to be positively associated with passage of licensure legislation for physicians, nurses, dentists, and pharmacists (Stigler, 1971) and physician assistants and psychologists (Graddy, 1991b), as well as the stringency of licensure regulation for dentists (Becker, 1986) and the stringency of other occupational regulation for optometrists (Begun, Crowe, & Feldman, 1981). The authors of these studies interpreted their results as evidence that “interest group” explanations of licensure hold true. However, it is unclear whether the proxy variables they used really measured the political power of health worker interest groups. For instance, Stigler (Stigler, 1971) and Becker (Becker, 1986) used the percentage of an occupation living in urban areas as a measure of political power. As Stigler argued, an occupation that is more concentrated in cities will be more powerful: “When the occupation organizes a campaign to obtain favorable legislation, it incurs expenses in the solicitation of support, and these are higher for a diffused occupation than a concentrated one” (Stigler, 1971). However, the concentration of health workers in cities may not matter for the political power of occupations in countries where most members of the occupation have access to communication technologies, such as mail or telephone, so that political support can be organized independent of physical distance between health workers.

Graddy (Graddy, 1991b) assumed that the political power of health worker groups (physician assistants and psychologists) increases with the ratio of group members “belonging to the
major professional association” to the total number of employed persons, and Begun, Crowe, and Feldman (Begun et al., 1981) assumed that the political power of optometrists increases with the ratio of optometrists to their competitors (ophthalmologists and opticians).

However, political power may decrease in these ratios beyond some point. Ceteris paribus, the ratios increased with group size. However, as group size increases, the cost of organizing political support among group members may rise, reducing the group’s political power (Peltzman, 1976):

“It is not enough for the successful group to recognize its interests; it must organize to translate this interest into support for the politician who will implement it. This means not only mobilizing its own vote, but contributing resources to the support of the appropriate political party or policy: to finance campaigns, to persuade other voters to support or at least not oppose the policy or candidate, perhaps occasionally to bribe those in office. While there may be some economies of scale in this organization of support and neutralization of opposition, these must be limited. The larger the groups that seeks the transfer, the narrower the base of the opposition and the greater the per-capita stakes that determine the strength of opposition, so lobbying and campaigning costs will rise faster than group size. The cost of overcoming “free riders” will also rise faster than group size.”

Note also that a number of variables hypothesized to proxy political power of interest groups in these studies were found not to be associated with licensure legislation and other regulation (for instance “the ratio of the occupation to the total labor force” in Stigler’s 1971 study (Stigler, 1971) and “the percentage of optometrists who belong to their state optometry association” and “whether or not the state’s optometry association has a legislative lobbyist” in the 1981 study by Begun, Crowe, and Feldman (Begun et al., 1981)). The absence of significant relationships between these variables and occupation regulation suggests that the variables either did not measure political power or that certain dimensions of political power did not influence occupational regulation.

Another problem with the above analysis belonging to the third category of licensure studies is that they did not control for reverse causality, so that their estimates of the effect of political power of interest groups on licensure legislation may be biased. It seems likely that in many cases the introduction of licensure or an increase in licensure stringency will reduce the costs of organizing political support and thus strengthen health workers’ political power. Licensed health workers are usually centrally registered (for instance, with a professional
chamber or council) and can thus be easily contacted in a political campaign. In addition, licensure defines the scope of practice of different types of health worker groups, ensuring a certain level of homogeneity of interests within the groups. Furthermore, the more stringent the licensure requirements, the smaller the size of the group of licensed health workers and the easier the political organization of the group.

In sum, while the literature on the relationship between political power of health worker interest groups and licensure suggests that political power does determine licensure legislation – i.e., that “interest group” explanations of the existence of licensure hold true – the studies in this category suffer from a number of limitations. In particular, it is unclear whether the proxy variables for political power chosen in these studies are valid and the analyses failed to control for reverse causality in the relationship between political power and professional licensure.

To sum up, we identify three categories of licensure studies: first, studies examining the level of competitiveness in the market for health services; second, studies analyzing the effect of changes in licensure legislation on health worker incomes or quantity, price, or quality of health services; and third, studies investigating the influence of political power of interest groups on the passage of licensure legislation. While several studies suggest that “interest group” explanation of the existence of licensure hold true, the evidence is not conclusive. Most of the studies we reviewed merely show that health workers have benefited from licensure legislation (studies in the first and second category) or that health worker groups have had some influence on licensure legislation (studies in the third category), but fail to demonstrate that consumers suffered a welfare loss. Thus, these studies cannot rule out that licensure legislation served the public interest. The studies of licensure effect on quality or quantity of health services are an exception. A decrease in the quality of health services rules out an increase in consumer welfare and an increase (decrease) in the quantity of health services consumed implies an increase (decrease) in consumer welfare. However, the few studies that investigated the effect of licensure on the quality or quantity of health services suffer from technical limitations in measurement of variables and effect estimation, so the evidence – again – is not conclusive.

An important shortcoming of the literature on health worker licensure is that almost all published studies were conducted in the US, even though health worker licensure regulations exist in most developed and developing countries. The US health services market has
characteristics that are rarely found in other countries. Some of these characteristics are likely to influence the effect of licensure on market outcomes. For instance, in the US, health workers may already have strong incentives to provide high-quality health services because of the high costs associated with health care malpractice suits. The marginal effect that licensure can have on market outcomes may thus be lower in the US compared to settings where malpractice suits are less expensive. Future studies need to investigate causes and effects of licensure regulation in other countries.

Two further criticisms apply to studies in the first and the second category. First, the real effects of licensure may differ from the effects that legislators or interest groups intend to achieve. Changes in licensure regulation are infrequent events. Legislators and members of interest groups will thus have little opportunity to observe the effects of licensure laws and to learn from failures. Hence, inferences about the behavior of actors in the regulatory process based on observed effects of regulation (as done in many of the studies discussed above) may be invalid. A case in point is a study on the regulation of physiotherapists in the US. Sass and Nichols (Sass & Nichols, 1996) found “that physical therapists have devoted considerable effort to obtain passage of laws that ultimately result in lower wages”. As an explanation of “this seemingly perverse result” the authors consider that physical therapists incorrectly believed that the professional regulation would increase their earnings (Sass & Nichols, 1996).

Second, analyses belonging to the first and the second category of licensure studies are concerned with structures and outcomes in the market for health services. Health services, however, are merely instruments to attain better health. It is therefore possible that consumers suffered a welfare loss in the market for health services, while overall consumer welfare increased. For instance, licensure regulation may have prevented incompetent health workers from providing a health service intended to reduce childhood mortality. The resulting inward shift in supply (due to the exclusion of health workers from the market) may have exceeded the outward shift in demand (due to the increase in average quality of the service), so that consumers suffered a welfare loss in the market for that health service. Nevertheless, more children may have survived (since the fewer children who did receive the service received a better service) and overall consumer welfare may thus have increased. In measuring the effect of licensure on social welfare, future studies should thus include the value of changes in population health outcomes due to licensure regulation.
IV. Third phase: The health worker as necessary resource

IV.1 Background

In the 1990s and 2000s, the lack of health workers became a major research focus in developed and developing countries. In contrast to the first phase (see above), health worker shortages in developed countries were local rather than nationwide, occurring in so-called medically underserved areas, which were most commonly rural or remote areas but included also certain areas in cities where predominantly poor people seek health care. In 2002, Brooks and colleagues described health worker maldistribution and the resulting health worker shortages in underserved areas in the US (Brooks, Walsh, Mardon, Lewis, & Clawson, 2002):

“While 20% of the population of the United States resides in rural areas, less than 11% of physicians practice in rural communities. Despite the fact that the supply of physicians has increased over the last 20 years, the percentage of physicians practicing primary care specialties has declined. Although the overall number of physicians practicing in rural areas has increased, the proportion relative to urban areas continues to decline. In sum, it continues to be difficult to attract and retain physicians in rural areas, particularly primary care physicians.”

Six years later, Rabinowitz and colleagues repeated, in essence, Brooks’ problem statement (Rabinowitz, Diamond, Markham, & Wortman, 2008):

“The persistent shortage of physicians in rural areas continues to have a major impact on access to care for those living in small communities. Although one of every five Americans (20%) lives in a rural area, only 9% of physicians practice there. The scope of this problem is substantial, because more than 20 million of the 60 million people residing in rural areas live in federally designated Health Professional Shortage Areas (HSPAs). This rural physician shortage has existed for more than 80 years, despite the fact that, in general, people living in rural areas have a greater need for medical care, being older, sicker, and poorer than their nonrural peers. Of greater concern, the future rural physician workforce is likely to decline

---

17 A medically underserved area is an area where the number of health workers falls below a target. As described in section II., such targets can be based on need, demand, or supply criteria (such as service targets or population ratios). Commonly, a mix of criteria is used in the definition of underserved areas (Bärnighausen & Bloom, 2009b).
even further, with only 3% of recent medical students planning to practice in small towns and rural areas.”

In developing countries, evidence emerged that the population density of health workers affected both population health outcomes (such as child mortality (Anand & Bärnighausen, 2004)) and health systems outcomes (such as childhood vaccination coverage (Anand & Bärnighausen, 2007)), emphasizing the need to increase health worker education and retention in developing countries (Joint Learning Initiative, 2004; WHO, 2006b). In addition, with the adoption of the MDGs by national governments and international organizations, health became a major focus of development policies. In particular, the fourth MDG (“reduce child mortality”), the fifth MDG (“improve maternal health”), and the sixth MDG (“combat HIV/AIDS, malaria and other diseases”) emphasized population health improvements as primary goals of development (United Nations, 2009). While these goals do not depend on health services delivery alone, they are unlikely to be achieved without substantial increases in service coverage in developing countries. The indicators used to monitor the progress towards the health MDGs include health service targets, e.g., the proportion of one-year old children immunized against measles (for the fourth MDG), the proportion of births attended by skilled health personnel (for the fifth MDG), and the proportion of population with advanced HIV infection with access to antiretroviral drugs (for the sixth MDG) (United Nations, 2009). It soon became apparent that such health service targets could not be achieved without substantially increasing the health workforce in developing countries (Joint Learning Initiative, 2004; WHO, 2006b, 2008). Access to antiretroviral treatment (ART) is a case in point. A number of studies modeling the need for health workers to provide universal access to antiretroviral treatment (ART) showed that at current health worker education and migration rates in developing countries it is very unlikely that sufficient health workers will be available to treat most patients needing ART in the foreseeable future (Bärnighausen & Bloom, 2009a; Bärnighausen, Bloom, & Humair, 2007). Even at current levels of relatively low ART coverage, many developing countries are struggling to find the health workers necessary to further expand coverage (Clark, 2006; IOM, 2007; Kober & Van Damme, 2004; MSF, 2007; Ooms, Van Damme, & Temmerman, 2007).

Of course, increases in the efficiency of health care delivery, e.g. through changes in health worker team composition and task shifting from more to less highly educated health workers, could decrease the gap between current health worker numbers and the numbers required to achieve the health MDGs. However, it is very unlikely that efficiency increases can eliminate the gap (WHO, 2008).
The shift in perspective from regarding health workers as economic agents to viewing them as a resource necessary to improve population health in underserved areas of developed and developing countries led to research on means to increase health worker supplies to such areas. In the following, we review two major topics in this third phase of research on the health workforce: first, programs aimed at attracting health workers to underserved areas in developed countries; and, second, health worker emigration from developing to developed countries (Aiken, Buchan, Sochalski, Nichols, & Powell, 2004). While the supply of health workers to underserved areas and regions is not only a function of health worker movements, but also of education rates (Bärnighausen & Bloom, 2009a), studies on health worker education are few (Eckhert, 2002) and thus do not warrant a review.

**IV.2 Developed countries: interventions to increase the supply of health workers to underserved areas**

Programs to increase the supply of health workers to underserved areas can affect health workers in different stages of their careers. Before the start of training to become a health worker, selective admission strategies attempt to increase the number health workers who will practice in underserved area by selecting those individuals into health care training who – given observable characteristics – are most likely to work in underserved areas after graduation. During training, curricula specific to health needs in underserved areas and exposure to practice in such areas attempt to increase graduates’ likelihood of choosing underserved practice by specifically preparing them for this type of service. Lastly, financial-incentive programs offer scholarships (during training) or loan repayments (after training) in return for service in underserved areas (Bärnighausen & Bloom, 2008). Table 3 shows an overview of studies extracted from three recent systematic reviews of four types of programs (selective admission to medical school, medical school training for practice in underserved areas, residency training for practice in underserved areas (Brooks et al., 2002; Rabinowitz et al., 2008), and financial-incentive programs for return of service in underserved areas (Bärnighausen & Bloom, 2009b)), supplemented by studies identified through additional literature searches by the authors of this chapter using the PubMed and Econlit databases.

We excluded from the overview in Table 3 studies that merely described program results and included only those studies that compared outcomes in program participants with outcomes in non-participants. All included studies measured an outcome in either one of two categories: provision of care – if the study compared all health workers enrolled in a program to all health workers – or retention – if the study compared only those enrolled and non-enrolled
health workers who at some point took up practice in an underserved area. Two types of retention outcomes can be distinguished: retention in the same underserved area (i.e., a health worker leaving a particular underserved area to work in another underserved area counts as non-retention) and retention in any underserved (i.e., a health worker leaving a particular underserved area to work in another underserved area counts as retention).

Our extraction and literature searches identified only one study reporting the effect of selective admission to medical school without additional medical school training to prepare students for practice in underserved areas (Basco, Buchbinder, Duggan, & Wilson, 1998). The study found – unsurprisingly – that students attending medical schools with selective admission targeting future generalists were significantly more interested in primary care and in rural practice than students attending medical schools that did not have such a selective admission policy. Other studies investigated programs that trained (future) health workers specifically for practice in underserved areas either during medical school (without or without an explicit selective admission policies) or during residency. These studies found that participants in these programs were significantly more likely than non-participants to practice in underserved areas (Bowman & Penrod, 1998; Brazeau, Potts, & Hickner, 1990; Fryer, Stine, Krugman, & Miyoshi, 1994; Rabinowitz, 1993; Rabinowitz, Diamond, Markham, & Hazelwood, 1999b; Rabinowitz, Diamond, Markham, & Paynter, 2001; Smucny, Beatty, Grant, Dennison, & Wolff, 2005), to intend to practice in underserved areas (Rosenthal, 2000), and to remain in the same underserved area (Pathman, Steiner, Jones, & Konrad, 1999; Rabinowitz, Diamond, Markham, & Rabinowitz, 2005) – with only one exception, which did not find any significant difference between participants and non-participants (Rabinowitz, Diamond, Hojat, & Hazelwood, 1999a).

Five of seven studies investigating the effect of participation in financial-incentive programs on retention in the same underserved found that participants were significantly less likely to remain in the same underserved area (G. M. Holmes, 2004; Pathman, Konrad, King, Taylor, & Koch, 2004; Pathman, Konrad, & Ricketts, 1992, 1994a, b), while one study did not report a significance level but found substantially lower retention among participants (Singer, Davidson, Graham, & Davidson, 1998), and another study did not find a significant difference (Jackson, Shannon, Pathman, Mason, & Nemitz, 2003).

On the other hand, 10 of the 12 studies investigating the effect of financial-incentive programs on provision of care or retention in any underserved area found that participants
were more likely to provide care to underserved populations (Brooks, Mardon, & Clawson, 2003; Inoue, Matsumoto, & Sawada, 2007; Matsumoto, Inoue, & Kajii, 2008; Pathman, Konrad, King, Spaulding, & Taylor, 2000; Probst, Samuels, Shaw, Hart, & Daly, 2003; Rabinowitz, Diamond, Veloski, & Gayle, 2000; Rittenhouse, Fryer, Phillips, Miyoshi, Nielsen, Goodman et al., 2008; Xu, Fields, Laine, Veloski, Barzansky, & Martini, 1997a; Xu, Veloski, Hojat, Politzer, Rabinowitz, & Rattner, 1997b) and to continue to practice in some underserved area (G. M. Holmes, 2004; Pathman et al., 2000). These differences were shown to be statistically significant in eight of the ten studies (Brooks et al., 2003; G. M. Holmes, 2004; Pathman et al., 2000; Probst et al., 2003; Rabinowitz et al., 2000; Rittenhouse et al., 2008; Xu et al., 1997a; Xu et al., 1997b), while two studies did not provide the results of significance tests (Inoue et al., 2007; Matsumoto et al., 2008) and two other studies reported that participants were significantly less likely than non-participants to remain in some underserved area (Matsumoto et al., 2008; Pathman et al., 1992).

In all of the above studies, the (future) health workers chose to participate in a program. It is thus difficult to distinguish whether the finding that program participants are more (or less) likely to provide care in underserved areas or to remain in an underserved area is a true effect of program participation or merely a selection effect. Selection issues are particularly apparent in the evaluation of programs with selective admission policies. All of the studies evaluating such programs compare participating health care students to their non-participating peers. However, such peers are unlikely to be a good counterfactual, because they will on average be less likely than the participants to have the characteristic required for participation in the program, e.g., rural upbringing or self-declared interest in primary health care. As the selection criterion was likely chosen because it was observed to be positively associated with practice in underserved areas, we would expect that observed differences in the likelihood to practice in underserved areas between participating and non-participating health care students can be (partially) explained by selective admission.

Indeed, the one study that does not find any significant difference in the provision of care in underserved areas between participants and non-participants in a program with selective admission (and medical school training for rural practice) controlled for rural upbringing in the analysis (Rabinowitz et al., 1999a) – while the program “recruits and selectively admits medical school applicants who have grown up in a rural area and intend to practice family medicine in rural and underserved areas” (Rabinowitz et al., 1999b).
Another study of the same program concludes that “the admission component of the PSAP [Physician Shortage Area Program] is the most important reason for its success” (Rabinowitz et al., 2001). Implicit in such a conclusion is the assumption that a large proportion of participants would not have studied to become health workers, had they not enrolled in the program. If this assumption were not met, the program would not have increased the supply of health workers to underserved areas and could thus not be called a “success”.

Of the 16 studies of program effect in Table 3, 14 controlled for additional variables in the comparison between participants and non-participants, such as sex, age, ethnicity, marital status, medical specialty (Bowman & Penrod, 1998; G. M. Holmes, 2004; Pathman et al., 2000; Pathman et al., 2004; Pathman et al., 1992, 1994b; Probst et al., 2003; Rabinowitz et al., 1999a; Rabinowitz et al., 1999b; Rabinowitz et al., 2001; Rabinowitz et al., 2000; Rittenhouse et al., 2008; Xu et al., 1997a; Xu et al., 1997b). However, even those studies that control for factors likely to be closely related to care provision and retention in underserved areas (such as growing up in an underserved area (Rabinowitz et al., 1999a; Rabinowitz et al., 2000; Xu et al., 1997a; Xu et al., 1997b) or “strong interest” prior to medical school to practice as a doctor in an underserved area (Rabinowitz et al., 2000; Xu et al., 1997b)) cannot rule out that participants selected into the program on unobserved characteristics related to the preference of working in underserved areas.

Only one study of program effect attempted to control for selection on unobserved variables by using a selection model (G. M. Holmes, 2004). In order to identify the program effect, the study used four medical school characteristics: “historical proportion of graduates specializing in primary care”, “quality of the school”, a “tuition index”, and a “public school indicator”. The use of these identifying variables assumes that they influenced selection into the program but did not affect the provision of care in underserved areas other than through their effect on program participation. However, the type of medical school that students attend is likely to be related not only to the decision to enroll in financial-incentive programs, but also – independent of program participation – to the decision to work in underserved areas. For instance, students with strong preferences to work in underserved areas may be more likely than their peers with weaker preferences for such care to select medical schools with a high “historical proportion” of graduates pursuing careers in primary care, because such schools are likely to focus on medical education relevant for underserved areas. This selection may determine work location decisions, independent of any effect the medical
school characteristic may have on participation in the programs. Thus the characteristics may not be valid variables to identify program effects.

In sum, studies evaluating programs aimed at increasing the supply of health workers to underserved area find that participants are more likely to provide care for underserved population and to remain in underserved areas in the long run (even if not in the same underserved area where they were initially placed). However, because the studies to date have not convincingly controlled for selection effects, the evidence to date does not allow the inference that the programs have caused increases in health worker supply to underserved areas.
### Table 3: Studies of programs aimed at increasing the supply of health workers to underserved areas

<table>
<thead>
<tr>
<th>Study</th>
<th>Program type</th>
<th>Program</th>
<th>Outcomes</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Basco et al., 1998)</td>
<td>Selective admission to medical school</td>
<td>All US medical schools “with premedical recruitment activities that targeted future generalists”</td>
<td>Self-reported interest in primary care and rural practice</td>
<td>Students attending medical schools with selective recruitment targeting future generalists were significantly more interested in primary care and in rural practice than students attending other schools.</td>
</tr>
</tbody>
</table>
| (Rabinowitz et al., 2001) | Selective admission to medical school  
Medical school training for practice in underserved areas  
Financial incentives | Physician Shortage Area Program (PSAP) of Jefferson Medical College (selective admission and medical school training for practice in underserved areas)  
Rural preceptorship  
NHSC (financial incentives) | Provision of care in any underserved area                                   | PSAP participation significantly increased the likelihood of a graduate of Jefferson Medical College graduates to practice rural primary care (adjusted odds ratio (aOR) 2.5, p < 0.001) net of rural preceptorship during medical school, NHSC participation, sex, expected peak income, rural family practice clerkship. Holding the other factors equal rural preceptorship and NHSC participation significantly increased the likelihood of rural primary practice (aORs 2.4 and 2.6, respectively, both p < 0.001). |
<table>
<thead>
<tr>
<th>Study</th>
<th>Program type</th>
<th>Program</th>
<th>Outcomes</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Rabinowitz et al., 1999a)</td>
<td>Selective admission to medical school</td>
<td>PSAP</td>
<td>Provision of care in any underserved area</td>
<td>PSAP participation was not significantly related to practice in rural area, when controlling for “size of community raised”, “size of community undergraduate college”, “father’s education”, “specialty plans as a Freshman”, “senior debt”.</td>
</tr>
<tr>
<td></td>
<td>Medical school training for practice in underserved areas</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Rabinowitz et al., 2005)</td>
<td>Selective admission to medical school</td>
<td>PSAP</td>
<td>Retention in the same underserved area</td>
<td>“After 11–16 years, 68% (26/38) of the PSAP graduates were still practicing family medicine in the same rural area, compared with 46% (25/54) of their non-PSAP peers (p = .03).”</td>
</tr>
<tr>
<td></td>
<td>Medical school training for practice in underserved areas</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Rabinowitz, 1993)</td>
<td>Selective admission to medical school</td>
<td>PSAP</td>
<td>Provision of care in any underserved area</td>
<td>“PSAP graduates from the classes of 1978 through 1986 were approximately four times as likely as non-PSAP graduates to practice family medicine (55 percent vs. 13 percent), to practice in a rural area (39 percent vs. 11 percent), and to practice in underserved areas (33 percent vs. 8 percent). They were approximately 10 times more likely to combine a career in family medicine with practice in a rural (26 percent vs. 3 percent) or underserved (23 percent vs. 2 percent) area.”</td>
</tr>
<tr>
<td></td>
<td>Medical school training for practice in underserved areas</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Program type</td>
<td>Program</td>
<td>Outcomes</td>
<td>Results</td>
</tr>
<tr>
<td>-------</td>
<td>--------------</td>
<td>---------</td>
<td>----------</td>
<td>---------</td>
</tr>
<tr>
<td>(Rabinowitz et al., 1999b)</td>
<td>Selective admission to medical school Medical school training for practice in underserved areas</td>
<td>PSAP</td>
<td>Provision of care in any underserved area</td>
<td>“PSAP graduates were much more likely than their non-PSAP classmates at JMC [Jefferson Medical College] to practice in a rural area of the United States (34% vs 11%; RR, 3.0), to practice in an underserved area (30% vs 9%; RR, 3.2), to practice family medicine (52% vs 13%; RR, 4.0), and to have combined a career in family practice with practice in a rural area (21% vs 2%; RR, 8.5).”</td>
</tr>
<tr>
<td>(Smucny et al., 2005)</td>
<td>Medical school training for practice in underserved areas</td>
<td>Rural Medical Education Program (RMED) of the State University of New York (SUNY)</td>
<td>Provision of care in any underserved area</td>
<td>“A greater percentage of former RMED students practiced in rural locations [22/86 (26%)] than did non-RMED students [95/1,307 (7%)]” (p &lt; 0.0001).</td>
</tr>
<tr>
<td>(Brazeau et al., 1990)</td>
<td>Medical school training for practice in underserved areas</td>
<td>Upper Peninsula Medical Education Program at Michigan State University</td>
<td>Provision of care in any underserved area Specialty choice of medical school graduates</td>
<td>Program graduates “chose rural practice and primary care specialties, especially family practice, more often than did their downstate colleagues” (whose medical school did not emphasize training for practice in underserved areas).</td>
</tr>
<tr>
<td>(Fryer et al., 1994)</td>
<td>Medical school training for practice in underserved areas</td>
<td>Statewide Education Activities for Rural Colorado’s Health (SEARCH) at the University of Colorado Medical School (UCSOM)</td>
<td>Provision of care in any underserved area Specialty choice of medical school graduates</td>
<td>UCSOM graduates who had participated in SEARCH were more likely to have chosen a primary care specialty (46.9% vs. 30.6%) and to establish practices in rural counties (16.4% vs. 9.6%) than graduates who had not participated.</td>
</tr>
<tr>
<td>Study</td>
<td>Program type</td>
<td>Program</td>
<td>Outcomes</td>
<td>Results</td>
</tr>
<tr>
<td>-----------------------</td>
<td>---------------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>(Pathman et al., 1999)</td>
<td>Medical school training for rural practice</td>
<td>All US medical schools or residency programs with rural rotations</td>
<td>Self-reported preparedness for rural practice</td>
<td>“Residency rural rotations predicted greater preparedness for rural practice (p = .004) and small-town living (p = .03) and longer retention (hazard ratio, 0.43, p = .003). Extended medical school rural rotations predicted only greater preparedness for rural practice (p = .03).”</td>
</tr>
<tr>
<td></td>
<td>Residency training for rural practice</td>
<td></td>
<td>Self-reported preparedness for small-town living</td>
<td>“The physicians' sense of preparedness for small-town living predicted their retention duration (hazard ratio, 0.74, p &lt; .0001), whereas their preparedness for rural medical practice did not predict their retention duration after controlling for preparedness for small-town living (hazard ratio, 0.92; p = .27).”</td>
</tr>
<tr>
<td>(Rosenthal, 2000)</td>
<td>Residency training for practice in underserved areas</td>
<td>Family medicine residencies with “rural training tracks”</td>
<td>Self-reported intent to practice in underserved areas</td>
<td>Graduates of family medicine residencies with rural training tracks were more likely to report that they intended to practice in rural communities than graduates of “small urban FP [family practice] residencies”, “midsize urban FP residencies”, and “all FP residencies” (76% vs. 39% vs. 18% vs. 24%).</td>
</tr>
<tr>
<td>(Bowman &amp; Penrod, 1998)</td>
<td>Residency training for practice in underserved areas</td>
<td>All US family residency programs that focus on training for rural practice</td>
<td>Provision of care in any underserved area</td>
<td>A one-month increase in the length of time of required rural training increased the proportion of graduates of a residency program practicing in rural areas by 0.82 percent (p &lt; 0.0001) net of “percentage of minority residents at program”, “full rural mission at program”, “partial rural mission at program”, “percentage of population of the state that is nonmetropolitan”, “emphasis on rural or OB [obstetrics] fellowship”, “director is rural contact person”, “other graduate training programs used at site”, “months of obstetrical training”, percentage of female residents at program”, “population of program site (city or metropolitan)”, “percentage of faculty with rural practice experience”, “sponsoring hospital of program is public”. Holding the other factors constant, having a rural mission statement increased the proportion of graduates practicing in rural areas by 12% (p &lt; 0.0001).</td>
</tr>
<tr>
<td>Study</td>
<td>Program type</td>
<td>Program</td>
<td>Outcomes</td>
<td>Results</td>
</tr>
<tr>
<td>-----------------------</td>
<td>----------------------------</td>
<td>-------------------------------------------------</td>
<td>---------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>(Inoue et al., 2007)</td>
<td>Financial incentives</td>
<td>Jichi Medical University, Japan</td>
<td>Provision of care in any underserved area</td>
<td>Participants and former participants were more likely to work in a rural area of Japan than non-participants.</td>
</tr>
<tr>
<td>(Matsumoto et al., 2008)</td>
<td>Financial incentives</td>
<td>Jichi Medical University, Japan</td>
<td>Provision of care in any underserved area</td>
<td>Former participants were about four times more likely to work in rural areas than non-participants (12.8% vs. 3.3% in 1994 and 10.7% vs. 2.6% in 2004).</td>
</tr>
<tr>
<td>(Pathman et al., 1992)</td>
<td>Financial incentives</td>
<td>National Health Service Corps</td>
<td>Retention in the same underserved area</td>
<td>Participants were significantly more likely to leave their practice of original placement than non-participants (adjusted hazard ratio (aHR) 1.98, p = 0.0002), when controlling for training in internal medicine and stated importance of living in a small community). Participants were significantly more likely to leave practice in underserved areas than non-participants (aHR 1.56, p = 0.02), when controlling for training in internal medicine and stated importance of living in a small community).</td>
</tr>
<tr>
<td>(Pathman et al., 1994a)</td>
<td>Financial incentives</td>
<td>National Health Service Corps</td>
<td>Retention in the same underserved area</td>
<td>Compared to non-participants, former participants were about three times less likely to remain in the same practice (13% vs. 44%, p &lt; 0.001) and about half as likely to remain in a non-metropolitan area (24% vs. 52%, p &lt; 0.001).</td>
</tr>
<tr>
<td>(Pathman et al., 1994b)</td>
<td>Financial incentives</td>
<td>National Health Service Corps</td>
<td>Retention in the same underserved area</td>
<td>Five years after starting work at a practice site, participants were significantly less likely to have remained at the site of first practice than non-participants (aOR 0.41, p = 0.01), when controlling for physician discipline (allopath vs. osteopath), physician specialty, initial plans to remain in the underserved area, percentage of minority patients in the practice, county population, type of county, per-capita income in the county, and physician population density in the county.</td>
</tr>
<tr>
<td>Study</td>
<td>Program type</td>
<td>Program</td>
<td>Outcomes</td>
<td>Results</td>
</tr>
<tr>
<td>-----------------------</td>
<td>--------------------</td>
<td>----------------------------------------</td>
<td>----------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>(Xu et al., 1997b)</td>
<td>Financial incentives</td>
<td>National Health Service Corps</td>
<td>Provision of care in any underserved area</td>
<td>Former participants were significantly more likely to practice in an underserved area ten years after graduating from medical school than non-participants (aOR = 3.7, p &lt; 0.0001, when controlling for age, sex, ethnicity, having grown up in an underserved area, family income as a child, strong interest in working in underserved areas prior to medical school, debt, medical school experience in an underserved area, and residency experience in an underserved area).</td>
</tr>
<tr>
<td>(Xu et al., 1997a)</td>
<td>Financial incentives</td>
<td>National Health Service Corps</td>
<td>Provision of care to poor patients</td>
<td>30% of former participants’ patients, but only 19% of non-participants’ patients, were poor.</td>
</tr>
<tr>
<td>(Singer et al., 1998)</td>
<td>Financial incentives</td>
<td>National Health Service Corps</td>
<td>Retention in the same community health center</td>
<td>After five years of work in a community health centre, 36% of participants, but only 17% of non-participants, still worked in the same centre.</td>
</tr>
<tr>
<td>(Rabinowitz et al., 2000)</td>
<td>Financial incentives</td>
<td>National Health Service Corps</td>
<td>Provision of care in any underserved area</td>
<td>“Participation in the NHSC is the only experiential factor related to caring for the underserved” (aOR 2.2, 95% CI 1.6-3.0, when controlling for sex, ethnicity, family income when growing up, childhood in inner-city/rural area, strong interest in underserved practice prior to medical school, clinical experience with the underserved during medical school).</td>
</tr>
<tr>
<td>Study</td>
<td>Program type</td>
<td>Program</td>
<td>Outcomes</td>
<td>Results</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-----------------------</td>
<td>------------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>(Brooks et al., 2003)</td>
<td>Financial incentives</td>
<td>National Health Service Corps</td>
<td>Provision of care in any underserved area</td>
<td>13% of rural primary care physicians, but only 3% of suburban and 3% of urban primary care physicians, had participated in the program.</td>
</tr>
<tr>
<td>(Probst et al., 2003)</td>
<td>Financial incentives</td>
<td>National Health Service Corps</td>
<td>Provision of care in any underserved area</td>
<td>Former participants were significantly more likely to be highly engaged in Medicaid(^{19}) inpatient practice than non-participants (aOR = 1.93 (95% CI 1.18-3.13), when controlling for physician’s sex, ethnicity, medical specialty, period of graduation from medical school, medical education in South Carolina, graduation from a non-US medical school).</td>
</tr>
<tr>
<td>(G. M. Holmes, 2004)</td>
<td>Financial incentives</td>
<td>National Health Service Corps</td>
<td>Retention in the same underserved area</td>
<td>Participants were less likely to remain in their first practice location than non-participants (beta coefficients between -0.248 and -0.272 across three graduation cohorts (all p &lt; 0.01), when controlling for age, sex, and ethnicity).</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Provision of care in any underserved area</td>
<td>Participants were more likely to serve in any underserved area than non-participants (beta coefficients between 0.528 and -0.745 across three graduation cohorts (all p &lt; 0.01), when controlling for age, sex, and ethnicity).</td>
</tr>
<tr>
<td>(Rittenhouse et al., 2008)</td>
<td>Financial incentives</td>
<td>National Health Service Corps</td>
<td>Provision of care in any community health center</td>
<td>Participants were significantly more likely to work in a community health centre than non-participants (aOR = 6.99, p &lt; 0.001, when controlling for sex, year of residency completion, private vs. public medical school, attendance of a medical school receiving Title VII funding(^{20})).</td>
</tr>
</tbody>
</table>

\(^{19}\) Medicaid is a means-tested health program funded by the US federal government and the states. It covers health care expenditures of some categories of low-income individuals, including children, parents, pregnant women, and children with disabilities (Centers for Medicare & Medicaid Services, 2009).

\(^{20}\) In the US, “Title VII grants are intended to strengthen the primary care educational infrastructure at medical schools and residency programs and to encourage physicians-in-training to pursue careers working with underserved populations” (Rittenhouse et al., 2008).
<table>
<thead>
<tr>
<th>Study</th>
<th>Program type</th>
<th>Program</th>
<th>Outcomes</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Pathman et al., 2000)</td>
<td>Financial incentives</td>
<td>National Health Service Corps, Indian Health Service Corps, State scholarships, State loan repayment programs, Practice and hospital-sponsored financial incentives</td>
<td>Provision of care in any underserved area</td>
<td>In comparison to non-participants, participants in financial-incentive programs were significantly more likely to practice in rural areas (33.3% vs. 6.5%, p &lt; 0.001) and to care for underserved populations (54.1% vs. 29.4%, p &lt; 0.001).</td>
</tr>
<tr>
<td>(Jackson et al., 2003)</td>
<td>Financial incentives</td>
<td>West Virginia Community Scholarship Program, West Virginia Health Sciences Scholarship Program, West Virginia Recruitment and Retention Community Program, West Virginia State Loan Repayment Program</td>
<td>Retention in the same underserved area</td>
<td>Retention in the first practice site was not significantly different between program participants and non-participants.</td>
</tr>
<tr>
<td>Study</td>
<td>Program type</td>
<td>Program</td>
<td>Outcomes</td>
<td>Results</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-----------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>(Pathman et al., 2004)</td>
<td>Financial incentives</td>
<td>State scholarship programs&lt;br&gt;State loan programs with service option&lt;br&gt;State loan repayment programs&lt;br&gt;State direct financial-incentive programs for medical residents&lt;br&gt;State direct financial-incentive programs for fully trained health professionals</td>
<td>Retention in the same underserved area</td>
<td>Participants were not significantly more likely to remain at their site of first practice than non-participants (aHR 0.75, ( p = 0.080 ), when controlling for age, sex, medical specialty, and marital status).</td>
</tr>
</tbody>
</table>
IV.3 Developing countries: health worker emigration

In the past decade, health worker migration from developing to developed countries has been a common topic of editorials (Bundred & Levitt, 2000; Chen & Boufford, 2005; Mullan, 2007; Pang, Lansang, & Haines, 2002), policy reports (International Pharmaceutical Federation (FIP), 2006; Joint Learning Initiative, 2004; WHO, 2006b), and scientific publications. Below, we summarize the research on international health worker migration regarding measurement (i.e., studies estimating the size of migration flows from developing to developed countries or the stock of health workers who have migrated), causes (i.e., studies investigating migration reasons), welfare impact (i.e., studies investigating how health worker migration affects social welfare in source or recipient countries), and programs and policies (i.e., publications discussing potential interventions to reduce migration or minimize any negative consequences of international health worker migration).

Numerous studies have described international migration flows or stocks of in- or out-migrated health workers for individual migration source countries (e.g., India, Zimbabwe (Chikanda, 2005), Ghana (Dovlo & Nyonator, 1999), India (Kaushik, Jaiswal, Shah, & Mahal, 2008), Malawi (Zijlstra & Broadhead, 2007), Philippines (Institute of Health Policy and Development Studies, 2005)) or from individual recipient countries (e.g., UK (Buchan, 2002; Buchan, Baldwin, & Mundo, 2008), USA (Akl, Maroun, Major, Chahoud, & Schunemann, 2007; Akl, Mustafa, Bdair, & Schunemann, 2007; Brush, Sochalski, & Berger, 2004), Italy (Chaloff, 2008), Canada (Dumont, Zurn, Church, & Le Thi, 2008), New Zealand (Zurn & Dumont, 2008)). While data on health worker migration published in these studies is useful to demonstrate the magnitude of health worker migration from developing to developed countries, it is insufficient for many research purposes. For one, the country-level data cannot be pooled for multi-country analyses, because different studies used different methods to estimate the size of migration flows and stocks. In addition, most of the studies covered only short periods of time. Finally, some studies measured only migration of health workers who graduated from specific institutions or regions in a country (e.g., (Zijlstra & Broadhead, 2007)).

Four different studies have published datasets of stocks of emigrated health workers from a number of countries (Clemens & Pettersson, 2008; Hagopian, Thompson, Fordyce, Johnson, &
Hart, 2004; Mullan, 2005; OECD, 2005; WHO, 2006b). All of the four datasets used numbers of health workers reported in recipient countries in order to estimate the magnitude of emigration from source countries. The stock estimates differ substantially across the datasets (see Table 4). These discrepancies are due to several factors, demonstrating some of the difficulties in estimating international health worker migration flows.

First, the four datasets used different sources of information on the numbers of foreign physicians in recipient countries. WHO/OECD (OECD, 2005; WHO, 2006b) and Clemens and Petterson (Clemens & Pettersson, 2008) used national census data, while Mullan (Mullan, 2005) and Hagopian and colleagues (Hagopian et al., 2004) used data from health workers’ professional organizations.

Second, in order to estimate emigration stocks the different studies counted health workers in different sets of recipient countries – WHO/OECD counted health workers in the UK, US, Australia, Canada, Finland, France, Germany, Portugal; Clemens and Pettersson in the UK, US, Australia, Canada, France, Portugal, Spain, Belgium, South Africa; Mullan in the UK, US, Australia and Canada; and Hagopian and colleagues in the UK and US.

Third, the different datasets used different definitions of migrant health workers. WHO, Mullan, and Hagopian and colleagues defined a migrant health worker as one who was trained in another country, while Clemens and Pettersson defined a migrant health worker as one born in another country. This distinction is important, because some of the countries with the worldwide lowest population densities of health workers (such as Lesotho and Swaziland (WHO, 2009)) do not have their own medical school and thus depend on physicians trained in other countries – in particular their own citizens – to staff their hospitals and clinics (Clemens & Pettersson, 2008). According to the definition of a migrant health worker as one who was trained in another country, these countries can never have any emigrated physicians and any physician working in these countries will always be an immigrant. Last, the four datasets estimate stocks of emigrated health workers for different years (Table 4).

---

21 The data reported in WHO 2006 (WHO, 2006b) is extracted from OECD 2005 (OECD, 2005).
Table 4: Estimates of stocks of physicians who emigrated from selected sub-Saharan African countries

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Year of data</td>
<td>2002</td>
<td>2000</td>
<td>2004</td>
<td>2002</td>
</tr>
<tr>
<td>Country</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>South Africa</td>
<td>12136</td>
<td>7363</td>
<td>3734</td>
<td>3788</td>
</tr>
<tr>
<td>Nigeria</td>
<td>4261</td>
<td>4856</td>
<td>3921</td>
<td>2281</td>
</tr>
<tr>
<td>Ghana</td>
<td>926</td>
<td>1639</td>
<td>NR</td>
<td>515</td>
</tr>
<tr>
<td>Ethiopia</td>
<td>335</td>
<td>553</td>
<td>NR</td>
<td>266</td>
</tr>
<tr>
<td>Uganda</td>
<td>316</td>
<td>1837</td>
<td>NR</td>
<td>175</td>
</tr>
<tr>
<td>Zimbabwe</td>
<td>237</td>
<td>1602</td>
<td>NR</td>
<td>101</td>
</tr>
<tr>
<td>Angola</td>
<td>168</td>
<td>2102</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Cameroon</td>
<td>109</td>
<td>845</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Tanzania</td>
<td>46</td>
<td>1356</td>
<td>NR</td>
<td>NR</td>
</tr>
<tr>
<td>Mozambique</td>
<td>22</td>
<td>1334</td>
<td>NR</td>
<td>NR</td>
</tr>
</tbody>
</table>

NR = not reported

One comprehensive dataset provides physician emigration rates for all countries worldwide for each year in the period 1991-2004 (Docquier & Bhargava, 2006, 2007). Figure 3 shows annual emigration rates (in % per year) for the year 2004 for all countries in sub-Saharan Africa covered in the dataset. According to the dataset, in 2004 emigration rates exceeded 10% in seventeen countries in the region and 20% in eight countries. Such panel on health worker emigration rates is extremely useful for a wide range of analyses, including on the causes and consequences of health worker migration. However, the dataset has a number of limitations. For one, the definition of migrant physician used for different recipient countries in the dataset differed across recipient countries (a physician born in another country, a physician with citizenship in another country, and a physician trained in another country) (Clemens & Pettersson, 2008). Moreover, the data on physicians who immigrated to a recipient country was available for all fourteen years from 1991 through 2004 for only 4 of the 16 recipient countries across which physician numbers were summed to estimate emigration flows from source countries. For all other recipient countries, the authors intra- or extrapolated some country-year values of emigrated physicians.
Figure 3: Estimates of health worker emigration rates in sub-Saharan African countries, 2004
The relatively recent arrival of the above datasets and the fact that each suffers from some limitations may explain why the literature on causes of health worker migration is largely descriptive and the literature on the welfare impact of the migration is small. Studies on the causes of health worker migration investigated reasons for migration decisions through surveys of health workers in both source and recipient countries. In these surveys, health workers consistently listed higher earnings in other countries as one of the most important reasons for migration plans or past migration (Awases, Gbary, Nyoni, & Chatora, 2004; Chikanda, 2005; Luck, Fernandes, & Ferrinho, 2000; Mackintosh, 2003; Martineau, Decker, & Bundred, 2002; Nguyen, Ropers, Nderitu, Zuyderduin, Luboga, & Hagopian, 2008). However, health workers also named a number of other factors as influencing their migration decisions, including job security, flexible leave policies (Mackintosh, 2003), opportunities for education and training (Awases et al., 2004; Mackintosh, 2003; Oman, Moulds, & Usher, 2009), opportunities for promotion (Bach, 2004; Buchan, Parkin, & Sochalski, 2003; Muula & Maseko, 2006), the management of health services (Bach, 2004; Chikanda, 2005; Kingma, 2006; Luck et al., 2000), workplace safety (in particular the risk of work-related HIV infection (Bossert et al., 2007; Institute of Health Policy and Development Studies, 2005; Muula & Maseko, 2006; Palmer, 2006)), family welfare (Chikanda, 2005; Oman et al., 2009), and country stability (Chikanda, 2005; Nguyen et al., 2008; Oman et al., 2009). Non-financial factors appear to be nearly as important as earnings differentials in motivating migration, but the patterns of important non-financial factors differ across countries (Awases et al., 2004).

Despite a well-developed economics literature on the welfare impact of migration (e.g., (Borjas, 1995; Goldfarb, Havrylyshyn, & Magnum, 1984; Grubel & Scott, 1966; Kwok & Leland, 1982)), only a few studies have examined the particular impact of international health worker migration on social welfare in source and recipient countries. Studies in Kenya (Kirigia, Gbary, Muthuri, Nyoni, & Seddoh, 2006) and Malawi (Muula, Panulo, & Maseko, 2006) estimated the financial loss due emigration of a health workers as the total costs of educating such a health worker (including tuition, study materials, accommodation, and living expenses during primary, secondary and tertiary education). The estimated losses of returns on investment were about half a million year-2005 US$ for a doctor in Kenya, more than 300 thousand year-2005 US$ for a nurse in Kenya (Kirigia et al., 2006), and between approximately 240 thousand year-2005 US$ (when an interest rate of 7% per annum is used) and 26 million year-2005 US$ (when an interest
rate of 25% is used) (Muula et al., 2006). While these financial losses seem large, two studies of remittances from emigrated health workers found that the remittances (of Filipino physicians (Goldfarb et al., 1984) and Tongan and Samoan nurses (Connell & Brown, 2004)) exceeded the initial investment in their human capital, suggesting that in some countries it may pay to train health workers for export.

Another study estimated the social net present value of a financial incentive that pays fully for a student’s education as a health worker in return for a few years of service in an ART program in sub-Saharan Africa. The study found that these financial incentives are highly cost-beneficial if they are effective in reducing health worker emigration. For instance, if the only effect of the financial incentive were to reduce annual health worker emigration from 12% to 5% for five years, the net present value of financing the education of a team of health workers sufficient to treat 500 ART patients would be more than 200 thousand year-2000 US$ (Bärnighausen & Bloom, 2009a).

A final set of publications on health worker migration has discusses potential programs and policies to reduce emigration of health workers from developing countries or to mitigate negative consequences of migration. Table 5 summarizes the proposed interventions. Some the programs and policies have already been implemented in developing countries (such as compulsory service (Reid, 2001) or financial-incentive programs (Ross, 2007) in South Africa or non-financial incentives in Zambia (Koot & Martineau, 2005)) or been adopted by developed countries (such as the UK’s “ethical recruitment policy” (Carlisle, 2004)). However, studies evaluating the effect of the implemented programs and policies on health worker migration – similar to those evaluating programs to increase the supply of health workers to underserved areas in developed countries – are as of yet lacking.

In sum, although the recent focus of development policy on population health has sparked research on health worker migration from developing to developed countries, the majority of studies are descriptive in nature, in part, because comprehensive data on international migration flows of health workers has not been available. Future initiatives need to emphasize further improvements in data on international health worker migration. Once validated data are available, researchers can draw on the extensive literature and theories of migration in economics
and other disciplines to further investigate the causes and consequences of health worker migration and to evaluate existing interventions to reduce migration rates.
Table 5: Programs and policies to reduce health worker migration from developing countries or to mitigate any negative consequences of such migration

<table>
<thead>
<tr>
<th>Program or policy</th>
<th>Definition</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Microlevel approaches</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Selective admission</td>
<td>Select students for health worker education who are likely to remain in the country of training</td>
<td>(Lehmann, Dieleman, &amp; Martineau, 2008)</td>
</tr>
<tr>
<td>Training for practice in developing countries</td>
<td>Focus health care education on the health care needs of developing countries</td>
<td>(Lehmann et al., 2008)</td>
</tr>
<tr>
<td>Increased health worker training in developed countries</td>
<td>Increase the national health worker education rate in countries that currently receive large numbers of health workers from developing countries</td>
<td>(Ahmad, 2005)</td>
</tr>
<tr>
<td>Compulsory service</td>
<td>Require all health workers to serve for some time in their country of training</td>
<td>(Metz, 1991; Reid, 2001)</td>
</tr>
<tr>
<td>Financial incentives</td>
<td>Offer financial incentives for return of service in the country of training</td>
<td>(Bärnighausen &amp; Bloom, 2009b; Ross, 2007)</td>
</tr>
<tr>
<td>Non-financial incentives</td>
<td>Improve the working and living conditions of health workers in developing countries</td>
<td>(Lehmann et al., 2008; Stilwell, Diallo, Zurn, Vujicic, Adams, &amp; Dal Poz, 2004)</td>
</tr>
<tr>
<td><strong>Macrolevel approaches</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visa restrictions</td>
<td>Potential recipient countries agree not to issue visas to health workers from certain potential source countries or to issue only temporary visas</td>
<td>(Ahmad, 2005; Stilwell et al., 2004)</td>
</tr>
<tr>
<td>&quot;Ethical recruitment&quot;</td>
<td>Potential recipient countries allow recruitment of health workers only if the source country consents to such recruitment</td>
<td>(Ahmad, 2005; Buchan &amp; Sochalski, 2004; Labonte, Packer, Klassen, Kazanjian, Apland, Adalikwu et al., 2006)</td>
</tr>
<tr>
<td>Compensation payments</td>
<td>Recipient countries make payments to certain migration source countries for every health workers they receive from the source countries</td>
<td>(Ahmad, 2005; Buchan &amp; Sochalski, 2004; Labonte et al., 2006)</td>
</tr>
</tbody>
</table>
V. Conclusions

In this paper, we have reviewed research on the health workforce since the 1960s. We have structured the review into three perspectives corresponding roughly to three chronological phases. During the first phase – which is associated with the perspective “health workforce planning” – shortages of health workers in developed countries triggered research on manpower planning methods to aid policy makers in ensuring adequate supplies of health workers (1960s through 1970s). Four main approaches to project future health worker requirements were developed, viz., approaches based on health need, demand for health care, health care service targets, and health worker-to-population ratios. We discuss the advantages and disadvantages of each of the four approaches and show that modified versions of the approaches, in particular the service targets approach, are experiencing a renaissance in current studies estimating health worker requirements to meet population health goals, such as the health MDGs, in developing countries.

During the second phase – which is associated with the perspective “the health worker as economic actor” – a perceived “cost explosion” in many health systems in developed countries shifted the research focus to the study of the influence of health workers’ behavior on allocative and technical efficiency in health systems (1980s through 1990s). We review the literature on one particular research topic during this phase, health worker licensure, which investigated whether licensure regulation was introduced to benefit health worker “interest groups” or to serve the “public interest”. We describe three categories of studies testing “interest group” against “public interest” explanations of the existence of health worker licensure. Studies in the first category, “market structure”, measured the level of competitiveness in the health care market, taking competitive structures as evidence against and monopolistic structures as evidence for “interest group” explanations; studies in the second category, “licensure effects”, examined the influence of licensure on health worker incomes, and the price, quality, and quantity of health services; studies in the third category, “political power”, analyzed the relationship between the political power of health worker interest groups and the passage of licensure legislation. While the findings of several studies suggest that “interest group” explanation of the existence of licensure hold true, the evidence is not conclusive because of technical limitations and conceptual problems in the analytical approaches.
During the third phase – which is associated with the perspective “the health worker as necessary resource” – local health worker shortages in developed countries and regional shortages in developing countries sparked research investigating the effectiveness of interventions to increase the supply of health workers to so-called underserved areas of developed countries and health worker migration from developing to developed countries (1990s through 2000s). Several studies showed that participants in programs aimed at increasing the number of health workers in underserved areas (such as selective admission to health worker education or financial-incentives for return of service) were more likely to provide care in such areas than their peers who did not participate. However, the studies could not conclude that the programs caused an increase in the supply of health workers to underserved areas, because they cannot rule out that selection rather than program effects was responsible for the differences in the behavior between program participants and non-participants. International health worker migration has recently emerged as a research topic (including studies on the causes and consequences of such migration), but has been hampered by the lack of validated data on health worker migration flows and stocks of migrated health workers.

Of course, the three research perspectives on the global health workforce are merely stylized descriptions of main research thrusts and do not imply that such foci were necessarily the best use of research resources during the different phases. Moreover, the three perspectives are not mutually exclusive. For instance, some of the studies undertaken from a perspective of the health worker as a necessary resource use health workforce planning approaches to estimate the size of the gap between available and needed health workers – even if the starting point of these studies is an observed health worker constraint in the delivery of certain priority health interventions. Other studies that are motivated by a perceived or measured lack of health workers clearly view health workers as economic actors that consider financial and non-financial incentives in their location decisions.

Opportunities for future research on the health workforce abound. While they are likely to include new approaches to the study, our review suggests that the transfer of existing approaches from developed to developing countries could lead to important insights. For instance, studies on health worker licensure were mostly carried out in the US. It seems likely that the strength of health worker “interest groups” relative to the power of the public differs across stages of
socioeconomic development and by type of health system. Evidence on the causes and effects of health worker licensure in different developing countries is thus likely to contribute substantially to the licensure debate. Another example where transfer of research topics and methods could lead to important insights is programs to influence health workers’ location decisions. Programs similar to those aimed at increasing the supply of health workers to underserved areas in developed countries have been implemented in many developing countries. Unlike their developed-country counterparts, however, they have yet to be evaluated in their impact on health worker movements from well-served to underserved areas and from developing to developed countries, suggesting the need for a coherent economic model of medical migration.
VI. References


Feldstein, M. S. (1967). An aggregate planning model of the health care sector. Medical Care, 5(6), 369-381.


Martineau, T., Decker, K., & Bundred, P. (2002). Briefing note on international migration of health professionals: leveling the playing field for developing country health systems. Liverpool: Liverpool School of Tropical Medicine.


Ministry of Health (1967). The system of public health services in the USSR. Moscow: Ministry of Health.


Most medical staffs have an oversupply of specialists, but don't plan to take action to correct this. (1995). *Med Netw Strategy Rep, 4*(9), 11.


The nursing shortage. (1967). *JAMA*, 200(9), 790.


