Assessing and planning for the future needs of the health care workforce

by

Mário Amorim Lopes

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in the

Department of Industrial Engineering and Management

of the

Faculdade de Engenharia da Universidade do Porto

February 2017

This research was partially supported by the PhD grant SFRH/BD/102853/2014 and by the project HHRPLAN, grant PTDC/IIMGES/4770/2014, awarded by the Portuguese Foundation for Science and Technology (FCT).

"The curious task of economics is to demonstrate to men how little they really know about what they imagine they can design."

— F. A. Hayek

Acknowledgements

Acknowledgment sections have always bewildered me. They always seemed like some obnoxious and assorted thank-you notes that you write just for the sake of it; a mandatory section that needs to be filled up with typical jibber-jabber. A platitude of dull gratitudes, in short.

Well, they aren't. When I look back in retrospective to the last three years I'm somehow compelled to really thank the people that have made my days brighter. Because if it weren't also because of them this three year marathon could've turned into quite a rocky ride.

First and foremost, I have to thank my wife, Maria. We started dating a couple of weeks after getting enrolled in my PhD, and somehow she provided the love and the warmness that makes your life greater, and that gives you the purpose and focus to achieve. We have married one year before concluding and submitting this PhD thesis—I guess winning the heart of the scientific academy just takes longer. Then, my parents. They have always stood behind me on all occasions, especially on life-changing decisions. «As long as you're happy»—they said. Well, I am. To my sister, that is about to embark in this adventure, and to whom I wish the best of luck. Also, there has to be some space left to mention Margarida, Rosinha and Mimi, our beautiful canine offsprings that drop a massive party every time we get home, even if we have just left and forgot something. You feel important. We *are* important to them.

Secondly, to the amazing people that have been sharing this ride all along. We share a magnificent open space where we talk, we laugh, we share, we celebrate—and we work, eventually. I have to thank Gonçalo for our endless and thought-provoking conversations. We've discussed economics, politics, philosophy and, invariably, international migration. I also have to thank Fábio Neves- (not a typo), a hell of a kind of a guy that has missed by an inch a fabulous career as a football player (great news for science); Pedro Amorim and Luís Guimarães, long time friends; Sofia Gomes, that has just embarked in this great adventure of doing a PhD; Maria João, Beatriz, Xavier (aka *Mestrando*), Prof Rocha, Elsa, António, Manu, and so many others I can only list a few.

Thirdly, to my students. Teaching has to be one of the most enriching yet challenging tasks ever. They contributed to my personal and academic development in a much more significant way than what I could deliver as their lecturer.

Fourthly, to the amazing team at Instituto Superior Técnico, Mónica Oliveira and Teresa Cardoso, that have always helped me out, and have also embraced this research problem. Also, to ACSS, IP, who have being very receptive and helpful.

Last but not least, to the two people that actually made this possible, my supervisors. To Álvaro Almeida, an amazing and inspiring professor that has thought me a lot about Economics and Health Economics, and with whom I had endless enriching talks in his office about economics and politics.

And to Bernardo Almada-Lobo. Words would do a poor job in telling how much I've gained from having the chance to work alongside with him. He's contributed to my professional, academic, but also to my personal development. From Bernardo I've learned way more than I'll ever be able to acknowledge, less so to pay back. He's an inspiring professor, a brilliant researcher and an outstanding negotiator. Here's the biggest compliment of all—you're the older brother I never had.

Thank you all. *Really*.

Abstract

Few problems are so serious as one's health. For ages, people have tried to help those in need by providing remedies capable of healing their maladies. They were once the healers, the curators and the magicians, but things have changed, and science has scrapped obscurant practices. Nowadays, they are highly trained professionals—the Health Human Resources (HHR), individuals, such as physicians and nurses, capable of delivering effective health care services.

HHR planning is then the critical venture that tries to ensure the right number of people with the right skills in the right place at the right time provide the right services to the right people. HHR planning has progressed significantly since the first manuscripts were published in the 50s. Several methodologies have been used, and several inputs were shown to influence HHR. However, the field has yet to find an established methodology capable of addressing the many facets of the problem.

Considering this, the present thesis was conceived so as to mitigate some research gaps and contribute to the field in three different but related ways: first, by providing models and methods that have several benefits over existing ones; second, by applying those methods to real cases, in particular to the Portuguese physician workforce; third, by providing insights to policy-makers, in this way bridging the gap between scientific research, practical application and policy making. The starting point was an extensive review of the literature in order to map methods and approaches. From this work became clear an integrated approached was needed. The ensuing publications were contributions to that integrated approach, but also covering one, two or all of the dimensions aforementioned. We developed an agent-based model to better forecast the supply of physicians; we extended that model with

microeconomic foundations and individual decision-making to study emigration and conduct policy experiments to inform on retention policies, if any; and finally, we have contributed with a methodology to improve cross-country comparisons, a tool frequently used in HHR planning. These models were tried out on data from the Portuguese physician workforce, so that practical insights could also be gained from the scientific inputs.

We hope this thesis may add real value to HHR planning by providing new methodologies and approaches useful to both the scientific community and policy makers.

Resumo

Poucos problemas serão tão sérios quanto a nossa saúde. Durante muito tempo, curandeiros e mágicos tentaram ajudar os necessitados, proporcionando-lhes curas e remédios, que supostamente tratariam das suas maleitas. A ciência veio trazer rigor à arte, dirimindo assim, ainda que parcialmente, práticas obscuras. Actualmente, aqueles que providenciam cuidados de saúde são indivíduos altamente especializados — são os recursos humanos para a saúde (HHR), médicos, enfermeiros, entre outros, capazes de providenciar cuidados de saúde eficazes.

O planeamento dos HHR é uma tarefa crítica que procura garantir que o número correcto de indivíduos com as devidas qualificações está no sítio certo a providenciar os cuidados de saúde a quem deles necessita. Esta área científica progrediu significativamente desde que os primeiros artigos foram publicados nos anos 50, tendo várias metodologias sido usadas, e *inputs* que influenciam a sua evolução identificados. Não obstante, ainda não foi possível encontrar uma abordagem consensual, capaz de incorporar as múltiplas dimensões do problema.

Por conseguinte, a presente tese foi desenvolvida com o propósito de mitigar algumas das lacunas da investigação, contribuindo para a área científica de três formas distintas, mas relacionadas. Primeiro, providenciando modelos e métodos que apresentam diversas vantagens em relação aos métodos em voga. Segundo, aplicando esses métodos a casos de estudo reais, em particular à força de trabalho médica portuguesa. Terceiro, procurando derivar implicações políticas e informando os decisores políticos, para desta forma aproximar investigação científica, aplicação prática e políticas públicas. O ponto de partida foi uma extensa revisão da literatura, por forma a mapear métodos e abordagens, tendo deste trabalho

emergido a evidência clara de que uma abordagem integrada era necessária. As publicações que derivam deste trabalho são, portanto, contribuições para essa abordagem integrada, que procuram dar uma resposta às lacunas identificadas. Assim, desenvolvemos um modelo estocástico de simulação por agentes para gerar projecções da oferta de médicos; estendemos esse modelo para incluir fundamentos microeconómicos e decisão individual por forma a estudar a emigração médica e a testar políticas de retenção, se alguma; e, finalmente, contribuímos com uma metodologia para melhorar as comparações internacionais de sistemas de saúde, uma ferramenta frequentemente empregue na área do planeamento de HHR. Os modelos foram calibrados com dados da força de trabalho médica portuguesa, procurando assim que esta tese encete também um contributo prático, que sirva tanto à comunidade científica como aos decisores políticos.

Table of contents

Chapter 1

Introduction

No healthcare system can operate without the health human resources (HHR) necessary to deliver health care services. Mismatches between supply and demand for HHR are an economic problem, insofar as efficient resource allocation goes, but may also have wide social repercussions. A surplus leads to inefficiencies and misallocated resources under the guise of unemployment, to downward pressure on wages, to emigration and to increasing costs due to supply-induced demand. A shortage, on the other hand, may potentially lead to lower quantity and quality of medical care and pressures for shorter visits; work overload, resulting in a sleep-deprived HHR, ultimately compromising patient safety; and queues and waiting lists, potentially causing avoidable patient deaths.

The situation is further aggravated because health labour markets are, in general, extremely rigid, preventing market-clearing mechanisms from working. A significant number of health systems, especially those based on a national health system, feature many rigidities. For instance, institutional, regulatory arrangements and government-fixed prices; imperfect market competition under the form of monopsony in the demand for health labour, and monopoly power in the supply of health care services; long and demanding licensure processes that delay access to the profession; etc. may all prevent self-adjustment. Under these circumstances, when the market is subject to a considerable amount of intervention and rigidities, and therefore unable to adjust automatically, imbalances are static, requiring further intervention to force adjustment.

Hence, effective planning of the HHR is critical to ensure that economic and social problems arising from an imbalance of clinical staff are avoided, and also to ensure that a health system can attain health gains, work efficiently, control its costs, and can respond to quick variations in the demand for healthcare. The problem can be stated as assessing the right number of people with the right skills in the right place at the right time to provide the right services to the right people. To restate its importance, HHR planning has been identified as the most critical constraint in achieving the well-being targets set forth in the United Nations' Millennium Development Goals. In fact, the *World Health Report 2006: Working together for health*, put forth by the World Health Organization (WHO), estimates that 57 countries, developing and developed, already face an absolute shortage of 2.3 million physicians, nurses and midwives. Estimates for 2015 pointed to a dramatic shortage of physicians in the WHO African region alone. On a global scale, WHO estimated that at least 15% more doctors were needed.

In the news, articles giving notice of health care facilities (especially those located in the interior) unable to fulfil medical positions or reporting surgical interventions that have to be postponed and put on hold up to a year or two abound. These news articles concern Portugal, but many other countries face similar challenges. To aggravate the situation, it takes years for a physician to obtain medical training, meaning that the problem may remain unaddressed for years—and in the time in between patients may die. Of course governments can always resort to hiring foreign doctors. Although voluntary international migration is in most cases a win-win situation for both the sending and the host country, helping the respective labour markets to clear, in extreme cases it may lead to a reverse shortage in the sending country. After having borne the training costs, as in most EU countries medical education is subsidised, they also have to face a shortage due to misguided planning in foreign countries. Becoming aware of this, the WHO pushed for worldwide adoption of the WHO Global Code of Practice on the International Recruitment of Health Personnel. This is an attempt, albeit limited, to contain the damage caused by migration flows.

In fact, the challenges posed by inadequate HHR workforces go well beyond tracing the evolution of the supply of and the demand for health care services; grasping the impact of international migration flows on the workforce; or assessing how shortages or surpluses may be mitigated. The topic has entered the agenda of national governments, European bodies and international institutions, such as the World Health Organisation, and with good reason. HHR planning has implications at many levels. First, it may disrupt operations in the health system, as health care services cannot be provided without HHR, at least for now. Second, it may cause social harm, as unmet care may eventually lead to the death of patients. Third, there are organisational and political interests at stake. Professional associations are well aware a limited and highly differentiated workforce provides plenty of bargaining power. Fourth, most health systems are run, at least partially, by the government, and any decision will certainly have an impact either on the government's budget, and therefore on the public's purse; or on the economy, and therefore on social welfare.

Therefore, the topic necessarily entails a broad and possibly never-ending discussion. For instance, one may ask why so many rigidities prevent the health labour market from self-adjusting, therefore requiring intervention that resembles old-fashioned central planning, and whether such rigidities should remain in place. In the case of Portugal, only public universities are licensed to train medical students; just a few private schools are allowed to have residency programs, despite the tremendous efforts from legitimate and reputable private universities for doing so; and residency vacancies are defined by the government based on repute assessed and vouched by the medical association, which therefore limits access to the profession. Indeed, no other profession is so strictly regulated. Moreover, one may wonder what are the economic consequences of having virtually no unemployment in the medical profession—does it mean the best and the worst professionals are both able to practice? Should there be a way of filtering those that despite having successfully completed training are clearly unfit for practice? Or should they be given a job simply because there are no substitutes? Is it *fair* that other labour markets have to endure an excess of supply because regulatory barriers set a limit to the health labour market supply?

As much as we would like to target some or all of these questions, as they certainly need sound answers, this thesis has its scope turned at the health system *as it is*, accepting both its advantages and cons. From an epistemological point of view, we therefore take a positivist approach to the problem, leaving the merit of legal, political and organisational choices unquestioned, working upon the current legal framework. Therefore, the answers we aim to provide to the research questions elicited throughout this thesis will not prescribe any structural changes to the health system.

Considering this, this thesis is oriented towards addressing the problems afflicting actual HHR planning, regardless of the health system idiosyncrasies. Ultimately, if regulatory barriers to entry exist and will continue to exist, political reforms are unlikely, or at least not imminent, and the health labour market is therefore unable to adjust, then HHR planning is surely needed, and it should be as accurate and rigorous as possible. Otherwise, large scale, strategic HHR planning would simply not be needed if that were not the case. HHR planning is precisely where this thesis is situated.

Having narrowed the scope of this thesis, we can identify three strands of research that relate to HHR planning and upon which the present work lays. The first has to do with operations research. At the most fundamental level HHR planning is an operational problem. Defining the optimal workforce size and adjusting input variables to achieve the desirable number subject to the many constraints that apply is an operational problem. However, since we are dealing with capacity planning at a very strategic level, economic questions also arise. Economic incentives, individual decision-making and utility, production and input requirements functions, demand for health care services, and many more economic factors all come into play and need to be incorporated. Finally, HHR planning is absolutely ineffective if the guidelines and insights resulting from the work developed do not enter the policy makers' agenda. Therefore, there are clear implications for public policies, even if no structural change is being suggested.

In terms of methods to address both the operational and the economic facets of HHR planning, the scientific literature has not yet reached a consensus on the best way to forecast future workforce requirements. Given the breadth of research involved, it may be the case that several competing methodologies will continue to exist, each having its own unique set of features and limitations. In general, these methodologies tend to focus on a single part of the problem, be it supply, demand, needs, or resort to static worker-to-population ratio comparisons. Notwithstanding, it is critical to understand, if one is to make advances in the field, how HHR planning has evolved since its early stages of research, the competing methodologies, and whether more recent research has refocused its methods on new apparatus. A truly comprehensive review of the literature is probably impossible, and beyond the scope of this thesis, but a thorough review is most certainly appropriate, culminating in the current state of affairs and directions for future research. We have developed this theme further, attempting to ascertain where HHR planning currently stands.

Even a quick glance at the literature shows that simulation has been extensively used to generate forecasts on the evolution of HHR, in particular System Dynamics (SD), a systems-level approach. Simulation as a methodology is a typical choice for modelling complex and non-linear systems. However, SD faces challenges in modelling complex human behaviour or in dealing with spatial awareness, including that of physicians and nurses. Its mechanical architecture of stocks and flows makes it difficult and unnatural (but not impossible) to model complex human behaviour that dictates, for instance, the decision to emigrate, where the physician decides to practice or whether the physician opts for public or private practice. Furthermore, it has limited functionality for adding reactiveness or ability to learn, as no definition of agent or entity exists. In short, it makes it extremely difficult to add microeconomic foundations to economic agents, and therefore falls short on providing the tools to incorporate economic behaviour. All of these cases are relevant to the matter at issue. Considering this, we explored this topic, proposing an agent-based simulation model (ABM) that forecasts the physician workforce and applying it to the Portuguese health system. We also account for changes to demand due to demography, factoring in the impact of population ageing and population decline, and use the model to obtain forecasts for the Portuguese workforce.

Of the many drivers that affect the HHR workforce, in particular the physician workforce, it became clear emigration may potentially be one of the most pressing concerns. Emigration causes, on the one hand, an immediate downsize of the workforce; but, on the other hand, it is extremely hard to solve, since training fresh graduates to replace those who leave takes a long time. To draw upon a medical allegory, the pain is immediate, but the remedy takes years to take effect. We can see why if we view the health system as a whole, from the moment a student joins medical school to the moment professionals leave the health system, either due to emigration, mortality or retirement. It is easy to prevent people from entering medical schools, but it is extremely hard to *push* them to do so, since in a free, democratic country people cannot (and arguably should not) be forced to join a particular course; even if there are people willing to do so, the process is long. Consequently, if there is no slack in the workforce, the problem can hardly be solved without resorting to international recruitment, with all the known risks. All things considered, emigration should be addressed, and the reasons for doing so are threefold: (i) from the previous work we observed that even small changes to the emigration rate may affect the workforce size considerably, and therefore the effect is not negligible; (ii) emigration may play a fundamental role in helping the health labour market to clear in case of a surplus, but it may also cause shortages; (iii) as previously mentioned, international migration of doctors may pose a serious problem to sending countries, and therefore it is a critical matter. Hence, a policy analysis that identifies the circumstances in which policy intervention should occur and the impact different retention policies have on the physician workforce is long needed.

We also observed that many studies, in particular white and policy papers but also some scientific articles, use simple worker-to-population ratios to draw cross-country comparisons and find the midpoint, as if the average were the *right* number. Unable to obtain data in order to apply more advanced and powerful models, policy makers have resorted to these simple back of the envelope calculations to guide HHR planning. Although such comparisons may be an interesting starting point for setting the stage to foster and frame the discussion, imprudent use may take its toll. For example, consider the decision by the British National Health Service to increase the intake to medical schools by 60%, a resolution motivated by the observation that the physician-to-population ratio was low in comparison to other OECD countries. The policy was adopted without first evaluating for other criteria that may affect the performance of the medical staff, namely the skill mix and productivity. Subsequent research justified the lower ratio of physicians with a better distribution of the skill mix and increased productivity resulting from a more efficient task delegation when compared to other countries. Whether such cross-country analyses can be improved, lessening the impact potentially caused by fallacious comparisons, is certainly worthy of academic thought. In fact, such comparisons can be extremely useful to guide and benchmark policy reforms provided that the countries being compared are indeed comparable.

This thesis comes forth as an attempt to provide the methodology and the tools to properly assess and plan for the future needs of the healthcare workforce, in this way pushing the boundaries of this scientific field and guiding policy-makers towards better decisions. As a proof of concept, we apply the methods and tools springing from this research to the Portuguese health system, trying always to complement the methodological work with applied insights. We hope that bridging scientific research, practical application and informed decision-making at the policy level is a contribution relevant to the academy, to practitioners and to the public in general. And above all to patients in need of health care support.

1.1 Problem description

Providing the best healthcare services and equitable access is the ambition of the physicians and nurses that attend to the medical needs of the population on a daily basis. Practitioners are a scarce resource and so there is a price to it. In fact, the healthcare sector absorbs a significant share of the government budget in most of the developed countries. Adding to this, the wages of clinical practitioners represent a very significant portion of the overall costs. Thus, due planning and optimal allocation of human resources is of the utmost importance to ensure that precious healthcare time and money is put to good use.

At its core, HHR planning involves forecasting the evolution in the supply of and in the demand for healthcare services, the two main drivers of the health labour market in any of its possible forms: a decentralized and free-market health system; a social health insurance system with both privately- and publicly-run hospitals; or a centrally planned national health system. A more comprehensive analysis would also factor in the workforce skills, their geographical distribution, the impact of technological progress, the epidemiological needs of the population or the kind of services provided, but the underpinnings of sound HHR planning are credible and reliable forecasts. How to generate and actually generating accurate forecasts is the aim of most research in the field.

Such work is necessary because contrarily to a competitive market where supply and demand, assuming no hysteresis effects are occurring, tend to converge to equilibrium, with prices and wages adjusting accordingly and the market clearing, the healthcare sector exhibits rigidities that may preclude immediate or complete adjustment [24]. For instance, institutional and regulatory arrangements and government-fixed prices may prevent automatic price and wage adjustment, implying sticky prices; imperfect market competition under the form of monopsony, as in some countries hospitals are the single buyer of health care services, although empirical evidence is mixed and does not fully sustain this hypothesis [12, 34]; monopoly power of the health care providers and labour unions, especially physicians, may restrict supply and impose price-fixing [33]; access to medical and nursing schools, as well as residency training, is sometimes limited by *numerus clausus* set by the government; long and demanding licensure processes that create a delay in the time required to attain a license to practice; low or virtually no access fees that have no impact on the demand for health care services, preventing allocative efficiency; very limited "freedom of choice" in the public sector; etc. may all prevent self-adjustment. This implies that an imbalance in the health labour workforce, particularly a shortage, may be extremely difficult to solve without policy intervention. Table 1.1 lists some of the most commonly found market rigidities that may prevent the health labour market from clearing.

Table 1.1: Potential cases when the market may be unable to self-correct for shortages or surpluses of health resources.

Institutional and regulatory arrangements preventing price and wage adjustment
Market concentration in hospitals (monopsony)
Monopoly power of health providers and labour unions
Government-fixed *numerus clausus* restricting access to medical universities
Long and demanding training to be eligible for clinical practice

In practice, then, HHR planning has to do with the supply of HHR, and everything involved in the process; with the demand for health care services, i.e. the medical conditions that drive people to seek for medical attention; and, most importantly, to whether supply and demand match and gaps exist in the present or in the future.

Supply is a shorthand for all the practitioners and non-practitioners necessary to deliver health care. Analysing the supply of HHR thus involves looking at the factors that cause shifts in supply, that is, either increase or decrease it. For instance, nursing students that have just graduated are now allowed to practice. By entering the health labour market these fresh graduates will increase the supply of nurses. On the contrary, if a nurse retires, dies or emigrates the supply of nurses will necessarily decrease—less nurses will be available for hire in the health labour market. Moreover, the quantity of health care services these professionals can provide is severely influenced by how productive they are. If they have acquired new and relevant knowledge, use advanced machinery or organisational changes that enhance productivity were put in place these very same practitioners may treat more people in a shorter period of time, with no impact in terms of health outcomes. In the same way, delegating tasks between professionals may also affect the supply of HHR. If nurses are able to provide some kind of primary medical care, then physicians may be reoriented to perform other more differentiated tasks. Hence, addressing supply involves looking at the training process and estimating how many students graduate per year, accounting for the outflows that may reduce the workforce, and eventually also considering productivity and the skill mix. By eventually we mean that productivity, being extremely relevant in other sectors, is less so in the health care sector since it is plagued by Baumol's cost disease—despite the real wage growth enjoyed by health care professionals, their productivity has not increased in proportion. Hence, it would not stand as an obnoxious assumption that productivity stays about the same, although such assumption should always be questioned, and an adequate model should allow for cases when productivity increases or, less likely, decreases.

Supply is only one-half of the equation, with demand being the other half that together fully explain variations in the use of medical services. Facing a *need* for medical care, individuals may seek a doctor for several reasons: to perform routine exams and screening tests, checking if they are healthy and pursuing treatment in case they are not; for medical interventions related to aesthetic medicine, either because they were prescribed or chosen voluntarily; to abide to cultural practices and traditions; or because they suffer or have suffered from a condition of illness, either due to a chronic disease or due to an accident or injury. In general, we may categorize the factors that drive people to seek for medical attention as incidence of illness, cultural-demographic characteristics and economic factors [13].

Medical care has been traditionally framed as a "need", but individuals also make choices about medical care. And if they do, they are sensitive to factors such as income and prices, even if indirectly. When an individual postpones medical treatment, such as a surgery or a medical appointment, either because it faces a long waiting list; cannot afford to pay the intervention at a private hospital; or as a result of his own self-assessment he is manifesting will, even if limited. As so, there is an economic choice involved, implying that the demand for medical care is therefore an economic concept encompassing both the "needs" and the "wants" [13]. According to seminal models such as Grossman's health-capital model, health is a consumption good that requires continuous investment in order not to depreciate, or put in another way, in order to improve the stock of health, and therefore the longevity of the individual as well [15]. More recent models, such as Dalgaard and Strulik's health deficit model, take another stance at the problem, building on insights from biologists and gerontologists that have stressed that individual ageing is an event-dependent process rather than a time-dependent process, although at the population level age is the best predictor of ageing and death [10]. As a consequence, health depreciates continuously, with some events speeding up that process (unhealthy lifestyle), and others delaying it (exercise, healthy diet, healthy lifestyle). This non-negligible change is able to cope with Grossman's paradox of infinite lives. Notwithstanding the change to the underlying logic, the demand for health care does not change substantially, in the sense that individuals still look for health care services for the reasons aforementioned.

The ontological discussion over the concept of demand is relevant for assessing how we should approach it henceforth. Being then framed within the realm of economics, we can safely assume it is influenced by factors typically affecting the demand for any other goods, such as income or demographic growth. Income plays a role at the individual level, increasing the basket of goods an individual may consume, including health care. Demographic changes, however, hit the population at large. Population size and ageing both influence the demand for health care services, since if there are less individuals there will necessarily be less consumption in aggregate terms (even if individually it stays constant or even increases), *ceteris paribus*.

Finally, epidemiological conditions, or simply "needs", stand perhaps as the strongest driver behind health care demand. To many people the decision of seeking health care comes as an unexpected occurrence frequently associated with a perceived illness or an extreme injury requiring medical attention. Except for chronic diseases or for conditions already diagnosed, the onset of illness and the subsequent use of the health care facilities is usually not planned in advance. In more urgent cases it is not planned at all. Notwithstanding, illness, which sometimes does not result from previous individual choices, drives people to look for medical attention. These unexpected events may lead to fluctuations or random elements in the day-to-day demand of individual health care facilities such as hospitals, suggesting a stochastic nature. In aggregate terms, however, much of these error terms cancel themselves, and the demand generated by a chronic condition can be assessed with relative accuracy.

Therefore, incidence and prevalence of diseases, emerging chronic conditions, and how these diseases may evolve in the future play also a pivotal role. What poses as a challenge, in particular to epidemiologists, is ascertaining how these diseases will evolve in the future. For instance, the burden caused by ischemic heart disease and many other cardiovascular diseases is expected to continue rising mostly due to behavioural risks and unhealthy lifestyles. Type II diabetes has also been witnessing a growing trend. Standing from the side of economics and operations research our task is to quantify the impact brought about by the evolution of these diseases on health care utilization levels, leaving the disease burden forecasts to the epidemiologists.

1.1.1 Methodologies for HHR forecasting

HHR forecast models function by projecting supply, demand or both. To determine the evolution of supply, the initial stock of physicians, nurses or of any other health care practitioner is considered. The current headcount is then updated according to the evolution of the factors known to affect it, such as changes in mortality and retirement rates, migration flows, medical school intakes, etc. Forecasting demand is arguably more complex due to higher uncertainty over the estimate of the underlying parameters. Common economic factors such as the evolution of demography, income or GDP growth rate (used as a proxy to estimate both health expenditure and income growth) influence the demand for healthcare services, but epidemiological factors, or needs, also play a major role [4]. After estimating both the evolution of supply and demand, a gap analysis can then be conducted.

Different methodologies have been used to project both the supply of and the demand for health care services, and they can be loosely grouped into supply- and demand-based methods:

Supply-based: The focus of supply-based methodologies is on the medical/nursing training process so as to obtain a clear picture of how many new entrants will join the workforce, but also of attritions, such as mortality, migration, retirement and so forth, that may generate flows out of the workforce. More recently, attention has been also given to productivity, feminization and skill mix of the workforce [35, 39, 40]. In its most simple form, worker-to-population ratios have been used to establish a desired ratio for the workforce per unit of population, and having it compared against actual ratios or against other similar regions or countries [41].

Demand-based: Three methods have been commonly employed from a demand perspective [16]. Most of the methods build upon the definitions of biological needs or economic demand [22, 30], namely:

a) Needs-based (potential demand) estimates consist of determining the effect of epidemiological patterns, in particular present and future incidence and prevalence of diseases in the demand for health care services [21];

b) Economic (effective demand) estimates look at the services actually contracted by the population subject to the usual economic constraints that may limit the quantity effectively in demand;

c) Service-target estimates extend a needs-based approach by incorporating other measures like consumer preferences in order to establish service-target ratios to be attained;

These methods make use of different analytical tools to conduct the forecasts, falling into one of three categories: population-based methods, econometric models and operations research methods [26]. Population-based models rely on demography data, while econometric models look at trends in aggregates, exploiting causality [29]. Operations research methods include simulation [3, 19, 36, 37], linear programming [20], Markov chains [17] and Markov processes [7].

Recent publications have tried to combine both supply and demand into an integrated model that is able to cope with the many drivers that impact HHR planning [1, 36]. Figure 1.1 depicts a high-level functional diagram that combines these many drivers. On the supply-side, we have the current stock of workers along with the training process, so as to obtain an initial snapshot of the current workforce. The current stock, which may or may not be enough to tackle current demand, in which case an imbalance exists, is subject to in and out flows that may alter its number and composition. This given quantity of workers may provide more or less health care services depending on their productivity and skill mix, which influences the conversion from headcounts to full time equivalents (FTEs). Such conversion is critical to properly assess the health care workforce, as a significant number of physicians and nurses work part-time only.

On the demand-side, economic demand can be initially measured by analysing indicators of utilisation. How this demand will evolve in the future will then be subject to typical economic factors such as demography and the growth of income/GDP. In parallel, potential needs can be assessed by incorporating incidence and prevalence of diseases, and then mapping a given disease to an estimate of FTE requirements. Whether future supply forecasts should address all of the estimated needs is a decision left to the consideration of the policy

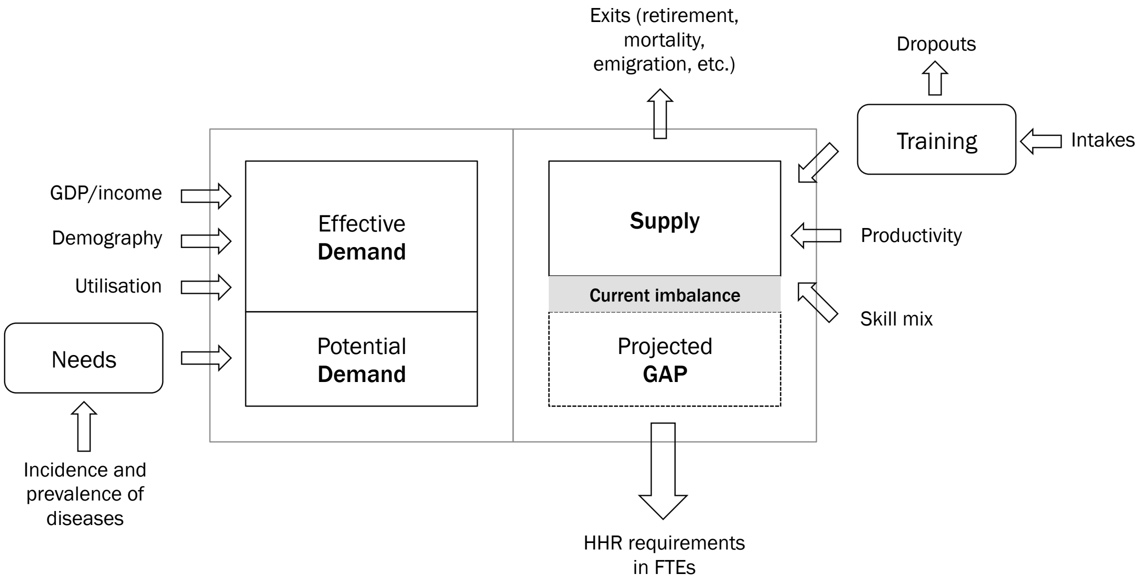


Figure 1.1: An integrated system that incorporates several methodologies to address the many facets of HHR forecasting.

maker, as this analysis does not incorporate financial constraints. Such an integrated approach is more complex, but not necessarily more difficult [40].

1.1.2 Assessing initial imbalances

Ideally, in case data exists, the starting point of HHR forecasting would be to assess the present situation by performing a gap analysis of the de facto supply and demand in order to find current imbalances, as these imbalances may considerably influence the projections (see Figure 1.2) [2]. From an economic point of view, an imbalance occurs when the quantity supplied by the present workforce of a given skill at a given market or government-stipulated price is insufficient to attend to the quantity demanded by patients. Such imbalance is in fact a disequilibrium between the demand for and the supply of labour in the health system.

Alternatively, imbalances may also be measured by value judgment or professional determination by experts in the field, extending the economic definition of demand effectively observed or measured with the perceived medical needs of the population [43]. Such normative stance is usually not bounded by any budget constraint of either the individual

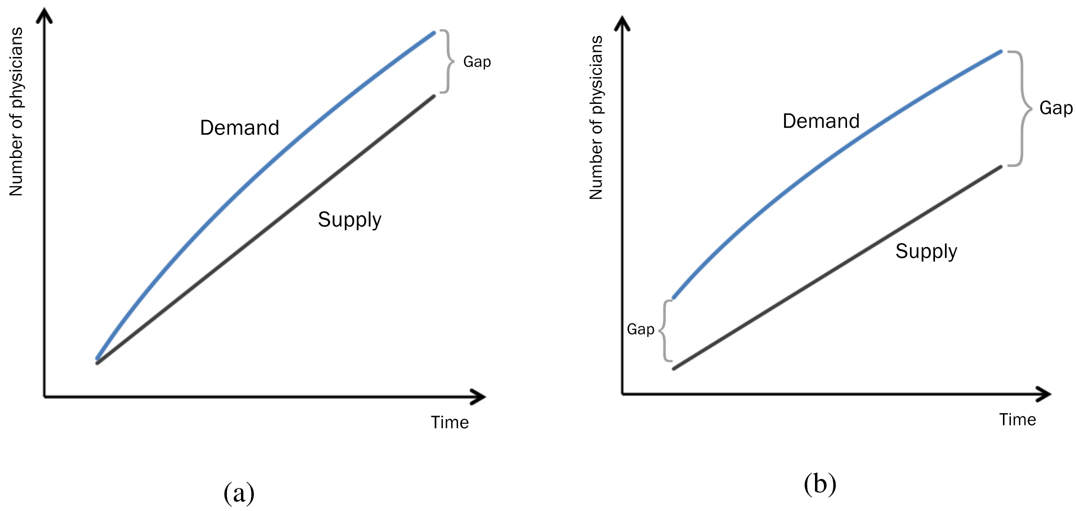


Figure 1.2: Comparison of the impact not accounting for the initial imbalance (a) and assessing current imbalances (b) can have on the forecasts.

agent, the insurance company or of the national health system, as the objective is to pinpoint unmet care needs.

Evidence in several countries suggests that some are already in the presence of shortages or surpluses, but despite this growing concern few HHR planning models actually consider the presence and magnitude of these imbalances [28]. Instead, most models assume that the market is in equilibrium, i.e. there is no shortage or surplus of HHR, and start from there, which will affect the estimated gap both in the baseline year and throughout the projection period. Some models perform the forecasting so as to guarantee that current worker-to-population ratios are maintained; other models assume that demand, indirectly measured through the current levels of utilisation, will stay constant, providing no statement or assessment whether such levels of utilisation already satisfy all the needs of the population. In most cases, however, not accounting for current imbalances is explained by the lack of reliable data and methods.

Typologies of imbalances

Imbalances may occur at several levels and are more complex than a simple mismatch in the headcount of professionals. Several categories can be considered [42]:

1. Profession/specialty imbalances cover disproportions between different professions, such as physicians and nurses, but may also expose shortages of a particular type of specialist;

2. Geographical imbalances reveal asymmetries between cities, regions or states, with urban-centres usually benefiting from a significantly larger workforce than the country-side. On average, the headcount may seem appropriate, but without a detailed analysis of how practitioners are distributed some regions may end up with an excess of human resources while others strive to fill vacancies;

3. Institutional and services imbalances reflect the differences in the endowments of health care facilities and services, with separate health service units providing disparate quantity and quality of service;

4. Public/private imbalances are those related to divergences in the distribution of practitioners between private and public sector. With disparaging wages between the two sectors, health labour tends to flow to the sector capable of providing higher wages, creating a potential deficit in the other sector;

5. Gender imbalances relate to disparities in the male/female distribution of the workforce, which may be a fundamental criterion in countries where sex ratios differ considerably. This is relevant not only for gender equality reasons. It has been shown that female doctors generate better outcomes in a consistent manner, as their practice patterns tend to be more evidence-based, sticking more closely to clinical guidelines [38]. Better outcomes lead to less visits to the doctor, and therefore less pressure on the health system.

Measuring imbalances

If labour shortages or surpluses are critical determinants of the accuracy of HHR planning, the question of how to measure these imbalances then arises. Theoretically, such exercise should be straightforward. In practice, there is no single indicator capable of providing an instant snapshot of the health market, and so several proxies have to be used and weighted against each other to achieve an estimate of the current gap.

Some countries analyse current imbalances by incorporating data on the available vacancies for physicians in the hospital sector. Others consider also the distribution of general practitioners in rural areas, anticipating a potential imbalance in the geographical distribution [28]. Alternatively, a survey from the main health care providers, hospitals, private-run clinics and long-term care facilities can be conducted to obtain information on potential difficulties in filling in vacancies [28]. In other studies, the demand for a given speciality is estimated using three indicators: increases in office visits in previous years; waiting times in comparison to other specialties; and benchmarks from health maintenance organisations (HMO) on the future requirements for that particular specialty [32]. Also, benchmarking the current situation against a needs-based estimate is also common practice in models where present imbalances are accounted for.

In fact, the combination of indicators to use will mostly depend on the typology of the market and the type of health system. Countries with a large private sector can use economic indicators like real wage rates and market rates of return as signs of abundance or scarcity, since prices in a free-market signal resource availability. Highly paid specialties may signal a small supply, and therefore the lack of market-driven competition; on the other hand, low wages combined with high unemployment rates may suggest an excess of that specific profession given current demand. Complementarily, operational indicators obtained from hospital units and other health facilities may also provide insightful information. Unfilled vacancies, waiting lists, surveys to the hospital's management board, hours of overtime work, etc. are all suggestive of a mismatch in the health labour market. Additionally, simple benchmarks between regions or countries using physicians and nurses to population ratios may also be used, although this a very unreliable comparison, as experience suggests [5].

Table 1.2: List of indicators that can be used as a proxy to assess current imbalances in the healthcare market.

Indicator	Advantages	Disadvantages
I. Economic indicators		
Occupational unemployment rate	Almost all countries collect employment and unemployment statistics	Does not disclose the specialty
Workforce/population growth rate	Can be applied to any health profession and health system	Difficult to assess if the growth of the workforce is a response to a previous shortage
Real wage rate	Easy to obtain for state-run healthcare systems	Hard to calculate and to convert into number of personnel
Rate of return	Provides information about the number of suppliers and the competitiveness of the market; may also be used to compare relative wage differentials with other labour markets	Hard to calculate and to convert into number of personnel
II. Operational indicators		
Vacancies (unfilled position)	Easy to measure; widely used	Information may not be available for private practitioners
Waiting lists	Easy to measure	Difficult to convert into supply numbers
Overtime work	Easy to measure and quantify	May be a response to a spike in demand; it may reflect a deliberate policy; understanding whether it is a continuous practice is required
Turnover rate	Easy to measure	Turnover may depend on other factors other than lack or abundance of professionals
Temporary workers	Easy to measure	Like the overtime work, it may be a response to a surge in demand, or a way to avoid work contracts and reduce costs
Level of substitution	Provides information on what professionals are doing what	Vertical and horizontal substitution may be a management decision, not a lack of professionals; sometimes it may be hard to measure who is qualified for which task
Foreign doctors	Resorting to foreign doctors is usually a government-enacted policy, so it is easy to get an exact number	May be misleading if confused with immigrant professionals

Table 1.2 provides a list of economic and operational indicators along with the advantages and disadvantages of each approach. They are a critical asset to uncover potential imbalances, but complications may arise when using multiple indicators. For instance, two indicators may exhibit negative correlations between them, yielding conflicting conclusions. Furthermore, the same indicator may present negative autocorrelations between observations in different moments in time, complicating its interpretation. Finally, indicators may also be changing as a response to cyclical adjustments or a temporary peak in demand, and as a result may not correctly portray the structural trend in the long term.

In the end these indicators do not provide a definite answer to the problem, but rather a predictor on the likelihood of a shortage or a surplus—the HHR planner's canary in a coal mine. Continuous monitoring and calibration may enhance the quality of the assessment. In fact, the repeated incidence may in fact suggest a distortion in the market, originating more robust results. Consider, for instance, a country where the following is observed at the same time in the health sector: rising wages; low unemployment; high or persistent vacancies; increasing use of overtime work; increasing use of temporary workers; waiting lists; increasing use of non-traditional workers (retirees, foreign doctors, etc.). These relevant indicators are highly suggestive of a shortage [31].

It is also worth noting that there is a positive correlation between health worker density and positive health status and outcomes [8]. Such empirical evidence helps to support the proposition that a shortage and a surplus have different degrees of importance, as we have already stressed in the beginning of this chapter. In this sense, if an exact balance is difficult or impossible to attain, it is better to err on the side of caution and have a skilled and available workforce to attend to the needs of the population at any time rather than a shortage of workers. That is even more so in the extreme case of pandemics, where all the assistance from health care professionals is critical.

1.2 Research objectives and methodology

1.2.1 Objectives

Contrarily to some other research fields, where a set of somewhat stable methodologies is used to solve varying instances of conceptually similar problems, in HHR planning no preferred approach or technique has been established yet. This may be explained by several factors, including an extremely long planning horizon, which makes it difficult to validate previous forecasts; the high degree of uncertainty and complexity of the problem; the market rigidities found in the health labour market, which move it away from standard Walrasian equilibrium economics; or policy decisions and interventions that are extremely hard to model, and that may invalidate previous forecasts.

Some methods have stood the test of time and others have been discarded. It is not obvious whether some have become obsolete on the grounds of technical limitations, or simply because there was not enough computational power at that time. Also, given the interdisciplinary nature of the problem, it may be the case that a robust approach requires an integrated model, combining different methods. Therefore, we need an extensive literature review to come to grips with the extent of methods in use in the field, since existing literature reviews are both narrow and outdated. This leads us to the first objective, arising from the following research question:

Q1 Which methodologies exist and are still used for HHR planning?

A preliminary glance at the literature highlighted the fact that methodologies are usually supply-, demand- or needs-based. The trend has been towards greater integration, as both supply and demand impact the health labour market. Of all the methods usually employed, simulation has been perceived as the most adequate to address the problem [26], since it is capable of handling all the relevant dynamics without requiring relaxing important assumptions [14]. Although simulation, in particular System Dynamics, has been used extensively, it lacks important features required to pursue further integration. By working with aggregates it makes it extremely difficult, if not impossible, to model complex individual behaviour. In such case, incorporating microeconomic fundamentals in the agents is impossible, rendering it useless to study problems at the micro level, but the health labour market is in fact a congregation of individual, institutional and business agents. Agent-based simulation (ABM), despite being a newer technology, does not suffer from this. However, no ABM model has been proposed to forecast HHR. This gave origin to the next research question covered in this thesis:

$\boxed{\textbf{Q2}}$ How can HHR planning be modelled using agent-based simulation?

From both the analysis of the literature and some exploratory modelling where all the relevant inputs were factored in, it came to light that the uncertainty over some inputs may generate very disparate forecasts. For instance, emigration may give origin to large variations in the workforce projections. The government has the possibility of stepping in in order to prevent imbalances from sudden emigration. However, these questions remain understudied. Which public policies to endorse, if any, and their impact on the workforce, either to attain a balanced workforce or to retain practitioners, are questions worthy of attention. From this observation another research question was raised:

$\boxed{\textbf{Q3}}$ Should government promote retention policies to mitigate or prevent the emigration of practitioners, and what is the impact of such policies on the workforce?

Finally, the literature review also brought to evidence that cross-country comparisons are used routinely for HHR planning, perhaps due to their simplicity and low data requirements. It is very frequent to find raw comparisons between worker-to-population ratios, a practice that obfuscates all the complexity and heterogeneity of health systems. The aforementioned case of the UK, where a policy intervention was enacted based solely on the observation that the domestic physician-to-population ratio was low compared to the OECD average, illustrates the damage a misleading analysis may cause [6].

Some attempts have been made to group countries by similarity, in this way reducing the error-proneness of random comparisons or comparisons to the average [25, 27]. In one case, countries are grouped by efficiency scores, according to a measure of a mean score by health outcome, i.e. countries obtaining similar outcomes (life expectancy, amenable mortality rate, etc.) are grouped together. In another case, countries are grouped based on institutional typology and health policies, i.e. countries that share broadly similar health system architectures in terms of provision, financing and regulation. None of these studies, however, group countries based on their inputs, both in terms of physical and human resources, and also of medical needs of the population. We then arrive at the last research question:

Q4 How can cross-country comparisons be conducted in a robust way, so that benchmarking can inform decision-makers properly and help elaborate policies?

1.2.2 Methodology

HHR planning is a broad area, requiring inputs from several research fields. On the supply-side, a snapshot of the workforce over time is necessary to understand whether it is expanding, contracting or remains constant. This forecast will provide a rough estimate of how many practitioners are integrated in the workforce, but also of how many may be available for hire. Furthermore, how the workforce actually delivers health care services, the pace of technological evolution, the underlying organizational structure and clinical practices all impact both productivity and outcomes, which should be also assessed.

On the demand-side, economic growth, demography and *needs* may all influence the demand for health care services. Economic growth is relevant since the demand for health care services, or for any other good or service, changes following variations to disposable income. Demography is also critical at two levels: first, to gauge whether the population will expand or decrease; second, to account for population ageing, which puts additional stress on the health system. Finally, *needs* somehow reflect the population's morbidity, in particular incidence and prevalence rates, the evolution of chronic diseases, etc. On top of all this, health systems have a significant presence of the State, with institutional stakeholders, the government and several other economic agents exerting an influence over power, which requires also studying the problem from a public policy perspective.

These subproblems are large and relevant enough to have their own scientific agenda. When we pin down each of the scientific areas and methods pertaining to HHR planning, as depicted in Figure 1.3, it becomes clear its broadness requires an interdisciplinary approach to capture all its dimensions.

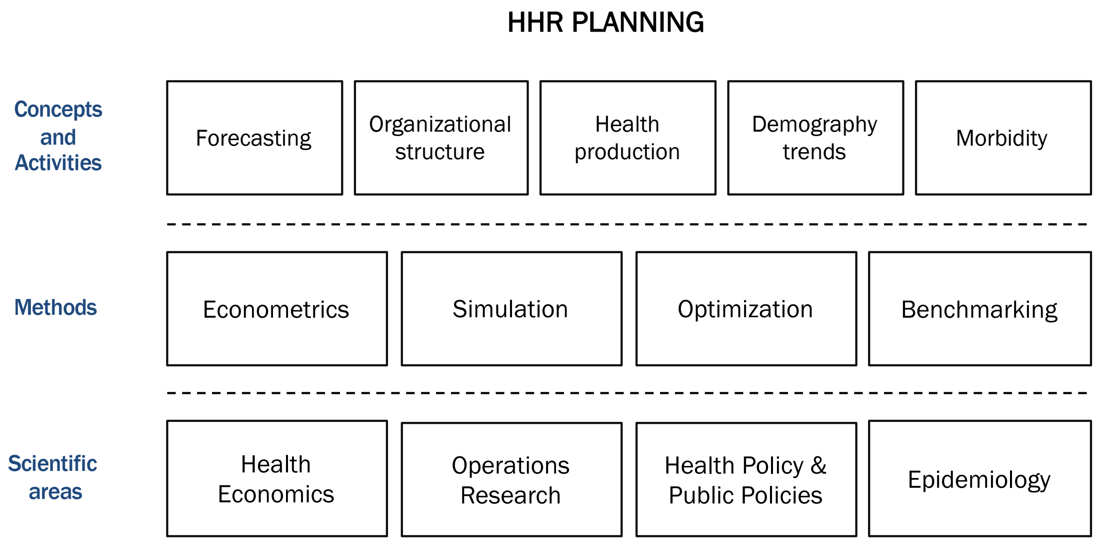


Figure 1.3: An integrated system that incorporates several methodologies to address the many facets of HHR forecasting.

First and foremost, HHR planning concerns the allocation of scarce resources, matching supply and demand, and therefore it is an economic problem. Thus, it requires knowledge stemming from economics and health economics. Models on health labour markets and health systems; health outcomes and efficiency; international migration; as well as microeconomic foundations for decision-making, preferences and choices; production functions; productivity; etc. are all critical concepts the economic field has been studying for long.

Operations research also plays a significant role, providing the tools to approach the problem. The complexity involved in forecasting the supply of and the demand for health care services is such that econometric models may be unable to incorporate all the non-linear and stochastic traits. Therefore, computational methods, such as agent-based simulation, are perhaps a better fit. In fact, this recent tendency to employ agent-based computational economics (ACE) has given origin to a scientific strand of its own.

Big Data tools and methods, such as data mining, can also be extremely useful in the analysis of health systems, and in providing a solid methodology to group countries featuring very particular characteristics in the way health care services are delivered, so that further cross-country comparisons may then be drawn. Actually, these tools are nowadays broadly used for making sense of the huge amounts of health data that is becoming available, both from a clinical perspective, which falls outside the scope of this work, and from an health systems perspective, as it can help in exploring knowledge contained in data from hospitals or any other health care facility.

Finally, all these research questions have policy implications. We therefore require an apparatus from a public policy perspective to frame both the problems and potential interventions. Health and public policy are fields that may also contribute method-wise.

1.3 Thesis synopsis

We have chosen to elaborate this thesis as a collection of papers in publication and papers in review by international peer-reviewed journals instead of a monograph. This decision entails pros and cons. On the one hand, it reports material already in print that has been subject to peer review and improvements. On the other hand, some introductory parts may be repetitive. With this in mind, each of the remaining chapters correspond to a paper.

Figure 1.4 provides an overview of how all the chapters collate. Chapter 2 reports the results of the literature review that gave origin to a functional proposal of an integrated model (depicted in Figure 1.4), as well as some new and interesting research questions and gaps. The literature review uncovered some research gaps, and helped in making sense of the plethora of methodologies employed in HHR planning. From supply- to demand- to needs-based methods, all are in fact required to obtain a good estimate of workforce imbalances. Our contribution was therefore synthesising the approaches reported in several publications, providing a functional description of a conceptual integrated model.

From this introductory work followed Chapter 3, an actual contribution of a model to forecast physicians' requirements using agent-based simulation, and also to assess the

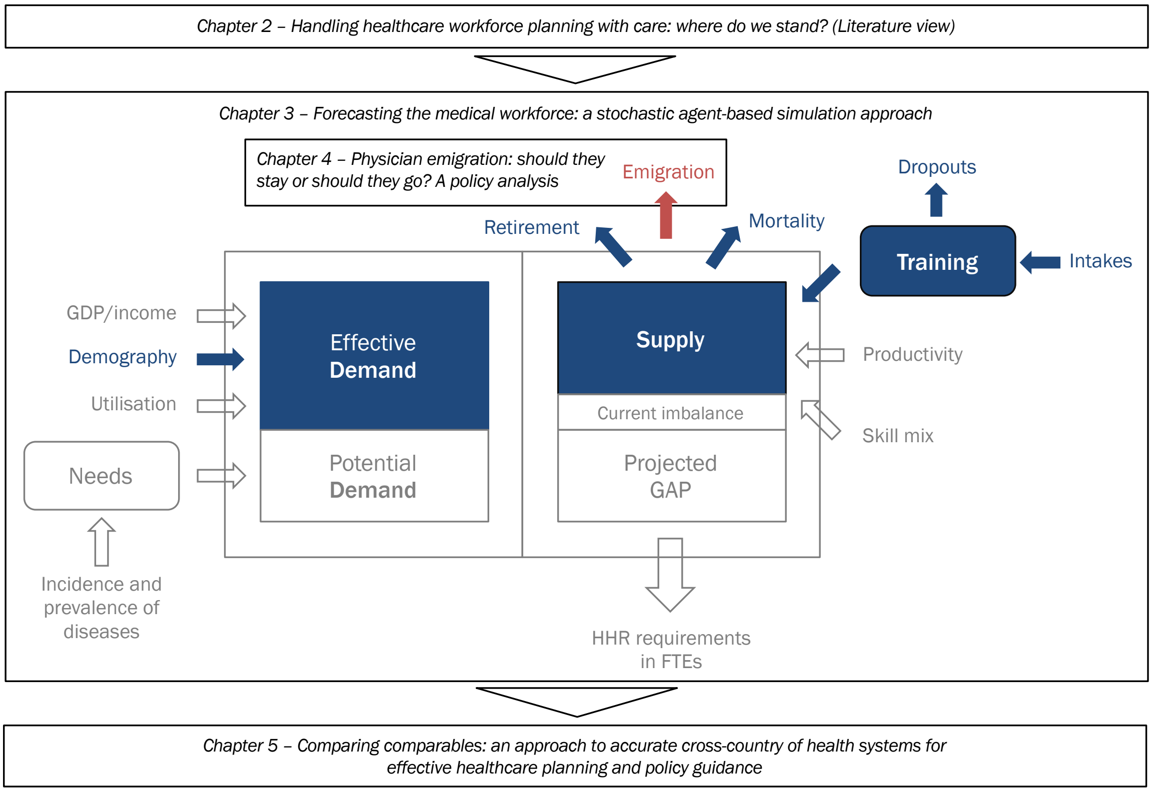


Figure 1.4: Thesis synopsis. Each chapter gave origin to a publication.

impact of demographic changes. The model was used to forecast the Portuguese physician workforce, including its 52 specialties. The contribution was twofold: first, an implementation of an agent-based model to forecast the physician workforce; second, actual projections of the Portuguese physician workforce based on real data, and also of the impact expected demographic changes may bring.

Chapter 4 derived directly from the work developed in the previous chapter. It came to light that emigration may influence the projections significantly, and it may be the case that intervention is required to prevent further “brain drain”, especially in case of a shortage. Consequently, we conduct some policy experiments on a concise model of the health labour and migration markets that emerge from individual agents with decision-making capabilities. Our contribution is in the field of agent-based computational economics applied to health care, where we provide a model of a non-Walrasian health system to study the impact of policy interventions on the workforce.

Finally, Chapter 5 attempts at tapering a research gap brought to light in the literature review, that is, the sometimes flighty way cross-country comparisons are conducted. We propose a methodology to group countries sharing similar traits in terms of their health systems, so that *ex post* comparisons may then be drawn in a more robust way. The contribution is methodological, where we show how to use data mining tools, such as k-means and hierarchical clustering to group similar countries, and illustrate its application by grouping OECD countries.

1.4 Conclusion

HHR planning still has a long way to go data- and methodology-wise. This thesis was a small contribution to push its frontier, supplying new methodologies to approach the problem in its many dimensions, and helping to bridge the gap between academic research and policy-making.

In summary, we have contributed with a literature review that tries to synthesise the best academics and practitioners have produced. Although methodologies abound, there is still no definite approach to address HHR planning. This work gave origin to a functional description of an integrated model that combines supply-, demand- and needs-based methods, which we hope may lay the foundations for further work. Based on this we implemented a model to actually forecast the supply of physicians using stochastic agent-based simulation, contributing with both a robust method and a practical application. The model was used to project the Portuguese physician workforce, and also to estimate the impact demography may have on the demand for health care services, and it constitutes a solid starting point for incorporating microeconomic foundations. In fact, we did so in the next publication, where we studied the impact emigration may have on the physician workforce. We implemented a simple economic model using agent-based computational economics to study and experiment with policies. The contribution was again both methodological and practical, as the model was applied to the Portuguese physician workforce, providing relevant insights. Finally, we also contributed with a framework to group countries that share similar health systems input-wise, i.e. human and capital resources, incidence and prevalence of diseases, etc. This contribution may help in conducting more robust cross-country comparisons, ensuring *comparable* countries are being compared.

We believe this thesis is a contribution to the scientific community and to the advancement of methodologies and tools available for conducting sound HHR planning, but also a direct contribution to society.

1.4.1 Contributions

The work underlying this thesis instigated several scientific contributions at various levels. The research agenda and the results obtained gave origin to four scientific papers published and to be published in international peer-reviewed scientific journals; two working papers; one book chapter; over ten communications at international conferences; a participation in the EU Joint Action on Healthcare Workforce Planning and Forecasting; and a project research grant, HHRPLAN (Grant No PTDC/IIMGES/4770/2014), developed in co-authorship and financed by FCT (Fundação para a Ciência e Tecnologia) that builds upon the foundations laid by some of the work developed, and bringing together several national and international institutions, such as Universidade do Porto, Instituto Superior Técnico, Administração Central do Sistema de Saúde (ACSS) and Cardiff University.

In terms of the thesis herein presented, in Chapter 1 we provided a brief overview of the research field, in particular the challenges involved in HHR planning, the research questions and the several methodologies employed to address the problem. One of the critical points identified was in assessing the current imbalances the workforce may face, as ignoring an existing gap may lead to that imbalance being propagated throughout the projections. We conducted a brief analysis of how these imbalances may be detected using both economic and operational indicators from health care facilities. This work resulted in a book chapter published in the handbook of the EU Joint Action on Healthcare Workforce Planning and Forecasting, an EU-funded research project on the topic:

1. Amorim Lopes, Mário, Álvaro Santos Almeida, and Bernardo Almada-Lobo. 2015. "Assessing the Current Situation." In *Handbook on Health Workforce Planning methodologies across EU countries*, 296–306. EU Joint Action Health Workforce Planning and Forecasting. ISBN 978-80-89825-00-4

Chapter 2 contributed with a literature review, covering a very large period that goes as far as 1950. We elaborated a list of methodologies that have been used in HHR planning since its early stages as a research field, highlighted a growing tendency to use integrated approaches that bring together supply-, demand- and needs-based methodologies, and based on this we proposed a functional representation of an hypothetical integrated model. This preliminary work paved the way for future research, uncovering gaps. For instance, it showed that existing methods to model the supply of HHRs typically follow a systems-level approach, therefore being unable to model at the individual level. Also, it brought to evidence how frequently *naif* approaches are used to inform policy makers. Comparisons against the average, a very typical benchmark tool, are not only highly misleading, but may also be detrimental to the health system, as the documented case has shown. This Chapter culminated in the following publication:

2. Amorim Lopes, Mário, Álvaro S Almeida, and Bernardo Almada-Lobo. 2015. "Handling Healthcare Workforce Planning with Care: Where Do We Stand?" *Human Resources for Health* 13 (1). BioMed Central Ltd: 38.
doi:10.1186/s12960-015-0028-0.

Chapter 3 is an actual proposal of a stochastic agent-based simulation model to project the evolution of the physician workforce. The contribution was both methodological, as we have described and implemented a stochastic ABM capable of forecasting the physician workforce; and practical, as we have used the model to project the supply of Portuguese physicians. We also took into account the impact on the demand for health care services demography may bring, and addressed potential sources of uncertainty by running Monte Carlo simulations that subject parameters to random variations. Results have shown that although the supply of physicians remains fairly constant, *ceteris paribus*, demography may play a significant role. On the one hand, the population is expected to decline, which could suggest less health care services to be required. On the other hand, however, ageing takes its toll, and our results suggest it outweighs the size effect. Hence, despite declining, ageing plays a larger effect, leading to an increased demand for health care services at current productivity levels. The work lead to an article being published in the Journal of Health Care Management Science:

3. Amorim Lopes, Mário, Álvaro Santos Almeida, and Bernardo Almada-Lobo. 2016. "Forecasting the Medical Workforce: a Stochastic Agent-Based Simulation Approach." *Health Care Management Science*. Springer US, 1–24. doi:10.1007/s10729-016-9379-x.

The previous chapter also highlighted a relevant research problem related to emigration. The Monte Carlo simulations conducted before brought to evidence that emigration is indeed a source of uncertainty, as small variations in the emigration rate may lead to large variations in the projections. The problem is further aggravated if the health labour market is short of medical practitioners, in which case emigration can further add to the problem. Traditional equilibrium models of international emigration barely apply to a market as rigid and atypical as the health labour market, and thus analysing the impact of retention policies to mitigate the impact of emigration requires a novel approach. To this end we have devised an agent-based computational economics model, with heterogeneous individual agents with decision-making capabilities, from which an health labour and a migration market emerge, which we have described extensively in Chapter 4. We used this model to experiment with different policy levers, gauging their impact on the Portuguese physician workforce. The model and the ensuing results gave origin to a working paper, later adapted and submitted to an international, peer-reviewed journal for publication:

4. Amorim Lopes, Mário, Álvaro Almeida, and Bernardo Almada-Lobo. "Physician emigration: should they stay or should they go? A policy analysis." *Submitted to an international journal for publication. In review.*.

Chapter 5 is the last scientific contribution of the thesis. It proposes a methodology for grouping countries based on several characteristics of the health systems, and applies it to the OECD countries. This work was born out of a research gap identified previously in the literature review (Chapter 2). It became clear that simple cross-country comparisons against the average, or even against individual countries, are very error-prone, potentially giving origin to misleading conclusions. Therefore, a more robust approach was needed to perform cross-country comparisons and derive more robust insights from such a useful tool. Note that this is extremely relevant for the topic of this thesis, as most of these country benchmarks were performed to gain insight on HHR planning, in particular on whether the number of physicians or nurses deviates much from OECD's average—a misleading comparison, given the differences in health systems. The resulting work was first published as a working paper and then submitted for publication to the Health Systems journal:

6. Amorim Lopes, Mário, Carlos Soares, Álvaro Almeida, and Bernardo Almada-Lobo. 2016. "Comparing Comparables: an Approach to Accurate Cross-Country Comparisons of Health Systems for Effective Healthcare Planning and Policy Guidance." *Health Systems*, February. doi:10.1057/hs.2015.21.

1.4.2 Further work

Much more remains to be done. In the course of this thesis we faced many challenges, and the work developed unleashed a great number of research questions that remain unanswered either due to time constraints or to lack of data. For instance, there are no historical records available for the stock of Portuguese physicians, with the existing data going no further than 2013. Moreover, the information available reports on the public sector only. Since the training process takes at least 10 years to complete (6 years at medical school plus 3 to 5 years at a residency program), running a backwards simulation in order to fully validate the model is not feasible. Using data from other countries is also not possible, as that would require changes to the model due to the particularities of each health system, and thus the model being validated would be a different one. Furthermore, some input data, such as the emigration rate, the number of physicians practicing in the private sector or the number of physicians accumulating with the private sector, is simply unknown. When we developed the model we tried to overcome these limitations by first recognizing there is a lot of uncertainty surrounding these inputs, and then by running Monte Carlo simulations, which provide an insight on how variations in these unknown input parameters may impact the forecasts. In fact, we aimed at providing several possible outcomes depending on how these parameters may vary, other than obtaining a definite projection of the physician workforce. We are currently conducting expert elicitation workshops with key stakeholders in order to gauge the interval within each these inputs may likely vary. Such insights will help to improve the calibration of the model.

Another limitation of the present work has to do with regional asymmetries, which have not been addressed. We are well aware a gross number reporting the physicians headcount, even when the number is arguably large, may hide huge disparities within the health system, meaning that in some places there may be a surplus and in others a shortage. Physicians seem to prefer to practice in urban, more dense areas, either because they were born there, had to move there in order to pursue a medical career, or because opportunities to practice in the private sector are more common. Regardless of the motives it leaves the interior underserved. There is a lot of literature on policies to mitigate some of these effects and encourage physicians to settle and practice in the interior, but less so to forecast how these regional imbalances may evolve under these policies. Adding to this, recent evidence has shown that nurses with advanced studies seem to be an effective replacement to general practitioners at primary care centres, with no discernible impact on health outcomes, on the contrary [18, 23]. If that is the case, physician-nurse task shifting could help to alleviate some of the stress put on the health system, or address some missed care due to the lack of human resources. Skill mix is therefore a very relevant driver that also needs to be accounted for in the projections. In such case, independent projections of both the physician and the nursing workforce are pointless—an integrated forecast, where skill mix is taken into account, is therefore required.

Considering this, a model to forecast the nursing workforce would be needed so it can be further extended to study skill mix. We have already illustrated how to use ABM to conduct such forecasts for the physician workforce. Adapting the model to the nursing workforce should require a minimal effort. It could then be used to study the conjoint evolution of the physician and the nursing workforce, and how the nursing workforce could compensate a potential shortage of physicians, all things considered. In fact, in some places graduated nurses have been replacing general practitioners with very successful health outcomes. If that trend continues and even intensifies, a decreasing physician workforce may then not be an issue.

Regardless of whether the future holds a surplus or a shortage of HHR, some intervention may be required to overcome the market rigidities that prevent the health system from self-adjusting. Being that the case, this work would most certainly benefit from an optimisation model capable of adjusting policy levers that will bring the stock of physicians, nurses or any other health professional back to desirable levels—when the gap between supply and demand is nil. Several approaches could be followed, with a Mixed Integer Programming model topping the list of candidates. However, many constraints apply, since the problem goes well beyond adjusting *numerus clausus*, as the work on emigration highlighted. Even if that were the case, vacancies cannot simply be opened or closed, running the risk of disrupting medical schools. The transition should be smooth, and the increase or decrease distributed throughout a larger time horizon. The problem is not only operational, but also economical, political and social. A reflection on how all these different schools of thought can be integrated is also well needed.

Some of the issues arising from this thesis and mentioned previously will be targeted by a research project funded by FCT that will further extend the work developed, encompassing the models developed in a larger, broader decision support system (DSS) capable of informing policy makers. Figure 1.1 provides some hindsights on the most significant points and Figure 1.5 details the features the DSS should have. The integrated simulation model is the backbone upon which several other features will be added. In particular, the DSS should be able to provide scenario analysis and foresight using inputs from experts in the field. For instance, whether the burden of a particular pathology will increase or decrease may influence how demand evolves, and therefore how supply should match the demand for health care services. Also, and since policy objectives are met using instruments, a MIP-based optimisation model is to be incorporated in order to adjust policy levers such as the number of vacancies. Despite constituting an adequate tool for assisting technical discussions, other economic and political decisions still stand, and therefore a political stance is not to be dismissed. It can, nonetheless, help in promoting evidence-based public policies.

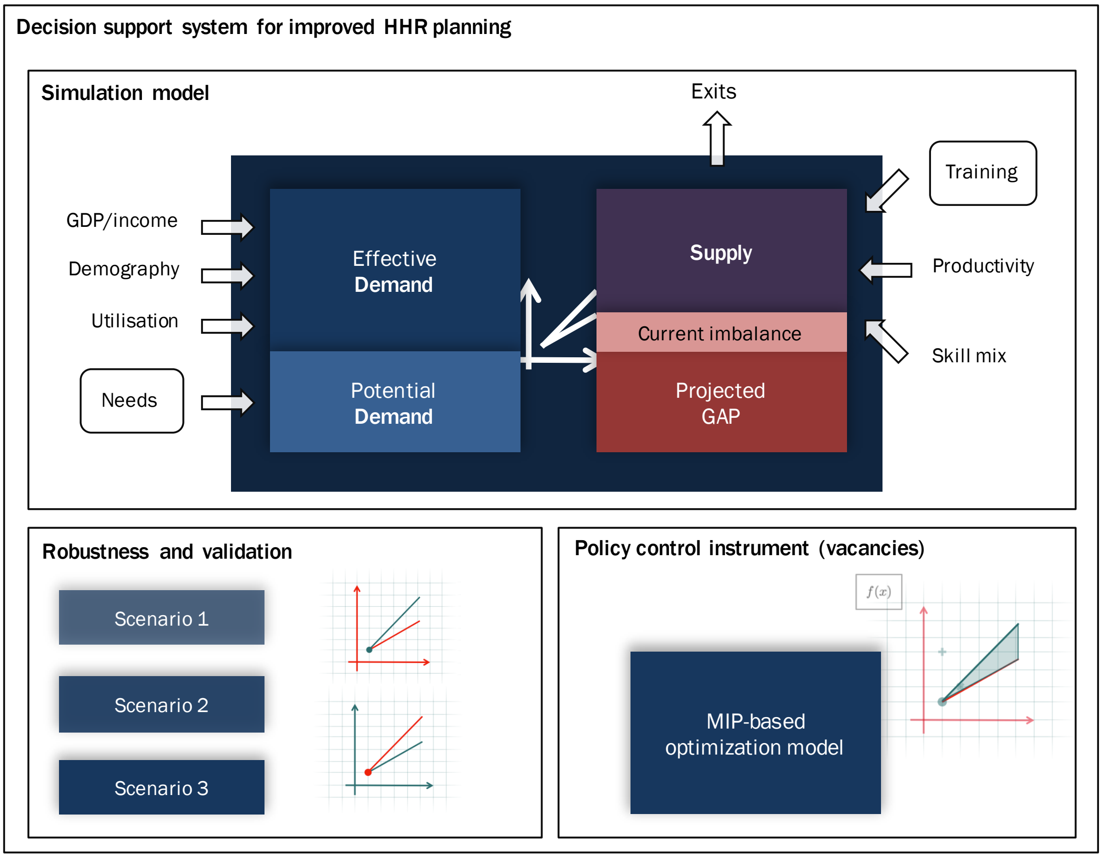


Figure 1.5: A decision support system for HHR planning.

Another point deserving appropriate attention and research has to do with productivity. Although the health care sector seems to suffer from Baumol's cost disease, as real wages do not grow pari-passu with labor productivity growth, disruptive technological changes will eventually affect HHR productivity, and in some cases replace them. Robot surgeons, Big Data applied to disease diagnosis, and so on, will definitely impact the way health care is delivered. It already has. Very recently, researchers have built an artificial intelligent algorithm based on neural networks capable of outperforming 21 dermatologists on diagnosing skin cancer [11]. If computer-based diagnosis is the future, perhaps specialties such as dermatology will be less needed. Coincidently, our forecasts show that the number of dermatologists may increase in the future. More generally, the internet revolution (one single human cannot compete with loads of data reporting diagnostics from thousands to millions of individuals distributed throughout the world) and these recent events may be the long needed remedy to cure Baulmol's disease. All this add to the already difficult task of forecasting HHR, but are critical for assessing how many health care services a physician or a nurse can deliver. Sensitivity analysis and scenario building may be a starting point, but perhaps more foresight analysis is needed. Again, the forecasts may be a good starting point for fostering discussion.

Finally, the demand for health care services has been overlooked. Although we have incorporated demographic projections to the model detailed in Chapter 3, which have allowed us to project how current demand levels may be affected by the population ageing, *needs* also play a very critical role, as they are the main driver that push people to seek for health care services. If people, elders included, did not have a condition requiring medical intervention, they would most certainly require less attention from health professionals, bar some routine exams. Addressing this requires dealing with two critical questions: first, how to quantify by age and pathology the health care services being delivered, as this will help in building a demand curve; second, how to estimate the HHR needed to provide those specific health care services at current productivity levels (in short, estimate a production function; since there are multiple inputs and multiple outputs, the most likely approach is to estimate an input requirements function). We can then invite epidemiologists to forecast how incidence and prevalence of certain diseases will evolve, and based on that inform HHR planning accordingly. A real example helps to illustrate the challenge. In China, type II diabetes now affects 10% of the population, compared to 1% 10 years before. Arguably more endocrinologists will be needed, but how many? This could well be the birth of a whole new thesis.

1.5 Bibliography

[1] Mário Amorim Lopes, Álvaro S Almeida, and Bernardo Almada-Lobo. Handling healthcare workforce planning with care: where do we stand? *Human Resources for Health*, 13(1):38, May 2015.

[2] Mário Amorim Lopes, Álvaro Santos Almeida, and Bernardo Almada-Lobo. Assessing the current situation. In *Handbook on Health Workforce Planning*, pages 296–306. EU Joint Action Health Workforce Planning and Forecasting, September 2015.

[3] John P Ansah, Dirk De Korne, Steffen Bayer, Chong Pan, Thiyagarajan Jayabaskar, David B Matchar, Nicola Lew, Andrew Phua, Victoria Koh, Ecosse Lamoureux, and Desmond Quek. Future requirements for and supply of ophthalmologists for an aging population in Singapore. *Human Resources for Health*, 13(1):86, November 2015.

[4] Stephen Birch, Thomas Mason, Matthew Sutton, and William Whittaker. Not enough doctors or not enough needs? Refocusing health workforce planning from providers and services to populations and needs. *Journal of Health Services Research & Policy*, 18(2):107–113, April 2013.

[5] Karen Bloor and Alan Maynard. Approaches to human resources in health: an international comparative review. *Canadian Health Services Research Foundation*, pages 1–34, April 2003.

[6] Karen Bloor, Vivien Hendry, and Alan Maynard. Do we need more doctors? *Journal of the Royal Society of Medicine*, 99(6):281–287, June 2006.

[7] Teresa Cardoso, Mónica Duarte Oliveira, Ana Barbosa-Póvoa, and Stefan Nickel. Modeling the demand for long-term care services under uncertain information . *Health Care Management Science*, 15(4):385–412, July 2012.

[8] Lincoln Chen, Timothy Evans, Sudhir Anand, Jo Ivey Boufford, Hilary Brown, Mushtaque Chowdhury, Marcos Cueto, Lola Dare, Gilles Dussault, Gijs Elzinga, Elizabeth Fee, Demissie Habte, Piya Hanvoravongchai, Marian Jacobs, Christoph Kurowski, Sarah Michael, Ariel Pablos-Mendez, Nelson Sewankambo, Giorgio Solimano, Barbara Stilwell, Alex de Waal, and Suwit Wibulpolprasert. Human resources for health: overcoming the crisis. *The Lancet*, 364(9449):1984–1990, November 2004.

[9] Carl-Johan Dalgaard and Holger Strulik. The Economics of Health Demand and Human Aging: Health Capital vs. Health Deficits .

[10] Carl-Johan Dalgaard and Holger Strulik. Optimal aging and death: understanding the Preston curve. *Journal of the European Economic Association*, 12(3):672–701, June 2014.

[11] Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, January 2017.

[12] R Feldman and D Wholey. Do HMOs have monopsony power? *International journal of health care finance and economics*, 1(1):7–22, March 2001.

[13] Paul J Feldstein. Research on the Demand for Health Services. *The Milbank Memorial Fund Quarterly*, 44(3):128, July 1966.

[14] Gonçalo Figueira and Bernardo Almada-Lobo. Hybrid simulation–optimization methods: A taxonomy and discussion. *Simulation Modelling Practice and Theory*, 46: 118–134, August 2014.

[15] Michael Grossman. On the Concept of Health Capital and the Demand for Health. 80 (2):223–255, January 1972.

[16] T Hall and A Mejia. Health Manpower Planning: Principles, Methods, Issues . pages 1–146, February 1978.

[17] Mylene Lagarde and John Cairns. Modelling human resources policies with Markov models: an illustration with the South African nursing labour market. *Health Care Management Science*, 15(3):270–282, November 2011.

[18] Miranda Laurant, David Reeves, Rosella Hermens, Jose Braspenning, Richard Grol, and Bonnie Sibbald. Substitution of doctors by nurses in primary care. *Cochrane Database of Systematic Reviews*, (2), 2005.

[19] Caroline O Laurence and Jonathan Karnon. Improving the planning of the GP workforce in Australia: a simulation model incorporating work transitions, health need and service usage. *Human Resources for Health*, 14(1):1, April 2016.

[20] Mariel S Lavieri, Sandra Regan, Martin L Puterman, and Pamela A Ratner. Using Operations Research to Plan the British Columbia Registered Nurses' Workforce. *Healthcare Policy*, 4(2):e117–35, November 2008.

[21] J N Lavis and S Birch. The answer is ...., now what was the question? Applying alternative approaches to estimating nurse requirements. *Canadian journal of nursing administration*, 10(1):24–44, January 1997.

[22] L B Lerer, A D Lopez, T Kjellstrom, and D Yach. Health for all: analyzing health status and determinants. *World health statistics quarterly. Rapport trimestriel de statistiques sanitaires mondiales*, 51(1):7–20, 1998.

[23] Nahara A Martínez-González, Ryan Tandjung, Sima Djalali, and Thomas Rosemann. The impact of physician–nurse task shifting in primary care on the course of disease: a systematic review. *Human Resources for Health*, 13(1):55, July 2015.

[24] Kostas Mavromaras, Josh Healy, Sue Richardson, Peter Sloane, Zhang Wei, and Rong Zhu. A System for Monitoring Shortages and Surpluses in the Market for Skills. pages 1–72, February 2013.

[25] João Medeiros and Christoph Schwierz. Efficiency estimates of health care systems. *Economic Papers 549*, June 2015.

[26] Linda O'Brien-Pallas, Andrea Baumann, Gail Donner, Gail Tomblin Murphy, Jacquelyn Lochhaas-Gerlach, and Marcia Luba. Forecasting models for human resources in health care. *HEALTH AND NURSING POLICY ISSUES*, pages 1–10, January 2001.

[27] OECD. Health care systems: Getting more value for money. Technical Report 2, 2010.

[28] Tomoko Ono, Michael Schoenstein, and Gaetan Lafortune. Health Workforce Planning in OECD Countries. Technical report, June 2013.

[29] P A Prescott. Forecasting requirements for health care personnel. *Nursing Economics*, 9(1):18–24, January 1991.

[30] P Raina, B Bonnett, D Waltner-Toews, C Woodward, and T Abernathy. How reliable are selected scales from population-based health surveys? An analysis among seniors. *Canadian journal of public health = Revue canadienne de sante publique*, 90(1):60–64, January 1999.

[31] Susan Sue Richardson. What is a skill shortage? 2009.

[32] Robert A Rizza, Robert A Vigersky, Helena W Rodbard, Paul W Ladenson, William F Young, Martin I Surks, Richard Kahn, and Paul F Hogan. A Model to Determine Workforce Needs for Endocrinologists in the United States Until 2020*.

[33] Barry J Seldon, Chulho Jung, and Roberto J Cavazos. Market power among physicians in the U.S., 1983–1991. *The Quarterly Review of Economics and Finance*, 38(4): 799–824, January 1998.

[34] Douglas Staiger, Joanne Spetz, and Ciaran Phibbs. Is There Monopsony in the Labor Market? Evidence from a Natural Experiment. July 1999.

[35] Sabine Stordeur and Christian Leonard. Challenges in physician supply planning: the case of Belgium. *Human Resources for Health*, 8(1):28, December 2010.

[36] G Tomblin Murphy, A MacKenzie, R Alder, S Birch, G Kephart, and L O'Brien-Pallas. An Applied Simulation Model for Estimating the Supply of and Requirements for Registered Nurses Based on Population Health Needs. *Policy, Politics, & Nursing Practice*, 10(4):240–251, February 2010.

[37] Gail Tomblin-Murphy, Stephen Birch, Adrian MacKenzie, and Janet Rigby. Simulating future supply of and requirements for human resources for health in high-income OECD countries. *Human Resources for Health*, 14(1):853, December 2016.

[38] Yusuke Tsugawa, Anupam B Jena, Jose F Figueroa, E John Orav, Daniel M Blumenthal, and Ashish K Jha. Comparison of Hospital Mortality and Readmission Rates for Medicare Patients Treated by Male vs Female Physicians. *JAMA Internal Medicine*, December 2016.

[39] Malou Van Greuningen, Ronald S Batenburg, and Lud FJ Van der Velden. Ten years of health workforce planning in the Netherlands: a tentative evaluation of GP planning as an example. *Human Resources for Health*, 10(1):1–1, August 2012.

[40] Malou Van Greuningen, Ronald S Batenburg, and Lud FJ Van der Velden. The accuracy of general practitioner workforce projections. *Human Resources for Health*, 11(1):1–1, July 2013.

[41] J P Weiner. Forecasting the effects of health reform on US physician workforce requirement. Evidence from HMO staffing patterns. *JAMA : the journal of the American Medical Association*, 272(3):222–230, July 1994.

[42] World Health Organization. *The World Health Report 2006: Working Together for Health*. The world health report. World Health Organization, 2006.

[43] Pascal Zurn, Mario R Dal Poz, Barbara Stilwell, and Orvill Adams. Imbalance in the health workforce. *Human Resources for Health*, 2(1):13, 2004.

Chapter 2

Handling healthcare workforce planning with care: where do we stand?[1]

[1]Published in the Journal of Human Resources for Health (BioMed Central) with DOI 10.1186/s12960V015V. URL http://www.human-resources-health.com/content/13/1/38/abstract

Amorim Lopes *et al. Human Resources for Health* (2015) 13:38
DOI 10.1186/s12960-015-0028-0

HUMAN RESOURCES FOR HEALTH

REVIEW Open Access

Handling healthcare workforce planning with care: where do we stand?

Mário Amorim Lopes[1*], Álvaro Santos Almeida[2] and Bernardo Almada-Lobo[1]

Abstract

Background: Planning the health-care workforce required to meet the health needs of the population, while providing service levels that maximize the outcome and minimize the financial costs, is a complex task. The problem can be described as assessing the right number of people with the right skills in the right place at the right time, to provide the right services to the right people. The literature available on the subject is vast but sparse, with no consensus established on a definite methodology and technique, making it difficult for the analyst or policy maker to adopt the recent developments or for the academic researcher to improve such a critical field.

Methods: We revisited more than 60 years of documented research to better understand the chronological and historical evolution of the area and the methodologies that have stood the test of time. The literature review was conducted in electronic publication databases and focuses on conceptual methodologies rather than techniques.

Results: Four different and widely used approaches were found within the scope of supply and three within demand. We elaborated a map systematizing advantages, limitations and assumptions. Moreover, we provide a list of the data requirements necessary to implement each of the methodologies. We have also identified past and current trends in the field and elaborated a proposal on how to integrate the different methodologies.

Conclusion: Methodologies abound, but there is still no definite approach to address HHR planning. Recent literature suggests that an integrated approach is the way to solve such a complex problem, as it combines elements both from supply and demand, and more effort should be put in improving that proposal.

Keywords: Review, Health-care workforce planning, Supply, Demand, Needs, Health policy

Introduction

Health-care human resources (HHR) planning has been identified as the most critical constraint in achieving the well-being targets set forth in the United Nations' Millennium Development Goals [1]. Moreover, the effective use and deployment of personnel is paramount to ensure an efficient service delivery in terms of cost, quality and quantity [2]. Failure to do so may result in an oversupply or shortage of clinical staff. While the former may lead to economic inefficiencies and misallocated resources under the guise of unemployment [3] or inflated costs through supplier-induced demand [4], the latter is linked to a more extensive list of negative effects, including but not limited to the following: lower quantity and quality of medical care as few resources exist to provide the necessary services and the visits are shorter [5]; work overload of the available physicians and nurses, resulting in sleep-deprivation, ultimately compromising patient safety [6]; and queues and waiting lists resulting from insufficient medical staff, causing avoidable patient deaths [7].

Another argument supporting HHR planning is the recent rise in health-care expenditure, both in *per capita* spending on health and as a proportion of *per capita* domestic product in real terms [8]. The average annual growth rate of health-care expenditure in a selection of 18 countries that are part of the Organisation for Economic Co-operation and Development (OECD) was 3.0 % between 1980 and 1990 and 3.3 % in the decade after [8]. Recent studies confirm the rising trend, with health spending growing at an average of 3.8 % in 2008 and 3.5 % in 2009 [9], well above the growth rate of the gross domestic product. Health worker wages account for

*Correspondence: mario.lopes@fe.up.pt
[1]INESC TEC, Faculdade de Engenharia, Universidade do Porto, Porto, Portugal
Full list of author information is available at the end of the article

about 50 % of total public and private health expenditure across several countries [5], meaning that cost containment and efficiency improvements will necessarily require the involvement of the workforce.

In sharp contrast to other scientific areas where a set of well-defined methodologies and techniques is generally adopted and refined to solve a given problem, in HHR planning, *methodologies* (the conceptual scope of analysis) and *approaches* (the techniques applied upon a particular method) abound, and there is still no commonly accepted or favoured procedure to accurately forecast physician requirements [3, 10]. The methodologies followed by countries vary significantly, in some cases with no long-term strategic HHR planning at all, but a wide array of options does not seem to be a determining factor in improving the accuracy of forecasting [11]. Despite the lack of focus, the accuracy of the projections appears to be making progress in some cases, as a review reporting the case of The Netherlands shows [12], an encouraging sign to the ongoing research.

A definite approach to the problem, or at least a stable starting block, will require a comprehensive overview of how the problem has been tackled since its inception. For this purpose, we provide a thorough analysis of the field, to lay down the foundations for future research, coupled with a historical perspective on the development of the HHR literature, analysing how the field has evolved and what methodologies have emerged and continue to be employed. Secondly, we analyse the strengths and pitfalls of each of the methodologies and provide a data requirement framework containing all the variables and data that need to be taken into account in order to address the problem thoroughly. The review is selective as it focuses primarily on articles that seem to have had a clear impact on the evolution of the field, although broad in scope as it attempts to extensively describe all known methods. Finally, it describes where we stand and the road ahead, providing a brief overview of new and emerging approaches to the HHR planning problem.

To the best of our knowledge, the last comprehensive academic paper on the subject dates back to 1978 [13]. Literature reviews exist but tend to either focus on a particular period or on a subset of the methodologies or techniques [11, 14] or to be framed as technical reports aimed at a wider readership, such as the OECD's extensive review of 26 projection models used in 18 countries [9] or WHO's policy recommendations to the EU [15]. The literature reviews can also consist of a technical report targeting a country in particular [16]. In fact, some authors point out that more systematic reviews, assessments of potential interventions and further research to aid policy makers are highly needed [17]. This paper aims to narrow this gap by being a starting point both for academics and policy makers.

Literature search method

We carried out an extensive literature review, including academic research papers and technical reports from institutions such as the OECD or WHO. Selected papers date between 1951 and 2013, and the results were reported in a chronological and evolutionary way so as to clearly identify methodologies that are still in use to this day. The search methodology can be summarized as follows: after selecting a set of search terms and generating reliable combinations, we used electronic research databases to search for related articles. We then selected a maximum of 20 papers for each combination of search terms, including the 10 most cited, the 5 most recent and 5 that were randomly chosen. A backward/forward search was conducted, and the abstract was analysed to ensure that the papers met the search criteria. Papers that failed to meet any of the search criteria were excluded.

To identify search terms, we consulted the available literature reviews and technical reports [5, 10, 11, 13] so as to a obtain a list of key terms frequently used in this research field. Table 1 displays the search terms more frequently employed in the literature. Multiple combinations were selected using these key search terms. For instance, all possible combinations of *health* and *healthcare* with (AND) *workforce*, *manpower*, *physicians*, *nurses* and (AND) *forecast*, *projection*, *planning*. Related subordinate queries such as *physicians supply forecast*, *nurses supply forecast*, *healthcare supply forecast*, *healthcare demand forecast* were also employed. These terms were then used on the online databases PubMed, MEDLINE, Embase, ProQuest, Healthstar, ABI/Inform, INSPEC, Google Scholar and Scopus to obtain a base set of the 10 most cited, 5 most recent and 5 randomly chosen papers. Of this initial selection, an abstract matching and backward/forward search was conducted to assess whether the topic covered was relevant. Publications that failed to verify these criteria were excluded. A total of 308 publications were retrieved, with 75 meeting at least 1 of the inclusion criteria using the combination of search terms and were thus included in this review. Table 2 describes our search methodology.

Scope

HHR planning is a comprehensive field far extending the number of physicians and nurses. Other health-care workers such as hygienists, therapists, managers, administrative assistants and other support staff also play a critical role, relieving the clinical staff of bureaucratic and time-consuming tasks. In fact, skill-mix studies show that proper task delegation is critical to ensure proper health-care delivery. Furthermore, a complete assessment may also require the analysis of the impact of other indirect stakeholders, such as workforce educators, regulators, funders and employers. Assessing how the training

Table 1 Key terms used to conduct the search

Keywords	Search queries
Health	Workforce planning
Healthcare	Healthcare forecasting
Workforce	Health human resources
Manpower	Health manpower
Physicians	Health planning
Nurses	Healthcare planning
Forecast	Health services
Projection	Health supply
Planning	Health demand
. . .	Healthcare needs
	Healthcare providers
	Physician forecasting
	Nurse forecasting
	Nursing staff
	Manpower
	Manpower planning
	Workforce forecasting
	Workforce projections
	Workforce management
	Staff levels
	Health staffing levels
	Shortage healthcare workers

is conducted (i.e. could the training time be reduced?; do medical schools have the capacity to train a given number of trainees?; are more medical schools necessary?), the impact of regulatory requirements (i.e. is the entry to medical school limited by government-fixed *numerus clausus*?) or financial and service constraints (i.e. can the existing hospitals and health-care units absorb a planned increase in the number of health-care professionals?) is a critical requirement for a well-guided policy.

Table 2 The search method applied in this review

Step	Search method
1	Identify common search terms from reviews, books and technical papers
2	Generate plausible combinations of terms to be used for search using the key search terms identified
4	Search for these terms on PubMed, MEDLINE, Embase, ProQuest, Healthstar, ABI/Inform, INSPEC, Google Scholar and Scopus
5	Select a base set for the results consisting of the 20 papers (10 most cited, 5 most recent and 5 randomly chosen)
6	Match the abstract and perform a forward and backward search to verify the relevance of the paper for the selected base set
7	Exclude papers that address none of the topics covered, that only make a brief reference to the subject at hand or that are not written in English

Without disregarding the importance of these other professions, in this paper, we will focus solely on reviewing the planning of the clinical staff that directly provide health-care services and, more specifically, on the physicians and nurses, along with references to related fields like dentistry. Obtaining reliable projections for the available and necessary human resources is an obligatory starting point. Moreover, the prominence will be in the spectrum of different methodologies that may be used to obtain forecasts for the number of physicians and nurses, with short references to the approaches or technical apparatus, commonly used to apply a given methodology [a]. Also, our concern is HHR planning only at the national and regional level. HHR planning at a local level (hospital or medical centre) is conceptually different, involving other methodologies and tools, and therefore, it is not inserted in this paper.

The remainder of this paper is organized as follows: in the "Background" section, we introduce the general and governing principles that characterize the health-care market. The background information provided is critical to equip the reader with the necessary concepts. In the "Evolution of the field" section, we proceed with an evolutionary and chronological description of the field, exposing the work and methodologies that have been shaping the research field. In the "Discussion" section, we discuss the current trends in this research area and the road ahead regarding future research directions. We also present a summary of all the findings, including a table with an overview of the methodologies and a data-requirement framework to understand which methodologies can be used based on the data available, as well as a proposal suggesting a way to develop an integrated approach. Finally, we finish with a brief summary and conclusion.

Background

HHR planning as a scientific area and topic of theoretical and applied research evolved significantly from non-existence into a remarkable and serious effort of private and governmental institutions, which tried to anticipate how many human resources, primarily physicians and nurses, will be necessary in order to maintain or even improve the quantity, quality, availability and effectiveness of the medical services provided. Improved life expectancy and changing demographics, epidemiological trends, improved socio-economic conditions and an ever-increasing world population may result in a rise in the expected demand for health-care services [18] and, therefore, further additions to the list of patients of an ageing medical workforce [19]. It then comes as no surprise that health workers are recognized as a critical resource for achieving population health goals [1], working at the front gate of the health-care sector.

The health-care sector is an intricate, albeit fundamental, part of ancient and modern societies, and it comprises a long list of agents, from the individual seeking health-care services to the medical staff providing them, all operating within a legal framework involving providers, consumers, insurance companies, government, medical schools and regulatory institutions. Regardless of the statutory system in place, either a Bismarckian-based or a Beveridgean-based organization, at its core, the health-care market is always composed of both suppliers of health services and patients demanding their services. On the one side is the workforce of physicians, nurses and remaining clinical staff trained and ready to assist those in need. On the other side stand the forces that drive the demand for medical services, strongly related to demographic, socioeconomic and epidemiological factors. Analysing these two market forces is a critical step in assessing whether the available health-care human resources are enough in quantity and skills to meet the current and future demand in due time and may lay solid foundations for further research, considering perhaps changes to the existing health policy framework.

Despite the similarities, the health-care market diverges from a traditional market of goods and services for several reasons [20]. A high degree and extent of uncertainty affects both supply and demand; asymmetric information between physicians and patients, restrictions on competition, strong government interference and supply-induced demand are some of the most glaring differences that can be pinpointed. These may be relevant when assessing the impact of any policy involving HHR planning.

Supply

Supplying human capital with the appropriate expertise so as to enable workers to perform and satisfy the demand for health care is no simple task. The time and effort required to equip HHR, especially physicians and advanced nurse practitioners, exceeds that of most other professions. In some particular health-care professions, the set of necessary skills to qualify for medical practice is acquired through extensive academic learning which involves the enrolment in long courses that may take up decades to complete due to a strict licencing process.

A considerable amount of HHR studies focus solely on this approach, basing their research on the estimation of the expected supply of physicians by accounting for the intakes, exits, migrations and population growth in order to maintain the present ratio of practitioners, using "stock-and-flow" models for that purpose [3]. The analysis of the medical *training* process is relevant but may be insufficient, as several other factors may affect the efficiency and effectiveness of the care services delivered.

Despite the limitations, some measures to overcome imbalances in the quantity (number) of physicians and nurses have already been identified in the health policy literature [17, 21], namely the following: increasing the number of domestic- and foreign-trained medical graduates or increasing the number of medical schools and classroom sizes; increasing the enrolment limits (*numerus clausus*); reducing the requirements for entry to medical schools; raising the wages of the medical staff, as well as the perspectives for their future career path; or reducing the costs of attending medical school, which may encourage potential students to enrol. In Table 3, we provide a more extensive list of policies to cope with a shortage in the number of health workers. These proposals are short-term measures to alleviate the immediate stress put on the health-care system triggered by an undersupply of personnel and may not be suitable for tackling long-term imbalances due to huge shortages or surpluses of medical staff.

Still within the scope of supply, other approaches for handling the problem of insufficient human resources have also been suggested, addressing the problem from

Table 3 Health policy options for targeting health workforce imbalances and alter health-care outcomes (adapted from [17] and [86])

Field	Policy option
Education	Increase numbers of new students
	Recruit foreign graduates
	Recognize previous learning
	Improve curriculum content
Regulatory	Recognize overseas qualifications
	Introduce temporary employment regulations
	Subsidized education for return of service
	Enhanced scope of practice
	Different types of health workers
Financial incentives	Increase trainee salaries
	Raise wages
	Provide non-wage benefits
	Introduce incentives for return of skilled migrants
	Establish retirement policies
	Employ lay health workers
Professional and personal support	Better living conditions
	Safe and supportive working environment
	Career development programmes
	Public recognition measures

an angle besides medical training. For instance, the composition of the core competences and activities of the physicians, the *skill mix*, may be reorganized to enhance the roles performed by the clinical staff, relieving them from tasks that could be safely assigned to other health-care professionals [22]. This strategy does not require a change in the number of physicians but the restructuring of the available human resources and medical competences. Complementarily, supporting policies and reforms that enhance the *productivity*, that is, the ratio of output per unit of input given a certain level of technology and methodology, of the medical staff may result in an increased outcome that also does not require a change in the quantity of labour workforce [23]. Assessing the productivity of the clinical staff is now quite common [24], and operations research applied to the improvement of patient flows, queueing, master surgery scheduling, ambulance fleet management and staff rostering may play a very important role in increasing current levels of productivity. In summary, the initial focus of supply-based methodologies was on the training process. As of late, more focus has been given to the productivity and to the skill mix of the labour workforce as well.

Methodologies for modelling supply

Training (entries and losses) The purpose is to model the training process so as to predict the number of entrants in each year. This way, and in combination with migratory flows, mortality, exit and drop out rates, it becomes possible to estimate the number of physicians and nurses available for each year, with everything else held constant.
Productivity The productivity of the medical workforce is not constant, as some professionals work harder or better than others or simply because there is an excess of bureaucracy to comply with. Without touching on the *quantity* of professionals, it is possible to reorganize services and incentives so as to promote increased productivity or implement lean and operations research recommendations to significantly improve the output and outcome of the workforce.
Skill mix Since a degree of interdisciplinarity exists between medical professionals, it is possible to reassess the tasks performed by each professional, relieving physicians from day-to-day bureaucratic routines or reviewing the competences of the nursing profession so as to broaden their scope of action. Horizontal substitution (between different medical specialties) and vertical substitution (between different working classes) can be used to improve the amount of health-care services provided.
Worker-to-population ratios This method establishes a desired ratio for the number of physicians and nurses per unit of population and compares it to the actual ratios. Policies to increase or decrease these ratios may then be pushed forward. Although simple and easy to apply as long as data is available, the method lacks the fine detail of such a complex system, ignoring other factors such as needs, demand or institutional frameworks that may have an influence on the productivity of countries or regions with similar worker-to-population ratios. Moreover, it abstains from exposing the causes for such asymmetries or from evaluating the efficiency of the available workforce.

Demand

Demand for health care is a derived demand [25], which means that people do not seek health care services as a final good for consumption but as an intermediate service allowing them to be healthy and to improve their stock of *health capital* (well-being). They want to improve their health, and to do so, they seek health-care services. As in other markets, the determinants of aggregate demand for health-care services are population size, income and preferences. Moreover, for countries where medical care is mostly an out-of-pocket expenditure, demand is restricted by the patients' ability to pay. If a patient requires medical attention and is unable to finance it, this *need* for health care will not translate into *effective* demand, despite its existence. Accounting for these cases is especially important in countries where health care is not publicly subsidized or where there are obstacles to entry other than the availability of resources.

The concept of *needs* in health care is not consensual in the health literature, with a semantic confusion arising from its use in health economics [13, 26]. While the economic or *effective demand* translates the actual, observed demand, usually measured in terms of service utilization ratios (such as bed occupancy rates, number of inpatients), the *needs* component tries to fully encompass the epidemiological conditions that characterize a given population, measured through morbidity and mortality rates or by the opinion of a panel of experts, and how that may translate into a given quantity of required health-care services. Therefore, we see that the classical concept of economic demand may not reflect the biological needs of the population, as it may leave out the necessities of the population regardless of their ability to pay. In the needs component, the emphasis is on the medical conditions that may lead to demand for health care, deriving from the evolution of chronic diseases, prevalence rates and overall morbidity patterns. This distinction is better illustrated in Fig. 1, where we present the case when all demand is met, at a given price, and equilibrium is attained. Theoretical demand, projected strictly in terms of biological needs without a budget constraint (either households' income or public budget), may not always correspond to the demand effectively observed. The reason being that the quantity sought is limited by the disposable income directed towards out-of-pocket health expenditure or by limits to the government budget that is allocated to health

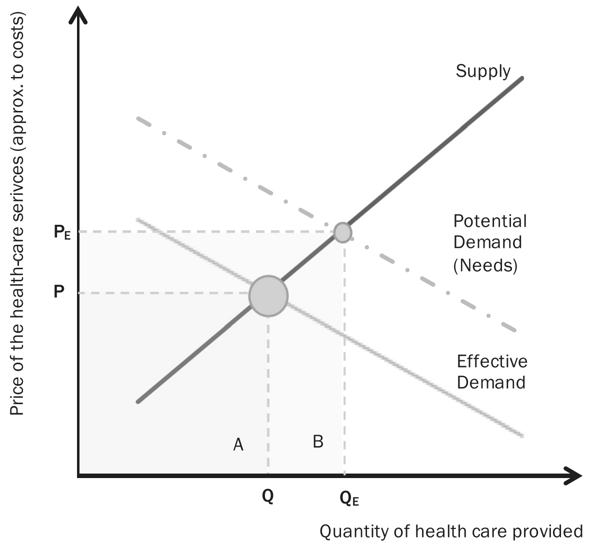


Fig. 1 Law of supply and demand applied to health services. The health-care market depicted in terms of supply and demand, with a tentative distinction between *potential* and *effective* demand

care. We draw the distinction by plotting both the curve of *needs* (potential demand), corresponding to a *no gap* scenario, and the *economic* (*effective*) *demand* that is actually observed.

Although needs is a fundamental concept, it should not be decoupled from economic demand, as it should not ignore the budget constraints of the economy. In fact, the country may not have the ability to provide all the health-care services presumed to fully satisfy needs. If the area delimited by B (cf. Fig. 1) is larger than the domestic product of the economy, it will be impossible to meet all the perceived health-care needs of the population. Like any other problem involving scarce resources, a serious analysis should not abstain from recognizing the existence of financial impediments. Conversely, it should try to quantify needs, serving as a theoretical benchmark for the future.

This has not always been the case. Some studies estimate demand solely based on the current level of service in relation to future projections of demographic profiles [27, 28], thereby leaving out an important determinant of demand, the epidemiological needs [29, 30]. When and how disease trends evolve is critical to properly anticipate the needs of the population, a proxy to the expected future demand. For instance, chronic diseases have been increasing globally [31]. China, a country usually not associated with overweight and obesity problems, has experienced an upsurge in type two diabetes. According to the data reported, in 1980, less than 1 % of Chinese adults had diabetes, but by 2008, the prevalence of the disease had already reached 10 % of the population [32]. As a result, it is expected that more endocrinologists will be necessary to assist with the treatments. The raw definition of needs is not subject to any boundaries other than those set by epidemiological constraints and medical advances.

A substantial part of the studies targeting supply hold current demand constant, thereby leaving out a proper analysis of what drives demand for health care. In fact, a change in the factors that influence demand or the emergence of new health conditions in a population may require a reorganization in the quantity, composition and skill mix of the medical workforce to ensure that all supply meets demand. This suggests that targeting the right number of people and the right skills depends as much on the health conditions and epidemiological characteristics of a given population as on the supply of physicians and nurses [33].

In summary, three methods are commonly used to analyse HHR planning from a demand-based perspective [13]. Most of the methods build upon the definitions of needs and effective demand, and some overlap in their scope of application. Contrarily to the approaches found in supply-based methodologies, where the object of study remains the same and alternative analytical methods are employed, in demand, opting for a different method may change the scope of the analysis.

Methodologies for modelling demand

Needs (or potential demand) This method determines the effect of health diseases, epidemiological patterns and overall mortality and morbidity rates in the demand for health services and obtains an approximate number of personnel hours required to cover those needs. Needs are usually assessed by a panel of experts in epidemiology and may not match the services that the public wants.

Economic (or effective demand) In this method, we look at the services actually contracted by the population, subject to the usual economic constraints that may put an upper bound on the quantity solicited. In sharp contrast to the first method, effective demand may not imply a healthy population, especially for poor countries without a subsidized health-care service since the general citizen lacks the means to obtain health-care services. The method ignores needs or wants and assumes that all the remaining variables remain constant, although that requirement may be relaxed by complementing the results with other methods.

Service targets Service targets extend a needs-based approach by incorporating other measures, such as consumer needs, in order to establish service-target ratios to be accomplished. Service-target approaches decouple the multiple areas of health-care services and proceed with an independent analysis of each subsystem, with the main advantage being a more detailed proposition of the changes required, with separate recommendations for distinct areas.

Evolution of the field

Although the health workforce has long been a concern to policy makers, including those of ancient Rome [34], the first academic research articles discussing manpower planning in general, and health-care workforce planning in particular, date back to the 1950s. This was a natural response to both the creation of national health-care systems and universal insurance schemes.

A universal health-care system with no exclusion based on preconditions and with no restrictions on access, an idea put forward by Bismarck in the compulsory social insurance form, and promoted by Beveridge as a national health service [35], requires a well-prepared and readily available team of physicians, nurses and administrative staff. To ensure that services are in fact provided, public medical universities were created along with subsidized access to medical training. These reforms resulted in the emergence of a national ecosystem of health-care suppliers and a pool of patients, a significant change from the decentralized network of health-care providers. The ubiquity of access required providers to be distributed evenly so as to satisfy the needs of the population.

After this period of sustained and prolific economic growth, a period of crisis followed. Expectably, the economic slowdown put the focus on efficiency, towards a better use of the available resources. During this period, many developed and developing countries experienced shortages of health-care providers, mostly nurses [36], justifying the growing interest in this newborn academic research field.

This was the period when the first articles on health-care workforce planning emerged. We separate the analysis of the unfolding of HHR planning into three separate stages, corresponding to the evolution of how the health-care worker is perceived as an object of study [37]: (a) the health worker as a production factor, (b) the health worker as an economic factor and (c) the health worker as a necessary resource. This structure is helpful in the sense that it exposes the role given to the workforce, once studied as an inorganic fixed-input factor and more presently viewed as a complex and necessary resource with its own idiosyncrasies like any other economic agent.

First phase: factor of production

The first articles published on the subject date back to 1950, with HHR planning being perceived as a production function, where the labour workforce is an input factor. The research, triggered by general health worker shortages in developed countries [38, 39], led a growing and diversified body of research that diverged into different approaches. Not surprisingly, some of these articles are the result of initiatives promoted by governments and international organizations to address their own domestic shortages of physicians and nurses, while others are *ad hoc* contributions of attentive researchers keen on providing an insightful contribution. The techniques employed vary from descriptive to predictive or merely comparative techniques and usually involve econometric regressions, static tables, linear programming or benchmarking. These techniques are then applied to the areas of analysis previously described, either supply, economic demand, needs and service-target or worker-to-population ratios, which we will identify next.

A significant part of the research papers produced at that time are well-documented, with comprehensive lists and reviews of the models developed still available [40, 41]. Of these, we highlight those that are still cited in the literature and available online.

Supply-based methodologies

The very initial concern of those conducting HHR planning was estimating the necessary number (head count) of medical professionals to either maintain the current worker-to-population ratios or reduce/increase it if an imbalance was found. One of the first insights into the evolution of the supply of physicians was done by crossing the observed physician-to-population ratios along with the posited population growth in the United States of America, by that time impulsed by the "baby boom" and by an expected increase in the use of medical services. The people in charge of HHR planning evaluate the number of physicians required to maintain the ratios given those demographic and economic changes [42, 43]. In the report, the same criterion is used to estimate future manpower requirements for all the available medical specialties, nurses and miscellaneous professions necessary for due operation.

One way of doing so is to look at the current stock of professionals and factoring in negative and positive flows that affect the stock. Factors such as mortality, migration or retirement generate losses to the current workforce stock. Likewise, entries from medical schools and immigration increase the current level of professionals. Models that map this structure are commonly known as "*stock-and-flow*". Despite not using this specific terminology, models created at the time already incorporated the idea of increases and decreases in the current stock due to exogenous factors and then used that information to obtain projections [44–46].

Focusing particularly on the supply of nurses in the United States of America, other papers proceed with an analysis of the economic factors, namely the hourly wage and the wage of the nurse's spouse and the effect on the supply of nursing professionals [45, 47]. Evidence suggested that hospitals exercise monopsony power, which has an impact on how a supply gap may be tackled. Moreover, results also suggest that the cost of paying wage incentives to increase working hours is considerably smaller than the cost of training additional professionals, something to take into consideration when evaluating HHR reforms.

The product of this novel research was tested in the field. For instance, in the analysis of the health-care workforce in Taiwan, estimates for the supply were generated on the basis of retirement, migration and death rates applied to graduations. They incorporate the training process and its effect on the supply of physicians [48].

Methodologies: Training (entries and losses) [42–46, 48], Productivity [45, 47], and Worker-to-population ratios [42, 43].

Demand-based methodologies

One of the first publications in the field of HHR planning starts by differentiating the aforementioned dimensions of workforce planning [49]. Klarman et al. argue that, although medical *needs* could form the basis for determining workforce requirements, it cannot be decoupled from economic costs, an active constraint to the extent, scope and applicability of reformist policies. A forecast of the necessary supply of physicians is not provided, but it is suggested that the shortages in the specialty areas may be a sign of an overall supply shortage.

Another way of predicting the necessary future hospital beds is by extrapolating from a set of factors assumed to drive the demand for health care, namely socio-economic factors and biologic needs, measured through morbidity rates [50]. This approach was also used to estimate hospital bed requirements, providing both empirical works on real data for the United States [51] and theoretical frameworks with hypothetical parameters [52]. In some cases, the approach of forecasting bed requirements would be extended to other health-care units such as primary medical care, nursing home care, consultant medical care (medical care provided by a physician with specialized training), hospital care or domiciliary care [52].

Methods for estimating the number of professionals required (head counts) from a demand perspective also started emerging at around this time. For instance, in one case, estimating the number of necessary physicians for the future was done by calculating the number of professionals necessary to close the gap between observed and unattended demand, where demand is measured in terms of utilization. In this case, using service-level indicators again for the United States [53].

In other studies targeting the U.S.'s health system, the influence of exogenous variables such as age, income and urbanization is used to extrapolate the effect of dependent variables on health policy and HHR planning, including the number of persons with health insurance, the number of general practitioners, medical specialists, available short-term general hospital beds, admissions and mean duration of stay per case [54]. This approach is also similar to the one used in two other models, the first using data aggregates to facilitate HHR planning at national, state and substate levels and the second going to the level of detail of the individual and his interactions with professionals and institutions [46].

More comprehensive approaches to estimate economic (effective) demand were also addressed. Some papers suggested incorporating indicators such as an increase in population, economic development, improved education, a change of supply, age distribution and other unpredictable factors. Simple calculations, such as the ones used in the former Soviet Union, could be performed by extrapolating based on observed norms of practice regarding the number of patients attended and then complemented with basic biological needs by incorporating data about morbidity and mortality rates [44]. Methods like this were then applied to countries such as Taiwan, characterizing current public and private sector demands for health services [48].

Another option for measuring demand also elaborated during this time consisted of using other indirect indicators, namely short-stay services, services of nervous and mental hospitals, physicians' services outside hospitals, dental services and other health services. The data is then fed into a model that tries to minimize the gap between the number of individuals employed in medical services that attend to the demand for personnel in that occupation [53]. Estimates were generated for the United States.

Finally, it should be noted that attention was constantly being drawn to the importance of prevailing morbidity, a basic indicator for assessing medical manpower based on a needs-based approach. Some authors stress that it is the hospitals and their internal need for residencies that actually determine the number of specialties [55]. This may not reflect with accuracy the actual needs of the population since patients could potentially remain unattended or in long waiting lists, but it is an insightful indicator if waiting lists are also factored in. Finally, they also consider the specialty of the physicians' role, warning that general practitioners fulfil key medical functions and should not be relegated to second place. The concept of *skill mix*, despite not formally and explicitly defined, is here put in evidence.

Methodologies: Needs (potential demand) [44, 46, 48–50, 55], Economic (effective demand) [44, 46, 49–54], and Service targets [46, 53].

Second phase: economic agent

The first phase of HHR planning was characterized mainly by an aggregate analysis of the health-care market, with independent and/or cross-analysis of supply and demand. Reviews produced at that time refer essentially to needs-and demand-based approaches, as well as simple worker-to-population ratio benchmarks [56]. The phase that starts in the late 1970s and goes onward through the 1980s and 1990s redefines the role of the HHR, previously seen as an homogeneous input factor, into a complex economic agent [37]. The adoption of such perspective broadens the scope of analysis, namely by assuming that health-care workers react to economic incentives.

The deepening of the analysis is done through the application of microeconomic theory to the study of health labour workforce, thereby exposing dimensions that had gone unnoticed when looking only at the aggregates, although a macroeconomic analysis continued to take place [57]. It was triggered by two macroeconomic observations occurring at this time [37]: a perceived oversupply of physicians and nurses [58–60] and an upsurge in health-care expenditures [8]. During this phase, attention was given to topics such as health worker licensure [37, 61], information asymmetry distortions [62] and its potential repercussion as an unnecessary increment in demand induced by health suppliers [63] and health worker performance and productivity [64]. Furthermore, HHR planning became a major concern in related fields, such as dentistry [65].

Supply-based methodologies

Although the previously mentioned topics are of notable relevance, some have no direct utility in the elaboration of projections and forecasts of future health-care needs, serving only for policy guidance. For that reason, we will concentrate our efforts on the performance and productivity of health workers, a method fully within the umbrella of supply. In terms of policy, it is less demanding to put in practice as it does not require structural changes to the training process or to medical schools. In theory, more people can be served with the exact same amount of human resources if only their productivity increases. Improving the efficiency of the available pool of resources is therefore an attractive methodology.

This is the line of research followed in a paper where a microanalysis of the factors that may influence the output (and therefore productivity) of the health workers is conducted, in particular nurses in the United States [47]. Sloan et al. found that there is a strong supply response to the hourly wage. Raising the hourly wage is, in fact, their proposal to respond to a short-run supply shortage, arguably a quicker response than changing the number of intakes to nursing schools. Taking another route to reach the same goal, one study tries to undercover job satisfaction indicators and perceived productivity in 24 hospitals for a staff nurse population [66]. The purpose is to understand the factors that may raise productivity but also to find a connection between job satisfaction and the quality of care provided. Similarly, waiting and distance times can also be used to assess the physicians' productivity, a study conducted using data from the United States [67].

In the same line of research, some authors conducted an observational study of 56 physicians in order to uncover the factors that may influence productivity, measured as the ratio between the number of patients seen per physician and the time spent with the patient [24]. The main research question was understanding which factor contributed the most to the variance in productivity: the patient or the physician. Results suggest, according to the study conducted in a Veteran Affairs' medical centre in the United States, that the individual physician explains the variations in productivity observed, with the actual patient playing a minor role. Similarly, in another study also conducted in the United States, the productivity of physician assistants and nurse practitioners and their role in the health-care workforce is analysed [68]. Scheffler et al. find that these two categories of health workers could have a significant influence on the future health-care workforce if some vertical and horizontal substitution occurs and tasks are delegated. Note that the change of setup hereby suggested tackles productivity from a different angle: instead of raising the output, the inputs are altered.

Methodologies: Productivity [14, 24, 47, 64, 66–68] and Skill mix [68].

Demand-based methodologies

Studies focusing solely on the demand side produced during this phase are considerably less common than in the first phase. The ones that do so are more concerned with the lack of attention given to the importance of biological needs. It is interesting to note that, at the turn of the decade and in subsequent years, a lot of emphasis is again put on the *needs* of the population. Some authors suggest a needs-based evaluation as a requirement to produce accurate forecasts [29, 56]. This option contrasts with that of other authors, which propose using benchmark as a viable alternative to potential or effective demand projections [69]. The work developed consisted of comparing the number of active physicians *per capita* in the United States, adjusted for population differences between similar locations, without uncovering the causes for the given asymmetries.

Assessing the needs of the population was also the method of choice in the dentistry field to calculate oral health workforce requirements. In particular, needs were projected by the amount of oral care, including preventive, special group care, surgical, orthodontic, periodontal, restorative and prosthetic, that different age cohorts would require [70]. Then, the time necessary to treat each of these conditions is estimated, and the number of dentists to perform those tasks is derived. Also applied to dentistry but with a focus on the skill-mix distribution, productivity changes are estimated by examining role substitution in dentistry [71], helping to conduct evidence-based scenario analyses in The Netherlands.

Methodologies: Needs [29, 70], Skill mix [71] and Worker-to-population benchmarking [69].

Integrated methodologies

A new strand of the literature also emerged during this phase covering supply while at the same time considering projected changes to demand. In a review of supply projections conducted both in Canada and in the United States [14], the authors argue that the traditional supply projection methodology that characterizes the licensure cycle and productivity metrics is incomplete if unmet *needs* of the population are not defined and included as a clear research goal, as well as economic, financial or infrastructure resource constraints.

The integrated approach is also present, for instance, in the implementation of the "System for Health Area Resource Planning" (SHARP) [72]. This analytical framework combines all the major methodologies: it includes the socio-economic factors that drive economic demand, morbidity and the remaining epidemiological factors that drive needs, the formation process of the health-care supply of workforce and utilization rates in order to incorporate the current use of health-care services. The framework was successfully used to support HHR planning in Canada, especially in the province of Ontario, reinforcing the idea that an integrated or systems approach, combining the multiple facets of the problem, is the way to go in the future.

Methodologies: Integrated [14, 72].

Third phase: fundamental resource

In this phase, the notion of health labour workforce is reformulated, this time viewing it as a necessary resource. From the 1990s onto the 2000s, the emphasis is on the regional asymmetries in the placement of the workforce and in the migration flows from developing to developed countries [37]. All models proposed include both supply- and demand-based methodologies to tackle the problem.

Integrated methodologies

Methodology-wise, the trend observed is a continuation of the second phase, with the call for a holistic approach to the problem. HHR planning must be addressed from an integrated perspective, including when analysing all the blocks of the functioning system so as to calculate the current and future *gap* between supply and demand [73]. The authors' proposal is in line with the SHARP framework: modelling key demand (economic and epidemiological) and supply inputs. Furthermore, it is continuously stressed that the epidemiological drivers of the need for health-care services should always be part of HHR planning [30, 74].

When looking at the research literature produced at the turn of the century, this trend becomes clear. Summing up the results achieved so far, we can see that health-care workforce planning is a complex endeavour, and it becomes necessary to identify all the relevant variables to accurately forecast the necessary resources for the future [75]. Again, these variables relate to supply and needs methodologies. A practical work conducted in Lithuania to forecast family physicians for a 10-year timespan employs this approach [76]. Firstly, this approach calculates the supply of physicians through the usual process of modelling the training of physicians. Moreover, it crosses the supply forecasts with three different projections for demand: firstly, the requirements established by a panel of experts using a Delphi technique; secondly, the resources necessary to increase the number of visits; and thirdly, an upper bound placed on the worker-to-population ratio so that one family physician serves no more than 3 000 inhabitants. The conclusions reached suggest that the well-informed panel of experts elaborated the most accurate projection of demand for family practitioners and that none of the supply projections was right on target. Similarly, in a forecast analogous to the nursing profession in Germany, the analysis is extended from the usual supply and demand to include the effects of occupational flexibility and employment structure. Adding these two elements to the analysis has a relevant influence on the projections [77]. Notably, this pensiveness with the organizational role, where the HHR is more than an aggregate number but rather a dynamic and complex sum of individuals, is clearly gaining traction.

In the same line, some researchers suggest a needs-based analytical framework that incorporates input from four separate elements: demography, epidemiology, standards of care and provider productivity [30], again falling in the realm of integrated approaches. Alternatively, needs can be decoupled in a functional form so that service targets can be defined and deployed [1]. Dreesch et al. claim that methods focusing strictly on the supply, on the demand or on both fail to address or recognize the effects of the skill mix (the potential of substitution) between health professions. The importance of a more integrated approach to HHR planning is also restated. With more or less variables, the trend is clear: recent models use information from both demand- and supply-based methodologies, including inputs as varied as demography, the training process, workers' productivity or biological needs in order to generate their forecasts [18, 78, 79].

Although the emphasis is fundamentally put on addressing the problem from an integrated perspective, new strands of literature were also developed during this phase. For instance, it is suggested that instead of addressing the problem from a quantitative perspective, either by adding to or subtracting from the stock of health workers, it should rather be addressed with internal reorganizations, redefining which tasks can be performed by whom [80]. Such internal substitution and activity delegation could be executed by transferring skills from the medical specialist and the general medical practitioner to other health professional roles, namely nurses with higher education (midwives) or by creating new roles. This methodology involves, therefore, playing with the *skill mix* of the health-care professionals. This was put in practice in Ireland by employing a model that targets both supply and demand, reflecting the concerns for including all parts of the system [28, 81]. Moreover, it tests four policy interventions, three of which related to supply and the last related to the skill mix: increasing vocational training places, recruiting professionals from abroad, incentivizing later retirement and increasing nurse substitution so that nurses can deliver more services. Similar studies, encompassing the workforce supply, demand and the skill mix, were also conducted in the dentistry field during this phase [82]. In this case, workforce supply and demand for oral health needs are projected to study the impact of skill-mix reorganizations. To forecast future dentist numbers, a simple percentage increase based on previous yearly increases is considered. To estimate demand, demography evolution, rates of edentulousness, patterns of dental attendance and treatment rates of older people, as well as general dental service treatment times, are considered. The effect of the skill mix is then studied considering several scenarios of varying skill-mix use. Gallagher et al. find that widening the skill mix can be extremely helpful to build capacity for dental care.

Another concern that is raised during this phase is that of measuring the *outcome* as an important indicator for assessing the quality of the health-care services. The outcome is a fundamental indicator for HHR planning. In particular, equitable and timely access to health care are a precondition to a good outcome, which is the variable to be maximized [83].

In summary, it can be said that this stage was a phase of settling with methodologies, namely supply-, demand- and needs-based approaches, and of urging for a more integrated approach while paying attention to the roles of each health professional and the degree of substitution between professions. Furthermore, a concern about the *outcome* of health-care services was raised, where effectiveness and quality of the treatment is considered on par with the number of patients seen (productivity).

Methodologies: Integrated [18, 18, 28, 30, 73–79, 81, 82], Skill mix [1, 28, 77–82], Needs [30, 77], Service targets [1] and Productivity [77–79]

Discussion

Five decades of work in HHR planning fuelled by eminent global shortages of health professionals have contributed to establishing this research field as an important scientific area, decisive for achieving worldwide health-care targets [1]. Significant results have been attained. In particular, new methods and techniques were developed, and the accuracy of projections improved remarkably [23], and HHR planning became an area of prominent interest, with the number of publications in the field increasing over the years. Moreover, the literature evolved, replacing some approaches with others, paying more attention to the health-care workers and their productivity and to the delegation and distribution of skills. It prioritized integrated approaches and the role of epidemiology in addressing the problem. In fact, when we look through all the methodologies reviewed (Fig. 2), the emerging trend clearly supports this claim. Integrated approaches are gaining ground after decades of partial analyses turning to either a supply- or a demand-based approach and in its simplest form only resorting to worker-to-population ratio benchmarks.

In Table 4, we summarize the methodologies and describe the necessary assumptions for using each of the approaches, along with their advantages, limitations, how these limitations are overcome, requirements and the countries in which their usage was documented (according to [9]). In the past, this overview would probably help in choosing the methodology to adopt. With the call for more integration, it assists in showing how a methodology may fill in the gap towards a cohesive framework. Also, it serves to show that there is no perfect methodology capable of providing accurate forecasts without considerable pitfalls and that there is a trade-off between simplicity and completeness, where going for a simpler methodology may implicate leaving out important parts of the problem.

An integrated approach

The importance of a comprehensive, integrated approach is continuously emphasized throughout the period in review [3]. Although the need for an integrated approach had already been stressed in several past publications, it keeps on reappearing, suggesting that it might not have been fully addressed as of yet. This approach faces many challenges. A dynamic, system-level perspective covering key drivers of supply and demand that includes both manpower planning and workforce development is critical to overcome such challenges [81]. The importance of paying attention to needs is also continuously stressed, as changes in the health patterns of the populations take place [84]. In summary, integrated approach refers to a method that incorporates in its process projections of the workforce supply and the impact of microeconomic and organizational changes in productivity and in the skill mix, of the evolution of demand for health-care services and also of

	Supply				Demand		
Article	Training (Entries and Losses)	Productivity	Worker-to-population ratio	Skill mix	Needs	Economic demand	Service targets
Klarman (1951)					●	●	
US SG Consultant Group and Bane (1959)	●		●				
Beenhakker (1963)					●	●	
Rosenthal (1964)						●	
Baker (1966)	●				●	●	
Feldstein (1967)						●	
Maki (1967)						●	●
U.S. Bureau of Labor Statistics (1967)	●		●				
Baker and Perlman (1967)	●				●		
Health (1968)	●				●		
Navarro (1969)						●	
Altman (1971)	●						
Yett et all (1972)	●				●	●	●
Reinhardt (1975)		●					
Sloan and Richumpan (1975)		●					
Sloan (1977)		●	●				
DeFriese (1982)		●	●			●	
Lomas (1985)	●	●			●		
Kramer (1989)		●					
Birch and Eyles (1991)					●	●	
Pathman (1991)			●		●	●	
Smith (1995)		●					
Goodman et al (1996)			●				
Scheffler (1996)	●	●		●			
Birch (2002)	●				●	●	
Joyce (2004)	●	●			●	●	
Dreesch (2005)	●	●		●	●	●	●
Starkiene et al (2005)	●		●		●	●	
Maynard (2006)		●		●			
Birch (2007)	●	●			●	●	
Scheffler (2008)	●				●	●	
Tomblin Murphy et al. (2009)	●	●		●	●	●	
Barber (2010)	●				●		
Masnick et al (2010)	●			●	●	●	●
Sotrdeu and Leonard (2010)	●			●	●	●	
Teljeur et al (2010)	●			●		●	
Van Greuningen et al. (2012)	●	●		●	●	●	
Maier and Afentakis (2013)	●	●		●	●		

Fig. 2 Identification of the conceptual methodologies found in some of the literature for the period of 1950–2013

the evolution of health diseases and its potential impact on the health system.

Notwithstanding, integrating all the pieces may be a puzzling task. To assist with the task, in Fig. 3, we provide a high-level functional diagram with a proposal for how methodologies could be coupled so as to turn it into a seamlessly integrated system. On the supply side, we have the current stock of workers along with the training process so as to obtain an initial snapshot of the current workforce. The current stock, which may or may not be enough to tackle current demand, in which case an imbalance exists, is subject to positive and negative flows that may alter its number and composition. This given quantity of workers may provide more or less health-care services depending on their productivity and skill mix, and that influences the conversion from head counts to

Table 4 The methodological approaches established during the first phase of research

Methodology	Description	Assumptions	Advantages	Limitations	Overcoming limitations	Requirements	Documented usage[a]
Supply							
Training	Projects the availability of health-care professionals based on the current stock of clinicians, the training process (entries and dropouts), migration flows, attritions and retirement rates	Demand for medical services is assumed to remain constant and the projections are used to reduce the supply gap	Predictions for the future supply can be obtained in a fairly simple and immediate way	Demand for medical services is assumed to remain constant, which may not be true No critical assessment of the adequacy of current service levels	Incorporate a model of demand: economic or needs-based (or both) Evaluate current level of service through waiting lists, overtime hours, foreign workers, etc.	Accurate and up-to-date accounting of the current stock of physicians and nurses, migration rates, entry and drop out rates and expected retirees Service usage levels from the health-care sector	Australia, Belgium, Canada, Chile, Denmark, Finland, France, Germany, Ireland, Israel, Japan, South Korea, Norway, Switzerland, The Netherlands, United Kingdom, USA
Productivity	Reorganize services and/or economic incentives to promote higher productivity. Work harder or work smarter	Physicians and nurses act as rational agents and react to economic incentives like wage increases	Does not require a change in the quantity of human resources. Can be implemented immediately	Productivity improvements may not be enough to accommodate large gaps in the supply of professionals	Do not preclude from evaluating the number of professionals necessary given different productivity levels	Operational indicators like the number of patients served with a given number of FTEs (or head counts)	Australia, Canada, Japan, Korea, Netherlands, Norway, Switzerland, United Kingdom, USA
Skill mix	Delegate certain tasks to other health professionals. Substitution can be horizontal (between medical professions) or vertical (between physicians and nurses)	Professionals can assume new roles and perform new tasks	Does not require a change in the quantity of human resources. Can be implemented immediately	Enforcing such changes can be a political challange. Does not solve large gaps in the supply	Providing success stories to involved stakeholders, health authorities and medical associations	Education schools that can provide advanced education to the existing workforce	Netherlands, United Kingdom
Worker-to-population ratios	Specifies desirable worker-to-population ratios based on direct comparison with another region of country	Regions and/or countries can be directly compared	Extremely easy to understand and apply Useful for providing baseline comparisons	Does not take into account the intrinsic differences between regions and countries, the productivity and skill mix of the available workforce	Does not take into account the intrinsic differences between regions and countries, the productivity and skill mix of the available workforce	Records of the current workforce to population ratios	Chile, France, Ireland, Israel, Switzerland, United Kingdom

Table 4 The methodological approaches established during the first phase of research *(Continued)*

Demand							
Economic	Estimates future requirements by projecting the effect of demographic and socio-economic factors on the current level of service	Current level of service is adequate. Skill mix and distribution of health service is appropriate Demographic profile of the population and its effect on health-care demand can be accurately forecasted	Conceptually easy to understand and to apply Allows decoupling of the various components of demand and their influence on the overall aggregate demand	Tends to produce estimates of HHR demand that exceed practical limits No critical assessment of the adequacy of current service levels Ignores the real demand, focusing instead on the effective demand	Take financial constraints into consideration Evaluate current level of service through waiting lists, overtime hours, foreign workers, etc. Include a needs-based evaluation	Accurate and long-term demographic estimates Service-usage levels from the health-care sector Macroeconomic indicators and statistical data crossing income and usage	Australia, Belgium, Canada, Denmark, Finland, Germany, Japan, Norway, South Korea, Switzerland, The Netherlands, USA
Needs	Considers the effect of epidemiology on the demand for health-care services Projects age- and gender-specific needs based on morbidity epidemiological trends	All health-care needs can and should be met Resources are used in accordance to needs	Allows for a fine-grained analysis of the requirements of each medical specialty Is independent of the current service-utilization ratios Easy to understand	Absence of economic/efficiency considerations may render the projections unattainable Dependent on epidemiological projections which may not be obvious Does not consider the current level of provision nor the capacity of the country to deliver health care	Consider an upper bound for a practical result Consider projections of the most common health patterns Incorporate economic considerations in the model	Demographic estimates that are accurate Service-usage levels from the health-care sector	Belgium, Canada, Germany, United Kingdom
Service targets	Defines normative targets for the production of health-care services, which are then converted to HHR requirements	Assumes that established service targets are achievable in terms of financial and physical capital resources	Easy to define, interpret and understand Facilitates cost estimation Requires modest data and planning capabilities	May originate unrealistic assumptions Ignores financial and other active constraints	Incorporate economic considerations in the model	Current level of service	

[a]OECD Report
Source: adapted from Hall and Mejia [13], O'Brien-Pallas [11] and Dreesch [1]

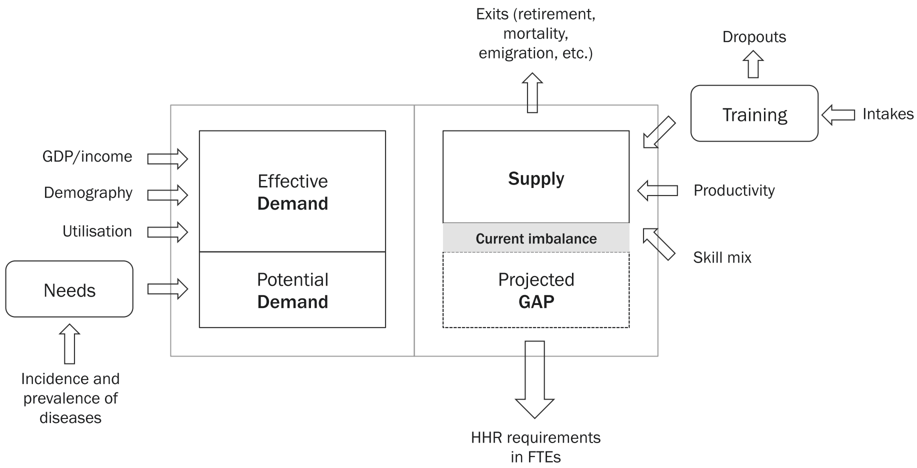


Fig. 3 An integrated system that incorporates several methodologies to address the many facets of HHR planning

full-time equivalents (FTEs). Such conversion is critical to properly assess the health-care workforce, as a significant number of physicians and nurses work part-time only. For this reason, FTE is a more accurate measure as it normalizes the head counts. On the demand side, economic (effective) demand can be initially measured by analysing utilization indicators. How this demand will evolve in the future will then be subject to typical economic factors such as demography and the growth of the income/GDP. In parallel, potential needs can be assessed by incorporating incidence and prevalence of diseases and then mapping a given disease to an estimate of FTE requirements. Whether future supply forecasts should tackle all of the estimated needs is a decision left to the consideration of the policy maker, as this analysis does not incorporate financial constraints. Such an integrated approach is more complex, but not necessarily more difficult [12]. In fact, policy-making cannot abstain from factoring in financial and service planning considerations in a *post hoc* analysis, since there may not be enough resources to accommodate for a sudden increase in the number of professionals. Such analysis is not limited to a money perspective, to the financial burden inputted on the system for educating and hiring these medical professionals or to the installed capacity in terms of medical schools, university hospitals, hospital beds, primary care facilities and others, in order to absorb planned increases in the health-care services labour market.

Data requirements

None of these methodologies can be applied without the adequate data to feed the model. A bare minimum of information regarding the available medical workforce is always required. Table 5 summarizes the most important indicators for conducting a proper forecast. It is not strictly necessary to possess all the information listed, but the availability of the data increases the probability of a more comprehensive projection.

Simpler approaches require fewer data. Worker-to-population ratio benchmarks require a head count of the number of licensed medical professionals, usually made available by the government, medical and nurse associations or by unions. Service targets use the current level of service, which can be obtained from the hospitals' operational key performance indicators. Needs (potential) and economic (effective) demand, on the other hand, require a more extensive set of indicators. For needs, it is necessary to assess and validate current and future incidence and prevalence of diseases and how that may convert into necessary resources. Both tasks are not straightforward and usually require acclaimed experts in epidemiology to step in and provide both the estimates, as well as an accounting of the resources that will be necessary. Effective demand makes it necessary not only to obtain metrics similar to those indispensable for a service-target analysis (such as the number of inpatients and outpatients, number of occupied hospital beds, average length of stay) but also demography and socio-economic projections and how they affect demand. Finally, modelling supply is also a challenging task in terms of data requirements. Unless evidence is found showing that the worker-to-population ratios will remain constant for a long period of time, a supply-based analysis must be factored in. In such a case, it is necessary to know the current stock of licensed providers, as well as the number of intakes, exits and annual attritions, which makes it necessary to model the training of medical professionals.

Assuming that developing countries are in possession of fewer data and that developed countries have more information available, methodologies that require an extensive set of data will be difficult to implement in developing

Table 5 Data requirements for making use of each of the different documented methodologies

Methodology	Indicators	Data requirements
Supply	Stock of licensed providers Baseline stock, age/sex distribution, growth projections Annual additions to licensed stocks Graduates, in-migration (foreign-trained, immigrantes, on temporary work permits), returned to profession Education/training programmes Number of programmes and students enrolled, attrition rates, years to complete programme, number of graduates, costs Annual attritions to licensed stocks Retirements, mortality, career changes, emigration, abroad	High
Productivity	Labour market Occupational participation rates, occupational employment rates, employment projections, vacancy rates, turnover rates, wage rates, productivity growth, cyclical factors, alternative career options Employment status Full-time, part-time, casual, full-time equivalent (FTE), average hours worked, direct patient care hours, no longer practising, not licensed in jurisdiction	High
Skill mix	Government policy variables HHR education funding, alternative delivery modes, licencing regulations, professional roles/deployment, recruitment/retention strategies, immigration policy, remuneration rates/types, HHR capacity-building	High
Worker-to-population ratios	Health labour workforce Number of active and employed physicians and nurses	Low
Economic	Population demographics Total population, age/sex distribution, births/deaths, population projections Socio-economic variables Disposable income, GDP growth projections, ethnic factors	High
Needs	Population health status Age/sex mortality, morbidity, acuity Epidemiology Incidence and prevalence rates, hospital discharges, health patterns of the population	High
Service targets	Utilization patterns Number of occupied beds, number of inpatients and outpatients, number of surgeries/screenings/consultations performed, etc.	Low to high

countries. Therefore, such countries may start by using simple techniques such as the worker-to-population ratio or service-based benchmarks to tackle their present imbalances. Developed countries should continue collecting data and enhancing their models, adding less tangible and yet relevant dimensions, such as productivity or skill mix if they are not present already.

Conclusion

In this paper, we reviewed over 60 years of publications in HHR planning. While doing so, we observed the evolution of the field, when and how methodologies emerged, how they have been applied and the robustness of the results, and we also identified the current trends in the field. This work was called for because there is still no accepted methodology to address HHR planning. Given the rampant costs in the health-care sector and the overall influence that health care has on the general welfare of society, as well as the potential impact of shortages on the worldwide supply of medical professionals, an assessment of what has been done and achieved and what remains to be done was necessary to properly guide further developments in this relevant field. Moreover, when we contemplate the complex training process required to earn a licence as a practitioner, we understand that a shortage in medical professionals cannot be accommodated fast enough by decree, either by increasing the number of intakes to medical schools or by inviting more foreign-trained doctors or nurses.

Despite the abundance in approaches and techniques to determine supply and need for professionals, none of the methodologies has ultimately proved to be superior [85]. Recent studies testing current forecasting models show that there is still plenty of room for improvement given the gap between projected and actual results [12].

It becomes even clearer that workforce planning should be accurate and performed in due time given the attritions and the delays in enacting policies in the health-care sector. Adapting medical schools, altering legislation and changing roles is an effort that may take years to bring forth. Therefore, planning has to target a long enough time horizon if it is to be useful and applicable and has to be done pre-emptively.

It now seems obvious that, like any other complex problem, all the determining pieces of the system and their interdependent relationships must be duly accounted for. Therefore, pressing for integrated approaches is still a valid and up-to-date concern. Furthermore, envisioning the health worker in its entire complexity makes it possible to address the problem more comprehensively, leaving room to improvements in productivity and in the distribution of work without having to directly interfere with the training process or with the health providers. Operations research and lean management are particularly relevant in this area. This strategy may be, in fact, a first attempt to solve the lack of professionals.

The results of our review point in one clear direction: accurate HHR planning requires an approach that is both integrated and flexible, featuring supply and demand (potential and effective) and incorporating less tangible factors, such as skill mix and productivity. The road to accurate HHR planning cannot abstain from this.

Endnote

[a]Henceforth, the term 'approach' is used loosely to refer to the conceptual methodology employed rather than to the technical and scientific apparatus used to obtain a projection or forecast.

Abbreviations
HHR: Health-care human resources; OECD: Organisation for Economic Co-operation and Development; WHO: World Health Organization.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
MAL proceeded with the literature review and drafted the paper, with BAL and ASA providing guidance, critical assessment and peer review of the writing. The three authors read, reviewed and approved the final manuscript.

Acknowledgements
We are extremely grateful to all the reviewers for their insightful comments and contributions, as they significantly contributed to the improvement of this paper. Also, we would like to thank several members of the European Operations Research Society and the scientific committee of the EURO Operational Research applied to Health, which provided insightful ideas and feedback on the ongoing work.

Author details
[1]INESC TEC, Faculdade de Engenharia, Universidade do Porto, Porto, Portugal. [2]Faculdade de Economia, Universidade do Porto, Porto, Portugal.

Received: 10 November 2014 Accepted: 2 May 2015
Published online: 24 May 2015

References

1. Dreesch N. An approach to estimating human resource requirements to achieve the Millennium Development Goals. Health Policy and Planning. 2005;20(5):267–76.
2. Ozcan S, Taranto Y, Hornby P. Shaping the health future in Turkey: a new role for human resource planning. The International Journal of Health Planning and Management. 1995;10(4):305–19.
3. Roberfroid D, Leonard C, Stordeur S. Physician supply forecast: better than peering in a crystal ball? Human Resources for Health. 2009;7(1):10.
4. Birch S. A re-examination of the meaning and importance of supplier-induced demand. J Health Econ. 19931–22.
5. WHO. The World Health Report 2006: working together for health. Geneva: WHO; 2006.
6. Williamson AM, Feyer AM. Moderate sleep deprivation produces impairments in cognitive and motor performance equivalent to legally prescribed levels of alcohol intoxication. Occup Environ Med. 2000;57(10): 649–55.
7. Steinbrook R. Private health care in Canada. N Engl J Med. 2014;354(16): 1661–4.
8. Huber M. Health expenditure trends in OECD countries, 1970-1997. Health Care Financ Rev. 1999;21(2):99–117.
9. Health at a Glance 2013 OECD Indicators: OECD Indicators, 2013edn. Health at a Glance 2013 OECD Indicators. OECD Publishing; 2013.
10. Ono T, Schoenstein M, Lafortune G. Health workforce planning in OECD countries. Technical report OECD, France. 2013;62:131.
11. O'Brien-Pallas L, Baumann A, Donner G, Murphy GT, Lochhaas-Gerlach J, Luba M. Forecasting models for human resources in health care. Health and Nursing Policy Issues, 1–10. 2001 21.
12. Van Greuningen M, Batenburg RS, Van der Velden LF. The accuracy of general practitioner workforce projections. Human Resources Health. 2013;11(1):1.
13. Hall T, Mejia A. Health manpower planning: principles, methods, issues. 19781–146.
14. Lomas J, Stoddart GL, Barer ML. Supply projections as planning: a critical review of forecasting net physician requirements in Canada. Soc Sci Med. 1985;20(4):411–24.
15. Dussault G, Buchan J, Sermeus W, Padaiga Z. Assessing future health workforce needs. 20101–46. http://www.euro.who.int/__data/assets/pdf_file/0019/124417/e94295.pdf.
16. Dall T. The physician workforce: projections and research into current issues affecting supply and demand. 20081–111. http://bhpr.hrsa.gov/healthworkforce/reports/physwfissues.pdf.
17. Chopra M, Munro S, Lavis JN, Vist G, Bennett S. Effects of policy options for human resources for health: an analysis of systematic reviews. The Lancet. 2008;371(9613):668–74.
18. Scheffler R. Forecasting the global shortage of physicians: an economic- and needs-based approach. Bull World Health Organ. 2008;86(7):516–23.
19. Schofield DJ, Fletcher SL, Callander EJ. Ageing medical workforce in Australia - where will the medical educators come from? Hum Resources Health. 2009;7(1):82.
20. Folland S, Folland S, Goodman AC, Goodman AC, Stano M, Stano M. The economics of health and health care, 2007th ed. Pearson Prentice Hall, USA; 2007.
21. Lakhan S, Laird C. Addressing the primary care physician shortage in an evolving medical workforce. Int Arch Med. 2009;2(1):14.
22. Maynard A. Medical workforce planning: some forecasting challenges. Aust Econ Rev. 2006;39(3):323–9.
23. Van Greuningen M, Batenburg RS, Van der Velden LF. Ten years of health workforce planning in the Netherlands: a tentative evaluation of GP planning as an example. Hum Resources Health. 2012;10(1):1.
24. Smith D, Martin D, Langefeld C, Miller M, Freedman J. Primary care physician productivity: the physician factor. J General Internal Med. 1995;10(9):495–503.
25. Grossman M. On the concept of health capital and the demand for health. J Pol Econ. 1972;80(2):223–55.
26. Culyer AJ, Wagstaff A. Equity and equality in health and health care. J Health Econ. 1993;12(4):431–57.
27. Al-Jarallah K, Moussa M, Al-Khanfar KF. The physician workforce in Kuwait to the year 2020. The International Journal of Health Planning and Management. 2010;25(1):49–62.

28. Teljeur C, Thomas S, O'Kelly FD, O'Dowd T. General practitioner workforce planning: assessment of four policy directions. BMC Health Serv Res. 2010;10(1):148.
29. Birch S, Eyles J. Needs-based planning of health care: a critical appraisal of the literature. CHEPA Working Paper Series. 1991;91(5):.
30. Birch S, Kephart G, Tomblin-Murphy G, O'Brien-Pallas L, Alder R, MacKenzie A. Human resources planning and the production of health: a needs-based analytical framework. Can Public Policy. 2007;33(s1):1–16.
31. Nugent R. Chronic diseases in developing countries. Ann NY Acad Sci. 2008;1136(1):70–9.
32. Hu FB. Globalization of diabetes: the role of diet, lifestyle, and genes. Diabetes Care. 2011;34(6):1249–57.
33. Murphy GT, O'Brien-Pallas L. How do health human resources policies and practices inhibit change? A plan for the future. 20021–36. http://qspace.library.queensu.ca/bitstream/1974/6884/11/discussion_paper_30_e.pdf.
34. Rosen G. A history of public health. A Johns Hopkins paperback. USA: JHU Press; 1959.
35. Burau V, Blank RH. Comparing health policy: an assessment of typologies of health systems. J Comp Policy Anal Res Prac. 2006;8(1):63–76.
36. Aurousseau P. Distribution of case in Europe - severe shortage of nurses. Can Hosp. 1971;48(10):91–2.
37. Barnighausen T, Bloom DE. Changing research perspectives on the global health workforce. NBER Working Papers. 20091–84.
38. Guerra M. The shortage of physicians. Jornal do medico. 1965;57(173):643.
39. Hale T. Why the nursing shortage persists. N Engl J Med. 1964;270: 1092–1097.
40. Vector Research Inc, Bonder S, Development U. S. B. o. H. R. Health manpower models: an analysis of health manpower models. DHEW. 1974;1:1–220.
41. Doyle TC. An inventory of health manpower models volume II. DHEW. 1975;2:1–313.
42. US SG Consultant Group, Bane F. Physicians for a growing America: report, 1959th ed. USA: Public Health Service, U. S. Dept. of Health, Education, and Welfare; 1959.
43. Statistics U. S. B. o. L. Health manpower, 1966-75: a study of requirements and supply, 1967edn. Report. USA: U.S. Bureau of Labor Statistics; 1967.
44. Baker TD. Dynamics of health manpower planning. Medical Care. 1966;4(4):205–11.
45. Altman SH. Present and future supply of registered nurses. MD: US DHEW, 1–168 (1971). J Hum Resources. 1975;10(3):403–406.
46. Yett DE, Drabek L, Intriligator MD, Kimbell LJ. Health manpower planning: an econometric approach. Health Serv Res. 1972;7(2):134–47.
47. Sloan FA, Richupan S. Short-run supply responses of professional nurses: a microanalysis. J Hum Resources. 1975;10(2):241–57.
48. Baker TD, Perlman M. Health manpower in a developing economy: Taiwan, a case study in planning, 1967th ed. USA: Johns Hopkins Press; 1967.
49. Klarman HE. Requirements for physicians. Am Econ Rev. 1951;41(2): 633–45.
50. Beenhakker HL. Multiple correlation-a technique for prediction of future hospital bed needs. Oper Res. 1963;11(5):824–39.
51. Rosenthal GD. The demand for general hospital facilities, 1964edn. Hospital monograph series, no. 14. USA: American Hospital Assoc; 1964.
52. Navarro V. A systems approach to health planning. Health Serv Res. 1969;4(2):96–111.
53. Maki DR. A forecasting model of manpower requirements in the health occupations, 1967edn. USA: Industrial Relations Center; 1967.
54. Feldstein MS. An aggregate planning model of the health care sector. Medical Care. 1967;5(6):369–81.
55. Health CP. Estimating need for physicians. Bull NY Acad Med. 1968;44(8): 1068–1084.
56. Pathman DE. Estimating rural health professional requirements: an assessment of current methodologies*. J Rural Health. 1991;7(4):327–46.
57. Lipscomb J, Kilpatrick KE, Lee KL, Pieper KS. Determining VA physician requirements through empirically based models. Health Services Res. 1995;29(6):697–717.
58. Grayson MA. Medical educators told to reduce enrollments,. Hospital Med Staff. 1978;7(12):37–40.
59. Schroeder SA. Western European responses to physician oversupply: lessons for the United States. JAMA: J Am Med Assoc. 1984;252(3):373–84.
60. Iglehart JK. From physician shortage to patient shortage: the uncertain future of medical practice. Health Affairs. 1986;5(3):142–51.
61. Gaumer GL, National Center for Health Services Research. Regulating health professionals: a review of the empirical literature, 1984edn. USA: U.S. Department of Health and Human Services, Public Health Service, Office of the Assistant Secretary for Health; 1984.
62. Blomqvist Å. The doctor as double agent: information asymmetry, health insurance, and medical care. J Health Econ. 1991;10(4):411–32.
63. Evans R. Supplier-induced demand : some empirical evidence and implications. Economics of Health and Medical Care. 1974162–73.
64. Reinhardt UE. Physician productivity and the demand for health manpower: an economic analysis, vol. 1, 1975 edn. USA: Ballinger Pub. Co.; 1975.
65. DeFriese GH, Barker BD. Assessing dental manpower requirements: alternative approaches for state and local planning. Issues in dental health policy. USA: Ballinger Pub. Co. 1982.
66. Kramer M, Hafner LP. Shared values: impact on staff nurse job satisfaction and perceived productivity. Nursing Res. 1989;38(3):172–7.
67. Sloan FA. Access to medical care and the local supply of physicians. Medical Care. 1977;15(4):338–46.
68. Scheffler RM, Waitzman NJ, Hillman JM. The productivity of physician assistants and nurse practitioners and health work force policy in the era of managed health care. J Allied Health. 1996;25(3):207–17.
69. Goodman DC, Fisher ES, Bubolz TA, Mohr JE, Poage JF, Wennberg JE. Benchmarking the US physician workforce: an alternative to needs-based or demand-based planning. JAMA: J Am Med Assoc. 1996;276(22):1811–7.
70. Bronkhorst EM, Truin GJ, Batchelor P, Sheiham A. Health through oral health; guidelines for planning and monitoring for oral health care: a critical comment on the WHO model. J Public Health Dent. 1991;51(4): 223–7.
71. Brenninkmeijer OP, Kuitenbrouwer RKJM, Nelissen AMH, Burgersdijk RCW, Netherlands, Ministerie van Welzijn VeC1, et al. Scenario Committee on Dental Health Care: future scenarios on dental health care : a reconnaissance of the period 1990-2020 : scenario report. Norwell, Mass.: Kluwer Academic Publishers; 1993.
72. Denton F, Gafni A, Spencer B. The SHARP way to plan health care services: a description of the system and some illustrative applications in nursing human resource planning. Socio-Economic Planning Sci. 1995;29:1–13.
73. Joyce CM, McNeil JJ, Stoelwinder JU. Time for a new approach to medical workforce planning. Med J Aust. 2004;180(7):343–6.
74. Birch S. Health human resource planning for the new millennium: inputs in the production of health, illness, and recovery in populations. Can J Nurs Res. 2002;33(4):109–14.
75. Ros K, Drzymala L. Planning physician services: is there a method to the madness? Econ Health Policy. 2002;79:1–6.
76. Starkiene L, Smigelskas K, Padaiga Z, Reamy J. The future prospects of Lithuanian family physicians: a 10-year forecasting study. BMC Family Prac. 2005;6(1):41.
77. Maier T, Afentakis A. Forecasting supply and demand in nursing professions: impacts of occupational flexibility and employment structure in Germany. Hum Resour Health. 2013;11(1):1.
78. Tomblin Murphy G, MacKenzie A, Alder R, Birch S, Kephart G, O'Brien-Pallas L. An applied simulation model for estimating the supply of and requirements for registered nurses based on population health needs. Policy, Politics, & Nursing Pract. 2010;10(4):240–51.
79. Masnick K, McDonnell G. A model linking clinical workforce skill mix planning to health and health care dynamics. Hum Resour Health. 2010;8(1):11.
80. Duckett S. Health workforce design for the 21st century. Australian Health Review. 2005;29(2):210–10.
81. Stordeur S, Leonard C. Challenges in physician supply planning: the case of Belgium. Hum Resour Health. 2010;8(1):28.
82. Gallagher JE, Kleinman ER, Harper PR. delling workforce skill-mix: how can dental professionals meet the needs and demands of older people in England? Br Dental J. 2010;208(3):6–6.
83. Astrid Guttmann ECCM. Outcomes-based health human resource planning for maternal, child and youth health care in Canada: a new horizon for the 21st century. Paediatr Child Health. 2009;14(5):310.
84. Tomblin Murphy G, Kephart G, Lethbridge L, O'Brien-Pallas L, Birch S. Planning for what? Challenging the assumptions of health human resources planning. Health Policy. 2009;92(2–3):225–33.

85. Ricketts TC. The health care workforce: will it be ready as the boomers age? A review of how we can know (or not know) the answer. Ann Rev Public Health. 2011;32(1):417–30.
86. WHO. Increasing access to health workers in remote and rural areas through improved retention: global policy recommendations: World Health Organization; 2010. 72 p.

Chapter 3

Forecasting the medical workforce: a stochastic agent-based simulation approach[1]

[1]Published in the Journal of Health Care Management Science with DOI 10.1007/s10729-016-9379-x. URL http://link.springer.com/article/10.1007%2Fs10729-016-9379-x

Health Care Manag Sci
DOI 10.1007/s10729-016-9379-x

Forecasting the medical workforce: a stochastic agent-based simulation approach

Mário Amorim Lopes[1] · **Álvaro Santos Almeida**[2] · **Bernardo Almada-Lobo**[1]

Received: 13 February 2016 / Accepted: 22 August 2016

Abstract Starting in the 50s, healthcare workforce planning became a major concern for researchers and policy makers, since an imbalance of health professionals may create a serious insufficiency in the health system, and eventually lead to avoidable patient deaths. As such, methodologies and techniques have evolved significantly throughout the years, and simulation, in particular system dynamics, has been used broadly. However, tools such as stochastic agent-based simulation offer additional advantages for conducting forecasts, making it straightforward to incorporate microeconomic foundations and behavior rules into the agents. Surprisingly, we found no application of agent-based simulation to healthcare workforce planning above the hospital level. In this paper we develop a stochastic agent-based simulation model to forecast the supply of physicians and apply it to the Portuguese physician workforce. Moreover, we study the effect of variability in key input parameters using Monte Carlo simulation, concluding that small deviations in emigration or dropout rates may originate disparate forecasts. We also present different scenarios reflecting opposing policy directions and quantify their effect using the model. Finally, we perform an analysis of the impact of existing demographic projections on the demand for healthcare services. Results suggest that despite a declining population there may not be enough physicians to deliver all the care an ageing population may require. Such conclusion challenges anecdotal evidence of a surplus of physicians, supported mainly by the observation that Portugal has more physicians than the EU average.

Keywords Healthcare workforce planning · Agent-based simulation · Health economics · Monte Carlos analysis

✉ Mário Amorim Lopes
mario.lopes@fe.up.pt

Álvaro Santos Almeida
almeida@fep.up.pt

Bernardo Almada-Lobo
almada.lobo@fe.up.pt

1 INESC-TEC Porto, Faculdade de Engenharia da Universidade do Porto, Porto, Portugal

2 CEF-UP, Faculdade de Economia da Universidade do Porto, Porto, Portugal

1 Introduction

Health care delivery is a labor-intensive sector, composed of physicians, nurses, clinical assistants and other administrative staff. Imbalances in the supply of human resources delivering health care services are known to cause severe economic and social harm [4], and may ultimately lead to patient deaths [52]. Concern has been growing over a future lack of human resources for health [49], and the United Nations has included this pressing problem in their Millennium Development Goals [21]. It is not an isolated problem, as a lack of medical staff in developed countries may divert scarce resources from developing countries. As a result, several health authorities and academic institutions have been addressing this problem. The work hereby presented is part of one such initiative, the EU Joint Action on Healthcare Workforce Planning and Forecasting. One of the main goals is to develop a robust model to forecast the evolution of the medical workforce. The first step was to consult the literature. After reviewing over 60 years of research in the field to understand how it evolved in terms of methodologies and its current state [4], we noted that no

Published online: 03 September 2016

best or preferred solution exists, but there is a clear trend pointing to an integrated approach. Best practices suggest that such model should integrate both drivers of supply and demand for health care services, as well as epidemiological needs, productivity changes and skill mix [4]. The latest papers published on the subject employ methods that target most if not all of these dimensions [36, 39, 53, 56, 57]. Figure 1 exemplifies one such conceptual model. For the purpose of implementing the forecasting models, simulation has been used extensively, in particular System Dynamics (SD) [18]. Simulation as a methodological approach is a typical choice for modelling complex and non-linear systems. Several past works employ SD to generate forecasts of the physician, nursing or dentistry workforce [9, 33, 39]. However, SD faces challenges in modelling complex human behavior or in dealing with spatial awareness, including that of physicians and nurses. Its mechanical architecture of stocks and flows makes it difficult and unnatural (but not impossible) to model complex human behavior that dictates, for instance, the decision to emigrate, where the physician decides to practice or whether a physician opts for public or private practice. Furthermore, it has limited functionality for adding reactiveness or ability to learn, as no definition of agent or entity exists. For instance, adding an heterogeneous utility function to model the reaction to a wage increase (or decrease) is extremely complicated to achieve with SD, and even more so when such utility function may adapt with time. All of these cases are relevant to the matter at issue. Notwithstanding the merits of SD, we are of the view that an agent-based model (ABM) is best suited to address this problem. ABM has been used in the healthcare sector [10, 15, 38, 43] to study epidemiological phenomena [40], user-payer-provider interactions [32], the effects of patients derivation policies in emergency departments [54], how new drugs are diffused through social interaction [45], and many other applications. However, we found little to no use in the healthcare workforce planning literature [18], despite having several interesting benefits: i) it allows incorporating behavior rules and microeconomic foundations into the agents, adding several degrees of freedom to the model; ii) the agents logic can be modelled using state charts that are easy to visualize and understand, assisting in an external validation by experts in the field; iii) it makes it easy to define dynamic agent interactions, a feature particularly useful to study horizontal (between specialties) and vertical (between physicians, nurses or any other agent) substitutability (skill-mix); iv) being a more intuitive process, policy-makers are more likely to understand results and policy implications. In this work we present a stochastic agent-based model to address the supply side of the integrated system depicted in Fig. 1. As a case study, we have used it to forecast the supply of physicians (all of the 52 specialties) of the Portuguese health care workforce until 2050. To the best of our knowledge no such work exists, and the literature reviews consulted found no application of ABM to workforce planning, and very few exist to health care planning in general [4, 18]. The model is calibrated with real data reported by the Portuguese health authorities. We also ran sensitivity and scenario analysis to assist in the validation of the model. The remaining chapters are organized as follows. Section 2 provides a literature review of the apparatus typically used in the healthcare research field, exposing merits and pitfalls of each approach. Section 3 thoroughly describes the ABM model used to forecast physicians, including state diagrams, parameters, calibration and how uncertainty was factored in, and how we modelled the demographic impact on demand. Section 4 presents the results, as well as sensitivity and scenario analyses to help interpret and validate the results. Finally, Section 5 discusses the results and provides directions for future work and Section 6 concludes.

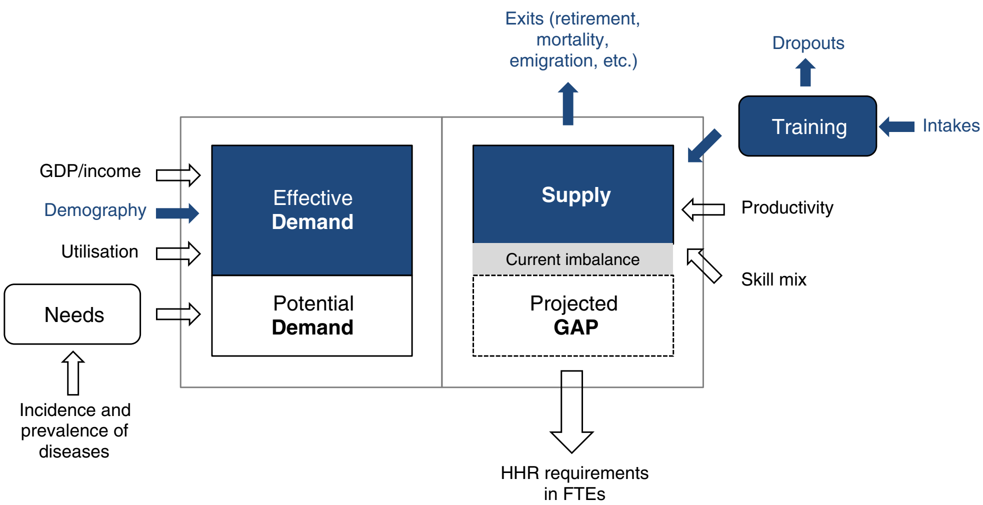


Fig. 1 A functional description of a desirable model for healthcare workforce planning. The part of the model covered in this manuscript is highlighted in blue. Adapted from: [4]

2 Literature review

In the very early stages of research in the field, when health human resources were modelled as input factors of a production function [4], the methods and techniques used to forecast the evolution of health professionals usually involved econometric regressions. That is the case of the work by Yett and colleagues [61], who employed a micro- and macroeconometric model to project physicians' requirements. Alternatively, Delphi panels of experts were formed to provide their own forecasts based on intuition, knowledge or both [34]. When data were lacking, researchers usually resorted to simple worker-to-population ratios, solving for the physician output required to maintain the physician to population ratio at base year [42].

As models grew more complex in order to incorporate further socio-economic variables, the methods had to follow suit. Approaches such as linear programming, Markov chains or simulation became more common. To illustrate the usage of linear programming for planning the healthcare workforce, Lavieri et al created a model to plan the registered nurses' workforce in British Columbia, Canada [30, 31]. More recently, Lagarde and Cairns demonstrate how to simulate and assess immediate and long-term effects of policy interventions in the South African nursing labour market using Markov chains, and how to use the model to inform policy-makers on how to reduce staff shortages [29].

Notwithstanding, the method of choice for modelling the healthcare workforce, especially the supply-side, appears to be System Dynamics simulation, which can be found in several publications [7, 9, 33, 39]. SD takes a system-level approach, decoupling the system into modules (supply, demand, needs, etc.) that can be further decoupled. The system is then thought of and modelled in terms of stocks (levels) and flows (rates) affecting the system behaviour. The flows need not to be constant. How the system develops may reinforce or weaken some of these flows, and in some particular cases internal feedback loops may arise. Consider, for example, the stock of physicians undergoing medical training. Each year there will be a flow of fresh students entering medical school and a flow of students exiting (either finishing, dropping out, emigrating, etc.) medical school. These flows will influence the stock of physicians undergoing medical training. If we simulate all these flows for a given period of time we are able to generate forecasts. Implementing the whole system is then a matter of including all the stocks, flows and feedback loops that best describe how things work in real life.

Methodological issues apart, more recent research has been taking a different angle at human resources for health. In lieu of dealing with physicians or nurses as input factors, they became economic agents driven by incentives [5, 11]. Such view implies that physicians, nurses, dentists or any other health professional are more than static entities. Agents may have their own behaviour rules, or microeconomic foundations, and thus behave differently from each other. Capturing heterogeneity across individuals is particularly relevant when researchers want to study the decision-making process of health professionals. Although technically possible, incorporating such features into a SD model can be extremely complex. Agent-based simulation modelling was created in order to tackle these insufficiencies, and has been drawing a lot of attention. In Nature's opinion column *"The economy needs agent-based modelling"*, from Farmer and Foley, the authors make a good case for using ABM to model the economy, urging economists to adopt it [22].

System dynamics and agent-based modelling represent two different approaches that remind of the separation between macro and microeconomics prior to its synthesis, before macro incorporated microeconomic foundations. Likewise, while SD deals with aggregates (stocks), the functional relationships between these aggregates (flows) and the causal loops resulting from these interactions, ABM models a system starting from the bottom to the top (bottom-up), simulating agents with decision-making capabilities, behavior rules and interactions with other agents and the environment. Global behaviors and trends emerge from this interaction [32].

Agent-based modelling has been used extensively in areas of science as vast as land-use and agricultural policy [13, 25], economic policy [19], traffic control [59], epidemiology [40] or education [51], and have become critical tools to inform policy-makers. Arguably, some authors have decided to resort to ABM precisely because traditional approaches, such as statistical regressions, are unable to capture the dynamic processes. Brailsford and Schmidt, for instance, claim that incorporating human behavior in models of health care systems is a more accurate method for modelling a broad class of problems related to healthcare, and propose a simulation model for that purpose [17]. Auchincloss et al propose ABM for understanding place effects on health, stating that classical methods have constrained both the research questions and the theoretical formulations developed [8]. Other authors highlight the stochastic nature of ABM models, a feature most relevant when the research question targets a distribution of possible outcomes, as frequently is the case of research carrying policy implications, rather than the output from a single run of a deterministic model [46].

Despite the call for action, few publications have used ABM in the field of health services research and related areas. In fact, it has been the case that only recently have articles featuring ABM been published, notwithstanding its merits for health policy planning being long established [44]. As expected, articles resorting to ABM typically need to model the behaviour of individuals. For instance, Day and colleagues use ABM to study patients with diabetic retinopathy, a case in which ABM allows studying interactions between individuals and the environment [20]. Gorman et al, in a similar fashion, employ ABM to examine the drinking behavior, showing that drinking behavior is linked to the social influences emerging from contacts between drinkers [24]. In a more operations research oriented study, Jones et al briefly describe how ABM can be applied to scheduling emergency department physicians in order to avoid overcrowding [50].

However, when we refocus on healthcare workforce planning, we were unable to find any publications employing ABM to forecast future physicians' or nurses' requirements. In order to obtain a very detailed forecast, answering to a question such as *"How many physicians will we need in the future?"* may involve studying their spatial location in the territory, as the physician workforce is not uniformly distributed throughout the country; the interactions between specialties and other health professions, such as nurses (i.e. studying the skill mix); or the impact of physician migration. ABM is a very adequate tool to answer these questions.

3 Methods

For a problem that is largely dependent on human resources and in which individual preferences and decision-making may influence the outcome, ABM offers several advantages over SD. Surprisingly, we have not found an ABM approach to medical forecasting, and neither have other authors [18], despite the numerous advantages this method presents. This work uses a fully-fledged ABM model to generate forecasts of the physician workforce. In Appendix A we also provide an alternative mathematical formulation for the ABM model, which may be easier to follow through for those having no background in ABM simulation, but possessing some expertise in SD or in differential equations.

3.1 Supply model

The supply-side consists of an ABM, with agents representing physicians and an entity representing the government. This entity manages decisions outside the reach of the medical staff, such as the number of vacancies to medical school, the number of years required to complete the course, the number of employment contracts, the wage level, etc. These decision variables have a large impact on the model, and may be used to optimize the system at a later stage. Agents representing physicians try to reproduce their expected behavior throughout life, which includes university attendance to earn a license to practice, going to public or private practice and finally retiring. The life stages are modelled using state diagrams. Transitions between states are stochastic, and follow statistical distributions that best model the probability of jumping from one state to another. For instance, in some countries medical school requires a minimum of 6 years to complete, but some students may take longer. To incorporate this effect we assume the time required to transition follows an exponential decay distribution calibrated as close as possible to available statistical records, which we validate running a goodness of fit test model.

3.1.1 Physician agent

The physician agent represents a physician throughout the various stages required by Portuguese law to obtain a diploma and be entitled to practice medicine. The lifecycle is as follows: upon finishing high school, he enters medical school (state "Med School"), which takes at least 6 years to complete. Afterwards, he is required to practice for a year as an intern (state "IAC"). When finished, the student can proceed to the specialty exam, and depending on the grade obtained choose from a given list of specialty vacancies distributed throughout the country (state "Specialty School"). If he is unable to obtain a vacancy, a transition to state "No specialty" occurs. The student will have another chance the next year. Entering a given specialty is subject to the number of training places available and to the grade obtained in the specialty exam. After some years, which varies according to the specialty chosen, he finally graduates (state "Graduation"). At any given moment he may dropout from university. In case he does not, he graduates. Joining the national health system (state "Public Practice") is then dependent upon the number of contracts available for that year, but he may choose to go directly to private practice (state "Private Practice") or transition between them. In both cases he is part of the composite state "Workforce". When the physician is approximately 66 years of age he qualifies for retirement (state "Retired"). All the transitions that may occur during the lifecycle are represented in Fig. 2.

Immigration of physicians is handled differently, since it requires creating new agents. It is not a state in the physician lifecycle, but rather the creation of a new agent. It works as follows. Each year, the model summons a given number of foreign doctors, with the number of physicians generated varying according to historical trends [47]. Age and

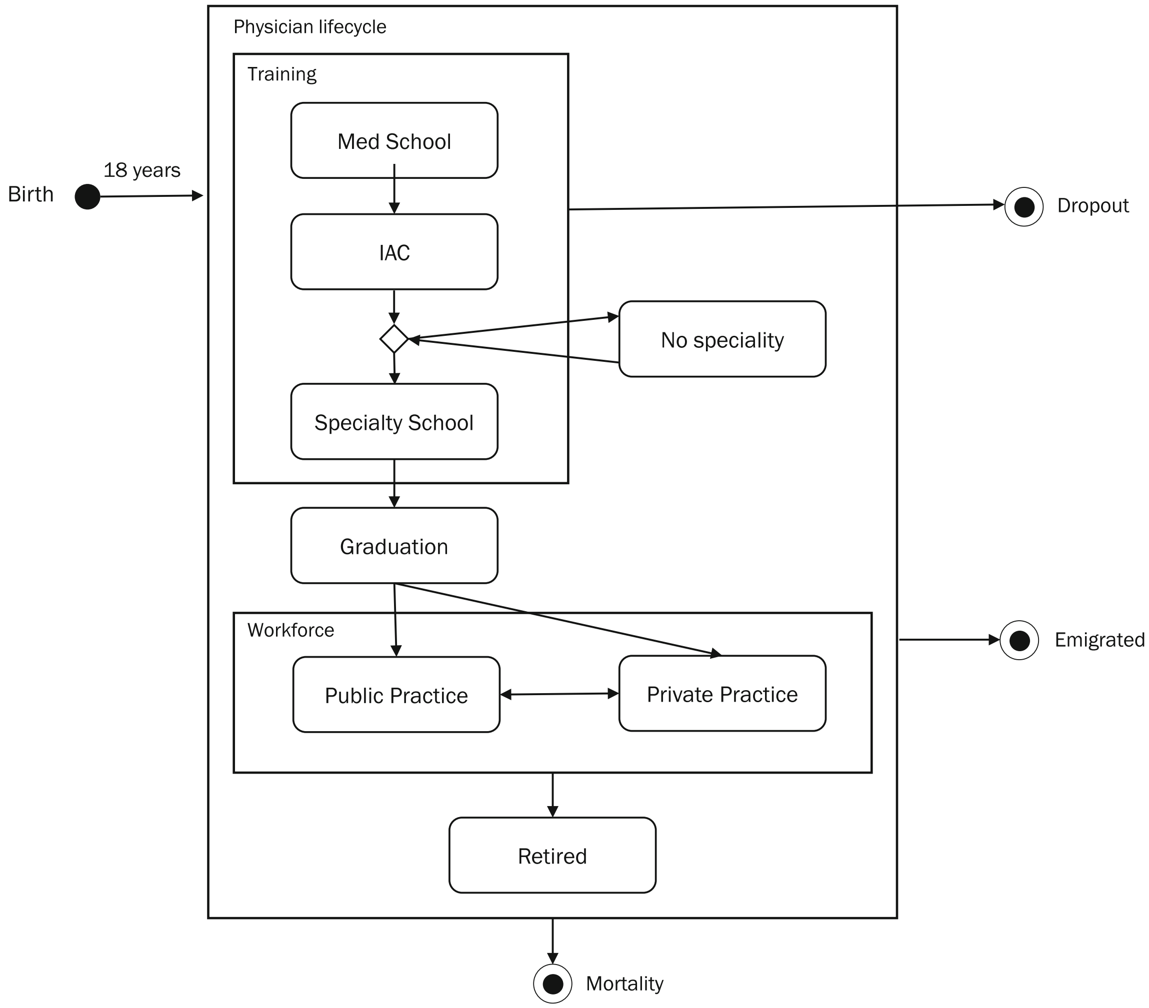


Fig. 2 The state chart defining the inner logic of the agent representing a physician

specialty are attributed randomly, taking into consideration the typical profile of a foreign doctor. For instance, foreign doctors are on average below 40 years of age.

Sources of uncertainty There are several sources of uncertainty surrounding the life of a physician, which were incorporated in the simulation model. First, the number of years the physician stays in each state. Students take 6 years (mode) to complete medical school, but they may take longer. Likewise, the time required to complete specialty school and graduate, which varies according to the specialty chosen, is also subject to some uncertainty. Moreover, physicians may emigrate, dropout from school or suddenly pass away. These delays and accidental actions may weight on the forecasts, thus affecting the supply of physicians. Considering this, we have modelled the transitions using statistical distributions that best purport the real life events so that the effects of randomness can be properly accounted for. In Table 1 we provide a description of all the transitions, and how uncertainty affects them in each case.

3.1.2 Government

The Government is not an agent (i.e. it is stateless), but rather a model class executing routine actions that need to be performed every year (but it could be modelled as such). In particular, in some countries the Government or third-party institutions with a mandate are responsible for setting the number of intakes both to medical school and to residency training, hire physicians to the medical workforce, hire foreign doctors, give medical practice license to students and eventually decide on the wage levels. This class is able to perform all of these actions. These parameters may be changed in order to understand the impact of policy changes using *what-if* scenarios and cost-benefit analysis.

Table 1 Sources of uncertainty affecting the Physician agent

Transition	Description
Med School to IAC	The minimum time required to complete medical school is 6 years at any domestic and accredited university, but some students may take longer. The majority take 6 years to complete.
IAC to Specialty	The internship is a one-year period, but it may take longer if the student fails the specialty exam and is not admitted to residency training.
Specialty to Graduation	Most of the medical specialties take between 3 and 6 years to complete, but some students may take longer due to unforeseen circumstances.
Graduation to Public/Private Practice	A certain number of physicians are given a job at the public sector each year, while others go directly to private practice. The number of contracts is decreed by the Government (no uncertainty) and the rate of physicians absorbed by the private sector is an estimate.
Public to Private Practice (and vice-versa)	At any time, physicians may move between public and private practice.
Workforce to Retired	The retirement age is 66 years, but a significant number of physicians ask for early retirement. On average, it is 63.4.
Mortality	Currently, life expectancy in Portugal is 77 years for men and 83 for women born in 2001. Most of the physicians currently employed are over 40 years of age.
Dropout	According to official statistics, approximately 2 % of the students quit university during the first years. Dropouts are picked randomly.
Emigration	No official statistics exist, but according to the Medical Association "hundreds"? (approximately 200) of physicians are asking for permits each year, but may or may not follow through and actually emigrate. We assume all emigrate, which may not be exact as there are no records whether they have actually moved abroad.

For instance, what would be the impact of increasing the vacancies to pediatrics? Moreover, trade-off analysis can also be undertaken. If we incorporate tuition costs, analysts and policy makers become better equipped to decide whether it is preferable to increase the *numerus clausus* or increase wage and attract more foreign doctors while decreasing emigration.

3.1.3 Environment

Similar to the Government, the environment is not an agent. The environment has no impact on the projections. Its sole purpose is to load the initial stock of students and actual physicians, load parameters and other data, record statistics and write the results to an external file, reproducing the current state of affairs as accurately as possible. The data loaded includes the current stock of physicians by age cohort (0 to 29, 30 to 39, 40 to 49, 50 to 59, 60 to 64, 65 or over) and specialty; the number of students currently enrolled in a medical course (total number); the number of physicians by age cohort doing internship and residency; the number of students that have graduated recently and are still applying for a license; and the number of physicians currently applying for specialty, distributed by specialty and age cohort. Also, parameters that may be changed by the analyst are loaded into the environment. Some of these parameters are the *numerus clausus* for entering medical school, the number of vacancies available for each specialty, the minimum duration of each specialty, the number of physicians by specialty hired to the national health system each year, the number of foreign doctors arriving each year, the dropout rate, the emigration rate, the percentage of physicians employed each year by the private sector, and the percentage of physicians moving to private practice each year.

3.2 Demand model

Answering whether a given number of physicians is sufficient to provide all the necessary health care requires also knowing how many health care services will be needed in the future. Such question is easy to pose, but hard to answer. Figure 1 shows why. Demand is dependent on several factors, such as the evolution of demography (population size and age structure), the GDP growth (health care spending, and hence health care services is highly correlated to GDP) and the overall morbidity of the population, since different pathologies put a different burden on the health system.

At this stage, due to the lack of data it is not possible to take all these variables into consideration. One can, nonetheless, study the isolated effect of demography on the demand for health care services, as demographic projections do exist, and use it to project the ratio of physicians to population size through the time horizon:

$$\text{ratio}_t = \frac{\text{physicians}_t}{\text{population}_t}. \tag{1}$$

By doing so we are implicitly assuming there are currently no imbalances, and hence the present ratio of physicians to population is a good approximation to a balanced workforce [5]. The number of physicians for each year is estimated by the ABM model, while for the population growth we use the official projections provided by the National Institute of Statistics (INE) [27]. The institute elaborates two scenarios other than the baseline case. Currently, the Portuguese population is estimated to be 10.5 million people, and in all scenarios the population is expected to decrease by 2060. In the baseline scenario the population is expected to decrease to 8.5 million people; in the low (pessimistic) scenario the population decreases to 6.3 million people and in the high (optimistic) scenario the population size is expected to slightly decrease to 9.2 million people.

Also important, there is a significant shift in the age structure of the population, as Fig. 3 shows. In all scenarios, the Portuguese population ages significantly, which has a strong impact on the demand for health care services. According to several studies [14], the quantity of health care services tends to increase past a certain age, especially near the end of life. In fact, health care spending by age group, a proxy to the health care services provided, confirms this empirical evidence. Figure 4 shows the public health care spending by age group for the Portuguese population according to OECD data [37].

The population decline in the three projections could lead one to inadvertently conclude that the ratio of physicians to population, in case the number of physicians stays relatively constant, would increase significantly, and therefore a surplus in the physician workforce would occur. Arriving at this conclusion based solely on such observation and without accounting also for the ageing of the population would be extremely weak. We can approximate the effect of the population's age structure by multiplying the health care spending by age group by the number of people in each age group. We then obtain a given quantity Q of health care services by computing:

$$Q_a = \text{population}_a \times f(\text{health\_spending}_a) \tag{2}$$

where a is the age ($a = 0, 1, \ldots, 95+$) and f is a rescaled function of health care spending by age group, with the minimum value (10-14) set to 1.0. Figure 5 illustrates the impact of incorporating the ageing of the population into the demographic forecasts. In this case, only in the low scenario there would be a decrease in the demand for health care services. Both in the baseline and in the high scenario the demand for health care services would continue to increase.

By using Eq. 2 and indexing it to time we can then compute a more accurate proxy to demand for health care services:

$$\text{ratio}_t = \frac{\text{physicians}_t}{\sum_{a=0}^{95+} Q_a(t)}. \tag{3}$$

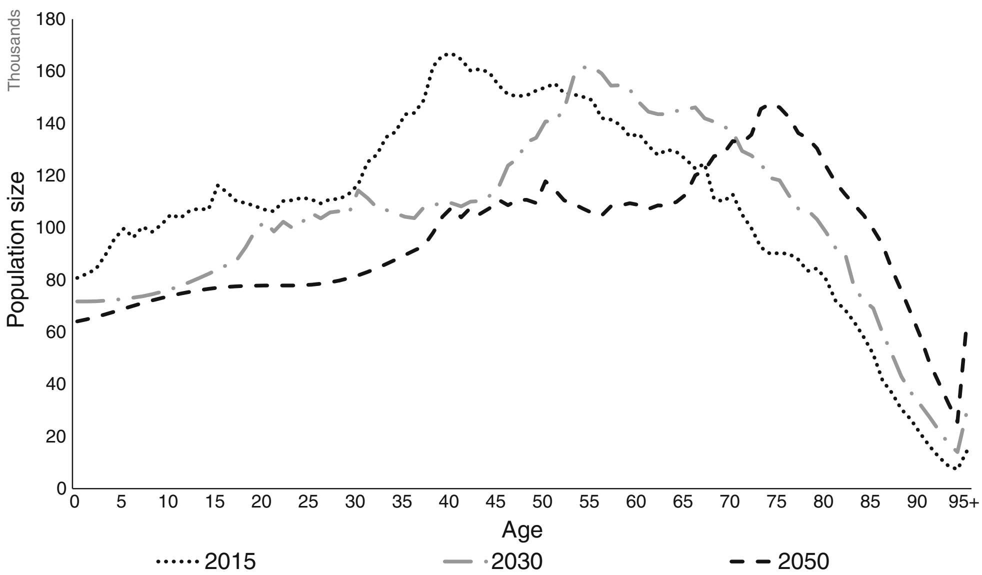


Fig. 3 Baseline projections for the Portuguese population. Source: [27]

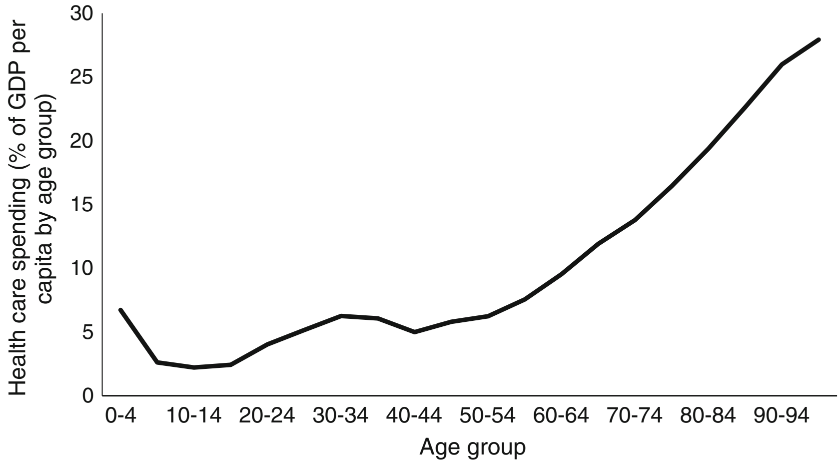


Fig. 4 Public health care spending by age group as a % of GDP per capita for Portugal. Source: [37]

3.3 Model implementation

The simulation model was used to forecast the evolution of the physician (including the 52 medical specialties) workforce in Portugal until 2050. Portugal is an interesting and relevant case study. It has improved remarkably in the last decades in terms of health status of the population. The mortality rate has declined significantly, while life expectancy has increased and converged to the rates of other Western European countries [12]. At the heart of these improvements are a National Health System (SNS) and a private sector composed of physicians, nurses and other medical and administrative staff that provide health care services to the population. With 3.77 physicians per 1 000 population, Portugal has one of highest rates of physicians in the EU27, which has an overall average of 3.28. In terms of nursing personnel, the country has one of the lowest rates, 534 per 1 000 population, low when compared to the EU27 average of 7.95 [60], but similar to Spain or Italy, while exhibiting similar mortality and morbidity levels [6]. It also registers one of the lowest ratios of nurses per physician (1.4 nurses for each physician, compared to 3.9 nurses per physician in Switzerland). Notwithstanding, doubts persist whether there may be a shortage of physicians in the present, and whether it will be enough to cover the future demand for health care. Technical reports conducted by the Portuguese health authorities report a shortage for a significant number of medical specialties, and anticipate a decline in the short-term as a result of the high number of physicians expected to retire soon [1] (see Fig. 6). According to this study it is estimated that there is a current deficit of 145 stomatologists or of 245 general practitioners.

The model was calibrated using real data made available by the Portuguese health authorities (ACSS) and INE. Information regarding the current size of the workforce and

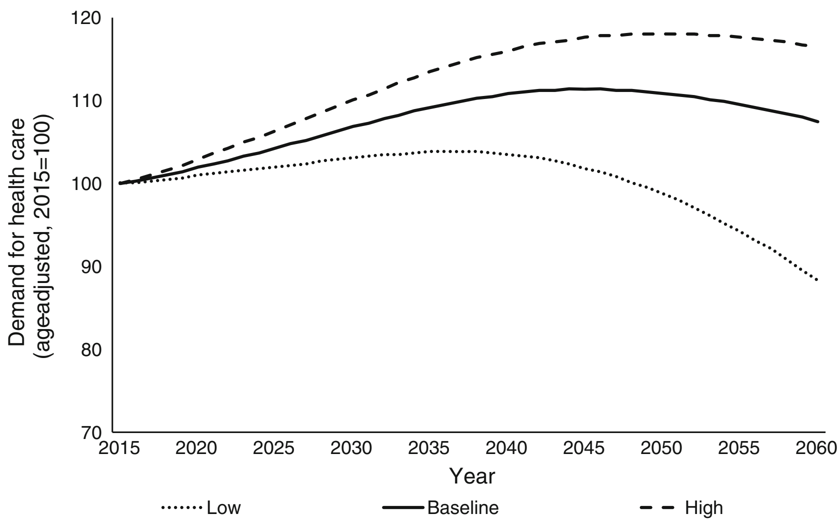


Fig. 5 A projection of the health care services demanded by age group. Source: author's calculations based on projections by INE and on OECD data

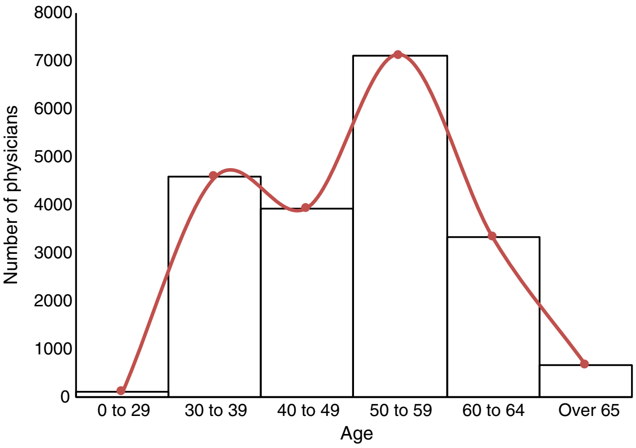


Fig. 6 Age profile of the current medical workforce and the distribution used to extrapolate the age of the initial stock of physicians

students enrolled at each stage of training were inputted into the model (Table 2).

Parameters were calibrated using statistical distributions that best fit the empirical data available or using official records when available (Table 3). When such information is not available, rates are inferred. For instance, the emigration rate was set so that the number of emigrated physicians falls within likely boundaries. Likewise, the number of foreign doctors hired each year was approximated to a 10-year average growth rate, as calculated by Ribeiro et al [47]. This parameter is in fact a decision variable, although at this point it is not used as such, and thus it is assumed it just follows the historical trend (see Fig. 7).

It is safe to assume the time required to complete each stage of the training phase can be approximated by a simple exponential function of time, for the reason that there is a minimum number of years required to complete training and after some years of inactivity the college membership is suspended. Using official records from one of the leading Portuguese medical schools [23], we fit an exponential distribution to the number of times (years) students have to enrol until they graduate. The distribution was calibrated using maximum likelihood estimation, followed by goodness-of-fit tests to verify the adequacy of the statistical distribution. Since no records exist for the remaining

Table 2 Initial data loaded into the model. Source: Portuguese health authorities [2]

Data	Calibration
Physician workforce	Number of physicians by specialty and age cohort
Vocational training graduates	Number of students in medical school, internship or specialty residency (physicians)

transitions (i.e. from IAC to residency and from residency to full graduation), we assume similar exponential functions of time adapted to the duration of each phase (i.e. IAC is expected to only take one year, so we set the minimum at 1).

Also, students move from IAC to a residency program if there are vacancies. In real life, medical students have to take an exam, and based on the grade obtained they choose from a list of specialities, with the best students having priority over the others. In case they are unable to enrol in the residency program they aspired for, they end up going for the next in the list. In the simulation model we do not require such a complex procedure. We just need to draw a random residency program for each physician based on the number of vacancies available. If there are more candidates than vacancies those unable to obtain a spot will have to wait for a year until they can apply again.

With respect to the retirement age, there are no specific statistics for the physician professional class. We therefore assume general retirement statistics also apply to physicians. Citizens may qualify for early retirement or postpone their retirement, thus the normal distribution is a suitable candidate to approximate the empirical data. As it currently stands the legal age is set at 66 years and 3 months, but the mean is approximately 63.4 years of age due to the effect of early retirement. The parameters used to calibrate the distributions for modelling transitions are listed in Table 4.

In the case of agents still in the training phase that were imported into the model another step is required. Since we are unable to determine whether they have just enrolled or are about to complete medical school or specialty residency (statistics obtained only report the aggregate number of students enrolled), we use uniform distributions to estimate their age and, therefore, the year they are at. This implies that we assume students are uniformly distributed throughout the graduation years.

Finally, the agent-based simulation model was implemented using AnyLogic v7.2 (XJ Technologies, Chicago, IL), a leading software tool for simulation.

3.3.1 Data sources and limitations

Data was retrieved from several sources, as a comprehensive database is not available (Table 3). In the public sector (SNS), Portuguese health authorities keep an up-to-date database for payroll purposes, containing all the relevant information about the professionals, including headcount, age, sex, specialty, location, etc. The information was collected from the annual reports for the year of 2014 [2]. Health authorities in partnership with the Ministry of Education also keep track of information regarding the number of students enrolled in medical courses, vacancies to medical schools, average time required to complete the degree,

Table 3 Parameters included in the supply workforce model

Variable	Value	Source
Foreign doctors	Between 20 and 30 arriving each year. Age: $\mathcal{N}(30,\ 5)$.	ACSS and [47]
Dropouts	2 %	Ministry of Higher Education
Retirement age	$\mathcal{N}(63.4,\ 4.3)$	Labour Force Survey (INE) and [35]
Mortality rate	Lifetable. Not adjusted to profession.	INE
Numerus clausus	1645	Ministry of Higher Education
Residency vacancies	1496 (distributed through 52 specialties)	ACSS
Yearly contracts for joining the public sector	1934	ACSS
Specialty duration	4-6 years depending on specialty chosen	ACSS
Emigration	0.9 % (estimated) (rate that approximates an emigration level between 100 and 200 physicians/year)	Medical Association

etc. Public universities are also required by law to publish drop out rates. Residency vacancies, which are set by the Medical Association and by the health authorities, are also published annually.

In the private sector, the Medical Association keeps track of the stock of physicians. Unfortunately, their records do not provide information whether registered professionals are still active, retired, dead or have emigrated; the type of activity they perform; or if they practice in the private or in the public sector [47]. The OECD claims these numbers overestimate the workforce size by 30 % [41]. For this reason, this work covers the public sector mainly. Note that such decision has no impact in the model per se, as we still account for the net number of physicians potentially moving to the private sector, and generate projections for the hypothetical evolution of the private sector according to the same principles governing any other physician. Whenever the initial stock of doctors practicing in the private sector is accurately determined, it can be added to the projections.

As for the number of physicians coming from abroad (immigration), historical data exists, but no forecasts are available to be incorporated in the model. The number of foreign doctors currently working in the public sector was retrieved from the work of Ribeiro et al [47], and using that data we ran a linear regression to obtain a long-term trend.

The only parameter for which no precise information exists is the number of physicians that have emigrated and the number of physicians moving from public to private practice. Regarding the emigration rate, the Medical Association claims that between 100 and 200 physicians emigrate per year. Although not all physicians asking for a best practices certificate from the Medical Association actually

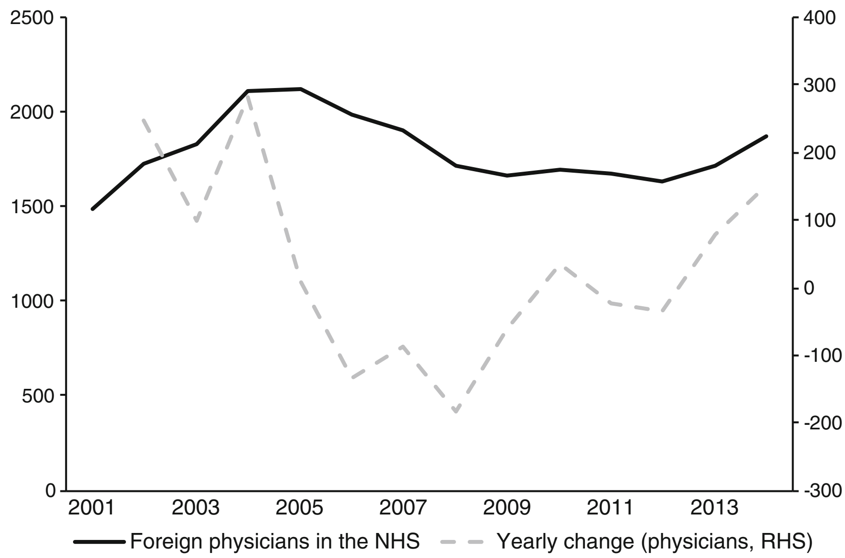


Fig. 7 Historical data of the stock of foreign physicians working in the SNS (LHS) and of the yearly change (RHS)

Table 4 Distributions used to model the time for moving between states

Transition	Description
Med School to IAC	Exponential (truncated, min 6, max 12, shift 0, stretch 2)
IAC to Specialty	Exponential (truncated, min 1, max 2, shift 0, stretch 0.4)
Specialty to Graduated	Exponential (truncated, min $specialty\_duration$, max $specialty\_duration + 4$, shift 0, stretch 1)
Workforce to Retired	$\mathcal{N}(63.4,\ 4.3)$

emigrate (which is how they estimate the number of emigrants), since no other information exists we approximate the emigration rate to best match these numbers. We then confirm in the results file (cf. Table 7) that the number of physicians emigrating varies within these boundaries. The net migration from public to private practice is less concerning, since physicians still exist in the system and are accounted for.

Finally, the physicians' gender, their exact age and the number of physicians that practice both in public and private practice is also missing, which makes it impossible to convert headcounts to full time equivalents (FTE). Similarly, feminization of the workforce is a relevant factor potentially influencing the conversion to FTEs [26].

3.4 Verification and validation

In order to verify and validate the model we have followed the best practices reported in the literature for ensuring the model is an accurate representation of the system being simulated and was implemented correctly [28, 48]. In particular, to ensure the model performs as intended (verification) we followed good programming practices, namely: (i) unit tests at a modular level allow us to verify whether a piece of code being tested acts as expected and is not broken by a new addition; (ii) observable tools such as graphical displays provide feedback of what the model is doing at any given time, a simple but powerful way of monitoring its execution; (iii) a simple *ex post* verification system double-checks the results in Excel (i.e. it sums all the inflows and outflows for a given year and compares them to the variation in stocks reported by the simulation model). Using all these tools at once we were able to detect a lot of unusual or strange behavior arising from badly implemented code.

In terms of validation, which consists in determining whether the conceptual simulation model accurately represents the real system [28], we used real-world data and designed experiments based on Monte Carlo sampling to test whether the model was behaving as expected. Due to the unavailability of historical records, it is not yet possible to perform a retrospective (backwards) simulation, which could be used as an additional validation step [58]. The inventories of human resources only go as far as 2013, which is insufficient for a model with a long time horizon. Also, using data from other countries is not feasible, since it would require several changes to the original model in order to adapt to the specificities of that country. Consequently, the model being validated would be a different one, rendering the validation useless. However, this does not imply the model is not generalizable, in the sense that it is possible to alter it according to the rules in practice in each country, calibrate it with the respective data and obtain forecasts. For instance, if there is no IAC (mandatory practice year between medical school and residency) in a given country, the researcher may simply delete that state out of the physician's lifecycle state chart and run the model. What is not possible is to validate the model as it currently stands using data from other countries. The model would have to be changed accordingly, and by doing so we would be validating a slightly different model.

Instead, we resorted to indirect means to ensure the model was correctly implemented. First, we elaborated experiments with predictable outcomes to verify if the model behaved according to the expectations. For instance, in a scenario where the number of training places is increased there should be an increase in the workforce numbers in comparison to a baseline scenario where no change is performed. Likewise, if the retirement age is raised, the number of retirees should decrease while the workforce size should increase.

We have also used other indicators to validate the model. A significant number of retirees is expected to take place in the next years, a result of the age structure of the current workforce (cf. Fig. 6), after which it should revert to historical retirement levels. Such effect must be perceptible in the projections. Between 2015 and 2017 there are almost twice the number of retirees than the average for the remaining years. The number of retirees then reverts to historical levels for the same workforce size, between 200 and 500 per year [3]. This behavior is consistent with the empirical data. Finally, some unit-level tests were performed on the physician agent to ensure the model executed all the relevant functions correctly, in particular the lifecycle of the agents. Deviations from expected behavior trigger a warning or an error depending on the degree of importance. For instance, if an agent does not die, if he gets stuck in a given state, etc., the model suspends execution. Similarly, the outcome of critical functions is verified for each simulated year. For instance, if a given number of training places are created, the number of agents taking those places cannot exceed the slots available. These unit-level tests helped in identifying bugs and implementation errors during development and testing phases, or when new features are implemented.

4 Results, validation and scenario analysis

A baseline analysis was conducted to forecast the supply of physicians. It was calibrated to match current policies and reproduce the factors affecting the health labor market, including the number of training places and friction rates (i.e. the number of physicians that leave the workforce in search of a new job, drop out from school, move to private practice, etc). This analysis is used as a benchmark against other scenarios in which policy levers (i.e. decision variables) are changed. In the current formulation any parameter can be defined as a decision variable, but in fact only *numerus clausus* and the number of contracts to join the public workforce, which is directly linked to the number of training places for specialty school, is used as a *de facto* policy lever. Retirement age, which in theory could also be changed, would affect the entire public sector workforce, deeming its use in practice unfeasible.

Hypothetical scenarios in which decision variables are changed were also conjectured. They serve two purposes. First, to analyze how uncertainty surrounding critical parameters may affect the model. Second, to qualitatively validate the model. For instance, holding everything else constant, an increase in the number of training places should lead both to an increase in the number of students and graduates versus the baseline benchmark. Similarly, increasing the minimum legal age for retirement should lead to an expansion of the workforce.

Furthermore, several Monte Carlo experiments were performed to analyze how forecasts may vary with randomness. In particular, how uncertainty in input parameters such as the emigration and the dropout rate may affect the projections. By running the simulations hundreds of times we can estimate the probability distribution of the possible future values of the projections, and elaborate a density forecast to analyze how large can the deviations be within a confidence interval. These density forecasts can also be used to perform a sensitivity analysis on the parameters, providing guidance for the adjustment of the decision variables so that the projection variation stays within a safety band [55].

4.1 Baseline model

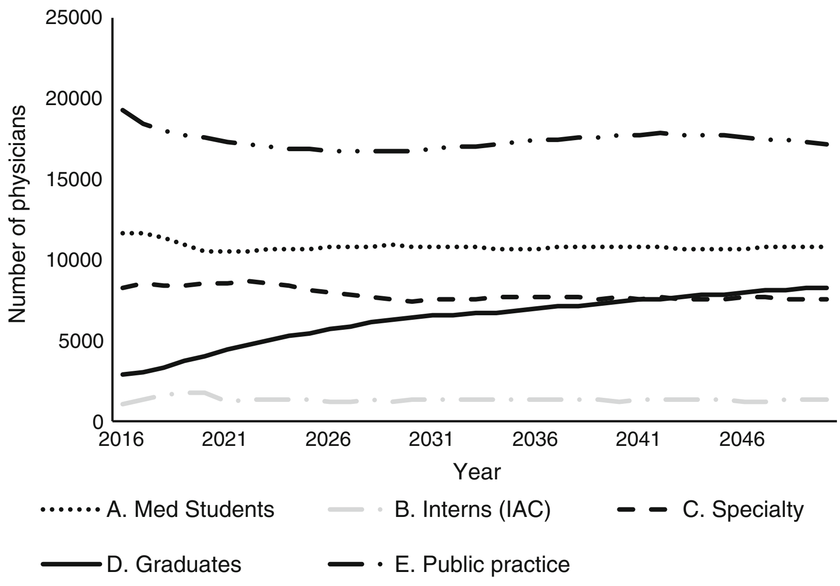
In the baseline model, all the data and parameters listed in Section 2.2 were left unaltered. Additionally, training places in medical schools were fixed at 1496. We assume there are always as many students wishing to join medical school as the maximum number set by the *numerus clausus*. This is supported by evidence, since every year there is a considerable number of students unable to enrol in Medicine, who end up emigrating or waiting for a year until they can apply again. The high grades for accessing medical school also support this assumption. Vacancies to specialty schools were set to 1441, distributed by 52 specialties according to official records; the number of job offers to join the public sector was set to 1934, the value observed in 2014, and contracts were distributed by specialty following the same distribution of training places at specialty school; the physician emigration rate was set to 0.40 % of the total population of physicians (except those in medical school, who typically emigrate before joining university; students who do emigrate during medical school are captured in the dropout rate); and the dropout rate to 2 %. The net migration rate from public to private practice was set to 2 %, and the direct hiring rate by the private sector (after graduation) was set to 3 %. Finally, the number of yearly job contracts for hiring physicians was set to 1441.

With the exception of the emigration rate and of the net migration rate between public and private practice, all the values were obtained from official records. The emigration rate and the number of physicians working in the private sector had to be estimated, since physicians are not obliged to report their status and therefore there are no records available. The Medical Association claims that between 200 and 300 physicians ask for a permit to practice abroad. However, not all physicians actually leave, so it is not possible to know the number with precision. We calibrated the model so that between 100 and 200 physicians emigrate per year when considering the initial stock of physicians (0.4 % emigration rate). Regarding the transfer between public and private sector, the lack of exactness does not affect the overall projections, since these physicians are actually practicing either in the public or in the private sector. It may affect a sectoral analysis, though, in case the aim is to look exclusively at one of the sectors. This is not the case of the present work.

4.2 Evolution of physicians

According to the simulations ran using the baseline scenario, the physician workforce in the public sector will decrease slightly in the first 15 years and recover afterwards, albeit to a level inferior than the one observed initially in 2015, as Fig. 8 (public sector, line E) and Table 7 show. The decrease in the physicians workforce (public sector) that takes place until 2030 can be explained by the age structure of the current medical population, in particular by the predominance of physicians over 50 years of age, which are expected to retire in the next ten to twenty years, and by a constant hiring rate that does not automatically compensate for the loss.

Fig. 8 Evolution of the physician workforce (students, graduates and public sector)



Nonetheless, such analysis would be incomplete for several reasons. First, it does not account for the number of graduates that practice neither in the public nor in the private sector, but are available for hire. The number of graduates yet to be hired provides a measure of the slack available in the health labour market. If no graduates were available there would be no way other than recruiting foreign doctors to solve a short term shortage of health human resources. Unfortunately, there is no reliable information on the initial stock of graduates, but the model is able to capture new students who graduate and are unable to find a job. Second, physicians doing the internship year (IAC) or in specialty residency are already practicing under supervision, although with very limited action.

Therefore, a better outlook is obtained if we incorporate both interns and specialty students in the analysis of the public sector workforce, and also account for graduates. Figure 9 shows the evolution of the workforce in the public sector including medical students already practicing. Incorporating medical students does not seem to alter the long-run trend pointing to a slight decrease in the workforce. When graduates are also considered there is an increase of approximately 9000 physicians by 2050, a better outlook than the one presented before. However, special precaution should be taken when interpreting these numbers, especially when converting the headcounts to FTEs. First, students practicing medicine are not capable of performing all the tasks of a graduated physician. Second, we are not accounting for the

Fig. 9 Evolution of the physician workforce in the public sector, including students already practicing

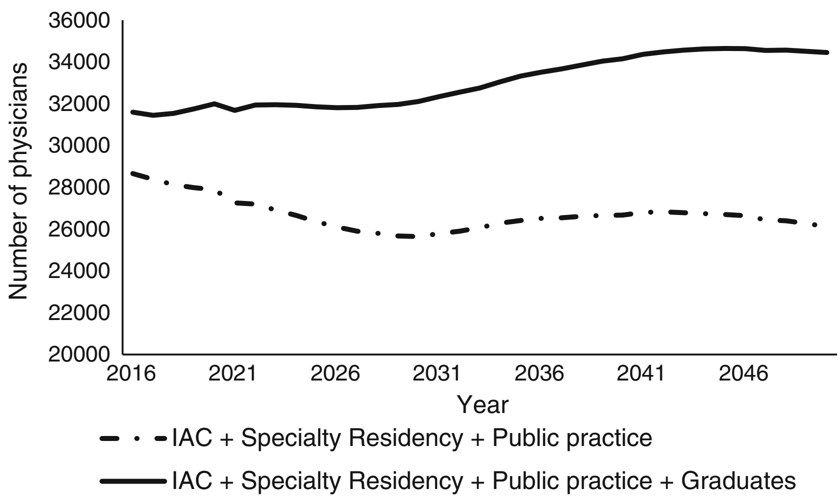
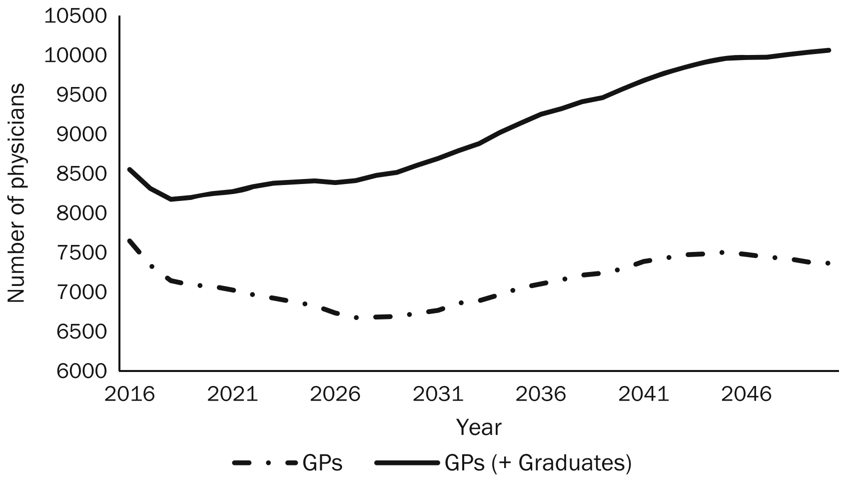


Fig. 10 Evolution of GPs



specialists training the students, hence unavailable to perform clinical practice (records with the ratio of supervisors to specialty students are not publicly available).

There is no information regarding the number of physicians currently working in the private sector, only crude estimates. Nonetheless, we are able to capture physicians absorbed (either moving or being directly employed) by the private sector. Such information is critical to calculate a total number of physicians practicing in Portugal. For this reason, we leave the private sector out of the analysis, with no loss of generality (the size of the private sector has no impact in this model). The private sector is therefore not included in any of the plots or tables. When we look in detail at each specialty we see considerable variations, both negative and positive. The increase in GPs was a deliberate policy enacted in recent years to ensure enough practitioners were available to work at primary care centers and to serve as gatekeepers. The increase in the number of vacancies for GPs in specialty schools indeed increases the number of GPs, as Fig. 10 demonstrates. On the other hand, there is a significant decrease in a lot of other specialties. In 2014, they account for almost 20000 professionals, while by 2050 there will be a 15 % decrease to around 17000 physicians if we ignore graduates, as Fig. 11 shows. Whether these will be enough to provide the quantity of care required will depend on other factors such as the expected demand, the change in epidemiological needs and the evolution of productivity, all outside the scope of the present work.

Moreover, specialties are projected to evolve in very disparate ways. While some specialties such as clinical pathology, internal medicine, oncology or stomatology see a large increase by the end of the forecasting horizon, other specialties will decrease significantly. In particular, neuroradiology will decrease 55 %, from 175 specialtists to 79 by 2050. Likewise, general surgeons, gynecologists and urologists will see a decrease of 25 %, 48 % and 43 %, respectively. In contrast, stomatology increases 50 %, immunochemotherapy 34 %, clinical pathology 53%, public health 61 %

Fig. 11 Evolution of all specialties except general practitioners

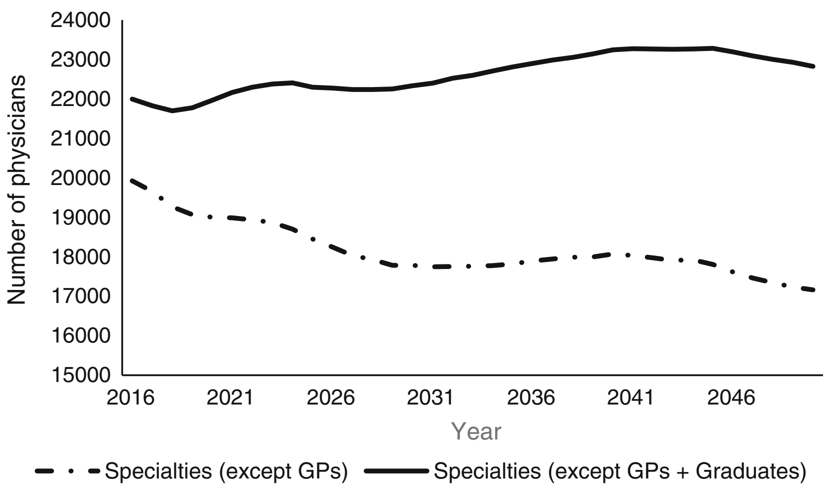


Table 5 Evolution of a selected group of specialties (public workforce + specialty students + graduates)

	2014	2030	2050	Δ
Anesthesiology	1784	1547	1458	-18 %
Cardiology	687	630	625	-9 %
Clinical Pathology	481	573	734	53 %
Endocrinology	221	233	252	14 %
Gastroenterology	427	346	298	-30 %
General Practitioner	8729	8661	10190	17 %
General Surgeon	1549	1240	1161	-25 %
Gynecology	1352	911	701	-48 %
Immunochemotherapy	264	273	353	34 %
Infectology	226	261	310	37 %
Internal Medicine	3167	4158	4874	54 %
Neurology	498	507	555	11 %
Neuroradiology	175	128	79	-55 %
Neurosurgery	184	137	125	-32 %
Ophthalmology	621	591	602	-3 %
Oncology	373	488	492	32 %
Orthopedics	1031	962	937	-9 %
Pediatrics	1874	1753	1594	-15 %
Pneumology	574	680	754	31 %
Psychiatry	901	1063	1267	41 %
Public Health	478	572	771	61 %
Radiology	654	606	578	-12 %
Radiotherapy	112	109	90	-20 %
Rheumatology	162	219	217	34 %
Stomatology	189	214	283	50 %
Urology	341	256	194	-43 %

and oncology 32 %. Table 5 lists initial and final projections for a selected group of specialties, accounting for the public sector workforce, specialty students and graduates, and Table 6 provides the raw estimates for all aggregates, including frictions affecting these aggregates.

Finally, the evolution of frictions that may affect the workforce confirms the expected upsurge of retirees (cf Table 6), a result of the age structure heavily dependent on physicians aged fifty or more. The retirement rate decreases as the aged workforce gets replaced by younger physicians. The dropouts are fairly constant, since the dropout rate is not affected by any internal or external factor, and the number of students is not expected to vary much. Emigration is not expected to change significantly, as there are virtually no physicians unemployed in Portugal, and so the emigration rate is fixed throughout the simulation horizon. Figure 12 provides a detailed account of all the frictions affecting the physician workforce.

4.3 Demand analysis

According to the OECD, in 2013 there were approximately 4.3 doctors per 1 000 population in Portugal [41]. However, as OECD also notes, these numbers may be misleading. Data refer to all doctors licensed to practice and registered in the Medical Association's database, which results in an overestimation of about 30 %. This means that the number of physicians reported is well above the number of physicians actually practicing. For this reason, OECD no longer reports numbers for Portugal in their "Practicing physicians (doctors)" database.

Since there are no data available on the number of physicians practicing in the private sector, the current ratio of physicians to population is fallacious. Therefore, we will use the number of physicians in the public workforce only (including graduates and those in residency training) to project the ratio against the three demographic scenarios. Figure 13 shows the evolution of the ratio.

In all scenarios, the ratio increases, and in the case of the low scenario, where the population decreases to 6.5 million by 2060, it increases significantly from approximately 3 physicians per 1 000 population to 4.2. However, looking only at the population size without accounting for the age structure would be insufficient, a point explained further in Section 3.2. Therefore, we also compute Eq. 3, which is a better proxy to estimate future demand. Figure 14 shows the evolution in the ratio.

As a visual comparison between Figs. 13 and 14 attests, the projections change dramatically when we extend the analysis with the age structure of the population. In all three demographic scenarios, the ratio of physicians per Q decreases, meaning that the shift in the age structure of the population plays a dramatic role. In comparison to the base year (2015), the ratio (Eq. 3) decreases in both scenarios, bar the low one, in which case it ends with the same ratio by the end of the projection period. In the two other scenarios, and assuming everything else remains constant (i.e. productivity, skill mix, morbidity, etc.), the number of physicians would be inadequate.

4.4 Scenario analysis

Most of the factors affecting the evolution of the medical forecast are out of reach of policy makers, but some can still be influenced. To analyze the influence of changes to policy variables and also to validate the model we have conceived three different scenarios with likely outcomes. In the first scenario, we consider a change to the minimum retirement age, delaying retirement until 66 and 67 years of age. The effect is predictable in comparison to the baseline scenario: less retirements should occur and the workforce will be marginally bigger. Moreover, the effect of 2 extra years

Table 6 Forecasts of the number of physicians (aggregated by GPs and others), both students, graduates and workforce, as well as frictions affecting the workforce

	2014	2016	2020	2025	2030	2035	2040	2045	2050
Stocks									
Students	12200	11721	10496	10674	10874	10736	10806	10676	10781
Internship + Residency (IAC)	1418	1111	1805	1336	1374	1382	1286	1361	1326
Specialty school									
GPs	1889	1986	2138	2012	1856	1977	1953	1977	1958
Other specialties	6400	6296	6406	6187	5655	5743	5731	5679	5690
Workforce									
GPs	6053	7650	7056	6794	6700	7097	7355	7520	7446
Other specialties	13718	11614	10475	10028	10058	10209	10350	10168	9691
Graduates									
GPs	787	903	1207	1638	1961	2097	2358	2577	2744
Other specialties	1756	2029	2904	2796	4501	4817	5119	5362	5595
Flows									
Emigration	-	121	161	164	149	163	174	173	185
Foreign hires	152	36	35	41	36	33	39	37	40
Dropouts	-	465	419	406	416	391	427	341	407
Retirements	-	831	498	488	352	281	361	493	517

should be larger than the effect of 1 extra year. In the second scenario we consider a 10 % increase and a 10 % decrease of training places, both in pre-grad medical school and in specialty school. An increase should lead to a surge both in the number of graduates and in the workforce. Conversely, the opposite should occur when the *numerus clausus* are reduced. Results show that, as expected, raising the minimum retirement age produces an overall positive effect in the supply of physicians (includes specialty students, graduates and the workforce working in the public sector). In particular, 1 extra year is responsible for a -0.58 % decrease in the number of GPs and a 1.14 % increase in all other specialties in comparison to the baseline scenario[1]. When 2 extra years before retirement are considered, GPs increase 1.82 % by 2030, while all other specialties see an overall increase of 3.40 % by 2050 (Table 7). These numbers are in line with the literature, which examined the effects of late retirement and concluded that it has a very limited impact in medium to long term projections [56]. Albeit limited, note that it has an immediate impact in the workforce exits.

At first, a change in the number of training places seems to produce a greater impact on the supply of physicians than postponing retirement. When the number of vacancies is increased by 10 % the number of GPs rises 8.58 % and all other specialists 7.48 %. On the other hand, a decrease is responsible for a decline of approximately -9.11 % of GPs and -7.88 % of all other specialists. However, if the retirement age were increased or decreased by 10 % similar results would be obtained, and so in practice they exhibit similar elasticity. Not surprisingly, an increase in the number of training places takes longer to produce an effect, since it requires up to 10 years for students to graduate and integrate the workforce, as Fig. 15 makes clear.

4.5 Uncertainty analysis and Monte Carlo simulations

In order to estimate the potential impact of uncertainty in key model inputs we conducted a sensitivity analysis. Being a stochastic model with parameters sampled from probability distributions, simulations need to be run several times to analyze a large set of outcomes. We have run 1000 simulations for each scenario. Monte Carlo simulations were run for a baseline scenario and for stochastic dropout and emigration rates. Since these rates are determined empirically and subject to much uncertainty, obtaining a confidence interval within which projections may vary is critical to ensure the model is robust. If a small change in one of these inputs causes a major deviation in the forecasts such effect must be duly accounted for by analysts and policy makers. To help visualize the results of the Monte Carlo simulations we built density forecasts, which are estimates of the probability distribution of the possible future values of that variable within a confidence interval [55]. The probability

[1]The -0.58 % decrease is not statistically significant as it is within the margin error for the baseline scenario (see Fig. 16a). Running the model again would most likely generate a positive increase. We left this *anomaly* as a reminder of how stochasticity may affect the model, and why Monte Carlo simulations are critical.

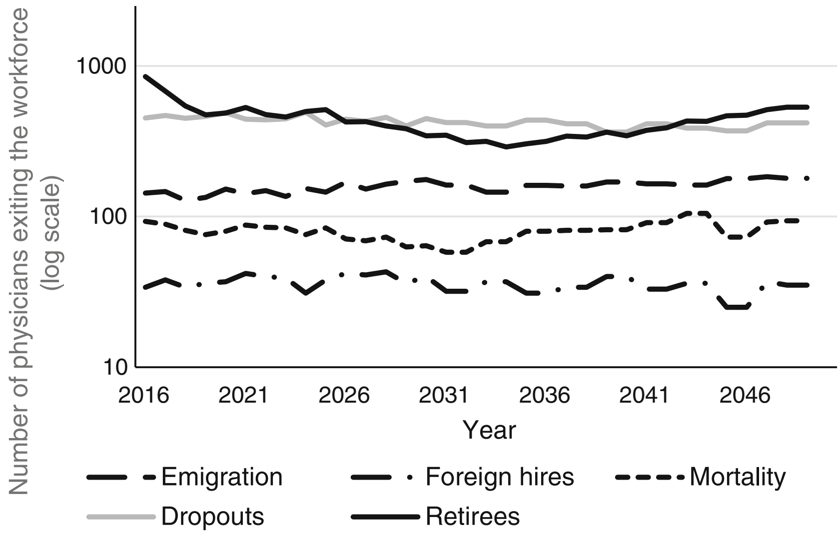


Fig. 12 Physicians exiting the workforce each year due to emigration, mortality or retirement, and foreign doctors (immigration) joining the workforce

distribution was extrapolated out of the histogram generated by running the model 1000 times. Figure 16a shows the density forecast of the workforce estimates for the baseline scenario and for each of the inputs that were subjected to variation. Figure 16a suggests that the uncertainty stemming from the model, such as the transition times in the state charts, does not introduce much variability. Despite varying, it does so within relatively narrow bands. The maximum mean absolute deviation is 1.42 %. Also, the maximum variation between the lower and upper estimate is 621 physicians, approximately 2.26 % of the median estimate. Furthermore, the green area around the median identifies the confidence level. The closer to the median the smaller the level of confidence. For instance, if the baseline scenario is correct then the number of physicians observed at the end of the forecasting horizon may vary between 28029 and 28585 for a 95 % C.I. When the emigration rate is allowed to vary following a normal distribution with mean 0.004 and standard deviation 0.002, the band between forecasts widens considerably. This implies a deviation in the emigration rate may have a large impact on the forecasts (Fig. 16b). The maximum mean absolute deviation is 6.43 %, but the absolute difference between extreme values may go as high as 10.84 %, or 3104 physicians. If a deviation of 600 is not easy to accommodate, fivefold the difference may pose a serious problem to a functioning health care system. The problem is aggravated further with the variability introduced by the dropout rate, another key input. If it follows a

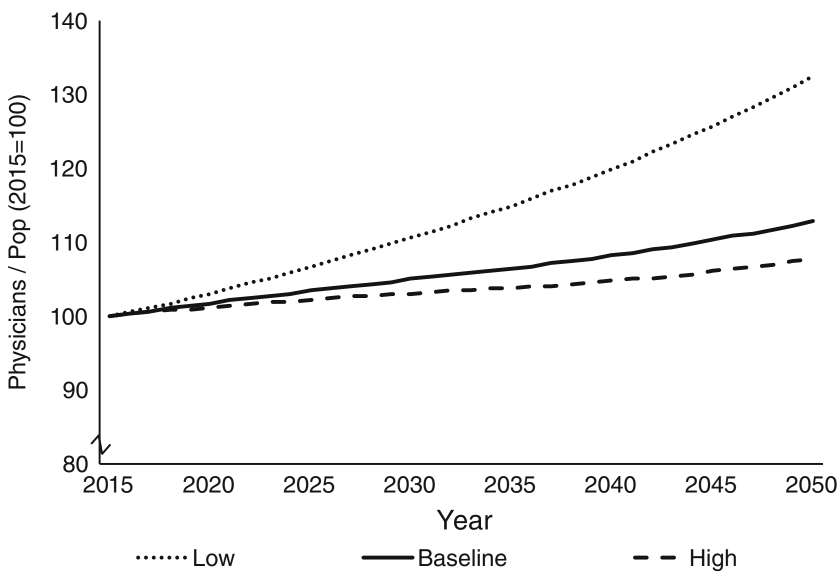
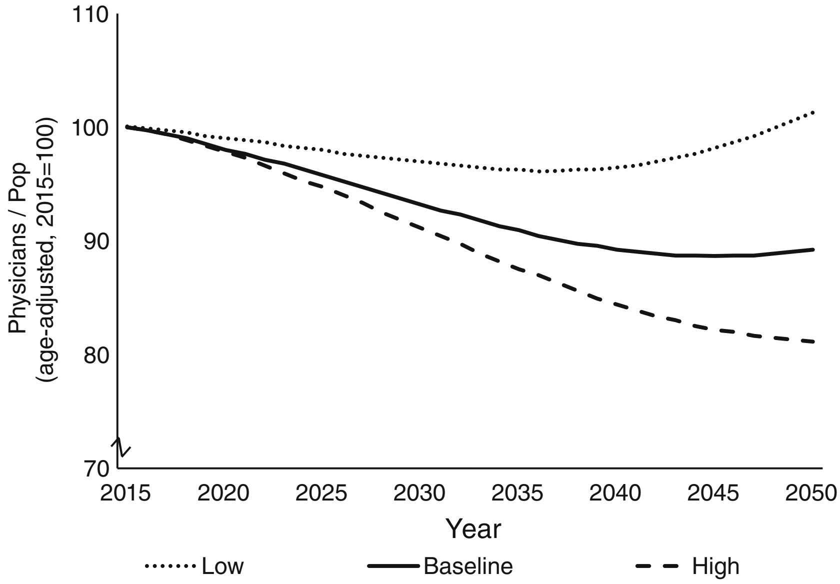


Fig. 13 Evolution of the ratio of physicians (all physicians except medical students attending university and private practice) to 1 000 population

Fig. 14 Evolution of the ratio of physicians (all physicians except medical students attending university and private practice) to Q



normal distribution with mean 0.02 (2 % dropout rate) and sigma 0.01 the forecasts may deviate up to 17.15 % from the absolute mean. In such case, there is a difference of 7729 physicians from the lower and upper bounds for a 95 % C.I (Fig. 16c). Although the coefficient of variation is high ($c_v = \sigma/\mu = 0.02/0.01 = 2$), it is not that abnormal to observe a change in the dropout rates of a magnitude of 1 percentual point.

The uncertainty analysis helps to understand how some variability in input parameters may alter the forecasts, but also to validate the model. The model reacts as expected when either the dropout or the emigration rates vary.

5 Discussion

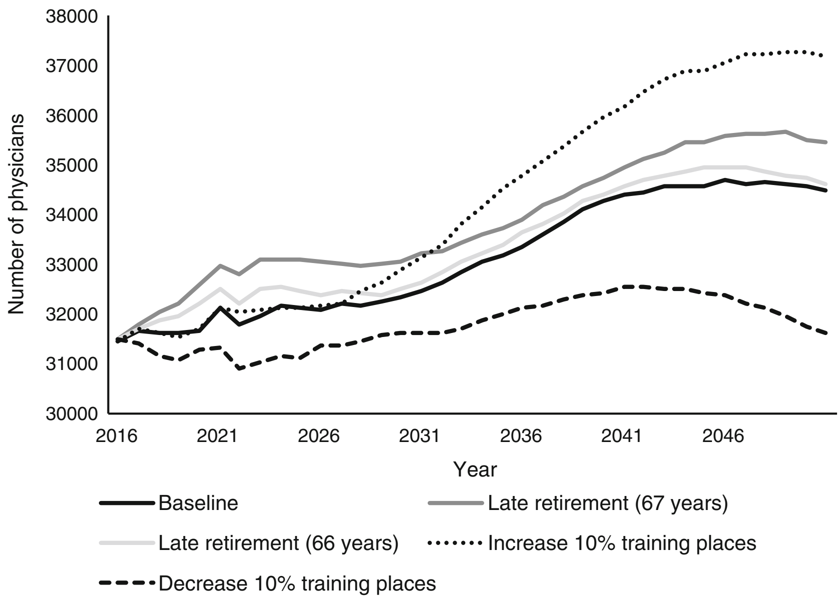
The purpose of this study was to present an agent-based simulation approach to forecast the physicians workforce. This supply model is part of a larger initiative aiming to conceive an integrated model, providing a comprehensive view of the health care market and the drivers influencing it. By simulating at the individual level we are able to collect detailed statistics of all actions that take place within the model. Moreover, it becomes possible to incorporate additional microeconomic foundations into the physicians to analyze complex behavior, such as their propensity to emigrate based on wage, and with such information contemplate different policy scenarios. For instance, consider a scenario where an increase in wages could be used to dissuade physicians from emigrating. Proceeding in a similar way, the model can be used to analyze vertical substitution between physicians and nurses, the impact of enhanced productivity or internal migrations from and to rural areas. In fact, the model may be particularly useful for studying the regional asymmetries that exist in Portugal.

Providing bullet-proof projections is not the aspiration of this work, but rather to exemplify how the model could be used with that aim. For that purpose, we have demonstrated its application to the Portuguese medical workforce. The model was conceived to replicate as best as possible the current working conditions, economic constraints and frictions affecting physicians in Portugal. Preliminary results

Table 7 Impact of the hypothetical policy interventions on the physician workforce

Scenarios		2030 GPs	Δ	Other specialties	Δ	2050 GPs	Δ	Other specialties	Δ
Baseline		10667	-	20470	-	12161	-	20988	-
Training places	Increase by 10 %	11045	3.54 %	20609	0.68 %	13205	8.58 %	22557	7.48 %
	Decrease by 10 %	10510	-1.47 %	19911	-2.73 %	11053	-9.11 %	19334	-7.88 %
Late retirement	66 years	10634	-0.31 %	20661	0.93 %	12090	-0.58 %	21227	1.14 %
	67 years	10838	1.60 %	21004	2.61 %	12382	1.82 %	21701	3.40 %

Fig. 15 Comparison of the evolution in the supply of physicians (interns + specialty students + public practice + graduates) under the difference scenario assumptions



suggest that in the absence of any policy intervention (the baseline model) there will be a slight decline in the number of physicians doing public practice. If we also account for the number of graduates there will be a slight increase of approximately 2000 physicians by 2050.

Results also show that there will be a significant increase in the number of GP graduates. At the current hiring rate, the additional 1400 GPs will not be immediately absorbed, which poses a risk. If they are unable to find work for too long, either in the private or in the public sector, they may emigrate. Current policies suggest an expansion of primary care centres and family physicians, a role typically performed by GPs. If such policies are to be pursued, the excess of 1500 GPs may easily be absorbed by either the private or the public sector. An accurate analysis requires an integrated model that includes all the drivers of supply and demand previously mentioned, and also accounts for any current imbalances in the workforce [5]. Nevertheless, this serves to reinforce the fact that an excess or a shortage per se may not be indicative of an imbalance in the health-care workforce, a conjecture already put forward when we incorporated demand forecasts in the analysis.

When we look in detail at the figures for each specialty (cf. Table 5), we can see some considerable changes. For instance, changes of up to 61 % in the number of Public Health specialists or decreases of -55 % in Neuradiologists.

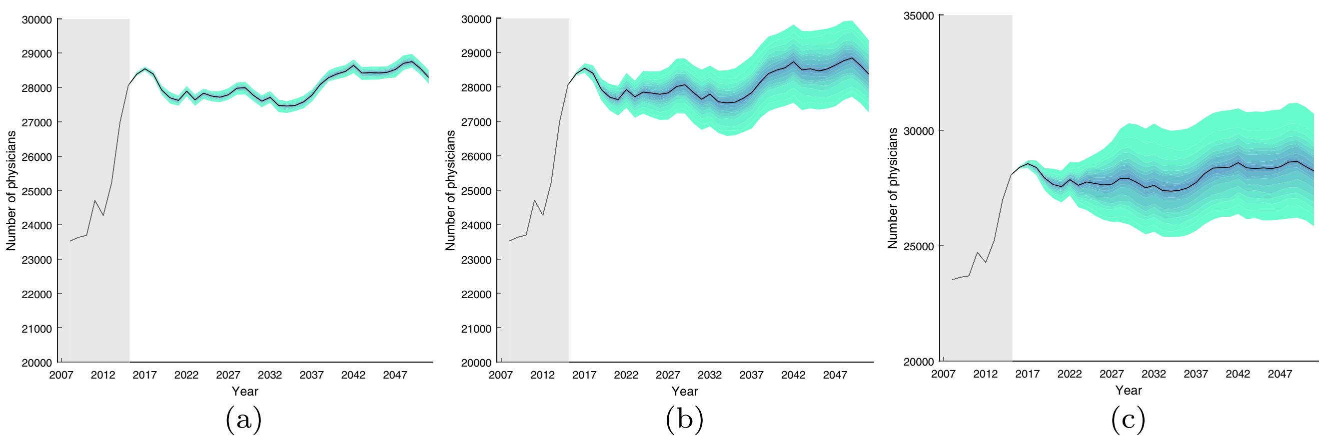


Fig. 16 Density forecast of the workforce estimates in the baseline scenario (panel a), when the emigration rate is given by a normal distribution with mean 0.004 and sigma 0.002 (panel b), and when the dropout rate varies according to a normal distribution with mean 0.02 and sigma 0.01 (panel c). Area in grey is historical data, the black line is the median and the gradient is the variation within confidence intervals limits

These changes are, in fact, expected. Recently, the Government increased the number of intakes to Public Health. That means that the future flow of new specialists to Public Health is actually expected to increase considerably. Observing the data, we count 315 Public Health physicians currently in practice, plus 100 in residency training that will be joining the workforce any time soon. In 2003 there were 36 new residency spots just for Public Health. In the time horizon of 35 years that figure comes to over 1260 fresh new Public Health physicians (ignoring the outflows from the stock). In the case of neuroradiology, the number of vacancies to residency training has been reduced to 3 for the whole country (increased to 5 in 2014 and expected to increase further in 2015). Since there are only 175 neuroradiologists currently practicing, of which 46 % are 50 years of age or older and 44 are still in residency training, it does not come as a surprise that the number falls to 79 by the end of the time horizon. Additionally, Monte Carlo simulations suggest that the effect of uncertainty on some key inputs is not negligible, and may in fact alter the trend of the projections, bringing an increase to a substantial drop (Fig. 16c). While a surplus may lead to unemployment and a bad allocation of resources, a shortage may put patients lives at risk [4]. Policy makers can account for this trade-off and decide accordingly, in particular whether to risk erring on the side of caution. Despite not providing a full description of the problem, as discussed previously, these results may still be useful for policymakers. For instance, we have seen that the increases in GPs and Public Health specialists or, contrariwise, the reduction in neuroradiologists were *expected*, in the sense that policies had been enacted to achieve such outcome. Policymakers can observe the projected numbers and realize whether the evolution of the remaining specialties is in accordance with what is expected. If not, they can experiment with the number of intakes to residency programs to adjust to a desirable level. Moreover, if new information arrives, pointing to an effective decline in the population, policymakers can simulate the impact of reducing the *numerus clausus* in the size of the workforce.

When we extend the analysis to incorporate demand profiles, in particular the impact of demographic trends, it is interesting to observe that the ratio of physicians to a proxy of total quantity of care actually decreases, despite a declining population. Population ageing, and thus a shift in the age structure, seems to explain this phenomenon, as elders and end-of-life care tend to put an additional weight on the health system. In fact, only in the pessimistic scenario, where the population size decreases by 40 %, does the ratio remain relatively stable through the forecast period. In the optimistic case, where population decreases *only* by one million persons, the ratio decreases from 0.87 to 0.73 physicians per total quantity of care by 2050. Note, however, that this analysis assumes the current ratio is adequate, which may be a frail assumption. Assessing whether current imbalances do exist would strengthen the analysis [5].

A comprehensive policy analysis wishing to inform decision-makers on the number of vacancies to open for each medical specialty should extend the analysis of demand, whenever data is available, incorporating the remaining drivers identified in Fig. 1, in particular needs. To do so, future research can start by estimating the production function for each medical specialty, i.e. the quantity of health care (surgeries, outpatient and inpatient contacts, emergencies, etc.) that can be delivered given the availability of physicians, nurses and capital. Next, it should obtain a precise account of the quantity of health care services sought by age group, identified by pathology. It can then combine this information with demographic changes, extrapolating which pathologies may be more frequent in each age group, and its impact in terms of physicians' requirements.

Other factors in addition to demand may also play a significant role. Productivity growth, for instance, may help in delivering the same care with fewer resources. Skill mix, in particular horizontal substitutability with nurses, may also leverage the work of the physician. Again, this interaction takes place at the micro-level, between agents, in a non-trivial way. ABM is therefore a very adequate method to conduct such studies.

5.1 Agent-based modelling vs System Dynamics

This paper served also to show that ABM is an adequate tool for healthcare workforce planning, both for researchers and policy-makers. It is an effective tool for generating forecasts of the evolution of health professionals, but also for studying other relevant economic behavior that may affect the workforce, presenting several benefits over SD simulation. Although the physician agent as it currently stands does not incorporate any microeconomic foundations to illustrate some of the advantages of using ABM, one may conceptualize several scenarios for which the ability to model individual behavior would be critical.

One illustrative example of a research question for which ABM is clearly more suitable is the case of handling emigration in healthcare workforce planning. Assume the physician emigrates whenever the utility of a proposal coming from abroad is better than the utility of working in the home country plus a fixed cost for moving abroad (dependent on age). There will be thousands of utility values for different physicians, depending on their age, wage level, etc. Implementing such feature using System Dynamics would be extremely complex, and aggregating the results into hundreds of stocks would be senseless, assuming it is doable.

For problems largely stemming from individual behavior, which is the case, ABM is clearly a better alternative. In fact, other pressing topics, such as regional asymmetries within a country, can only be studied by employing a method capable of handling spatial information, something ABM does. Extending this model to study regional asymmetries would require little effort. Doing so with SD would be extremely laborious.

The distinctive feature of ABM over SD is not generating better results. In fact, for problems where adding individual features to the agent, such as individual memory (history), agent communication, spatial awareness or heterogeneous utility functions is unnecessary, ABM would not present any advantages [16]. Some of the individual properties can be added to System Dynamics with the help of dimensional arrays that try to capture each property. For instance, there may be a dimensional array for the age inside each stock, another for the specialty, etc. However, as soon as the models grow in complexity, so do the array dimensions, making the model extremely slow [16]. ABM can deal with any number of properties without any significant hit in performance.

6 Conclusion

In this paper we have illustrated how agent-based modelling may be used to forecast the supply of physicians, and highlighted some of the advantages over other competing methods, such as System Dynamics. We specified and implemented an ABM that replicates the physicians' lifecycle, calibrated the model using available data and generated forecasts for the Portuguese workforce until 2050. The forecasts do not indicate a significant change to the size of the medical workforce, although substantial variations do take place for some particular specialities. When we factor in demand, in particular changes to the population size, a first analysis could lead to the (frail) conclusion that the number of physicians would be in excess of a declining population. However, when we also account for the shift in the age structure of the population, results suggest that the number of physicians may actually be insufficient to deliver enough care to an ageing population. Portugal could well have to deal with a shortage of physicians in the nearby future.

Faced with this, policy-makers may use several levers to reduce the imbalance. In this study we analysed the impact of two, a change to the *numerus clausus* and raising the retirement age. Others could have been considered. For instance, changes to the wage bill or other retention policies to reduce emigration. We do not address these policy options here, but the model developed may be a good basis upon which to study such options, and provide foresights to inform policy makers. ABM makes it easy to incorporate decision-making capabilities and reactive behavior into the agents, critical to assess the impact of enacting such policies. ABM should definitely become part of the researcher's toolkit.

Appendix

A Mathematical description of the model

The ABM proposed can be loosely approximated by a set of differential equations. Keep in mind, however, that the differential equations that define the stock and flow equations emerge out of the agents' actions, and are thus not hardcoded in the model. In fact, they provide an aggregate overview of the number of physicians at each state, and not a precise account for each individual physician. Under this formulation it is not possible, for instance, to know the age of an individual physician or if that particular physician is retired, only the aggregate number of physicians retired by year and speciality. It is also not possible, or extremely laborious, to incorporate microeconomic foundations. Notwithstanding, this alternative formulation may be of interest to those well versed in math or with experience in SD simulation, for whom it will be easier to follow through.

Let t be the time period ($t \in T = \{0, \ldots, T\}$), s the state the physician is in ($s \in \mathcal{S} = \{1, 2, 3, \ldots\}$) from the list of states described in Fig. 2 and i the physician specialty ($i \in \mathcal{I} = \{0, 1, 2, \ldots, 52\}$). The stock S of physicians by specialty i at time t and state s can be computed by integrating the following equation:

$$S_{is}(t) = \int_0^T [\text{Inflows}_{is}(t) - \text{Outflows}_{is}(t)]\,\text{dt} + S_{is}(0) \quad (4)$$

The *Inflows* function represents all positive flows of physicians with specialty i accruing at each period t in state s, while *Outflows* represents negative flows to the stock i at period t and state s. $S_{is}(0)$ represents the initial stock of physicians in state s with specialty i. Which inflows and outflows affect the stock S will depend on the state s the physician is in. For instance, dropping out of school only applies to agents still in training (see Fig. 2). The stock equations for each state s will then be an adaptation of Eq. 4.

For instance, let us consider the case of students undergoing training. These students are not specialized yet, and so we assume $i = 0$. The growth rate in the stock of students in medical school is given by:

$$\frac{S_1(t)}{\text{dt}} = M(t) + (x - d - m_a) \cdot S_1(t) - \delta \cdot S_1(t), \quad (5)$$

where $M(t)$ is the fixed number of fresh students joining medical school, x is the net migration rate (i.e. immigrations - emigrations), d is the dropout rate, m is the mortality rate at age a and $\delta \cdot S_1(t)$ is the share of already enrolled students moving to the next state (S_2). The mortality rate is obtained from a life table, more specifically q_a, the probability that someone aged exactly a will die before reaching $(a+1)$.

The stock equations for the remaining states are similar, except for one detail. Since "Med School" is the initial state, there are no agents transitioning from a previous state, only an initial stock and a constant positive flow M of fresh students arriving at each period t. However, subsequent states will have previous states. For instance, medical students transition from state "Med School" to "IAC". Hence, we should also include a term representing all the positive flows arising from the previous state as follows:

$$\frac{S_s(t)}{\mathrm{dt}} = (x - d - m_a) \cdot S_s(t) - \delta \cdot S_s(t) + \omega \cdot S_{s-1}(t), \quad (6)$$

where $\omega \cdot S_{s-1}(t)$ is the share of physicians moving from the previous state to current state s and $\delta \cdot S_s(t)$ represents the share of physicians moving onwards to the next state.

To make it clear, let us instantiate a particular case. Consider state S_2, corresponding to "IAC", the practice year between medical school S_1 and residency training S_3. We assume students are uniformly distributed throughout college years, and so the probability of completing medical school S_1 and moving to next state S_2 can be approximated by the probability of being in the last year of studies. However, the time required to complete studies is not deterministic. Although medical school takes no less than 6 years to complete, some students take longer. We model the probability of shifting from state by first considering that the number of years λ necessary to fulfil the requirements follows a probabilistic distribution. In this case, we assume the exponential distribution, as it exhibits the best fit to the data:

$$\lambda \sim \text{Exponential}\,. \quad (7)$$

Then, assuming all years are equally laborious, the probability of transitioning to the next state is equal to the probability of being in the final year, which is equal to:

$$\omega = p_{1,2} = \frac{1}{\lambda}, \quad (8)$$

and the share of physicians moving from the previous state S_1 to the current state S_2 is then approximated by:

$$\omega S_1(t) = \frac{S_1(t)}{\lambda}. \quad (9)$$

We also need to compute the probability of moving to state S_3 to obtain the complete equation. Transition to S_3 implies obtaining a vacancy at a residency program. There are several types of residencies corresponding to each medical specialty, and the probability of moving to *any* residency program can be approximated by the ratio between the number of residency vacancies available for year t, $n(t)$, and the number of students wishing to apply in that year, $\omega \cdot S_2(t)$:

$$p_{2,3} = \begin{cases} n(t)/\omega S_2(t) & \text{if } n(t) < \omega S_2(t) \\ 1 & \text{if } n(t) \geq \omega S_2(t) \end{cases} \quad (10)$$

Consequently, the number of physicians moving to residency, $\delta \cdot S_3$, can be approximated by $p_{2,3} \cdot S_3(t)$.

When the physician is already employed the equations change slightly. To calculate the number of physicians with specialty i in the public sector ($s = 5$) at time t we need to integrate the following equation:

$$\frac{S_{i5}(t)}{\mathrm{dt}} = (x - y - m_a - r) \cdot S_{5i}(t) + \omega \cdot S_{4i}(t), \quad (11)$$

where y is the net migration rate from the public to the private sector and r is the rate of physicians retiring at year t. This is required since, at any time, the physician may move back and forth between the public and the private sector. When calibrating the model we assume that the net transfer rate is negative with regards to the public sector. After working for some years the physician retires. The retirement age is drawn from a normal distribution:

$$R \sim \mathcal{N}(\mu, \sigma^2), \quad (12)$$

and calibrated according to the distributions specified in Table 4.

The remaining differential equations are trivial to build following the same procedure.

Acknowledgments The authors acknowledge Grant No SFRH/BD/102853/2014 and Grant No PTDC/IIMGES/4770/2014 from FCT - Fundação para a Ciência e Tecnologia that funded part of the research. Also, the authors would like to thank the anonymous reviewers, who made extremely valuable suggestions, helping to improve the paper considerably. Remaining errors are our sole responsibility.

References

1. (2011) ACSS: Actuais e Futuras Necessidades Previsionais de médicos (SNS), Tech. rep., Administração Central do Sistema de Saúde
2. (2015) ACSS: Inventário de Pessoal do Setor da Saúde–2014. Tech rep
3. Amaral J, Lemos I (2009) Aposentações no ministério da saúde 2007, Tech rep, Lisboa
4. Amorim Lopes M, Almeida ÁS, Almada-Lobo B (2015) Handling healthcare workforce planning with care: where do we stand? Hum Resour Health 13(1):38
5. Amorim Lopes M, Santos Almeida Ád, Almada-lobo B (2015) Assessing the current situation. In: Handbook on health workforce planning, pp. 296–306. EU joint action health workforce planning and forecasting
6. Amorim Lopes M, Soares C, Almeida Á, Almada-lobo B (2016) Comparing comparables: an approach to accurate cross-country comparisons of health systems for effective healthcare planning and policy guidance Health Systems

7. Ansah JP, De Korne D, Bayer S, Pan C, Jayabaskar T, Matchar DB, Lew N, Phua A, Koh V, Lamoureux E, Quek D (2015) Future requirements for and supply of ophthalmologists for an aging population in Singapore. Hum Resour Health 13 (1):86
8. Auchincloss AH, Roux AVD (2008) A new tool for epidemiology: The usefulness of Dynamic-Agent models in understanding place effects on health. Am J Epidemiol 168(1):1–8
9. Barber P, López-valcárcel BG (2010) Forecasting the need for medical specialists in Spain: application of a system dynamics model. Hum Resour Health 8(1):24
10. Barnes S, Golden B, Price S (2013) Applications of Agent-Based modeling and simulation to healthcare operations management. In: Handbook of healthcare operations management. Springer New York, New York, NY, pp 45-74
11. Barnighausen T, Bloom DE (2009) Changing Research Perspectives on the Global Health Workforce. NBER Working Papers:1–84
12. Barros PP, Machado SR, Simoes JDA (2011) Portugal. Health Syst Rev 13(4):1–156
13. Berger T, Troost C (2014) Agent-based Modelling of Climate Adaptation and Mitigation Options in Agriculture. J Agric Econ 65(2):323–348
14. Berhanu Alemayehu KEW (2004) The Lifetime Distribution of Health Care Costs. Health Serv Res 39(3):627–642
15. Bilge U, Saka O (2006) Agent based simulations in healthcare. Stud Health Technol Inform 124:699–704
16. Borshchev A, Filippov A (2004) From system dynamics and discrete event to practical agent based modeling: Reasons, Techniques, Tools
17. Brailsford S, Schmidt B (2003) Towards incorporating human behaviour in models of health care systems: An approach using discrete event simulation. Eur J Oper Res 150(1):19–31
18. Brailsford SC, Harper PR, Patel B, Pitt M (2009) An analysis of the academic literature on simulation and modelling in health care. J Simul 3(3):130–140
19. Dawid H, Neugart M (2011) Agent-based Models for Economic Policy Design. East Econ J 37(1):44–50
20. Day TE, Ravi N, Xian H, Brugh A (2013) an Agent-Based modeling template for a cohort of veterans with diabetic retinopathy. PLoS One 8(6):e66,812
21. Dreesch N (2005) An approach to estimating human resource requirements to achieve the Millennium Development Goals. Health Policy Plan 20(5):267–276
22. Farmer JD, Foley D (2009) The economy needs agent-based modelling. Nature 460(7256):685–686
23. (2016) Gabinete de Estudos Estrategicos e Melhoria Continua: Diplomados pela U.Porto de ciclos de estudos (cursos conferentes de grau). Tech rep, Reitoria da Universidade do Porto
24. Gorman DM, Mezic J, Mezic I, Gruenewald PJ (2006) Agent-Based modeling of drinking behavior: a preliminary model and potential applications to theory and practice. Am J Public Health 96(11):2055–2060
25. Heckbert S (2011) Agent-based modelling of emissions trading for coastal landscapes in transition. Journal of Land Use Science 6(2-3):137–150
26. Hedden L, Barer ML, Cardiff K, McGrail KM, Law MR, Bourgeault IL (2014) The implications of the feminization of the primary care physician workforce on service supply: a systematic review. Hum Resour Health 12(1):32
27. (2014) Instituto Nacional de Estatística: Projecções de Populacão Residente 2012-2060. Tech rep
28. Kleijnen JPC (1995) Verification and validation of simulation models. Eur J Oper Res 82(1):145–162
29. Lagarde M, Cairns J (2011) Modelling human resources policies with Markov models: an illustration with the South African nursing labour market. Health Care Manag Sci 15(3):270–282
30. Lavieri MS, Puterman ML (2009) Optimizing nursing human resource planning in British Columbia. Health Care Manag Sci 12(2):119–128
31. Lavieri MS, Regan S, Puterman ML, Ratner PA (2008) Using Operations Research to Plan the British Columbia Registered Nurses' Workforce. Health Policy 4(2):e117–35
32. Liu P, Wu S (2014) An agent-based simulation model to study accountable care organizations. Health Care Manag Sci:1–13
33. Lodi A, Tubertini P, Grilli R, Mazzocchetti A, Ruozi C, Senese F (2015) Needs forecast and fund allocation of medical specialty positions in Emilia-Romagna (Italy) by system dynamics and integer programming
34. Lomas J, Stoddart GL, Barer ML (1985) Supply projections as planning: A critical review of forecasting net physician requirements in Canada. Soc Sci Med 20(4):411–424
35. Machado CS, Portela M (2011) Hours of work and retirement behavior NIPE Working Papers
36. Maier T, Afentakis A (2013) Forecasting supply and demand in nursing professions: impacts of occupational flexibility and employment structure in Germany. Hum Resour Health 11(1):1–1
37. de la Maisonneuve C, Martins JO (2013) Public Spending on Health and Long-term Care 6
38. Marshall BDL, Galea S (2015) Formalizing the role of agent-based modeling in causal inference and epidemiology. Am J Epidemiol 181(2):92–99
39. Masnick K, Mcdonnell G (2010) A model linking clinical workforce skill mix planning to health and health care dynamics. Hum Resour Health 8(1):11
40. Meng Y, Davies R, Hardy K, Hawkey P (2010) An application of agent-based simulation to the management of hospital-acquired infection. J Simul 4(1):60–67
41. OECD (2015) 2015 OECD Indicators. Health at a glance, Health at a Glance. OECD Publishing, Paris
42. Pathman DE (1991) Estimating Rural Health Professional Requirements: An Assessment of Current Methodologies*. J Rural Health 7(4):327–346
43. Patlolla P, Gunupudi V, Mikler AR, Jacob RT (2004) Agent-Based simulation tools in computational epidemiology. In: Innovative internet community systems. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 212-223
44. Paul P, Maglio PLM (2011) Agent-Based models and systems science approaches to public health. Am J Prev Med 40(3):392–394
45. Pombo-Romero J, Varela LM, Ricoy CJ (2012) Diffusion of innovations in social interaction systems. An agent-based model for the introduction of new drugs in markets. Eur J Health Econ 14(3):443–455
46. Rahmandad H, Sterman J (2008) heterogeneity and network structure in the dynamics of diffusion: Comparing Agent-Based and differential equation models. Manag Sci 54(5):998–1014
47. Ribeiro JS, Conceição C, Pereira J, Leone C, Mendonça P, Temido M, Vieira CP, Dussault G (2014) Health professionals moving to... and from Portugal. Health Policy 114(2-3):97–108
48. Sargent RG (2012) Verification and validation of simulation models. J Simul 7(1):12–24
49. Scheffler R (2008) Forecasting the global shortage of physicians: an economic- and needs-based approach. Bull World Health Organ 86(7):516–523
50. Spencer S, Jones RSE (2008) An Agent Based Simulation Tool for Scheduling Emergency Department Physicians. AMIA Ann Symp Proc 2008:338

51. Spiro Maroulis EBLG, Wilensky U (2014) Modeling the Transition to Public School Choice. 2013:39:2 58(2):115
52. Steinbrook R (2014) private health care in Canada. N Engl J Med 354(16):1661–1664
53. Stordeur S, Leonard C (2010) Challenges in physician supply planning: the case of Belgium. Hum Resour Health 8(1):28
54. Taboada M, Cabrera E, Epelde F, Iglesias ML, Luque E (2013) Using an Agent-based Simulation for Predicting the Effects of Patients Derivation Policies in Emergency Departments. Procedia Computer Science 18:641–650
55. Tay A. S., Wallis KF (2000) Density forecasting: a survey. J Forecast 19(4):235–254
56. Teljeur C, Thomas S, O'Kelly FD, O'dowd T (2010) General practitioner workforce planning: assessment of four policy directions. BMC Health Serv Res 10(1):148
57. Van Greuningen M, Batenburg RS, Van der Velden LF (2012) Ten years of health workforce planning in the Netherlands: a tentative evaluation of GP planning as an example. Hum Resour Health 10(1):1–1
58. Van Greuningen M, Batenburg RS, Van der Velden LF (2013) The accuracy of general practitioner workforce projections. Hum Resour Health 11(1):1–1
59. Wang C, Ge J, Xu X Analysis of Air Traffic Flow Control through Agent-Based Modeling and Simulation. IEEE
60. WHO: (2011) WHO Evaluation Of the National Health Plan of Portugal. Tech rep, WHO Regional Office for Europe, Geneva
61. Yett DE, Drabek L, Intriligator MD, Kimbell LJ (1972) Health manpower planning: an econometric approach. Health Serv Res 7(2):134–147

Chapter 4

Physician emigration: should they stay or should they go? A policy analysis[1]

[1] Submitted to an international peer-reviewed journal. In review.

Physician emigration: should they stay or should they go? A policy analysis

February 6, 2017

Abstract

Physician emigration can either function as an escape valve to help the health labour market clear from a supply surplus, or aggravate the problem further in case of a shortage. Either way, policy-makers should be particularly aware and devise policies to minimize the occurrence of an imbalance in the physician workforce, which may require physician retention policies if barriers to entry and other market rigidities can not be removed. To this purpose we have developed an agent-based computational economics model to analyse physician emigration, and have used it to study the impact of potential short- and long-term retention policies. As a real case study we have calibrated it with data from Portugal, which features a very particular health system with many rigidities. Results show that all policies are capable of increasing the workforce size, but not all reduce emigration. Furthermore, the welfare impact of the policies varies considerably. Whether policies to retain physicians should be enacted or whether policy makers should let physicians go will depend on the type of imbalance present in the health system.

Keywords: Healthcare workforce planning; Health policy; Agent-based computational economics; International migration; Physician migration; International medical graduates

1 Introduction

An adequate medical workforce is a critical requirement to an healthy population, as the delivery of health care services is still highly dependent on physicians. In general, health authorities are aware of the consequences of a disproportion in the physician workforce, and put this issue on top of their political agendas (Ono et al., 2013). Notwithstanding, shortages of physicians or a geographical maldistribution of health human resources are still very common and affect all countries alike, from low- to high-income countries (Labonté et al., 2015; Gauld and Horsburgh, 2015), suggesting that, despite the frequent proclamations, no public policies have been enacted to address this issue or such policies have been ineffective in achieving their goal.

Several factors originating from both the demand- and the supply-side may cause an imbalance in the medical workforce (Amorim Lopes et al., 2015). Institutional and regulatory barriers, such as *numerus clausus*, licenses to practice or mandatory retirement age; service delivery arrangements that may condition both the skill mix and the productivity of the health care workforce; non-controllable factors, such as mortality; changes to demography, in particular to the size or age structure of the population; migration flows; epidemiological needs; and many other factors may cause a change in the medical workforce. Of all these factors, emigration is one of the most difficult to handle, as it exerts an immediate influence on the supply of physicians.

Physicians are no different from other people, and emigrate with the purpose of being better off (ex ante). The international migration literature lists a significant number of "push" and "pull" factors that may act as stimuli to leave or attract one to other countries, respectively. In the particular case of human resources for health, several authors have identified critical factors (Clark et al., 2006; Ribeiro et al., 2014), for instance: inadequate compensation, when compared to other sectoral wages, but also to wages in other countries; remuneration system not related to performance; heavy workloads; poor working conditions; political instability; or unmet demand for medical education due to *numerus clausus* or lack of residency positions are reasons frequently provided for wishing to move abroad. Regardless of the working and living conditions in the home country, foreign countries may also offer fringe benefits that may attract physicians. For instance, training opportunities; better compensation, in terms of wage differential; active recruitment; or higher demand for health professionals.

In the case of an oversupply of medical doctors, emigration may actually act as a safety valve in clearing the health labour market and minimizing the negative welfare impact arising from unemployment. However, if there is a shortage in the medical workforce, emigration aggravates the problem further. Actually, if it happens repeatedly, even small numbers of physicians moving abroad may be responsible for large workforce variations if no action is taken to mitigate it. Monte Carlo simulations run by Amorim Lopes et al. (2016) demonstrate that small variations in the emigration rate may cause large variations in the workforce size, which may put the health system under severe pressure.

Although in some cases retention policies may exhibit a limited effect on dissuading physicians from moving abroad (Hussey, 2007), empirical evidence seems to support the view that some may actually work. For instance, Okeke (2014) found evidence that a wage increase programme in Ghana reduced the foreign stock of Ghanaian physicians by approximately 10%. The same conclusion was reached by Labonté et al. (2015), who also found evidence from South Africa that wage incentives seem to work. In the same line, Ortega and Peri (2009) estimated that an increase in the wage differential between origin and destination countries of 1000 US $ increases the flow of migrants by 10-11%, with the corollary being that reducing the wage differential would decrease the flow of migrants. On the opposite direction, Vujicic et al. (2004) found little correlation between the supply of health care migrants and the wage differential between source and destination countries, and suggests that wage differentials may be so large that a small salary raise would produce little to not effect.

Theoretically, it is hard to argue against the role played by wages in the international migration flows. Either when the physician is modelled as a rational utility-maximizer agent that moves based on the wage differential, as it is commonly found in classic theories, or when migration is taken as part of an investment in human capital, an approach typically found in modern migration theories, the expected income is hardly negligible (Bodvarsson and Van den Berg, 2013). Even if other less tangible factors, such as working or living conditions, exert a bigger influence, there is theoretical ground to assume that income, and therefore wages, is a relevant factor as well.

Wage incentives, framed within a suitable incentive system, may then be an adequate way of targeting current physicians. Ono et al. (2014) points to this factor as a means to respond to imbalances in the geographical distribution of physicians, but it applies also to international migration. Overall, the authors identify three broad strategies available to policymakers: (i) targeting current physicians by devising an incentive system, mostly focused on financial incentives (a short-term strategy); (ii) targeting future physicians in order to maximize the pool of physicians available to practice, which implies increasing the number of qualified individuals (a long-term strategy, as earning a license to practice is a decade-long effort); (iii) do with less, which means accepting the current levels and optimizing the delivery of health care services to compensate for the lack of physicians.

Although it is quite clear by now that policy levers exist to mitigate the emigration of physicians, it is not always the case that we can grasp the full extent of their impact. In very flexible health labour markets featuring few rigidities and institutional barriers one can use analytical tools, such as dynamic stochastic general equilibrium (DSGE) models, to study the effects of such policies, contingent upon no severe nonlinearities being present. Otherwise, the model loses tractability and it is very hard, if not impossible, to account for second-order effects, and in some cases to first-order effects also. In such cases, the predictive power of the model is simply non-existent for setting policy (Bouchaud, 2008). In practice, some health labour markets can be extremely rigid, with uncertainty, asymmetric information, barriers to entry and a quasi-monopsonistic employer causing it to divert strongly from textbook Walrasian equilibrium. In fact, non-market clearing and Pareto inefficiency is the rule and not the exception.

For a real-case example consider the Portuguese health system. It features many rigidities: *numerus clausus* exist to limit access to medical schools, since education is subsidized; residency training is also limited by the Medical Association, who decides how many vacancies to open for each medical specialty; the health labour market conveys many features of a monopsony, with the government being by far the largest employer and with restrictions to supply reducing the short-run elasticity of supply; pay for performance is used seldom, with few incentives existing to adjust physicians' wage levels according to their productivity; wages in the public sector are very sticky, being fixed by the government through collective bargaining; wages in the public sector put an"anchor" on the salaries practiced by the private sector; the wage ceiling in the public system caps the quantity of health care services supplied; very limited "freedom of choice" in the public sector, with citizens unable to choose their family doctor; low access fees that virtually have no impact on the demand for health care services; etc. On top of this, Portugal features one of most rigid labour markets in the EU according to OECD (2013).

Despite not being suitable for a traditional analysis, studying emigration in an health system so far away from theoretical competitive equilibrium models, evaluating and quantifying potential retention policies and informing policymakers accordingly is still very relevant. In the end, the apparatus should be able to answer the question of how policymakers should deal with physician emigration, although a one-size-fits-all optimal policy is unlikely to exist. For this purpose we have developed a theoretical economy that reproduces all the traits of the Portuguese health system, calibrating it with real data. Using this model we simulate the evolution of the health system and conduct policy experiments to mitigate emigration, tracking the impact at the micro-level – at the level of detail of a single agent. We play with three policy levers: adjusting *numerus clausus*, the number of job offers to the public sector and physician wages, and we observe their impact on the simulated economy, as well as first- and second-order effects.

To implement the model we resorted to agent-based computational economics (ACE), as these models go beyond Walrasian equilibrium foundations, representative agents and the assumption of market clearing (Colander et al., 2008). In a non-Walrasian health labour market, as is most likely the case of Portugal for the reasons aforementioned, an agent-based simulation model (ABM) is a very adequate tool for the problem at hand. Despite not being ubiquitous, ABMs are gaining traction, and several authors are advocating for its use (Farmer and Foley, 2009). Fagiolo and Roventini (2012) provides several examples of the use of ABM in modelling economics and policy, comparing it to typical DSGE models. Neugart and Richiardi (2012) elaborates on the use of ABM to analyse the labour market, while Dosi et al. (2010) documents a very interesting application of ABM to devise a policy-friendly model that reproduces endogenous growth and business cycles.

The remainder of this paper is organized as follows. Section 2 presents the economic model and its components. Section 3 runs policy invariant simulations so as to obtain a baseline scenario. Section 4 experiments with some policy levers, Section 5 discusses the results and their implications and Section 6 concludes.

2 A model for physician emigration

Our simple economy features two types i of agents, physicians H and non-physicians L, a publicly funded health care system run by the government and a private sector that exists for accounting purposes only. The public sector is a highly regulated Beveridgean national health system with salaries fixed by collective bargaining between unions and the government. Given the size of the public sector, the wage level that results from the negotiations anchors the salary paid in the private sector.

The aggregate supply of physicians S^H is given by the current stock of physicians already practicing plus the medical students about to join the profession and any foreign doctors due to arrive in the future. We assume, and empirical data support this stance, that the *numerus clausus* puts a limit on the supply of physicians, and that the number of students willing to join medical school is higher than the cap imposed by the number of vacancies.

We differentiate between *potential* and *effective* demand for healthcare services, where by potential demand we imply demand effectively observed plus unmet needs. Due to the restrictions imposed on the health labour market, such as a wage anchor and *numerus clausus*, the quantity of healthcare services provided is below potential. In practice, this results in care not delivered, usually observable in the guise of waiting lists and long queues (Amorim Lopes et al., 2015). In this model, aggregate demand is assumed to be exogenous and constant[1].

2.1 Physicians

2.1.1 Life cycle

Before being allowed to practice, aspiring physicians have to join medical school in order to earn a license. A physician has to go through the following stages in life, illustrated

[1]For an analysis of the impact of demography on the demand for health care services in Portugal refer to Amorim Lopes et al. (2016).

in Figure 1:

1. Enter medical school and complete the mandatory 6 years of studies

2. Upon earning the medical license, junior doctors have to do an admission exam and join internship for a year before applying to a specialty residency

3. Junior doctors are ranked according to the grades obtained in the admission exam and apply for a specialty residency; if unable to enrol in a residency program, the doctor stays on hold until he can redo the exam and apply again the next year (in most of the cases he can still practice as a clinician); depending on the specialty, it may take between 3 and 6 years to complete

4. Upon graduating, physicians join the public or the private sector, switching between the two at any time, and practice until retirement

5. Retire at around the retirement age.

Additionally, a medical student may drop out from medical school, and any physician may emigrate or die at any time. The transition time between states is stochastic.

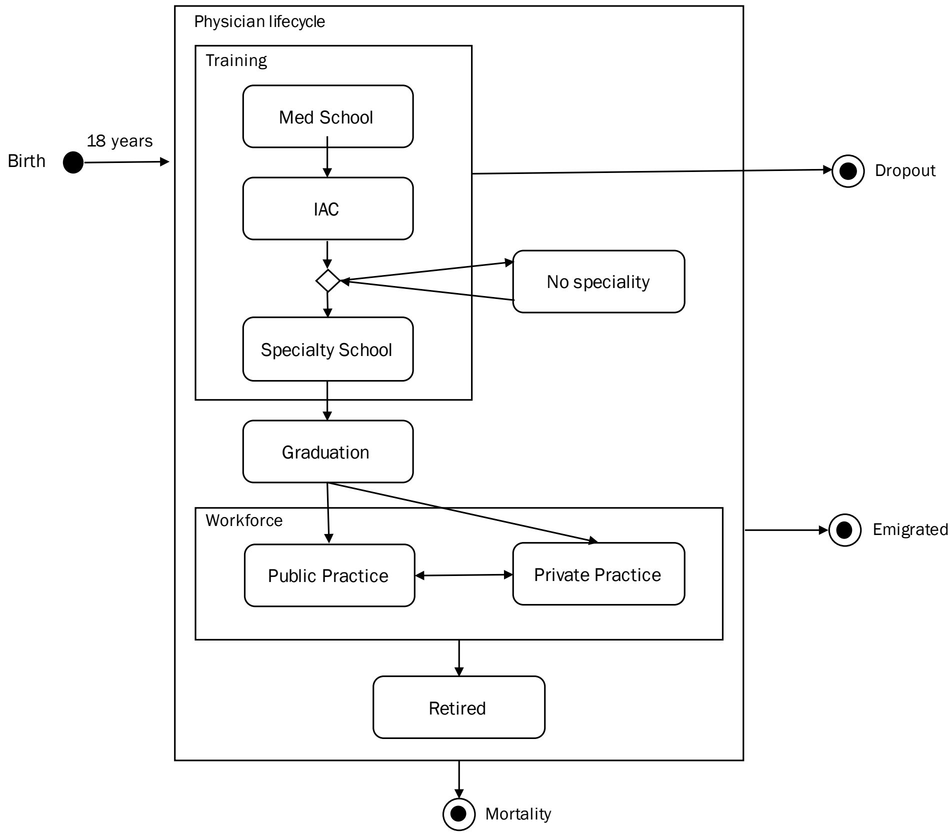


Figure 1: The life cycle of a physician since joining medical school.

We explicitly model the different stages of a physician's life cycle as they may influence significantly the internal decision-making of a physician. For instance, anecdotal data suggests it is more likely for a physician to emigrate in case he was unable to enrol in the specialty of his liking, or in case he is unemployed (Martins, 2015). In contrast, it is less likely for a physician to emigrate when he is at an older age, closer to retirement. In this way we can simulate precisely how external shocks and policy experiments unfold on our model.

2.1.2 Migration decision-making

An individual emigrates in the expectation of being better off by doing so. Such decision is similar to that of an investment – moving abroad entails an upfront cost, both financial and psychological; an opportunity cost, which are the earnings forgone, fringe benefits and family and friends left at the home country; and increased earnings or other fringe benefits obtained in the foreign country. The benefits, but also some of the costs, accrue over a period of time. This implies that the value of such wage and non-wage income and costs is discounted over time. Such assumption is consistent with the fact that emigration rates decline with age (DaVanzo, 1980).

To model the physician's internal migration decision we borrow from the labour migration literature (Borjas, 2001; Maloney and Gonzalez, 2005). We extend the typical definition by explicitly including the cost of being unemployed or unable to obtain a training place at a residency program, which we assume to be greater than the opportunity cost of not earning an income. The lack of vacancies at university hospitals is known to be one critical push factor in the health labour market, in particular mismatches between preferences and specialties (Ribeiro et al., 2014). Taken together, physicians emigrate according to the sign of the following index function:

$$I^* = U_k - U - C + Z, \tag{1}$$

where U_k is the utility of living in country k and U is the utility of living in the home country. C represents the disutility arising from migration, caused by leaving family and friends behind, and adapt to a new environment (Kennan and Walker, 2013). We model it as an exponential function of age, in line with (DaVanzo, 1980). Z is the disutility of being either unemployed or unable to enrol in a residency program.

The utility of either staying U or moving abroad U_k can not be determined empirically, as stated preferences for all the population are impossible to be obtained. Therefore, we can only observe the indirect utility V. Let $V(Y)$ be the indirect utility enjoyed by individuals as a function of their post-tax income:

$$V(Y) = Y^\alpha, \tag{2}$$

with $0 < \alpha < 1$ to ensure the concavity of the value function, that is to say we assume diminishing marginal returns. Moreover, we also incorporate time preference based on the assumption that people prefer present to future consumption, and therefore apply a discount factor until retirement age R. We therefore compute the discrete present value according to:

$$PV = \sum_{t=a}^{R} \frac{V(Y)}{(1+\rho)^t}, \tag{3}$$

where ρ is the discount factor and a the physician's age at the time of decision making. In the migration literature it is common to also define working conditions, which include non-wage job characteristics people may value, such as safety, lack of stress, prestige, housing or transportation (Vujicic et al., 2004). We assume, without any loss of generality, there is a monetary equivalent to the living conditions, and therefore wage and living conditions are perfect substitutes.

All things considered, the model posits that the physician emigrates if, at any given period t, I^* in Equation 4 is greater than zero:

$$I^* = PV_k - PV_i - C + Z. \tag{4}$$

The probability of a physician emigrating is then given by:

$$\Pr\{I^* > 0\} = \Pr\{PV_k - PV_i - C + Z > 0\}. \tag{5}$$

If, at any time, $I^* > 0$ is true for any physician, then the migration market is in disequilibrium. In fact, the market will always be in disequilibrium until all physicians searching for a job find a desirable match.

2.2 Health labour market

The health care labour market is not Walrasian, as too many frictions exist, preventing it from clearing. In particular, we assume real wages are set below their equilibrium level, structural unemployment for the sector does not exist and labour rationing is the rule. These assumptions are far from being unrealistic in the case of Portugal. We assume the aggregate demand for health labour D is exogenous and inelastic. The aggregate supply S^H is computed by running the ABM and obtaining the forecasts for the number of physicians available for practice. Finally, market and institutional factors may influence government decisions, but in the end the wage level is set by the government. We assume it is given exogenously by the salary rate schedule currently in use.

2.3 Migration market

We build a migration market by adapting a simplified version of Mortensen and Pissarides (1999)'s model of search in the labor market. Typically, the wage-setting relation establishes that wages equal the expected prices times a function of unemployment and a catch-all variable that includes factors at play in the labour market. In this case, we assume the wage-setting WS relation, or the supply of migrants curve, is a function of the worker's reservation wage, i.e. the minimum value such that

$$\text{WS} = \bigcup \arg\min_{W_k} \{I^* = 0\}. \tag{6}$$

Alternatively stated, it is the minimum value that compensates for the costs of emigrating, as defined previously:

$$\text{WS} = \bigcup \arg\min_{W_k} \{V_i + C + Z = V_k\}. \tag{7}$$

A change in the domestic wage causes a shift in the WS curve. If the salary increases (decreases), the reservation wage increases (decreases) as well, and the WS curve shifts to the left (right). Changes in foreign wages cause a move along the WS schedule. The higher the wage proposal, the more physicians will be willing to emigrate. Additionally, the unemployment rate indirectly affects the *WS* curve. The higher the unemployment rate in the health labour market, the greater the disutility Z brought by unemployment, which also causes a shift in the WS curve.

For the vacancy supply VS we assume it is exogenous. There is a constant stream of job offers produced by a stochastic process, with each physician receiving one random job offer per year. Foreign wages are sampled from a beta distribution, positively skewed, fitting a typical income distribution:

$$\text{VS} = \bigcup W_k \sim \text{Beta}(\alpha, \beta). \tag{8}$$

The interaction between the *WS* schedule (supply of migrants) and the VS schedule (demand for migrants) is insufficient to determine the number of physicians that emigrate. The mechanism that matches physicians and job proposals is imperfect and time consuming. Not all job proposals reach all physicians, and it may happen that a physician never receives a job proposal above its reservation wage. The typical approach is to invoke a matching function $m = f(v, u)$, where v is the vacancy rate and u the unemployment rate, which characterizes the aggregate meeting rate between employers and employees and is derived from the Beveridge curve. Formally, the matching process is described as a productive process, and modelled accordingly using a production function. We assume random job offers arrive deterministically at a constant rate. Each physician receives one job proposal per year, and the matching rate is equivalent to the probability of the value of the job offer received by the physician to be above the value of his reservation wage. In practice, and due to matching frictions, less physicians than the number the intersection between *WS* and *VS* suggests actually emigrate.

2.4 Market interactions in the economy

When we combine all the individual blocks, more specifically the microeconomic foundations behind migration decision-making, the health labor market and the migration market, we obtain a small functioning economy. In this economy the migration decision, and therefore the migration rate, is endogenous, but subject to the influence of policy levers such as changes to the domestic wage. Increasing the domestic wage reduces the incentives to emigrate, and decreasing it would generate the opposite effect. Furthermore, an increase in the stock of physicians caused by a change in the *numerus clausus* will affect the emigration levels, which is expected, but would also affect the growth rate of the medical workforce. The intuition is that since specialty vacancies and job proposals for joining the public sector are limited, more physicians will be awaiting for a spot. Waiting influences positively the decision to emigrate through the mechanism described before.

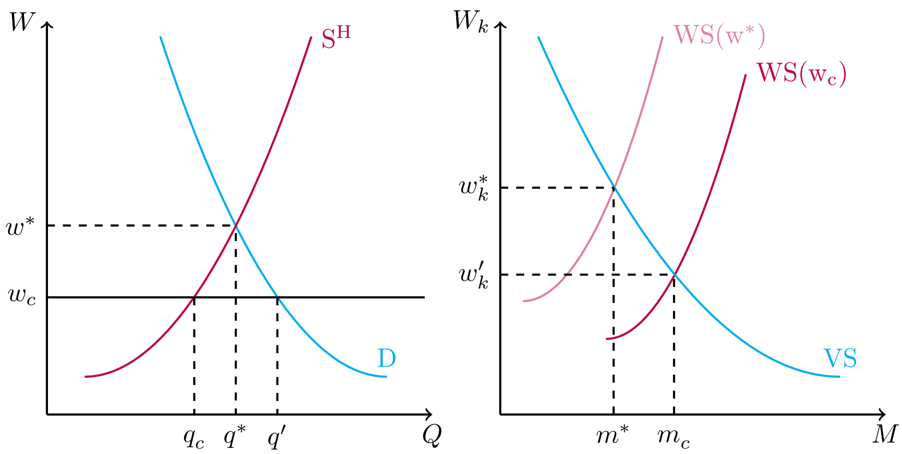


Figure 2: The interaction between the health and the migration market.

Figure 2 depicts some of the interactions in the economy. On the left there is the health labour market, where S^H is the aggregate labour supply of physicians and D is the demand for health care services, and on the right there is the migration market, which results from the interaction between the wage-setting curve *WS*, defined implicitly by the physicians' reservation wage, and the vacancy supply curve *VS*, i.e. the aggregation of all the foreign job proposals. Actual emigration levels are below the equilibrium point due to matching frictions in the migration market.

These two markets are always in interaction, and a change in one causes a change in the other. For instance, the wage anchor w_c puts a cap on the quantity of healthcare services provided, and so the actual quantity provided is q_c. Thus, it creates an hiatus in demand, reducing the quantity of healthcare services provided by $q^* - q_c$ units. But the level of the domestic wage w_c also influences the migration market by causing a shift in the wage-setting curve *WS*, in this case from $WS(w^*)$ to $WS(w_c)$. This increases the number of potential migrants by $m_c - m^*$. All things being equal, changes in the domestic wage will affect the migration market by influencing the decision-making process and the willingness to emigrate. The importance of this effect is considerable. Since wages in the public sector are set by the government, it can be used as a policy lever. However, increasing the wage bill puts an extra burden on public finances as well. Instead, if policy-makers decide to increase the *numerus clausus*, and therefore the number of available physicians, shifting the aggregate supply curve to the right, in the long run the number of new entrants will increase. However, that also shifts the *WS* curve to the right, implying that more people will emigrate also. Whether the wedge between the markets compensates the increase in migrants will depend on the elasticity of emigration with regards to wage. In the next two sections we will simulate the impact each type of policy may have in a real life setting.

3 Simulation results

Due to the highly stochastic nature of the model, in particular the life cycle governing the physician's various stages of life and the several nonlinearities arising from the agents' decision-making rules, it is not possible to obtain an analytical, closed-form solution. Also, since the health labour market departs too much from Walrasian assumptions an equilibrium is not attained, and therefore cannot to be computed.

We can, however, run computer simulations through the ABM to experiment with policy levers and observe and quantify their impact on the economy. To clear away any simulation variability we run Monte Carlo simulations, collecting the averages and standard-error bands, a procedure typically followed when running agent-based computational economic models (Dosi et al., 2010).

We start by devising a baseline model calibrated to replicate the *status quo* of the physician workforce. More specifically, we load the model with the active physicians working in the Portuguese health system, as well as the students attending medical school, internship or residency. Parameters such as the drop out rate, mortality rate or the *numerus clausus* are also fed into the model. Finally, the number of years until transitioning to another state in the physician lifecycle is drawn from distributions obtained empirically. Initial conditions and parameters used on the baseline model are presented in Table 1.

Although no data exists to empirically estimate the VS distribution and remaining

Table 1: Baseline parameters.

Description	**Symbol**	**Value**
Drop out rate		0.02*
Mortality rate (life table)	Q_x	$p = Q_x$‡
Numerus clausus		1496*
Residency vacancies		1637†
Job contracts (public sector)		1934†
Transition Med School to IAC (years)	$\exp(\gamma, \frac{1}{\lambda}, min, max)$	$[0, 2, 6, 14]$
Transition IAC to Specialty (years)	$\exp(\gamma, \frac{1}{\lambda}, min, max)$	$[0, 0.4, 1, 2]$
Transition Specialty to Graduated (years)	$\exp(\gamma, \frac{1}{\lambda}, min, max)$	$[0, 1, \text{duration}, \text{duration} + 6]$
Retirement age	$R \sim N(\mu, \sigma)$	$[63.10, 4.9]$
Discount rate	ρ	0.05
Utility function parameter	α	1.05
Emigration cost	C	$0.53e^{0.0492age} \cdot V_i$
VS distribution	$\text{Beta}(\alpha, \beta)$	$[1500, 16000]$
VS distribution support	$[\underline{x}, \overline{x}]$	$[0, 10]$

*Source: Ministry of Higher Education (DGEEC)
‡Source: National Institute of Statistics
†Source: Health System Central Administration (ACSS)

emigration parameters, such as the emigration cost or the shape of the utility function, the parameters were calibrated so that the endogenous emigration rate obtained follows the numbers claimed by the Portuguese medical association, i.e. a rate of between 100 and 200 physicians per year.

Running the baseline case we obtain the forecasts for an invariant policy scenario, the case when no policy interventions take place, and thus every exogenous parameter stays unaltered. Figure 3 shows how the different stocks of physicians evolve through the planning horizon. In the initial ten years there is a decrease in the number of physicians both in the public and in the private sector, which is expectable given the age structure of the physician workforce, as a considerable number of physicians are in their 50s and about to retire. This gets reflected on the high number of retirees in the first years (Figure 3c). Furthermore, the number of graduates, i.e. physicians that have completed residency training but have not yet joined either the public or the private sector, increases. This suggests that the current number of contracts is insufficient to absorb all the new graduates, and slack is created in the physician workforce (Figure 3b). Whether this slack is an actual surplus depends on whether demand stays constant and there is no shortage in the present. Previous studies suggest that due to the aging of the Portuguese population, demand for health care services is likely to increase (Amorim Lopes et al., 2016).

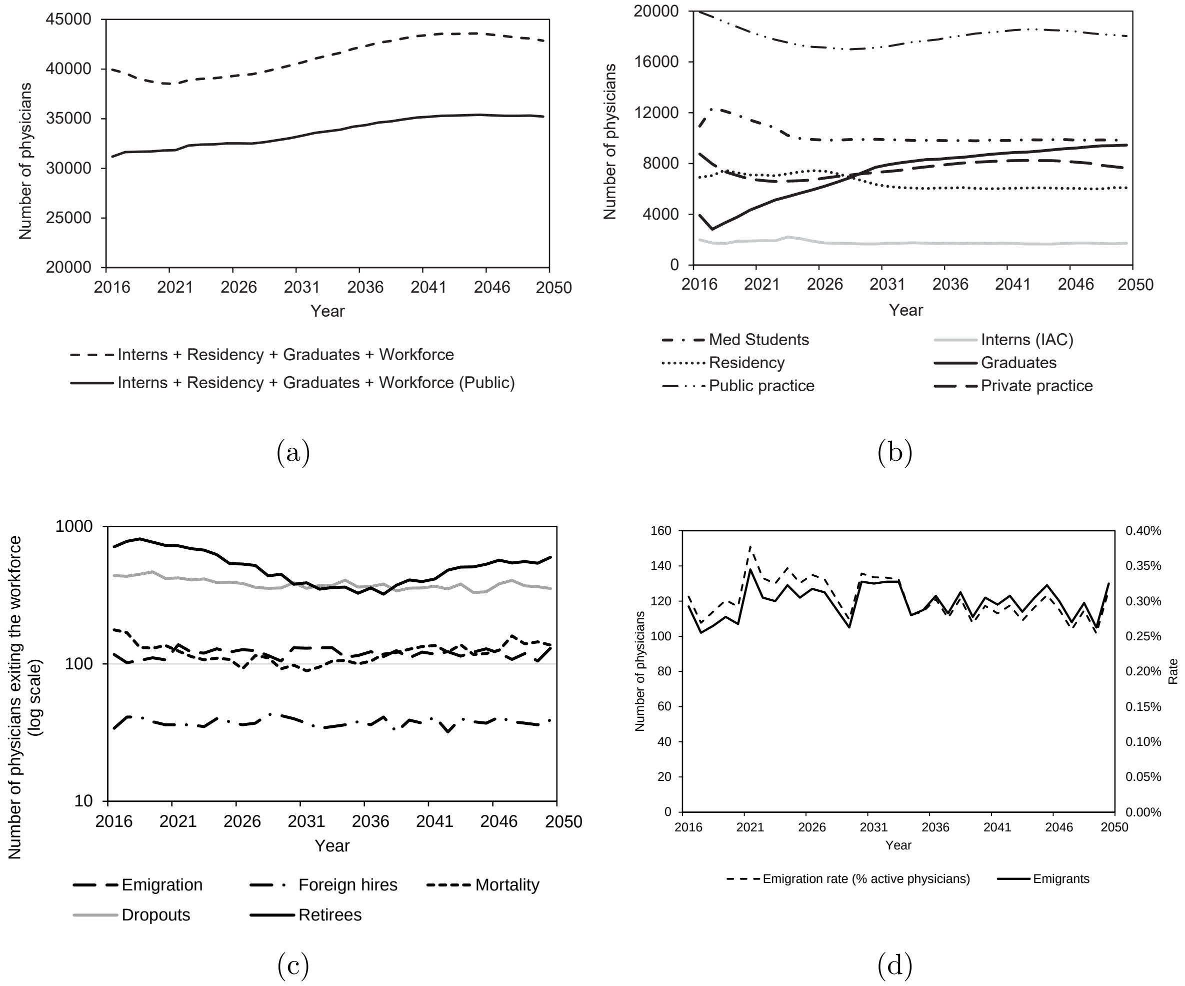


Figure 3: Forecasts generated by ABM for the baseline case. On panel (a) we have the forecasts for the aggregate supply of physicians, including interns and residents; on panel (b) the aggregate supply decomposed into each stock; on panel (c) are each of the frictions affecting the aggregates; and on panel (d) the number of emigrants and emigration rate.

Observing all the frictions affecting the health labour market (Figure 3c) we can identify the main factors driving physicians out of the workforce. With the exception of the number of retirements, which reflects the negatively skewed age structure of the physician workforce, all the outflows are fairly constant. In particular, when we look at the emigration rate we can observe a value varying between 0.25% and 0.35% of the workforce. Likewise, the number of emigrants varied between 100 and 140 per year (Figure 3d), in line with the values reported by the medical association.

As previously mentioned, stochasticity plays a major influence in the model. Therefore, Monte Carlo simulations are needed to clear away any long tail statistical effects. In Table 2 we report the averages and standard deviations of the number of emigrations and of the resulting emigration rate as a percentage of the total workforce when the model is run multiple times. Considering that no shocks hit the economy and the workforce size remains relatively stable it comes at no surprise that the emigration rate stays relatively

	2016	2020	2025	2030	2035	2040	2045	2050
Emigration (N)	124	117	122	124	120	122	121	124
	(13)	(10)	(10)	(11)	(11)	(10)	(11)	(11)
Workforce size (N)	35131	34767	37111	38648	40366	41639	41886	41138
	(43)	(81)	(81)	(103)	(94)	(116)	(122)	(125)
Emigration rate	0.35%	0.34%	0.33%	0.32%	0.30%	0.29%	0.29%	0.30%

Table 2: Average of the number of migrants and for the emigration rate (as a percentage of the total workforce size) for the forecasting horizon. Standard deviation for the Monte Carlo simulation runs reported in parenthesis.

stable at 0.29%-0.38% for the whole horizon, averaging 0.34%, with the number of physicians moving abroad ranging between 109 and 130, with an average of 120 (note that Table 2 only reports values every five years, but the values are calculated for all periods).

Figure 4 provides a more detailed account of the emigrants in our simulated economy. According to the simulations, most of the emigrants are either enrolled in a residency program or have graduated, but have not yet been hired. Physicians already under a contract either in the public or in the private sector represent a very small share (less than 8%) of all departures. Not surprisingly, the average age of those leaving while on or about to join residency training is around 30 years of age, while those that have already completed all medical specializations are usually older.

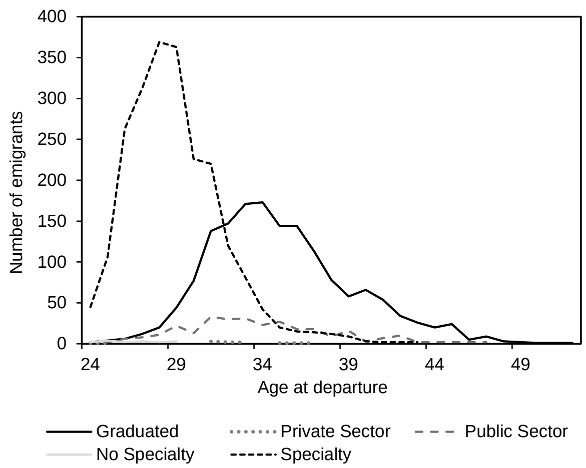
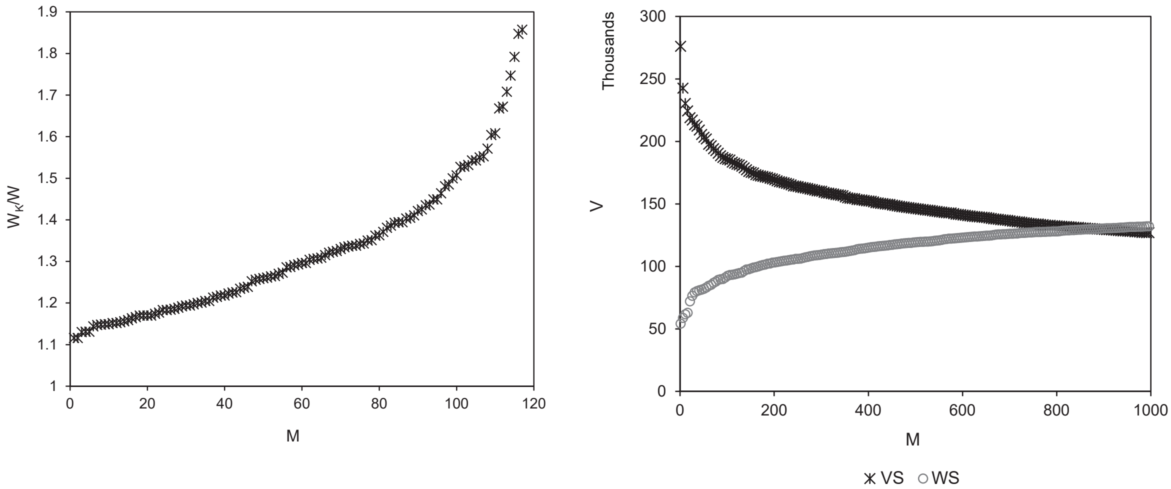


Figure 4: Age profile of the emigrants that have left during the simulation period (35 years).

We can obtain further information by looking at the migration market. Figure 5b provides an overview of supply (VS schedule) and demand (WS schedule). The quantity at which it reaches equilibrium is approximately 820 workers. However, due to mismatches in the migration market (i.e. not all physicians have access to the same job openings) the market does not clear, and the number of physicians that actually emigrate is considerably below the equilibrium point, as reported in Table 2. Figure 5a provides the wage premium for all the physicians that have moved abroad in one particular year. In our simulated economy half of the physicians that decide to emigrate do so for a wage differential equal or less than 30%. Additionally, 14% of the migrants, or 624 physicians, would not have emigrated had it not been for the burden of being unemployed ($I^* < 0$ if $Z = 0$).

The model we have obtained accounts for a large number of empirical regularities observed in the Portuguese health system: (i) the number of physicians emigrating falls within the boundaries reported in the grey literature and by the medical association; (ii) the average age at which physicians move abroad ranges between 25 and 40 years, targeting especially young doctors (Gauld and Horsburgh, 2015); (iii) most of the physicians that emigrate do so because they are not under a contract or could not join the desired specialty. This suggests the model is robust enough to serve as a test bed to study retention policies and obtain a qualitative measure of their impact. In the next section we will experiment with some policy levers and observe the impact on the model.



(a) (b)

Figure 5: On the left panel we have the ratio of external wage to domestic wage for all migrants on a random year. Half the migrants leave for a wage premium of less than 30%. On the right panel we can observe the migration market in action, with the short-term VS schedule aggregating all the job offers and the short-term WS schedule the reservation price of the physicians for a particular year.

4 Policy experiments

Barriers to entry and extensive regulation in the Portuguese health labour market make it difficult, if not impossible, for it to automatically adjust from an imbalance. Even if there is a clear sense that a shortage or a surplus of physicians exists, the size of the workforce is still severely conditioned by the number of vacancies to medical school, by the number of residency vacancies and by the number of job offers awarded by the largest employer, the national health system. Arguing about the economic sense of these market rigidities is necessary, but outside the scope of this work. We therefore assume government intervention through public policies is required to address an imbalance in the Portuguese health labour market.

It is also assumed the main policy objective is to ensure a balanced workforce, with a decrease in the emigration rate being a positive externality, but not a goal *per se*. Policies targeting the physician workforce typically fall into one of two categories (McPake et al., 2013): short-termed ones, producing immediate effects, such as raising wages, a salary top-up or hiring foreign doctors; long-termed ones, such as expanding the *numerus clausus* to medical school or increasing the number of job offers to the public sector. Regardless of the time required to produce an effect, public policies are usually gauged in terms of impact to either the level or the growth rate of income, or with regards to changes in the income distribution arising from these interventions. Policies that decrease the income level or rate, or increase inequality decrease social welfare. In the particular case of physician emigration, policy makers may face additional dilemmas. Some policies may help to tackle an imbalance in the physician workforce, but at the same time may also increase inequality. Others, such as hiring foreign doctors, may cause an imbalance in developing countries, as the "brain drain" typically flows from developing to developed countries (Clark et al., 2006). Theoretically, we can pin down the welfare impact of each policy. In practice, however, computing the precise outcome requires knowing the preferences of the population, which may not be feasible, and so the policy conclusions are frequently ambiguous and surrounded by uncertainty. Notwithstanding the welfare impact, each policy will have different first-order effects on the physician workforce. These effects can be quantified by subjecting the ABM previously devised to changes in these policy levers and observing the result in the workforce projections.

In this section we analyse the effect that three widely different policies exert on the physician population, which are: (i) raising wages; (ii) increasing job offers to the public sector; (iii) expanding *numerus clausus* to medical school[2]. A priori we can anticipate that increasing *numerus clausus* will expand the physician workforce, with the emigration rate increasing as well. Optionally, an increase in the number of contracts to the public sector reduces the chance of emigration due to unemployment, but it lasts only until there is a stock of graduates available. Finally, increasing wages has a positive effect both on the workforce size and on the emigration rate. The potential negative effects, however, are only visible through a welfare analysis, and will depend on whether there is a shortage or surplus in the supply of physicians.

[2]We do not consider hiring foreign doctors for three reasons: first, the economic and social impact on destination country is trivial to determine; second, we want to produce structural changes in the health labour market, not having to rely on an exogenous and uncontrollable source of labour to mitigate deficiencies in supply; third, the World Health Organization has established a code of conduct for the international hiring of physicians, aware of the harm emigration may cause on source, especially poor, countries.

4.1 Increase wages

Raising wages is a typical employee retention strategy, working through the channel described in Section 2.1.2. In any system governed by, even if partially, the laws of supply and demand, increasing prices ought to increase the quantity supplied. In the migration market increasing wages raises the utility of staying, and therefore reduces the incentives for moving abroad.

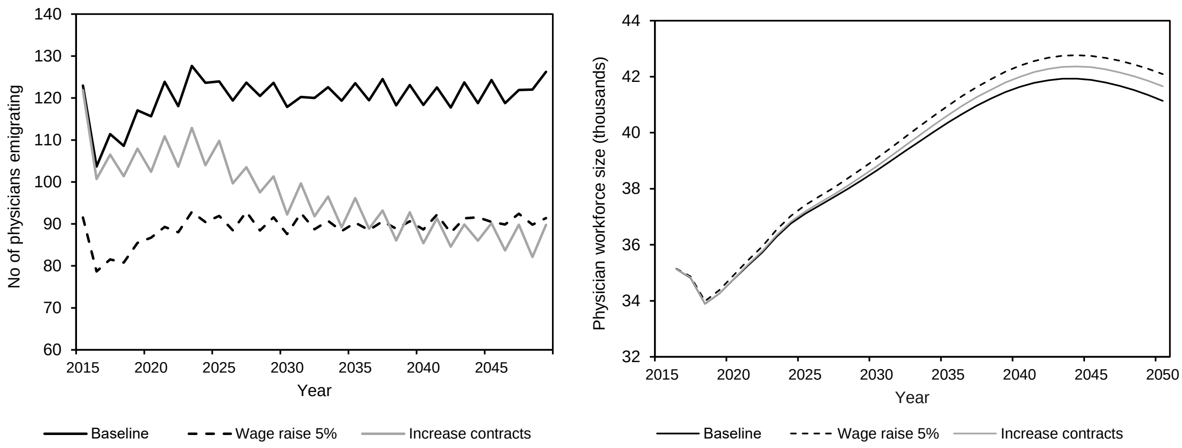
In the case of a private hospital or any other private health care institution, raising wages is a relatively straightforward managerial decision with costs and benefits, and can be enacted within a very short period. In a publicly funded health care system with collective bargaining and a workforce size of over 30 thousand physicians, increasing wages may have wide financial and economic repercussions. Understanding such consequences is critical for evaluating the adequacy of this policy.

The most immediate impact of increasing wages results from the fact that the government, like any other economic agent, is subject to a budget constraint. In the particular case of Portugal, to a very restrictive one. This means increasing public spending in one sector of the economy implies, at least in theory, reducing spending on others. Therefore, understanding the extent and impact on social welfare of such decision is critical.

There is one further nuance in our case study that complicates matters further. In Portugal, college fees are subsidized. In the particular case of medicine, government covers over 90% of the real cost in tuition fees. As a result, if a physician leaves the country before practicing the return on investment is negative from the government perspective. It has invested a substantial amount of resources in his or her education that could have been allocated elsewhere. However, it is a sunk cost, and it should not weight on the decisions to be taken henceforth. On the other hand, increasing wages has a direct cost and may be an additional burden to the taxpayer, which may aggravate inequality further, since physicians are already on the top end of the income distribution. In addition, this could lead to a "contagion" effect on other public servants, who would also push for wages increases.

We can simulate the impact of raising wages using the model. Not surprisingly, raising wages should decrease the number of physicians deciding to move abroad, which it does. Figure 6 provides an introductory overview of the effect raising wages by 5% has on the migration market, and hence on the domestic health labour market. The most relevant insight is the fact that a wage raise of 5% led to a decrease in the migration levels. This effect accrues over time, in the end exhibiting a non-negligible effect on the workforce numbers (cf. Figure 6b).

Looking in detail at Table 3 we can observe that, on average, the emigration rate decreases 0.09 p.p. when salaries are raised 5%. In absolute terms, there are approximately 30 physicians less emigrating per year, which corresponds to a decrease of approximately



(a) Number of physicians emigrating.

(b) Change to the overall physician workforce.

Figure 6: The effect on both the remaining physician workforce and on the number of emigrants of raising wages versus increasing the number of contracts to the public sector.

25% when comparing to the baseline case reported in Table 2.

4.2 Increase the intakes to the public sector

Figure 4 reveals the typical profile of the emigrating physician. First in line come those about to enter or that have already entered residency training. In second come the graduates that have already completed their training, but have not yet joined either the public or the private sector. This suggests that crafting a policy to retain these graduates might also be an adequate option.

With a similar impact on the public budget and as an alternative to raising wages in the public sector, we can also simulate the impact of increasing the number of job offers to be awarded to graduates. We now adjust the number of contracts to the public sector by calculating the number of physicians government can hire in alternative to raising wages. In rough numbers, a 5% wage increase will cost approximately 45 million Euros per year. Assuming an average yearly wage of 40 thousand Euros, government could offer approximately 1125 contracts more.

Table 3 reports the impact this policy has on the supply of physicians. In comparison to raising wages, it has two different effects. Firstly, it is not as effective in retaining emigrants, as it targets graduates mainly. Secondly, it takes a longer time to produce an effect on the medical workforce, which can be explained by the high number of graduates waiting for a position. In fact, the policy becomes ineffective after the initial slack of graduates is absorbed by the public sector, since the number of contracts awarded is higher than the number of physicians graduating each year, and the main reason being that access to medical school is limited by *numerus clausus*. Increasing the number of vacancies to medical school could also be an alternative policy, to which we turn next.

	2016	2020	2025	2030	2035	2040	2045	2050
				Raise wages by 5%				
Emigration	92	86	90	92	88	91	92	91
Δ baseline	-31	-32	-33	-32	-31	-32	-27	-35
	(10)	(9)	(9)	(9)	(10)	(10)	(9)	(10)
Workforce size	35147	34910	37413	39083	40969	42392	42741	42092
Δ baseline	16	143	302	435	603	753	856	954
	(39)	(75)	(81)	(90)	(105)	(123)	(126)	(120)
Emigration rate	0.26%	0.24%	0.24%	0.23%	0.22%	0.21%	0.21%	0.22%
Δ baseline p.p.	-0.09	-0.09	-0.09	-0.09	-0.08	-0.08	-0.07	-0.09
				Annual increase of 1125 contracts to the public sector				
Emigration	125	110	108	104	96	99	90	98
Δ baseline	2	-7	-16	-19	-23	-24	-29	-28
	(11)	(10)	(10)	(11)	(10)	(9)	(9)	(11)
Workforce size	35126	34779	37230	38933	40843	42293	42697	42108
Δ baseline	-5	12	119	285	477	654	811	970
	(42)	(76)	(85)	(95)	(92)	(96)	(110)	(113)
Emigration rate	0.35%	0.32%	0.29%	0.27%	0.24%	0.23%	0.21%	0.23%
Δ baseline p.p.	0	-0.02	-0.04	-0.05	-0.06	-0.06	-0.07	-0.07

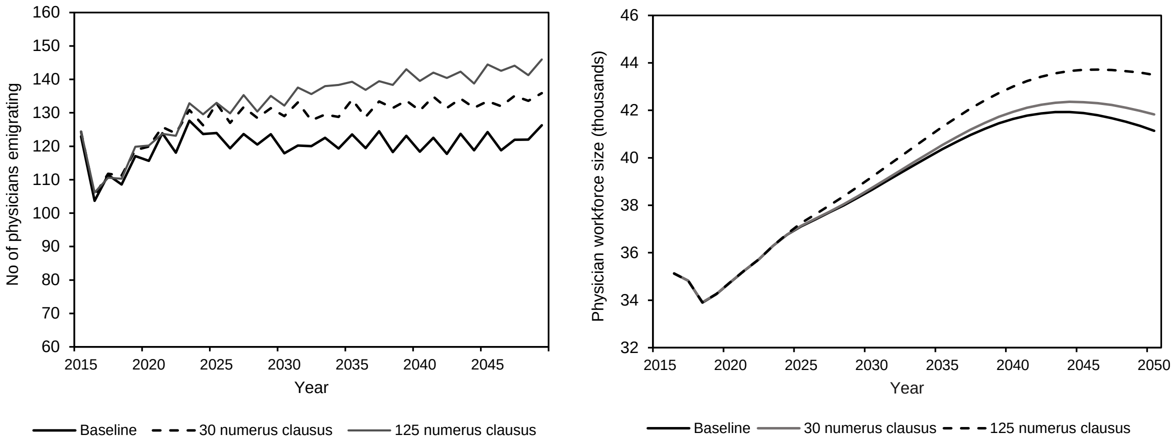
Table 3: Monte Carlo simulation runs average of the number of migrants, the workforce size and the emigration rate (as a percentage of the total workforce size). Standard deviations reported in parenthesis.

4.3 Increase numerus clausus

Increasing the *numerus clausus*, which acts as an effective barrier to entry in the health labour market, is a completely different approach to the problem. Instead of reducing the number of physicians willing to emigrate, the aim is to increase the workforce size so as to compensate for the losses. Notwithstanding its direct impact on emigration, inability to train and retain sufficient doctors may lead to an overdependence on foreign-trained doctors or to an outright shortage (Bidwell et al., 2013). Very restrictive *numerus clausus* may thus have a detrimental role on the size of the medical workforce.

As a matter of fact, the number of emigrants will most likely increase, albeit with the growth of the workforce size absorbing the leaving physicians. A growth in the medical workforce, notwithstanding the permanent losses to migration, has an additional benefit: more health care services can be provided if necessary, which seems necessary given the aging of the Portuguese population (Amorim Lopes et al., 2016). This is particularly relevant in case there is a shortage in the physician workforce, i.e. $H^S < H^*$.

To fully understand the effects in the health labour market we will increase the number of vacancies to medical school using two different metrics: (i) by considering the total number of physicians emigrating on the baseline scenario, which according to the results reported on Table 2 equals 120 per year, on average; (ii) by considering the number of physicians that have decided not to emigrate after the wage raise, which equals approximately 30, so that we can compare policies.



(a) Number of physicians emigrating. (b) Change to the overall physician workforce.

Figure 7: The effect of increasing the *numerus clausus* on both the remaining physician workforce and on the number of emigrants.

Overall, results presented in Figure 7 clearly indicate that increases in the *numerus clausus* bring changes to the physician workforce, helping to curb the alleged problem created by emigration. However, these changes take time to produce an effect, as medical students take 10 years, on average, to complete their specialized medical training. Therefore, this is a policy targeting the medium to long-term, but ineffective at solving short-term shortages. This is particularly noticeable in Figure 7b.

Results reported on Table 4 confirm some of the effects we predicted a priori: first, the workforce expands; second, the emigration rate does not decrease, and it actually grows, an expected consequence of the workforce expansion. Put differently, the emigration rate grows more than the growth rate of the workforce. Before addressing the actual numbers it is relevant to add that an increasing emigration rate is not necessarily adverse, unless the policy objective is to reduce emigration. Instead, if keeping a balanced physician workforce is the main goal, then changes to emigration levels are an acceptable outcome, regardless of the instrument used.

When *numerus clausus* are expanded to compensate for approximately all the physicians that emigrate in the baseline scenario, the workforce expands considerably. From a starting point of approximately 35 thousand physicians, the physician workforce expands to 43.5 thousand physicians by 2050 due to the additional inflows to medical college. As a result, the workforce expands by 2358 physicians with regards to the baseline projections, a 23% increase compared to the initial size observed in 2015, but only a 5.7% marginal

	2016	2020	2025	2030	2035	2040	2045	2050
Increase numerus clausus by 125								
Emigration	124	120	130	135	138	143	139	146
Δ baseline	2	3	6	11	19	20	20	20
	(12)	(11)	(12)	(12)	(13)	(12)	(10)	(14)
Workforce size	35123	34771	37255	39186	41304	43002	43713	43496
Δ baseline	-8	3	144	539	938	1363	1827	2358
	(43)	(81)	(78)	(96)	(104)	(111)	(114)	(137)
Emigration rate	0.35%	0.34%	0.35%	0.34%	0.34%	0.33%	0.32%	0.34%
Δ baseline p.p.	0.01	0.00	0.02	0.02	0.04	0.04	0.04	0.04
Increase numerus clausus by 30								
Emigration	124	119	126	131	129	134	132	136
Δ baseline	2	2	3	8	9	11	13	10
	(12)	(12)	(10)	(9)	(11)	(12)	(11)	(12)
Workforce size	35128	34772	37138	38732	40537	41936	42348	41826
Δ baseline	-3	5	27	84	172	297	462	688
	(43)	(81)	(78)	(96)	(104)	(111)	(114)	(137)
Emigration rate	0.35%	0.34%	0.34%	0.34%	0.32%	0.32%	0.31%	0.32%
Δ baseline p.p.	0	0	0.01	0.02	0.02	0.03	0.02	0.02

Table 4: Monte Carlo simulation runs average of the number of migrants, the workforce size and the emigration rate (as a percentage of the total workforce size). Standard deviations reported in parenthesis.

increase when compared to the baseline (cf. Table 2).

A less expressive increase of 30 vacancies, bringing the total number of students intakes to 1526 per year, produces an overall increase in the workforce of about 688 physicians by 2050, a marginal increase of approximately 1.6% when compared to the baseline scenario (see Table 2 - Workforce in 2050 vs Table 4 - Workforce in 2050). The increase in the emigration rate is also small, exceeding no more than 2 basis points.

5 Discussion

In the illustration presented here, all the policies devised are capable of counterbalancing, at least partially, the outflows in the physician workforce as a result of emigration. Table 5 provides an overview of the policies impact, measured against the forecast for the baseline scenario (described in Section 3). In case of a salary raise, the workforce growth rate increases 2.31% vis-à-vis the projected growth rate for the baseline scenario. As a side-effect, this policy is also capable of reducing the number of physicians that emigrate by approximately 1086. Considering that in the baseline scenario 4027 physicians are expected to emigrate until 2050, this accrues to a very significant 25% decline. In rough numbers, we expect this policy to cost approximately 45 million Euros per year.

Table 5: Relative impact of the different policies and approximate budget cost for enacting the policy.

Scenario description	Policy impact [a]		Approx. yearly cost (M€)	Δ to total loss from emigration
	Emigrants	Workforce		
5% salary raise	-25.81%	2.31%	45	-1086
Hire more 1125 physicians	-15.92%	1.27%	45	-670
Increase numerus clausus by 125	10.88%	5.73%	1.5	458
Increase numerus clausus by 30	6.77%	1.67%	0.36	285

[a] Policy impact of policy z is calculated as both the relative change in emigration and in the workforce with regards to the baseline scenario, $(P_z - P)/P$.

We also experimented with increasing the number of job offers to the public sector, adjusting the openings to a number not surpassing the cost of the previous policy. This policy also mitigates emigration and expands the medical workforce, although at a smaller rate. Emigration rate declines -15.92% and the workforce size sees an increase of 1.27% (already accounting for the physicians that have moved abroad). When measured strictly in terms of first-order effects to the supply of physicians, a salary raise may be more effective, in the sense that it produces immediate effects.

These first two policies are clearly tailored for the short-term, but are unable to properly address the problem if there is a structural shortage in the supply of physicians due to barriers to entry. One such case is *numerus clausus*, effectively putting a limit on the number of students that can enrol in medical school. We analyzed the impact of increasing *numerus clausus* as an alternative or as a complement to raising wages or increasing job offers. In the first case we increase the number of entries to equal the number of physicians no longer emigrating as a result of the wage increase, while in the second scenario we increase it to the point it completely offsets (ex ante) the number of physicians emigrating. As a result, the workforce size expands by 5.73% and 1.67%, respectively. The policy is effective at increasing the workforce size, but unable to mitigate the emigration rate, which increases. Considering that the policy-maker's primary goal is to ensure the workforce is balanced, such side effect is not necessarily negative. Otherwise, it would actually defeat the purpose, considering that, according to the results, the more the *numerus clausus* are increased, the more likely the emigration rate is set to increase[3].

This type of policy is considerably cheaper for the taxpayer's purse. On average, a medical student costs approximately 13 thousand Euros in tuition fees per year, with roughly 1 thousand being paid out-of-pocket. Increasing entries to medical school by 125

[3] One note is due: we do not account for students that are unable to enter medical college as emigrants. Anecdotal evidence suggests that a non-negligible number of students end up moving to another country so as to pursue their training in medicine. Accounting for these students could also provide a different, brighter picture regarding emigration numbers. However, no official records exists.

would then cost to the public purse some additional 1.5 million Euros per year. We are assuming universities are operating below their full potential, and therefore room exists to accommodate for an increase in the number of students without having to invest and expand fixed costs (i.e. build additional rooms, etc.). Under this scenario it is significantly cheaper to increase *numerus clausus*, although welfare effects may also arise, as noted previously. However, if this policy is not accompanied by an increase in the number of positions to the public sector, slack may be created in the physician workforce, which may ultimately lead to unemployment. This would aggravate both the social and the economic cost.

Wage elasticitities

Running a sensitivity analysis and varying the amount by which the salary is raised allows us to run a linear regression to estimate the elasticity coefficients of physician emigration with respect to wage. In our real-case example, we obtain an emigration elasticity of approximately -5 with respect to wage, implying that a marginal wage raise of 1% would decrease emigration by 5%. Such elasticity is extremely high compared to the value obtained by Okeke (2014) on his applied study to Ghana, where the author finds elasticities between -0.06 and -0.13. Unfortunately, to the best of our knowledge no further empirical studies exploiting the relationship between physician wages and emigration exist. Furthermore, when looking in detail at the results of the simulated migration market, we see that half of the physicians that move abroad do so for a wage differential of less than 30% (Figure 5a). This number is apparently low, as one would expect most of the migrants to emigrate for considerably higher wage differentials. This may suggest that other factors, such as being unemployed on unable to enrol in a specialty, play a bigger role than initially expected.

We raise three hypothesis to explain this apparent contradiction: (i) physicians are a very special case in the labour migration market; (ii) the utility functions and the *VS* curve are not calibrated properly; (iii) the number of emigrants reported by the institutions are not correct, which would induce the parameters to be erroneously calibrated. Note that there are no public records regarding the number of physicians that actually emigrate, only of those that request a medical practice certificate so they can practice abroad. Such request does not imply that they actually move abroad.

This is a limitation of the present study which we cannot address as of now, since no data exists to calibrate either the utility functions of physicians or the job offers they have received in the past. An empirical estimation of the physician migration market would be a very interesting addition to this work. We hope future research develops further and extends the model hereby presented, and can become a mainstream tool for conducting policy analysis.

6 Conclusion

This paper presents a simple yet powerful economic model for studying physician emigration, its impact on the supply of physicians and overall welfare implications. Using simulation we established grounds for the case that emigration may be a risk to a balanced physician workforce, potentially affecting the capacity of the health system to deliver health care services to a needing population. If that is the case, it is important to have a theoretical framework policy-makers can use to estimate the impact of public policies on any market, regardless of its idiosyncrasies.

The most innovative contribution of this work rests on the model being capable of predicting, in quantitative terms, the effects different policies exert on the supply of physicians and on the physician migration in a real-case setting that drifts too far from typical Walrasian conditions. The health system under study has too many rigidities, barriers to entry and regulations to be accurately represented through the lens of general equilibrium, and therefore we cannot rely on an analytical representation of the system. Hence, the model was implemented using agent-based simulation and calibrated using data from the Portuguese health system. ACE and ABM in particular are very powerful tools for implementing microeconomic foundations and then observing the impact on macro aggregates.

We illustrate the effect of typical short- and long-term retention policies on the Portuguese physician workforce. Using the model conceived we have shown that all policies are capable of expanding the medical workforce, in this way mitigating the effects of emigration. However, they incur widely different effects and costs. Salary raises and increased opportunities for joining the public sector were shown to increase the retention rate (with the emigration rate decreasing *pari passu*) and the size of the workforce, while raising the *numerus clausus* increased emigration, but this effect is partially offset in our results, as the policy also contributes to the growth of the medical workforce. If the policy-makers' top priority is to rebalance the workforce, an increase in the emigration rate may be an acceptable side-effect. Otherwise, policies that increase the workforce and decrease the migration rate would be more adequate. Such policies may present a significant cost to the public purse, though.

References

Amorim Lopes, M., Almeida, Á.S., and Almada-Lobo, B. (2015). Handling healthcare workforce planning with care: where do we stand? *Human Resources for Health*, 13(1), 38.

Amorim Lopes, M., Almeida, Á.S., and Almada-Lobo, B. (2016). Forecasting the medical workforce: a stochastic agent-based simulation approach. *Health Care Management Science*, 1–24.

Bidwell, P., Humphries, N., Dicker, P., Thomas, S., Normand, C., and Brugha, R. (2013). The national and international implications of a decade of doctor migration in the Irish context. *Health Policy*, 110(1), 29–38.

Bodvarsson, O.B. and Van den Berg, H. (2013). *The Economics of Immigration.* Theory and Policy.

Borjas, G.J. (2001). Does Immigration Grease the Wheels of the Labor Market? *Brookings Papers on Economic Activity*, 32(1), 69–134.

Bouchaud, J.P. (2008). Economics need a scientific revolution. *Nature*, 455(7217), 1181–1181.

Clark, P.F., Stewart, J.B., and Clark, D.A. (2006). The globalization of the labour market for health-care professionals. *International Labour Review*, 145(1-2), 37–64.

Colander, D., Howitt, P., Kirman, A., Leijonhufvud, A., and Mehrling, P. (2008). Beyond DSGE Models: Toward an Empirically Based Macroeconomics. *The American Economic Review*, 98(2), 236–240.

DaVanzo, J. (1980). Microeconomic Approaches to Studying Migration Decisions.

Dosi, G., Fagiolo, G., and Roventini, A. (2010). Schumpeter meeting Keynes: A policy-friendly model of endogenous growth and business cycles. *Journal of Economic Dynamics & Control*, 34(9), 1748–1767.

Fagiolo, G. and Roventini, A. (2012). Macroeconomic Policy in DSGE and Agent-Based Models. *EconomiX Working Papers.*

Farmer, J.D. and Foley, D. (2009). The economy needs agent-based modelling. *Nature*, 460(7256), 685–686.

Gauld, R. and Horsburgh, S. (2015). What motivates doctors to leave the UK NHS for a “life in the sun” in New Zealand; and, once there, why don’t they stay? *Human Resources for Health*, 13(1), 75.

Hussey, P.S. (2007). International migration patterns of physicians to the United States: a cross-national panel analysis. *Health Policy*, 84(2-3), 298–307.

Kennan, J. and Walker, J.R. (2013). Modeling individual migration decisions. 39–54.

Labonté, R., Sanders, D., Mathole, T., Crush, J., Chikanda, A., Dambisya, Y., Runnels, V., Packer, C., MacKenzie, A., Murphy, G.T., and Bourgeault, I.L. (2015). Health worker migration from South Africa: causes, consequences and policy responses. *Human Resources for Health*, 13(1), 92.

Maloney, W.F. and Gonzalez, P.A. (2005). Migration, Trade, and Foreign Direct Investment in Mexico.

Martins, M.J. (2015). Satisfação com a Especialidade entre os Internos da Formação Específica em Portugal. *Acta Médica Portuguesa*, 1–14.

McPake, B., Maeda, A., Araújo, E.C., Lemiere, C., El Maghraby, A., and Cometto, G. (2013). Why do health labour market forces matter? *Bulletin of the World Health Organization*, 91(11), 841–846.

Mortensen, D.T. and Pissarides, C.A. (1999). New developments in models of search in the labor market. *Handbook of Labor Economics*, 3, Part B, 2567–2627.

Neugart, M. and Richiardi, M.G. (2012). Agent-based models of the labor market. *LABORatorio R. Revelli Working Papers Series*.

OECD (2013). OECD Employment and Labour Market Statistics. OECD Publishing.

Okeke, E.N. (2014). Do higher salaries lower physician migration? *Health Policy and Planning*, 29(5), 603–614.

Ono, T., Schoenstein, M., and Buchan, J. (2014). Geographic Imbalances in Doctor Supply and Policy Responses. 69.

Ono, T., Schoenstein, M., and Lafortune, G. (2013). Health Workforce Planning in OECD Countries. Technical report.

Ortega, F. and Peri, G. (2009). The Causes and Effects of International Migrations: Evidence from OECD Countries 1980-2005.

Ribeiro, J.S., Conceição, C., Pereira, J., Leone, C., Mendonça, P., Temido, M., Vieira, C.P., and Dussault, G. (2014). Health professionals moving to… and from Portugal. *Health Policy*, 114(2-3), 97–108.

Vujicic, M., Zurn, P., Diallo, K., Adams, O., and Poz, M. (2004). PMC419378. *Human Resources for Health*, 2(1), 3.

Chapter 5

Comparing comparables: an approach to accurate cross-country comparisons of health systems for effective healthcare planning and policy guidance[1]

[1]Published in the Health Systems journal with DOI 10.1057/hs.2015.21. URL http://link.springer.com/article/10.1057/hs.2015.21

Health Systems (2016) 5, 192–212

www.palgrave.com/journals

ORIGINAL ARTICLE

Comparing comparables: an approach to accurate cross-country comparisons of health systems for effective healthcare planning and policy guidance

Mário Amorim Lopes[1],
Carlos Soares[2],
Álvaro Almeida[1] and
Bernardo Almada-Lobo[1]

[1]Faculdade de Engenharia da Universidade do Porto, Porto, Portugal; [2]Faculdade de Economia da Universidade do Porto, Porto, Portugal

Correspondence: Mário Amorim Lopes, INESC-TEC Porto; Faculdade de Engenharia da Universidade do Porto, Rua Roberto Frias, Porto, 4100, Portugal
Tel: +351914142332;
E-mail: mario.lopes@fe.up.pt

Abstract

With rising healthcare costs, using health personnel and resources efficiently and effectively is critical. International cross-country and simple worker-to-population ratio comparisons are frequently used for improving the efficiency of health systems, planning of health human resources and guiding policy changes. These comparisons are made between countries typically of the same continental region. However, if used imprudently, inconsistencies arising from frail comparisons of health systems may outweigh the positive benefits brought by new policy insights. In this work, we propose a different approach to international health system comparisons. We present a methodology to group similar countries in terms of mortality, morbidity, utilisation levels, and human and physical resources, which are all factors that influence health gains. Instead of constructing an absolute rank or comparing against the average, the method finds countries that share similar ground, upon which more reliable comparisons can then be conducted, including performance analysis. We apply this methodology using data from the World Health Organization's Health for All database, and we present some interesting empirical relationships between indicators that may provide new insights into how such information can be used to promote better healthcare planning and policy guidance.

Health Systems (2016) **5,** 192–212. doi:10.1057/hs.2015.21;
published online 05 February 2016

Keywords: health systems; comparative analysis; data mining; clustering analysis; health policy

Introduction

Healthcare costs have increased sharply over the years, well above the average growth rate of the GDP (Chandra *et al*, 2013). The immediate impact is a decrease in the real income of the population, reducing the disposable income available for other living expenses (Auerbach & Kellermann, 2011). Wealthier countries may be able to accommodate such increases but others may be confronted with the sensitive situation of having to opt between drugs and treatments based on their cost, and not on their clinical merits alone. In the worst-case scenario, no treatment is offered at all, or rationing is imposed through waiting lists. In order to avoid facing such dilemmas, policy-makers must strive to manage healthcare resources efficiently, both physical and human.

Health human resources (HHR) planning has been identified as a fundamental tool for mitigating rampant healthcare costs while preserving the

Received: 21 June 2015
Revised: 18 November 2015
Accepted: 6 December 2015

quantity and quality of service provided (Dreesch, 2005). Briefly defined, HHR planning consists in assessing the right number of people with the right skills, in the right place at the right time, to provide the right services to the right people (Birch, 2002). There are multiple approaches to HHR planning, each with its advantages and drawbacks (Amorim Lopes *et al*, 2015). Benchmarking, simple worker-to-population comparisons and other comparative analysis techniques are approaches frequently used to draw international comparisons between healthcare systems, including assessing HHR needs. The techniques consist of identifying similar regions or countries in terms of demographic and epidemiological profiles but differing sharply in the cost structure and resource allocation (Roberfroid *et al*, 2009).

Although benchmarking and other comparative analyses may be useful tools to assess and compare health systems, including HHR resources, imprudent use may take its toll. For example, consider the decision by the British National Health Service to increase the intake to medical schools by 60%, a resolution motivated by the observation that the physician-to-population ratio was low in comparison with other OECD countries (Bloor *et al*, 2006). The policy was adopted without first evaluating for other criteria that may affect the performance of the medical staff, namely the skill mix and the productivity of HHR. Subsequent research justified the lower ratio of physicians with a better distribution of the skill mix and increased productivity resulting from a more efficient task delegation (Bloor & Maynard, 2003). Health system performance comparisons and composite indices have also been subject to heavy criticism, in part due to some methodological fragilities (Richardson *et al*, 2003).

Provided that a valid methodology is used to support cross-country comparisons, comparative analyses may continue to be an easy and straightforward way of gaining quick insight into the health system performance of a country by observing the best practices, especially when few data exist to conduct more advanced analyses (Castelli *et al*, 2011). There are performance profiles and typological classifications of health systems in the literature that identify countries with similar systems (for a review of typologies of health systems, see Burau & Blank, 2006). Previous studies have mapped European healthcare systems according to a subset of indicators on healthcare expenditure, healthcare financing, healthcare provision and institutional characteristics, and then establishing a relative performance index between groups (Wendt, 2009). In other studies, several other dimensions, such as acquisition of human, financial, technological or material resources, health outcomes, risk factors or equity of access to healthcare were used to elaborate an absolute performance index, which is then used to group countries according to their performance profiles (Tchouaket *et al*, 2012). International comparisons have also been conducted using non-parametric techniques like Data Envelopment Analysis (DEA) (Bhat, 2005), techniques typically used for micro-level service efficiency measurements, such as hospital units. DEA measures the efficiency of health systems by calculating the ratio between health outcomes and healthcare spending.

In this work, we propose a new methodology to perform cross-country comparisons. As an alternative to constructing composite indices or absolute performance rankings, we start by creating clusters of countries that have similar results in several reference indicators, including those usually associated with demand for healthcare services (mortality and morbidity-based indicators, and utilisation statistics), and with the supply of healthcare services (physical and human resources available). This allows for intra- and inter-group local comparisons, avoiding attractive and yet inconclusive global performance rankings of substantially different health systems that have generated discord (Richardson *et al*, 2003), if not outright criticism (Bronnum-Hansen, 2014).

We then apply the method to data from the World Health Organization's (WHO) Health for All database (HFA-DB). Clusters are generated for each indicator and then intersected against each other to obtain groups of countries with similar features in more than one dimension. With this bottom-up methodology, future research using benchmarking or DEA can build upon a reliable basis of countries exhibiting similar trends in parts of their health systems. These indicators can then be used within HHR planning models, or any other area of healthcare planning and health policy, to improve the forecasts and projections to assist decision and policy-makers.

The remainder of this paper is organised as follows: in the next section we describe the methodology proposed, and after that we apply that methodology to WHO's HFA-DB database and present the results. A discussion of the results and some empirical insights are provided in the subsequent section. Finally, the last section concludes this paper with a brief summary and future research topics.

Methodology

Our methodology consists of grouping countries that are similar to each other in different dimensions of a health system. In this particular case, we consider the following: mortality-based indicators, morbidity-based indicators, utilisation indicators, physical resources and human resources. We then analyse which countries feature in the same cluster in more than one dimension, which makes it possible to generate a similarity matrix. Comparisons and cross-country performance analyses can then be conducted within the cluster, comparing countries with the local benchmark serving as a reference, or between clusters. By deliberately narrowing the scope, we ensure that cross-country comparisons are performed between countries with similar characteristics.

We use a sample of the WHO's HFA-DB, targeting countries belonging to the European Union, as European countries share common ground and have health systems that derive from either Bismarckian (i.e. social health insurance) or Beveredgian (i.e. publicly funded national health service) models (Wendt *et al*, 2009). Notwithstanding, this study can be applied interchangeably to any set of countries provided that data are available.

Data sources

The main source of data was the HFA-DB, last updated in April 2014. This database contains a selection of core health statistics covering basic demographics, health status, health determinants and risk factors, healthcare resources, utilisation and expenditure in the 53 countries in the WHO European Region (Europe W.R.O.F.). The data are compiled from various sources, including WHO's and United Nations' European delegation offices, OECD and Eurostat.

Selection of indicators

For this study, we have selected indicators that characterise supply of and demand for healthcare services. Also, since utilisation indicators may reflect both the delivery of healthcare services (supply) and the demand for healthcare services, we have a separate category with indicators of healthcare utilisation. These indicators are also typically used to assess input/output and outcome in efficiency studies. On the demand side, we have selected mortality-based, morbidity-based and healthcare utilisation indicators. These indicators are sometimes also used as indicators to measure output efficiency (Varabyova & Schreyögg, 2013). For the supply side, we resorted to healthcare resources, both physical and human, also commonly used as proxies to measure input efficiency. These indicators were selected due to the availability of data, and also because they share a list of desirable features: valid, communicable, effective, reliable, objective, available, contextual, attributable, interpretable, comparable, remediable and repeatable (Pringle *et al*, 2002). This is critical for any subsequent expert validation of the clusters formed.

On the demand-side, mortality-based indicators are provided as Standardised Death Rates (SDRs) by group of disease defined by International Classification of Diseases codes. Morbidity-based indicators describe the general health status of a population and incidence of diseases. Due to the significant amount of unreported data regarding these indicators, we resorted to hospital discharges by disease type as a proxy to health status (incidence and prevalence data were incomplete for a large number of countries). Finally, we also include the available sample of healthcare utilisation statistics, measured in terms of bed occupancy rate, inpatient care discharges and average length of stay.

On the supply-side, the selection includes indicators that describe both the physical and human resources available. Physical resources are both the hospitals and their capacity, measured in terms of number of beds available. HHR are accounted for by the number of health professionals (we consider only physicians, nurses, dentists and midwives).

Data treatment, standardisation and aggregation

The HFA-DB reports values to 2013. In some instances, data from 2013 were unavailable. In such cases, one of three procedures was followed: (1) if a clear linear trend could be found, a linear regression was run to estimate the missing value; (2) if no trend line exists, data from previous years were used up to 3 years back; (3) if a considerable amount of countries did not report the data, the indicator was excluded from the analysis.

Thereafter, we run a multicollinearity bivariate analysis to identify correlated indicators. If no action is taken to fix multicollinearity, an overrepresentation of a particular dimension may occur, thereby potentially biasing the cluster formation (Ketchen & Shook, 1996). The literature reports no unique upper bound, but rather a range of satisfactory values for the correlation coefficient (Hair, 2013). Since we do not have a large enough sample to accurately estimate the correlation coefficients, we define the cut-off point for indicators as those exhibiting a correlation coefficient equal to or greater than 0.85. Statistical significance was set to a global level of $P \leq 0.05$ (2-tailed distribution). To select the indicator to be retained, we take one of two possible actions: if one indicator encompasses the other, we select the most complete one; otherwise, we select the indicator that best differentiates countries by choosing the one with the highest coefficient of variation.

Data standardisation was applied on a case-by-case basis. In some instances, maintaining the absolute difference between indicators was intentional and useful. For example, it is relevant to retain the difference between the numbers of SDRs caused by different diseases, as the impact on the healthcare system will be notably different (although it may not be in a linear way as different diseases put different levels of stress on the system). In contrast, indicators using different measurement units require scaling to remove the effect of different scales. For instance, the number of hospitals and the number of hospital beds cannot be compared directly. To remove this effect a scale change is used.

It should be noted that the distortion caused by the age composition of a given population is already removed from indicators measured in SDRs. For this reason, no further action is required to standardise data.

Finally, and whenever appropriate, data were aggregated by summing the indicators. For instance, the SDRs caused by each group of diseases were summed to a grand total. Similarly, the number of hospital dispatches by group of diseases was also added up. This allows for a direct comparison between mortality and morbidity levels between groups of countries. The methodology adopted is summarised in Figure 1.

Clustering algorithm

We employ a two-stage clustering algorithm to group similar countries in each indicator. The first stage is exploratory. We apply an agglomerative Hierarchical

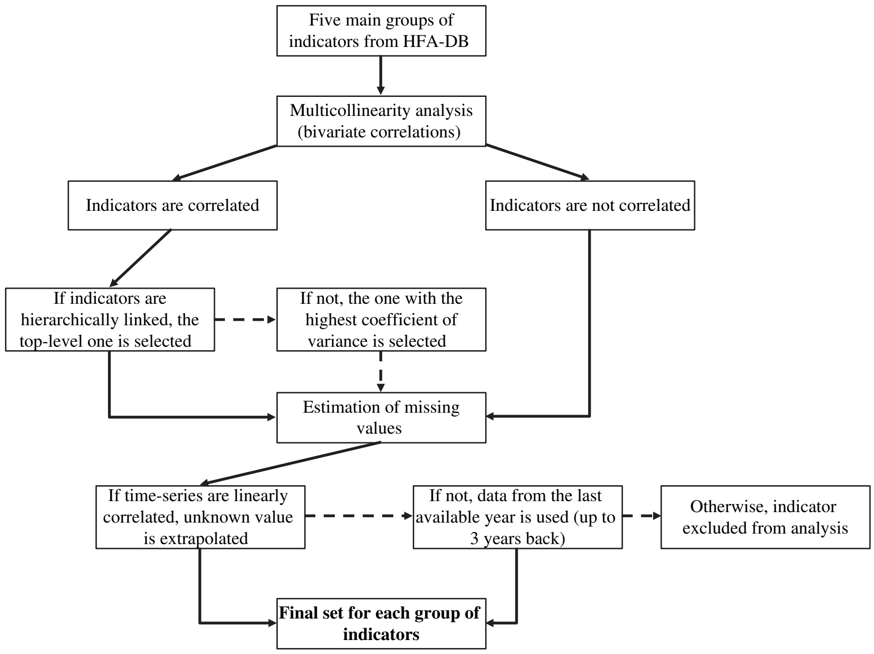


Figure 1 Methodology adopted for obtaining the final selection of indicators to be used for clustering countries.

Clustering Algorithm (HCA) with Ward's method (Everitt *et al*, 2001) with a squared Euclidean distance to analyse possible cluster formations and then decide on the number of clusters. The second stage is explanatory. We use the cluster centres defined previously using HCA with Ward's method, and run a *k*-means algorithm to obtain descriptive statistics about the clusters, including an ANOVA table detailing which indicators were more relevant in grouping the countries.

In more detail, the first stage is an iterative process. Initially, each country is part of one single cluster. An agglomerative HCA procedure is then applied for merging clusters. Very similar clusters are combined. Similarity is measured in terms of distance, and there are several ways of calculating this distance. Single linkage calculates the shortest distance between any two members in the two clusters. Complete linkage looks for the longest distance between any two members. Average linkage and centroid are also common approaches, but since the dataset does not include outliers, we resort to Ward's method. In this procedure, the single clusters are merged if the merger results in the minimum merging cost, with merging cost being defined as the increase in the sum of squared errors. This procedure favours the formation of a cluster with very similar members. The result of this process is a distance matrix between clusters that can be used to draw a dendrogram. Dendrograms are tree diagrams used to illustrate the arrangement of clusters. The size of the branches is a meaningful measure, depicting the distance between clusters, and the further apart, the greater the heterogeneity between clusters. A decisive step in the clustering process is defining the *right* number of clusters. Again, the dendrogram provides relevant information to assist in this decision. Optionally, a scree plot can be used, which is in fact another way of representing the same information, exposing the distance within clusters as a new cluster is added. We are interested in a distinct break (elbow), after which the creation of one additional cluster does not create a significant distinction.

The second stage consists of running a *k*-means partitioning method based on the number of clusters and cluster centres previously obtained. *k*-means is a non-hierarchical procedure that does not require calculating distances. The aim of the procedure is to obtain descriptive statistics to assist in explaining the cluster formations. *k*-means will form *k* clusters using the cluster centres defined previously using the scree plot, by finding the point where the inclusion of an additional cluster does not significantly increase heterogeneity. This procedure will then originate the final cluster formations, and also an ANOVA table with *F*-tests for each indicator.

Validation and interpretation

The final step consists of validating and interpreting the clusters. External validation consists of comparing the clusters with the true partition, which in this case cannot be known *a priori*. Internal validation, on the other hand, deals with the intrinsic properties of the dataset regardless of any external information, and is captured through the descriptive statistical information provided by *k*-means. To make this process more robust, we also visually inspect the correlations of the two most significant indicators.

To interpret the results we need to understand what binds members together but also what separates them. To do so, we examine the cluster centroids, which are the clustering variables' average values for all countries in a given cluster. To help understand which indicators maximise intra-cluster similarity, we resort to a one-way ANOVA table that calculates *F*-tests for each variable. We want to test whether the clustering variables' means differ significantly across at least two of the *k* segments (where *k* is the number of clusters selected). Unless the null hypothesis is rejected, the indicator was relevant in the cluster formation. Besides validating the procedure, this is also a way of understanding the results obtained.

Results

In this section we apply the two-stage clustering algorithm to the five groups of indicators, namely: mortality-based, morbidity-based, utilisation, physical and human resources indicators. With the clusters formed based on each set of indicators, we validate and analyse the results obtained. Each step of the methodology followed is described thoroughly.

Demand-based indicators

Demand for healthcare services can be assessed in two conceptually different ways (Castelli *et al*, 2011). *Effective demand* measures health care delivered. Utilisation indicators are commonly used as proxies to estimate effective demand. Alternatively, medical needs are assessed based on the epidemiological conditions of the population, and then translated into a given quantity of healthcare services necessary to meet those needs (for an example see Harper *et al*, 2010). Since not all the needs may turn into actual demand due to financial constraints or lack of hospital room capacity, this is also referred to as *potential demand*. These concepts need not to be used separately, and can actually be combined. Utilisation indicators provide current usage levels, whereas health and disease patterns of the population may reflect unmet or future care needs. This is especially relevant in countries with extensive waiting lists, where healthcare services are being delayed due to lack of capacity.

HFA-DB provides both indicators for measuring potential demand (morbidity-based indicators) and for measuring effective demand (mortality-based and utilisation indicators). Our analysis targets both types of indicators. Note that these indicators are also commonly used to measure the level of output of health systems (Varabyova & Schreyögg, 2013).

Mortality

Selection of indicators Concerning mortality-based indicators, HFA-DB contains both actual death rates per age cohort, and SDRs that cross-reference the medical cause of death. We have focused exclusively on the disease-specific mortality indicators rather than on the crude death statistics, as the former carry more explanatory power. There were no missing values in the data collected, therefore no estimation technique had to be applied.

Data treatment, standardisation and aggregation Regarding collinearity, and considering that the list of mortality-based indicators contains both groups and subgroups of diseases, and groups of diseases aggregate the data of subgroups, it follows that indicators belonging to the same taxonomical group will naturally exhibit a strong correlation. For instance, the indicator that reports deaths caused by diabetes, 'SDR, diabetes', will most probably correlate to the indicator that encompasses this type of disease, 'SDR, endocrine, nutritional and metabolic diseases'. Likewise, 'SDR, cancer of the cervix uteri' will probably correlate to with a general indicator on the incidence of cancer, 'SDR, malignant neoplasms'. We have addressed this potential source of multicollinearity by choosing only the top-level indicators that represent the groups and already account for the subgroups.

Notwithstanding this case, multicollinearity may still occur. To mitigate this, we also ran a bivariate analysis on the pre-selected list of SDRs. As expected, risk-factor indicators such as 'SDR, selected alcohol-related causes' and 'SDR, selected smoking-related causes' are highly correlated to other SDRs, thus were removed. For the remaining variables, and given that a degree of correlation between diseases is in fact a standard medical occurrence (Jeon *et al*, 2005; Li *et al*, 2014), we set the absolute correlation threshold to a value close to the upper bound typically found in the literature (Dormann *et al*, 2013). For this case, none were removed. After applying the aforementioned procedure we obtain the list of indicators shown in Table 4.

Since SDRs are already partly standardised by age distribution and by one hundred thousand people, no standardisation techniques had to be applied. Moreover, scaling is also not applied, since the SDR indicators all report the number of deaths, and absolute differences between indicators are relevant, representing the relative impact each disease may put on the health system.

Clustering algorithm Next, we apply HCA using Ward's method. In Figure A1 (Appendix) we present the resulting dendrogram. In this particular case, it is immediate to see a notable separation between two large groups. One side features Western Europe countries; the other features countries from Eastern Europe, and the data are consistent with well-known economic differences. On average, Eastern European countries perform worse in all mortality indicators.

Of the two initial groups, further separation can be found one level down the tree. Although not as significant as before (note the horizontal distance between the branches), it is still relevant. The two major clusters composed of Eastern European and Western European countries exhibit the largest distance. Nevertheless, given the high level of heterogeneity within these two clusters, a better grouping can be obtained if we further split into two more clusters, thus increasing the homogeneity within the cluster and increasing the distance between clusters. Henceforth, introducing additional clusters marginally increases the distance, implying that most heterogeneity has already been explored. The scree plot, represented in Figure A2 (Appendix), reinforces this indication.

Table 1 Cluster compositions obtained for each of the group of indicators

Indicators	*Cluster*						
	1	*2*	*3*	*4*	*5*	*6*	*7*
Mortality-based	Austria, Finland, Germany, Greece, Malta, Slovenia, Sweden	Croatia, Czech Republic, Estonia, Poland	Belgium, Denmark, France, Ireland, Italy, Luxembourg, the Netherlands, Norway, Portugal, Spain, Switzerland, United Kingdom	Hungary, Latvia, Lithuania, Slovakia, Romania			
Morbidity-based	Austria, Germany	Belgium, Luxembourg, Norway, Slovenia, Sweden, Switzerland	Hungary, Lithuania, Romania	Croatia, Ireland, Italy, Malta, the Netherlands, Portugal, Spain, United Kingdom	Czech Republic, Denmark, Estonia, Finland, Greece, Latvia, Poland, Slovakia	France	
Utilisation	Austria, Germany, Lithuania	Belgium, Croatia, Estonia, Latvia, Luxembourg, Slovenia	Czech Republic, Hungary, Slovakia	Denmark, Greece, Ireland, Norway, Sweden, Switzerland, United Kingdom	Finland, France	Italy, Portugal, Spain	The Netherlands
Physical resources	Austria, Czech Republic, Greece, Hungary, Luxembourg, Poland, Romania, Slovakia	Belgium, Croatia, Malta, the Netherlands, Slovenia	Bulgaria, Finland, France, Germany, Lithuania	Estonia, Latvia, Switzerland	Italy, Norway, Portugal, Spain, Ireland, Sweden		
Human resources	Austria, Germany, Lithuania, Italy, Portugal	Luxembourg, Belgium, Finland, Norway	Czech Republic, Croatia, Bulgaria, France, Estonia	Denmark, Switzerland	Hungary, Romania, the Netherlands, Slovenia, Latvia, Spain	Poland, Slovakia, Malta, United Kingdom	Greece

With a pre-established number of clusters, we can now apply *k*-means clustering with the centroids (cluster centres) obtained with the HCA, and use the ANOVA statistical information to understand which indicators were most relevant in maximising intra-group similarity while minimising inter-group similarity. Running *k*-means for $k = 4$, we obtain four final clusters (Tables 1 and 2).

Validation and interpretation Table 5 provides detailed information on the statistical significance tests for each variable. With a significance level below 0.001, 'SDR, diseases of circulatory systems' had the largest impact on the definition of cluster centroids, followed by 'SDR, diseases of digestive system', and 'SDR, malignant neoplasms'. With a smaller significance but still below the significance threshold, 'SDR, motor vehicle traffic accidents' and 'SDR, mental disorders, diseases of nervous system and sense organs' were also relevant when generating the clusters (Table 3).

After determining the indicators that most contributed to form the cluster centres, it is possible to obtain several indicators in a single scatter plot, such as the most relevant SDRs, clusters and countries. Figure 2 is then useful to understand with visual guidance how the clusters were formed. Distance between points reflects intra-member dissimilarity, which is common to most clustering algorithms, including Ward's. In fact, the more homogeneous the members are, the most likely is the chance of them featuring together in a single cluster.

The separation between the two main groups featuring Western and Eastern European countries, quite evident in the dendrogram of Figure A1 (Appendix), is then explained by 'SDR, diseases of circulatory system', for which the absolute difference is evident. The other indicators were subsequently used to extract further heterogeneity within these two main groups. In fact, Eastern European countries feature, on average, twice the mean value in the number of deaths caused by this group of diseases. These countries are also the worst performers in the number of deaths due to diseases of the digestive system and to malignant neoplasms. In contrast, the Southern and some Nordic countries in cluster 3 have the best performance, registering a considerably lower number of deaths, in some cases half the total average.

Finally, we resort to a simple visual tool for quickly inspecting the quality and validity of the results obtained. We plot the two most significant indicators against each other and use the cluster cases as labels. Heterogeneous (badly formed) clusters are featured in a dispersed,

Table 2 Cluster and country relative status by group of indicators

Relative Status	*Demand*		*Utilisation*	*Supply*	
	Mortality-based indicators	*Morbidity-based indicators*		*Human resources*	*Physical resources*
Low	1 Austria, Finland, Germany, Greece, Malta, Slovenia, Sweden	2 Belgium, Luxembourg, Norway, Slovenia, Sweden, Switzerland	2 Belgium, Croatia, Estonia, Latvia, Luxembourg, Slovenia	3 Czech Republic, Croatia, Bulgaria, France, Estonia	2 Belgium, Croatia, Malta, the Netherlands, Slovenia
	3 Belgium, Denmark, France, Ireland, Italy, Luxembourg, the Netherlands, Norway, Portugal, Spain, Switzerland, United Kingdom	4 Croatia, Ireland, Italy, Malta, the Netherlands, Portugal, Spain, United Kingdom	3 Czech Republic, Hungary, Slovakia	5 Hungary, Romania, the Netherlands, Slovenia, Latvia, Spain	5 Italy, Norway, Portugal, Spain, Ireland, Sweden
			5 Finland, France	6 Poland, Slovakia, Malta, United Kingdom	
			7 the Netherlands	7 Greece	
High	2 Croatia, Czech Republic, Estonia, Poland	1 Austria, Germany	1 Austria, Germany, Lithuania	1 Austria, Germany, Lithuania, Italy, Portugal	1 Austria, Czech Republic, Greece, Hungary, Luxembourg, Poland, Romania, Slovakia
	4 Hungary, Latvia, Lithuania, Slovakia, Romania	3 Hungary, Lithuania, Romania	4 Denmark, Greece, Ireland, Norway, Sweden, Switzerland, United Kingdom	2 Luxembourg, Belgium, Finland, Norway	3 Bulgaria, Finland, France, Germany, Lithuania
		5 Czech Republic, Denmark, Estonia, Finland, Greece, Latvia, Poland, Slovakia	6 Italy, Portugal, Spain	4 Denmark, Switzerland	4 Estonia, Latvia, Switzerland
		6 France			

Table 3 Clusters formed by the intersection of the clusters obtained using mortality- and morbidity-based indicators

Super cluster	*Countries*
1	Austria, Germany
2	Finland, Greece
3	Estonia, Poland
4	Hungary, Lithuania, Romania
5	Ireland, Italy, the Netherlands, Portugal, Spain, and United Kingdom
6	Latvia, Slovakia

uncorrelated way. Homogeneous groups, on the other hand, can be clearly identified. Figure 3 depicts this.

Morbidity

Selection of indicators The HFA-DB contains three distinct types of morbidity indicators: incidence of diseases, prevalence of diseases and hospital discharges per disease. Each group of indicators portrays different information. Since a lot of missing blanks exist for incidence and prevalence of diseases, we will only consider the hospital discharges as this type of indicator is commonly used to measure system output (Varabyova & Schreyögg, 2013). The list of hospital discharge indicators can be found in Table 4. Similar to the case of mortality-based indicators, only groups of diseases were considered.

Data treatment, standardisation and aggregation In this particular case, it was verified that some indicators for 2013 were missing, which needs to be handled adequately, since cluster analysis cannot be conducted if data are missing. We have decided to use the last available year as not enough data were available to run a statistically significant regression.

Subsequently, the process followed to obtain the clusters using the morbidity-based indicators is very similar to the one previously applied to mortality rates. A multicollinearity analysis needs to be conducted in order to identify correlations between indicators that could potentially boost the importance of a particular category of indicators, thereby biasing the cluster formation. None have surpassed the threshold of 0.85 below the significance threshold, and therefore no indicators have been removed. As before, no standardisation or transformation technique will be applied as hospital discharges by disease are

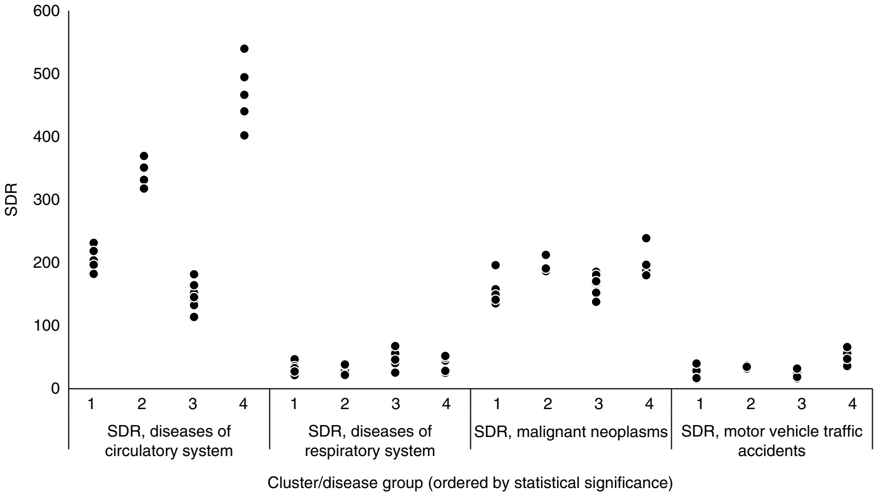


Figure 2 A scatter plot depicting the most significant SDRs used to form the clusters along with the distances between indicators for each country.

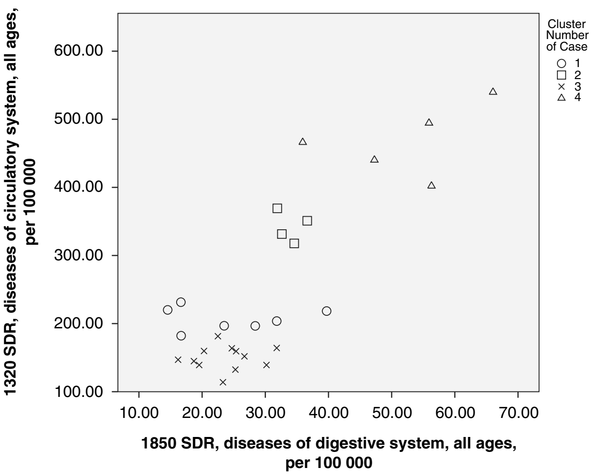


Figure 3 Scatter plot depicting the correlation between the two most significant indicators. Clusters can be clearly identified, suggesting an acceptable partitioning.

reported in ratios of 100,000 people, and the idea is to preserve the relative differences between diseases.

Clustering algorithm After running an HCA on the morbidity indicators we obtain a dendrogram to assist with the cluster formation. The tree obtained is more complex to interpret than the previous one (see Figures A2 and A3 in Appendix), suggesting that more indicators were used to calculate dissimilarity distances within and between clusters. First, there is a clear separation between the first group of eight countries and all the remaining ones. A quick look at the data reveals that these countries have, on average, the lowest number of hospital discharges registered. All the remaining groups exhibit higher hospital discharges, with Austria and Germany leading the chart with the highest number. The further expansion of the nodes suggests that further differences can be found within countries.

Before going deeper into the analysis, we again resort to a scree plot for determining the cutting point and defining the number of clusters (Figure A4). Contrarily to the previous case, there is no sharp elbow, but it is noticeable that after the sixth cluster the distance that can be increased is significantly reduced. We therefore define the number of clusters to six. It should be kept in mind that there is a trade-off in the aggregation process, where the distance is at its lowest value only when each country belongs to a single and unique cluster, but defeating altogether the purpose of the aggregation.

After fixing the number of clusters to six, we run the *k*-means algorithm to obtain a detailed description of the most relevant indicators to form the clusters (Table 4). The first thing we notice is that the algorithm was unable to group France in any cluster, implying that its values are so unique that the country is better classified as an outlier. Second, a clear geographical segregation between Eastern and Western countries is not as evident as with the mortality-based indicators, although it exists to a certain degree.

Validation and interpretation The analysis of the *F*-tests helps clarify on the relative importance of each variable when forming the clusters (cf. Table 5). Contrarily to the clusters generated with the mortality-based indicators, all the variables were used to form the groups, as they all were statistically significant at a level below 0.001. With the highest *F*-test score, '2520 Hospital discharges, digestive system diseases' was the most important indicator, followed by hospital discharges related to diseases in the circulatory system. As expected, these indicators exert a large influence since they represent the largest amount of hospital discharges. In contrast, infectious and parasitic diseases put less stress on the healthcare system when measured only in terms of number of persons discharged, and so were less critical when generating the clusters. Note that treatments to diseases differ in the amount of

Table 4 List of pre-selected variables for each indicator, and variables excluded for failing to meet all of the inclusion criteria

Code	*Name*	*Target*	*Sample*
Mortality-based indicators			
Included			
1320	SDR, diseases of circulatory system	all ages	per 100k
1520	SDR, malignant neoplasms	all ages	per 100k
1740	SDR, motor vehicle traffic accidents	all ages	per 100k
1820	SDR, infectious and parasitic diseases	all ages	per 100k
1830	SDR, diseases of respiratory system	all ages	per 100k
1850	SDR, diseases of digestive system	all ages	per 100k
1870	SDR, endocrine, nutritional and metabolic diseases	all ages	per 100k
1900	SDR, mental disorders, diseases of nervous system and sense organs	all ages	per 100k
1910	SDR, disease of genito-urinary system	all ages	per 100k
1920	SDR, symptoms, signs and ill-defined conditions	all ages	per 100k
1960	SDR, acute respiratory infections, pneumonia and influenza	< 5 years	per 100k
Excluded			
1340	SDR, ischaemic heart disease	all ages	per 100k
1360	SDR, cerebrovascular diseases	all ages	per 100k
1540	SDR, trachea/bronchus/lung cancer	all ages	per 100k
1560	SDR, cancer of the cervix uteri	all ages	per 100k
1590	SDR, malignant neoplasm female breast	all ages	per 100k
1840	SDR, bronchitis/emphysema/asthma	all ages	per 100k
1860	SDR, chronic liver diseases and cirrhosis	all ages	per 100k
1880	SDR, diabetes	all ages	per 100k
1890	SDR, diseases of the blood, blood forming organs and certain immunity disorders	all ages	per 100k
1930	SDR, tuberculosis	all ages	per 100k
1940	SDR, diarrhoeal diseases	< 5 years	per 100k
1970	SDR, selected alcohol-related causes	all ages	per 100k
1980	SDR, selected smoking-related causes	all ages	per 100k
Morbidity-based indicators			
Included			
2300	Hospital discharges, infectious and parasitic diseases	all ages	per 100k
2310	Hospital discharges, all neoplasms	all ages	per 100k
2450	Hospital discharges, circulatory system diseases	all ages	per 100k
2500	Hospital discharges, respiratory system diseases	all ages	per 100k
2520	Hospital discharges, digestive system diseases	all ages	per 100k
2530	Hospital discharges, musculoskeletal system and connective tissue diseases	all ages	per 100k
2540	Hospital discharges, injury and poisoning	all ages	per 100k
Excluded			
2460	Hospital discharges, ischaemic heart disease	all ages	per 100k
2480	Hospital discharges, cerebrovascular diseases	all ages	per 100k
Utilisation indicators			
Included			
6010	Inpatient care discharges	all ages	per 100
6100	Average length of stay, all hospitals	all ages	per 100k
6210	Bed occupancy rate (%), acute care hospitals only	all ages	% of total
6300	Outpatient contacts	all ages	per person per year
Excluded			
6020	Acute care hospital discharges	all ages	per 100
Physical resources indicators			
Included			
5010	Hospitals	–	per 100k
5050	Hospital beds	–	per 100k
5020	Acute (short-stay) hospitals	–	per 100k
5070	Psychiatric hospital beds	–	per 100k
5100	Nursing and elderly home beds	–	per 100k

Table 4: (*Continued*)

Code	*Name*	*Target*	*Sample*
Excluded			
5060	Acute care hospital beds hospitals	–	per 100k
Human resources indicators			
Included			
5250	Physicians	–	per 100k
5300	Dentists (PP)		
5320	Nurses (PP)	–	per 100k
5350	Midwives (PP)	–	per 100k
Excluded			
5290	General practitioners (PP)	–	per 100k

Table 5 ANOVA *F*-tests for the variables used to form the clusters. For descriptive purposes only as the clusters have been chosen to maximise the differences among cases in different clusters and the significance levels are not corrected for this

	ANOVA F-test results	
	F-test	*Sig.*
Mortality-based indicators		
1320 SDR, diseases of circulatory system	183.844	0.000
1520 SDR, malignant neoplasms	10.404	0.000
1740 SDR, motor vehicle traffic accidents	5.073	0.007
1820 SDR, infectious and parasitic diseases	1.959	0.147
1830 SDR, diseases of respiratory system	2.836	0.059
1850 SDR, diseases of digestive system	20.126	0.000
1870 SDR, endocrine, nutritional and metabolic diseases	0.331	0.803
1900 SDR, mental disorders, diseases of nervous system and sense organs	3.699	0.026
1910 SDR, disease of genito-urinary system	0.832	0.490
1920 SDR, symptoms, signs and ill-defined conditions	0.710	0.555
1960 SDR, acute respiratory infections, pneumonia and influenza	2.163	0.119
Morbidity-based indicators		
2520 Hospital discharges, digestive system diseases	31.863	0.000
2450 Hospital discharges, circulatory system diseases	28.524	0.000
2310 Hospital discharges, all neoplasms	19.504	0.000
2530 Hospital discharges, musculoskeletal system and connective tissue diseases	17.991	0.000
2540 Hospital discharges, injury and poisoning	12.204	0.000
2500 Hospital discharges, respiratory system diseases	11.550	0.000
2300 Hospital discharges, infectious and parasitic diseases	6.841	0.001
Utilisation indicators		
6010 Inpatient care discharges	17.747	0.000
6100 Average length of stay	7.895	0.000
6210 Bed occupancy rate (%), acute care hospitals only	8.855	0.000
6300 Outpatient contacts	17.968	0.000
Healthcare physical resources		
5010 Hospitals per 100,000	29.748	0.000
5050 Hospital beds per 100,000	21.278	0.000
5020 Acute (short-stay) hospitals per 100,000	24.331	0.000
5070 Psychiatric hospital beds per 100,000	5.082	0.005
Healthcare human resources		
5250 Physicians per 100,000	11.458	0.000
5300 Dentists (PP) per 100,000	11.566	0.000
5320 Nurses (PP) per 100,000	14.521	0.000
5350 Midwives (PP) per 100,000	12.516	0.000

human and physical resources required, and so the number of hospital discharges is an incomplete proxy to the stress put on the healthcare system. Although this has no impact on the methodological approach, it may have on conclusions drawn from the results. For that reason, it would be beneficial to have a proper balancing taking this into account.

With this information at hand, we can now proceed with a more precise analysis of the clusters. First, cluster 4 has, on average, the lowest number of hospital discharges for all diseases. In sharp contrast, cluster 1, composed of Austria and Germany, has the highest number of discharges, followed by cluster 3, which is mostly composed of Eastern European countries. France has the highest record in some of the indicators (e.g. '2520 Hospital discharges, digestive system diseases', where it has twice the average of discharges), but average or below average values in others. For this reason, it did not fit any of the clusters and was thus put into a single cluster (cluster 6). Finally, clusters 2 and 5 have similar number of dispatches, on average, although countries from cluster 2 have a lower number of dispatches. Overall, cluster 4 has the smallest number of hospital discharges and cluster 1 has the highest. Note that using only this indicator makes it impossible to ascertain whether this is a result of inefficiency or low incidence. This information is summarised in Table 2.

Utilisation indicators

Utilisation indicators can provide insights into both supply and demand. On the one hand, indicators such as inpatient and outpatient contacts or length of stay are outputs of the production of healthcare services. On the other hand, such indicators also reflect the demand for healthcare services, but only of those effectively provided. While mortality- and morbidity-based indicators provide us with an insight of the general healthcare *needs* of a population, which may or may not translate into effective demand (Castelli *et al*, 2011), utilisation-based indicators translate actual utilisation ratios of the healthcare facilities, such as hospitals or primary care centres. Health needs may not always translate into actual demand due to several impediments such as financial constraints arising from expensive out-of-pocket treatments, or due to long waiting lists resulting from inefficient health systems.

Selection of indicators The procedure followed does not differ significantly from the method previously used. We start by selecting the variables to be used. The set of available variables is limited, essentially reporting overall inpatient care discharges, bed occupancy rates, average length of stay and number of outpatient contacts per year. Table 4 lists all the indicators used.

Data treatment, standardisation and aggregation Note that the indicators are reported in different units, and so variable standardisation is required to remove this potentially misleading effect. We map the indicators to a 0–1 range, preserving relative differences between countries. We also run a multicollinearity analysis, in the end removing 'Acute care hospital discharges' as these discharges are already accounted for in the more general inpatient care discharges.

Clustering algorithm It is then possible to run the HCA and obtain a preliminary grouping. Complementing the dendrogram with a scree plot analysis (see Figures A5 and A6), we find no clear elbow where to define the cutting point. Nevertheless, and considering that decreases in the distance coefficient get smaller between the seventh and the ninth cluster, after which they are almost marginal, we fix the number of clusters to seven. Running a k-means cluster analysis with $k=7$, we obtain the final cluster formation reported in Table 1.

Validation and interpretation Again, the ANOVA table helps to explain and validate the cluster formations (Table 5). According to the F-tests, both 'Outpatient contacts' and 'Inpatient care discharges' were the most critical indicators for grouping the countries. Following the results already obtained for the morbidity indicators, Austria and Germany, and also Lithuania, exhibit the highest number of inpatient care discharges. As for the number of outpatient contacts per year, cluster 3, composed of Czech Republic, Hungary and Slovakia, tops the list. On the opposite side, countries from cluster 4 have the lowest number of outpatient contacts per person, and, therefore, a small percentage of ambulatory care. In contrast, this group of countries exhibits, on average, the highest rate of bed occupancy (in acute care hospitals). Cluster 5, on the other hand, has both a relatively low bed occupancy rate, and a low number of outpatient contacts. However, on average, it has the highest average length of stay. The Netherlands did not fit any of the pre-existing clusters due to its extremely low bed occupancy rate and low inpatient care discharges. Interestingly, Southern countries from cluster 6 (Portugal, Italy and Spain) exhibit the lowest number of inpatient care discharges and below-average bed occupancy rates.

In summary, these results help to divide the clusters as follows: countries with health systems more orientated towards inpatient care and with average outpatient contacts (clusters 1 and, to some extent, 2); and countries with a model devised towards outpatient contacts (cluster 3). We can also identify countries with high occupancy rates, average inpatient care discharges and low outpatient contacts (cluster 4), and countries with average to low inpatient care discharges but long stays at the hospital and high bed occupancy rates (clusters 5 and 6).

Supply-based indicators

Physical resources Indicators reporting the available physical resources can be used in several ways. First, when used together with the utilisation-based indicators, these indicators help to identify countries in a situation of under- or over-capacity of its physical resources. Second, they can shed a light on the capital intensity of the health system, that is, whether it is labour-intensive or capital-intensive. Finally, they can be used to identify countries with a similar infrastructure. This is highly relevant since infrastructure investment is a long-term decision that cannot be taken lightly, and therefore policy proposals that compare asymmetric countries without taking this into account will fail to provide realistic (applicable) suggestions.

Selection of indicators The indicators used to perform the analysis are identified in Table 4. Data reported either relates to the number of hospitals or hospital beds by specialty.

Data treatment, standardisation and aggregation The multicollinearity analysis identified a strong correlation between 'Hospital beds' and 'Acute care hospital beds', which is expected since the first indicator already encompasses the second. Therefore, it has been removed. Also, rescaling was applied to remove the effect introduced by the usage of different scales.

Clustering algorithm Running HCA followed by a scree plot analysis points to a cutting-point at 4, optionally at 5 clusters (cf. Figures A7 and A8). Fixing the number of clusters to 5 and running a *k*-means cluster analysis originates the cluster formations reported in Table 1.

Exhibiting high statistical significance, the number of hospitals, reported both through the indicators 'Hospitals' and 'Acute (short-stay) hospitals', played a significant role in forming the clusters. In fact, cluster 3 includes the countries with the highest number of hospitals, both for short and long stays. Cluster 4, composed of countries like Estonia, Latvia and Switzerland (cf. Table 1), is similar, but with a lower number of short-stay hospitals and hospital beds. Cluster 5, composed of Southern and some Nordic countries, exhibits some of the lowest numbers in terms of hospitals and hospital beds. Note that the capacity to deliver does not imply more health gains, as more resources do not necessarily amount to more healthcare services provided. In fact, it may actually mean that there is a suboptimal resource allocation, with too many (or too few) expensive physical resources. To clarify, we again resort to a brief analysis of the capital-intensity to identify low and high capital intensity in the health systems. Note that the indicators are already normalised by population size, hence removing the difference in that size. Results are reported in Table 2.

Validation and interpretation According to the ANOVA *F*-test results provided in Table 5, where we test for the variables explaining the differences between groups, all indicators bar the number of psychiatric hospital beds were highly significant in determining differences between health systems, and therefore in defining the clusters. The clusters formed reflect four different healthcare architectures: a low number of hospitals and hospital beds (cluster 5); a large number of hospitals but with few hospital beds (cluster 4); a small to average number of hospitals but with large amount of hospital beds (clusters 1 and 2); and health systems with both a high number of hospitals and hospital beds (cluster 3). Geographical distance does not explain these clusters, as both eastern and western countries, Nordic and southern countries are intermingled in different clusters. This fact is suggestive of an explicit choice in terms of the organisational model.

Human resources

Selection of indicators The indicators used to perform the analysis are identified in Table 4. Human resources are measured in terms of number of physicians, nurses, dentists and midwives employed, regardless of their specialisation, if any.

Data treatment, standardisation and aggregation Since 'General Practitioners' is a subcategory of 'Physicians', we anticipate a multicollinearity issue, which is confirmed by running the tests. We therefore discarded this indicator from the analysis as it was already incorporated in the parent set.

Since all indicators are reported in a pre-standardised way (number of professionals per 100,000 inhabitants), no further standardisation was required.

Clustering algorithm We repeat the exact same procedure, this time applied to the indicators portraying the availability of human resources (cf. Table 4). We ran HCA followed by a coefficient analysis through a scree plot to decide on the number of clusters. With the help of both the dendogram and the scree plot (Figures A7 and A8) we fix this number to seven in order to run *k*-means and obtain the final cluster composition.

Validation and interpretation The resulting clusters are reported in Table 1, and the ANOVA *F*-test statistics are reported in Table 5. The statistical significance tests seem to suggest that no particular indicator had a distinctive influence in the cluster formations. Despite this, the number of nurses was the most significant. Regarding how the countries were grouped, of the seven clusters formed, it is immediate to see that Greece is a clear outlier. It tops the chart in both the number of physicians and dentists, having twice the average number, while at the same time it is the country with the lowest number of nurses. With also a significant amount of physicians but a not so high number of nurses, although well below Greece, is cluster 1, composed of countries like Austria, Germany (again, these two countries are part of the same group), Lithuania, Italy or Portugal. In contrast, cluster 2, composed of countries like Luxembourg, Belgium, Finland and Norway, has the exact opposite characteristics: a relatively small number of physicians and a quite large number of nurses and midwives.

Cluster 4, composed of Switzerland and Denmark, also demonstrates an extremely high nurse-to-physician ratio, but with a considerably higher number of physicians, therefore not featuring on cluster 2.

We can then characterise some of the clusters according to their healthcare delivery model: countries with a low number of physicians and a low number of nurses (cluster 7); countries featuring a high number of physicians but a low number of nurses (cluster 1); countries featuring a low number of physicians but a high number of nurses (clusters 2 and 4); and countries with both a high number of physicians and nurses (cluster 3).

Discussion

In this article, we had several objectives: (i) to warn against the methodological flaws potentially posed by the usage of benchmarks or simple ratio comparisons drawn between different health systems, subject to disparate health patterns, asymmetric healthcare infrastructure, and several other contrasting characteristics; (ii) to propose a methodology to group countries according to their similarities in each health indicator, providing clusters of countries that share similar characteristics so that comparisons can be made in a (methodologically) robust way; (iii) to discern the profiles of the groups of countries for each indicator, providing preliminary insights and a similarity matrix ranking the most and the least similar countries. With this theoretical and empirical contribution, we aim to provide the underpinnings for more accurate comparisons between health systems, upon which performance analyses or benchmarking can be conducted. To some degree, the proposed approach avoids error-prone comparisons with the global average.

Relative performance analyses

Applying the methodology to the indicators provided by the OECD's HFA-DB provides some interesting insights that benefit from further discussion. After the clusters are generated, we perform an inter-cluster qualitative comparison and characterise the groups obtained in terms of relative performance. To do so, we start by calculating the average of the cluster for each variable considered, and then label countries as below average exhibiting a low performance, or above average exhibiting a high performance (note that *low* does not mean 'bad' or 'inefficient', but only that the measurement is lower comparatively to the average, and *vice-versa*). In terms of mortality, we then establish that, in relative terms, clusters 1 and 3 exhibit low mortality rates, while clusters 2 and 4 exhibit high mortality rates. Unsurprisingly, this reflects the Western/Eastern asymmetries mentioned previously, with former Soviet countries still trying to catch up in terms of healthcare gains in comparison with other health systems. In fact, this sharp difference is quite noticeable, with the worst performing clusters registering twice the average of SDRs.

Contrarily to mortality-based indicators, it would be fallacious to draw conclusions from the qualitative assessment on the relative performance of the clusters solely based on the number of dispatches, the proxy selected to characterise morbidity. Exhibiting a high or low number of hospital discharges does not characterise the system. The health system can be either highly efficient and effective because it can treat a large number of patients, or simply because the incidence and prevalence rates of those diseases are lower in some countries, which would naturally lead to less hospital entries, and hence less discharges. Nevertheless, this analysis provides a first insight into the comparative performance of the groups. We see that high numbers of hospital discharges are not a characteristic unique to less developed healthcare systems. In fact, the cluster composed of Austria and Germany report, on average, the highest number of hospital discharges.

Characterising demand

As previously discussed, a simple quantitative analysis of the morbidity rates measured in terms of hospital discharges would be misleading. High hospital discharges may not necessarily suggest that the health system is efficient. But if we compare the morbidity-based hospital discharges with the mortality SDRs, we can better characterise the potential demand (*needs*) for healthcare. *A priori*, it can be conjectured that countries exhibiting high mortality rates and low hospital discharges are not providing appropriate care. Likewise, countries exhibiting low mortality rates and high hospital discharges would suggest a high amount of care. Such information can be useful for characterising the stress that different healthcare systems have to endure, and therefore enhance the robustness of future comparisons.

Figure 4 shows how countries perform when both mortality-based indicators and morbidity-based indicators are considered. We also group those countries that were part of the same group both in the mortality- and morbidity-based cluster analysis. Despite reducing the number of countries in the clusters, this meta-clustering technique strengthens the cluster formations, bringing together countries that are similar in more than one set of indicators. The plot is divided into four quadrants. Countries exhibiting both low (high) hospital discharges and low (high) mortality rates are of no particular interest, as it is expected that if more people carry a disease, more hospital discharges and more mortality will follow, and *vice-versa*. The revealing cases are those countries that have either a high number of hospital discharges but low mortality rates, suggesting that these countries are making efforts for containing the diseases and treating patients, or a low number of hospital discharges and high mortality rates, implying that a considerable number of people are dying without proper treatment or inadequate care is being provided. The former is the case of Austria and Germany, which have a comparatively low mortality rate despite the large number of hospital discharges. Latvia and Slovakia

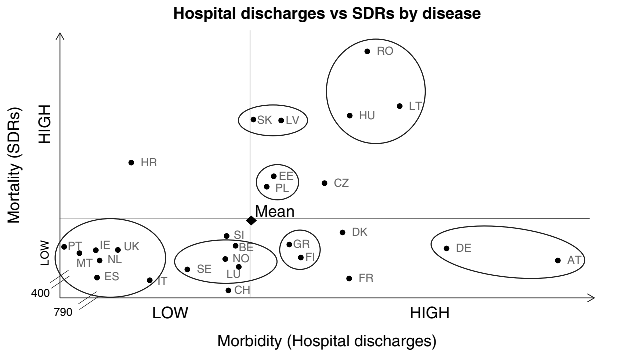


Figure 4 Hospital discharges *vs* mortality SDRs per country. The groups formed include countries that are featured in the same cluster in both criteria, mortality and morbidity rates.

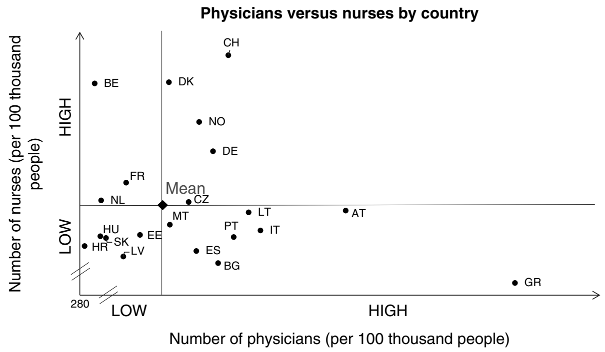


Figure 5 A two-dimensional graph depicting physician and nursing intensity in each country.

are on the opposite side, as they have extremely high mortality rates with a low number of hospital discharges, which can be interpreted as incapacity to deliver effective healthcare, or a healthcare system based on an outpatient contact approach that is not generating satisfactory results in terms of health gains.

With this information, we can form clusters of countries featuring similar morbidity trends, when measured in terms of outputs (hospital discharges), and effectiveness of the care services delivered, when measured in terms of mortality rates (Table 3).

With these super-clusters, it is possible to make comparisons at two different levels, between and within groups. For instance, if Romania would like to improve its death accruals, they should probably look to Hungary for guidance, as it has the same number of hospital discharges, and yet lower death records. Similarly, if Portugal wants to improve the death records, it should look to Spain or Italy, rather than the EU average or any other country outside its cluster. Comparisons between clusters are also possible using a relative performance index. In particular, clusters 1 and 5 in Table 3 both exhibit low levels of mortality. However, contrarily to Austria or Germany, countries from cluster 5 register significantly lower hospital discharges, implying that for each person discharged, holding everything else constant, more are dying due to a disease. This may be indicative of lack of proper or timely treatment.

Characterising supply

Output in terms of healthcare services is the result of the production using input factors needed to deliver care, both physical (capital) and human (labour). Physical resources encompass hospitals, hospital beds, screening and treatment technology, or drugs. We have focused only on the indicators provided in the HFA-DB, which include the number of hospitals and hospital beds. Human resources include physicians, nurses, midwives, dentists, medical assistants, etc. Only those present in the HFA-DB and with no significant missing information were included in the analysis.

The data available on the HHR in each country seem to suggest a very different approach to healthcare delivery, focused essentially in the role of the physician in the first case, and on a higher delegation of tasks to the nursing profession in the second case.[1] To illustrate this, we draw a two-dimensional graph depicting HHR (Figure 5). As mentioned before, Greece is a clear outlier, with a number of physicians well above the EU average, while having the lowest record of nurses. This is also visible in countries such as Austria, Italy or Lithuania. On the other extreme in terms of healthcare model are countries such Belgium, Denmark, Norway, Luxembourg, Finland or Switzerland, which register a high number of nurses but a low number of physicians. It is interesting to note that this trend is common to Northern European countries, apparently favouring a healthcare workforce with a high prevalence of nurses.

Another conclusive analysis consists of plotting both human and physical resources, which makes it possible to obtain a degree of capital and labour intensity of the healthcare system in each country. To assess labour intensity, we considered an unweighted sum of the number of physicians and nurses. Without disregarding the importance of other clinical actors, these are the core human resources of any healthcare system. As for capital intensity, we consider only the number of hospital beds. The scatter plot in Figure 6 provides visual guidance to understand these health models.

The most interesting cases are those countries exhibiting asymmetries between capital and labour, implying a radically different approach to healthcare delivery. Interestingly, countries such as Denmark, Norway or Switzerland appear to be using plenty of human resources, especially nurses, while keeping the number of hospital beds relatively low. A tentative explanation could lie in the fact that these countries are able to treat patients more quickly, and hence free physical resources (hospital beds). Considering that in our case capital intensity portrays the number of hospital beds and not the entire set of technologies used to

[1]Alternatively, this may be the result of disparities in the way the number of nurses are reported to WHO, since in some countries some auxiliary professions count as nursing.

treat patients, this is the only hypothesis standing. On the opposite side are the countries employing a significant amount of capital with a low labour intensity. Almost all Eastern European countries exhibit this trend. This apparently high availability of resources may help with a surge in demand, but it says nothing about the capacity to deliver effective healthcare services. While a physician or a nurse may always deliver services in better or worse conditions regardless of the infrastructure available, that is not the case with an empty bed. Either way, validating any of the hypotheses would require a thorough and methodologically rigorous approach, outside the scope of this work.

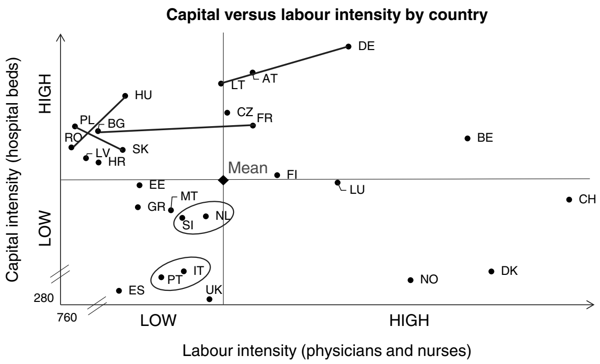


Figure 6 Capital *vs* labour intensity by country. Lines and circles group countries that were part of the same cluster both in the human and physical resources clustering.

Similarity matrix

Finally, and gathering all the clusters obtained for each of the indicators, we can construct a similarity/dissimilarity matrix, pinpointing countries that are part of the same group, and hence similar to each other, in a given number of indicators. Figure 7 shows that while some countries are similar to each other in several dimensions, others differ significantly. For instance, Portugal and Italy are part of the same cluster for five out of five indicators. With a similarity of four, countries like Portugal and Spain, Hungary and Romania, Belgium and Luxembourg, or Austria and Germany, feature in the same cluster in four out of the five indicators. Also important to international comparisons may be to identify countries that share no common ground at all. Although this may be due to different stages of development, which appears to be the case, for instance, of Austria and Hungary, it may also reflect substantially different approaches to healthcare delivery. For instance, Austria and Ireland, countries with an almost identical GDP per capita, are extremely different, and so they do not feature in the same cluster in any of the indicators. Note that due to the lack of data, it was not possible to include all countries for all the indicators, which means that this matrix may be under-represented.

Conclusion

In this paper we have proposed a new approach to cross-country comparisons, which avoids the pitfalls of benchmarking against the average or of absolute rankings that sometimes lead to disputable conclusions if not outright

	AT	BE	BG	HR	CZ	DK	EE	FI	FR	DE	GR	HU	IE	IT	LV	LT	LU	MT	NL	NO	PL	PT	RO	SK	SI	ES	SE	CH	UK
AT					1			1		**4**	2	1		1		2	1	1			1	1	1	1	1		1		
BE				2		1	1	1	1				1	1	1		**4**	1	2	**3**		1			**3**	1	1	2	1
BG					1		1	1	2	1						1													
HR		2	1				**3**		1				1	1	1		1	2	2		1	1			2	1			1
CZ	1		1	2			**3**	1	1		2	2			1		1				**3**		1	**3**					
DK		1			1			1	1		2		2	1	1		1		1	2	1	1		1		1	1	**3**	2
EE		1	1	**3**	**3**	1			1		1				**3**		1				2			1	1			1	
FI	1	1	1		1	1	1			2	2				1	1	1	1		1	1			1	1		1		
FR		1	2	1	1	1	1	2					1	1		1	1		1	1		1				1		1	1
DE	**4**		1					2	1					1		**3**		1				1			1		1		
GR	2				2	2	1	2		1			1		1		1	1		1	2		1	2	1		2	1	1
HU	1				2						1				2	2	1		1		1		**4**	**3**	1	1			
IE		1		1		2			1		1						1	1	2	**3**		**3**				**3**	2	2	**3**
IT	1	1		1		1			1	1			**3**			1	1	1	2	2		**5**				**4**	1	1	2
LV		1		1	1	1	**3**	1			1	2					1		1		1		2	2	2	1		1	
LT	2		1					1	1	**3**		2		1	1							1	2	1					
LU	1	**4**		1	1	1	1	1	1		1	1	1	1	1				1	**3**	1	1	1	1	2	1	1	2	1
MT	1	1		2				1		1	1		1	1							1	1		1	2	1	1		2
NL		2		2		1			1			1	2	2	1		1	2				2	1		2	**3**		1	2
NO		**3**				2		1	1		1		**3**	2			**3**		1			2			1	2	**3**	**3**	2
PL	1			1	**3**	1	2	1			2	1			1		1	1					1	**3**					1
PT	1	1		1		1			1	1			**3**	**5**		1	1	1	2	2						**4**	1	1	2
RO	1				1						1	**4**			2	2	1		1		1				1	1			
SK	1				**3**	1	1	1			2	**3**			2	1	1	1			**3**		2						1
SI	1	**3**		2			1	1		1	1	1			2		2	2	2	1			1				2	1	
ES		1		1		1			1			1	**3**	**4**	1		1	1	**3**	2		**4**	1		1			1	2
SE	1	1				1		1		1	2		2	1			1	1		**3**		1			2	1			1
CH		2				**3**	1		1		1		2	1	1		2		1	**3**		1			1	1	2		
UK		1		1		2			1		1		**3**	2			1	2	2	2	1	2		1		2	1	2	

Figure 7 Similarity matrix obtained by counting the intersections between clusters for the five indicators. High numbers indicate high similarity, and *vice-versa*.

erroneous policy choices. Instead of finding winners and losers, this methodology establishes similarities between countries. Performance and efficiency analyses can additionally be conducted, but at a different granularity level. From a policy perspective, it is better to inform and improve based on the experience of a country that shares common ground rather than on the most efficient and yet structurally different country.

Our methodology consists in applying state-of-the-art clustering techniques to group similar countries according to several indicators. We applied this methodology to WHO's HFA-DB, a comprehensive database containing indicators related to mortality, morbidity, utilisation, and physical and human resource indicators. We were able to group countries in each of the five dimensions, and explain the cluster formations. Moreover, and based on these results, we constructed a similarity matrix. The results obtained and discussed serve the primary purpose of demonstrating the methodology, but also allow for a preliminary comparative analysis of the health systems. In the future, we expect this methodology to be applied to a more comprehensive set of data so that more insightful and robust conclusions can be drawn to guide policy reforms.

Another interesting application of this methodology is the ability to directly compare countries. From the demand analysis, we highlighted that Hungary and Romania were part of the same cluster, but Hungary featured a considerably smaller mortality rate within the group when compared to Romania. If we add the dimension of supply, characterised by both labour and capital, we notice that Hungary has both a higher capital and labour intensity in comparison to Romania. Although a more rigorous econometric analysis that controls for other factors would be advisable to validate the hypothesis, this seems to suggest that improved health results may be due to more physical and human resources. This result can also be observed between Portugal and Italy, part of the same group in both dimensions. Italy has a smaller mortality rate than Portugal, but also more resources to deliver healthcare, both in terms of physicians and nurses, and in hospital beds. But in comparison with Spain, Portugal exhibits a higher mortality rate compared with Spain, despite employing more capital and more labour. Given the similarities and the results achieved, Spain may be reference country to guide future policy decisions in Portugal.

Notwithstanding, comparisons can also be drawn in a more elaborate way. For instance, production functions based on empirical estimates can be used to aggregate inputs such as physical and human resources, as well as hospital beds, other administrative staff, etc. Clustering can then be conducted using the production function as a measure of distance. In a similar way, analyses can be complemented with further information. In the comparison previously established between Portugal and Spain, one could also have included the age structure in addition to the previous results. Given the similarity between the age structures of both countries, this would further reinforce the point.

Finally, it would also be interesting to apply a non-parametric tool like DEA to obtain the efficiency frontier within clusters. This may only work if the number of countries exceeds the number of outputs, or if the outputs are aggregated in order to ensure this condition. Such a limitation of DEA is typically found in comparisons within hospitals, where the number of outputs (e.g. measured in DRGs) far exceeds the number of hospitals (Newhouse, 1994). For this reason, simplified measures of outputs, such as inpatient contacts or length of stay, are frequently used to be able to apply DEA. Another option may be the use of explicit weights to combine several outputs into a single index, as proposed by Castelli *et al* (2015), in this way avoiding the need of DEA. Such methodology could also be used to group indicators within each category prior to performing the clustering analysis. In fact, the methodology proposed by Castelli provides strong foundations on how to group hospital inputs and outputs. For instance, statistics of inpatient and outpatient contacts are grouped into hospital inputs, and data of agency and medical labour as well as capital and intermediate goods and services are grouped as inputs in a health production function. The application of such aggregation methodology could complement the methodology presented here.

References

AMORIM LOPES M, ALMEIDA ÁS and ALMADA-LOBO B (2015) Handling healthcare workforce planning with care: where do we stand? *Human Resources for Health* **13(1)**, 38.

AUERBACH DI and KELLERMANN AL (2011) A decade of health care cost growth has wiped out real income gains for an average US family. *Health Affairs* **30(9)**, 1630–1636.

BHAT VN (2005) Institutional arrangements and efficiency of health care delivery systems. *The European Journal of Health Economics* **6(3)**, 215–222.

BIRCH S (2002) Health human resource planning for the new millennium: inputs in the production of health, illness, and recovery in populations. *Canadian Journal of Nursing Research* **33(4)**, 109–114.

BLOOR K, HENDRY V and MAYNARD A (2006) Do we need more doctors? *Journal of the Royal Society of Medicine* **99(6)**, 281–287.

BLOOR K and MAYNARD A (2003) Approaches to human resources in health: an international comparative review. *Canadian Health Services Research Foundation*, 1–34.

BRONNUM-HANSEN H (2014) Ranking health between countries in international comparisons. *Scandinavian Journal of Public Health* **42(3)**, 242–244.

BURAU V and BLANK RH (2006) Comparing health policy: an assessment of typologies of health systems. *Journal of Comparative Policy Analysis: Research and Practice* **8(1)**, 63–76.

CASTELLI A, LAUDICELLA M, STREET A and WARD P (2011) Getting out what we put in: productivity of the English national health service. *Health Economics, Policy and Law* **6(03)**, 313–335.

CASTELLI A, STREET A, VERZULLI R and WARD P (2015) Examining variations in hospital productivity in the English NHS. *The European Journal of Health Economics* **16(3)**, 243–254.

CHANDRA A, HOLMES J and SKINNER J (2013) Is this time different? The slowdown in healthcare spending. Brookings Papers on Economic Activity, Economic Studies Program, The Brookings Institution 47(**2**), 261–323.

DORMANN CF *et al* (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **36(1)**, 27–46, [WWW document] http://onlinelibrary.wiley.com/store/10.1111/j.1600-0587.2012.07348.x/asset/j.1600-0587.2012.07348.x.pdf?v=1&t=i30935yj&s=76baabe51dbab844389d31061588f7682f591e7a.

DREESCH N (2005) An approach to estimating human resource requirements to achieve the millennium development goals. *Health Policy and Planning* **20**(**5**), 267–276.

Europe, W.R.O.F. *European Health for All Database*. WHO Regional Office, Copenhagen, Denmark.

EVERITT BS, LANDAU S and LEESE M (2001) *Cluster Analysis*. John Wiley & Sons, Ltd., West Sussex, United Kingdom.

HAIR JF (2013) *Multivariate Data Analysis*, 7th edn, Pearson New International Edition, Essex, England.

HARPER PR, POWELL NH and WILLIAMS JE (2010) Modelling the size and skill-mix of hospital nursing teams. *Journal of the Operational Research Society* **61**(**5**), 768–779.

JEON YD, LEE JI, CHO BO, ROTARU H, KIM JC and KIM SG (2005) Statistical correlation between pharyngitis and temporomandibular joint disease. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology* **99**(**6**), 677–681.

KETCHEN DJ and SHOOK CL (1996) The application of cluster analysis in strategic management research: an analysis and critique. *Strategic Management Journal* **17**(**6**), 441–458.

LI J-M *et al* (2014) Inverse correlation between Alzheimer's disease and cancer: implication for a strong impact of regenerative propensity on neurodegeneration? *BMC Neurology* **14**(**1**), 211.

NEWHOUSE JP (1994) Frontier estimation: how useful a tool for health economics? *Journal of Health Economics* **13**(**3**), 317–322.

PRINGLE M, WILSON T and GROL R (2002) Measuring 'goodness' in individuals and healthcare systems. *British Medical Journal* **325**(**7366**), 704–707.

RICHARDSON J, WILDMAN J and ROBERTSON IK (2003) A critique of the world health organisation's evaluation of health system performance. *Health Economics* **12**(**5**), 355–366.

ROBERFROID D, LEONARD C and STORDEUR S (2009) Physician supply forecast: better than peering in a crystal ball? *Human Resources for Health* **7**(**1**), 10.

TCHOUAKET ÉN *et al* (2012) Health care system performance of 27 OECD countries. *The International Journal of Health Planning and Management* **27**(**2**), 104–129.

VARABYOVA Y and SCHREYÖGG J (2013) International comparisons of the technical efficiency of the hospital sector: panel data analysis of OECD countries using parametric and non-parametric approaches. *Health Policy* **112**(**1–2**), 70–79.

WENDT C (2009) Mapping European healthcare systems: a comparative analysis of financing, service provision and access to healthcare. *rszarf.ips.uw.edu.pl* **19**(**5**), 432–445, [WWW document] http://esp.sagepub.com/cgi/doi/10.1177/0958928709344247 (accessed 14 June 2014).

WENDT C, FRISINA L and ROTHGANG H (2009) Healthcare system types: a conceptual framework for comparison. *Social Policy & Administration* **43**(**1**), 70–90.

Appendix

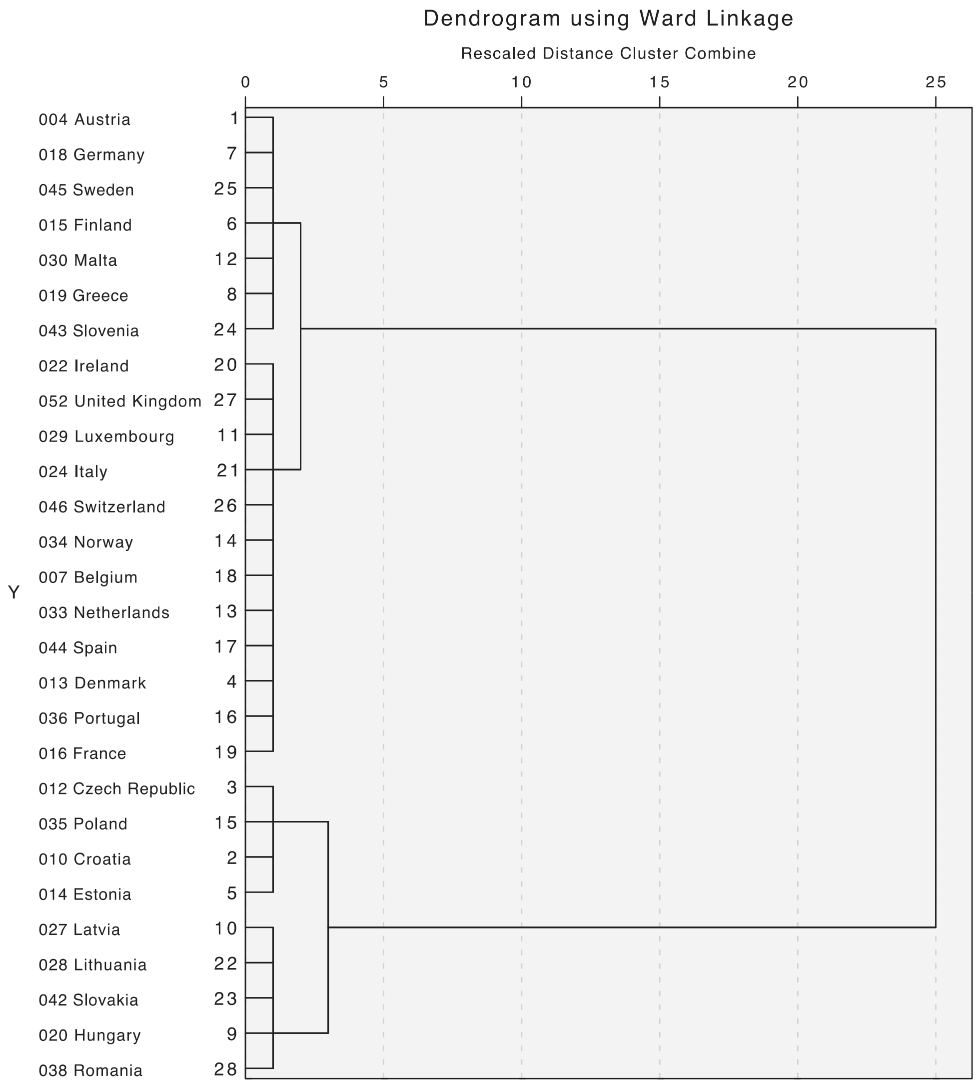


Figure A1 Dendrogram obtained from applying HCA with Ward's method to mortality-based indicators.

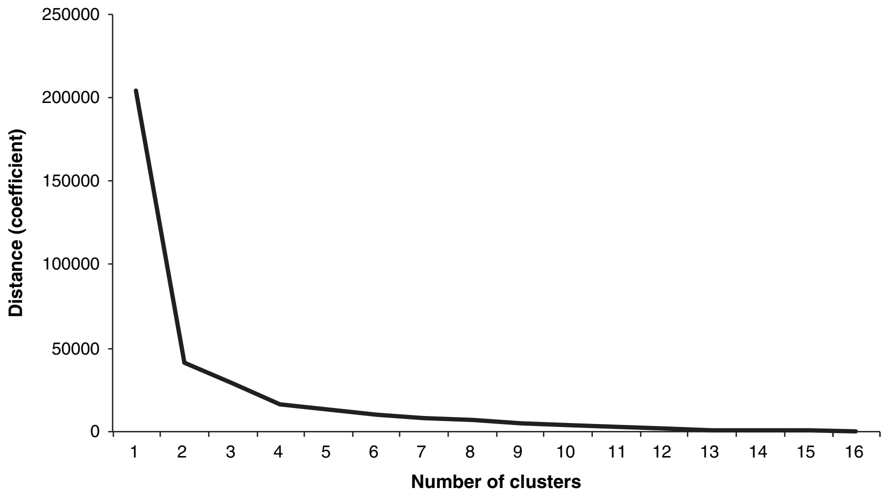


Figure A2 Scree plot for defining the elbow, that is, the cutting point in the number of clusters. The distance measured is within the cluster.

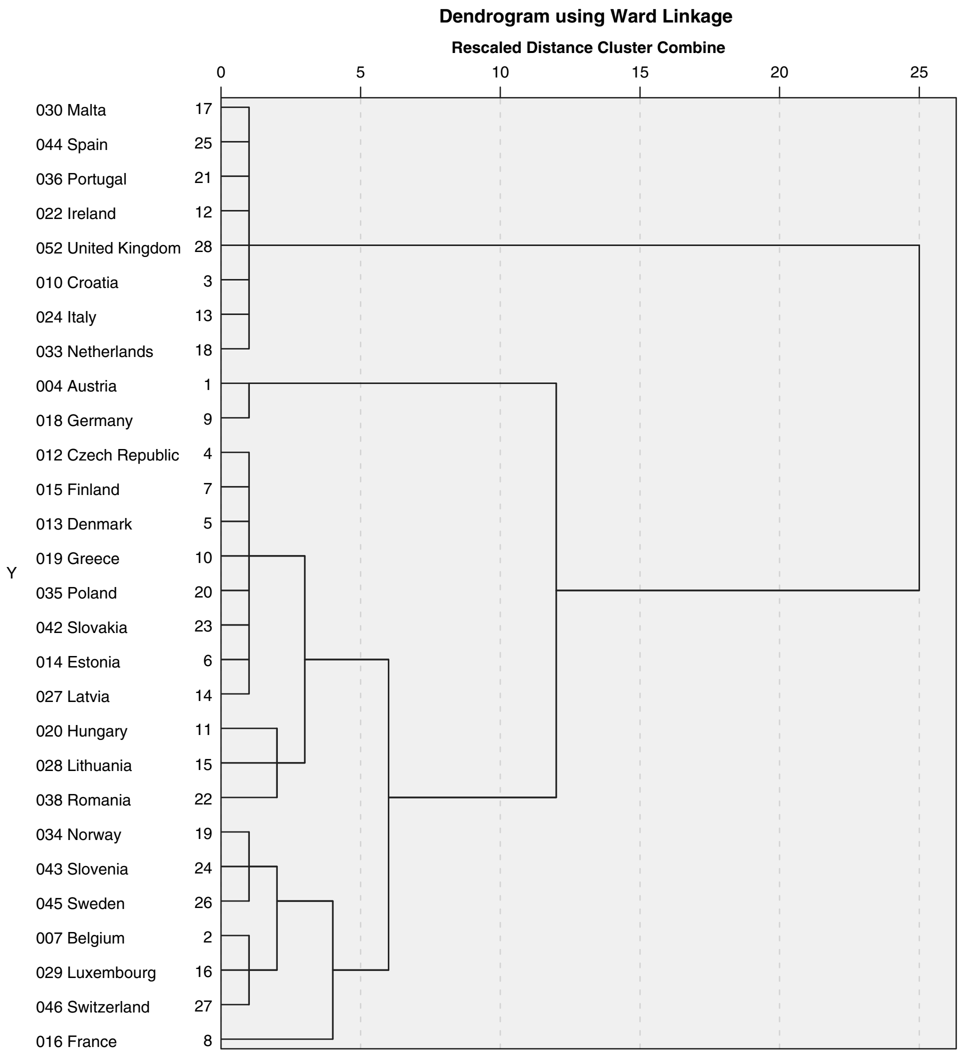


Figure A3 Dendrogram obtained from applying HCA with Ward's method to morbidity-based indicators.

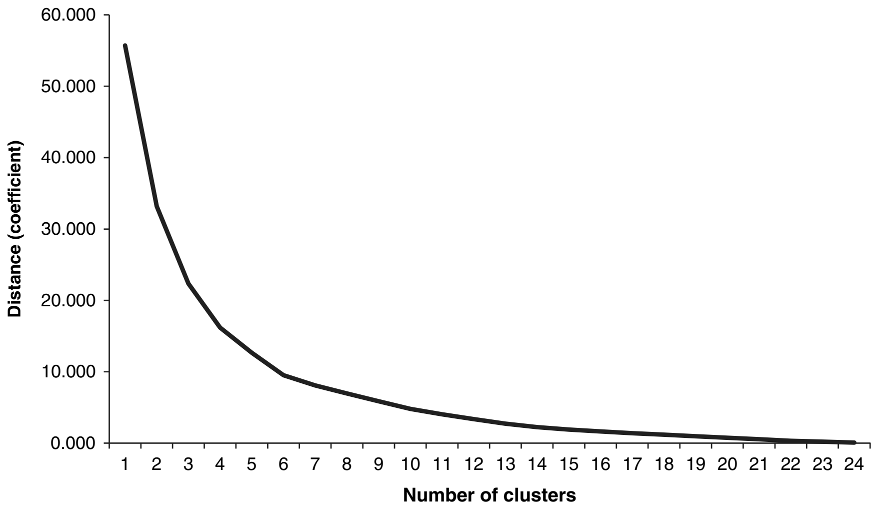


Figure A4 Scree plot for defining the elbow, that is, the cutting point in the number of clusters. The distance measured is within the cluster.

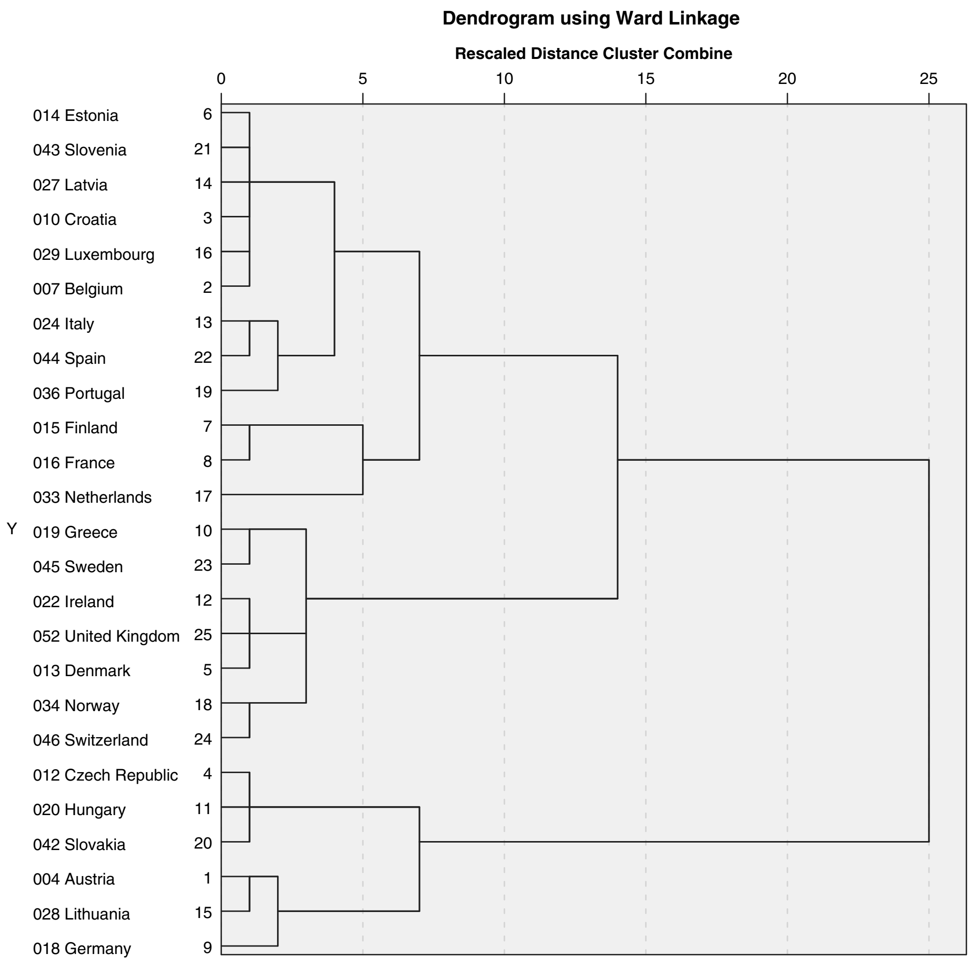


Figure A5 Dendrogram obtained from applying HCA with Ward's method to utilisation indicators.

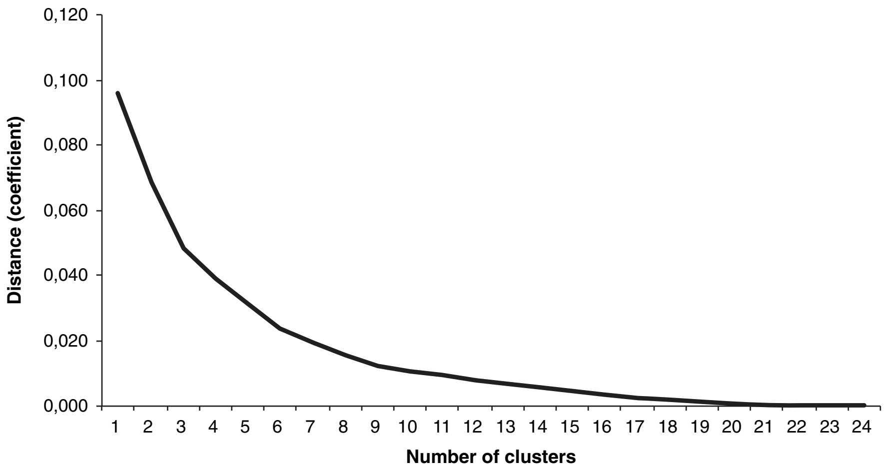


Figure A6 Scree plot for defining the elbow, that is, the cutting point in the number of clusters. The distance measured is within the cluster.

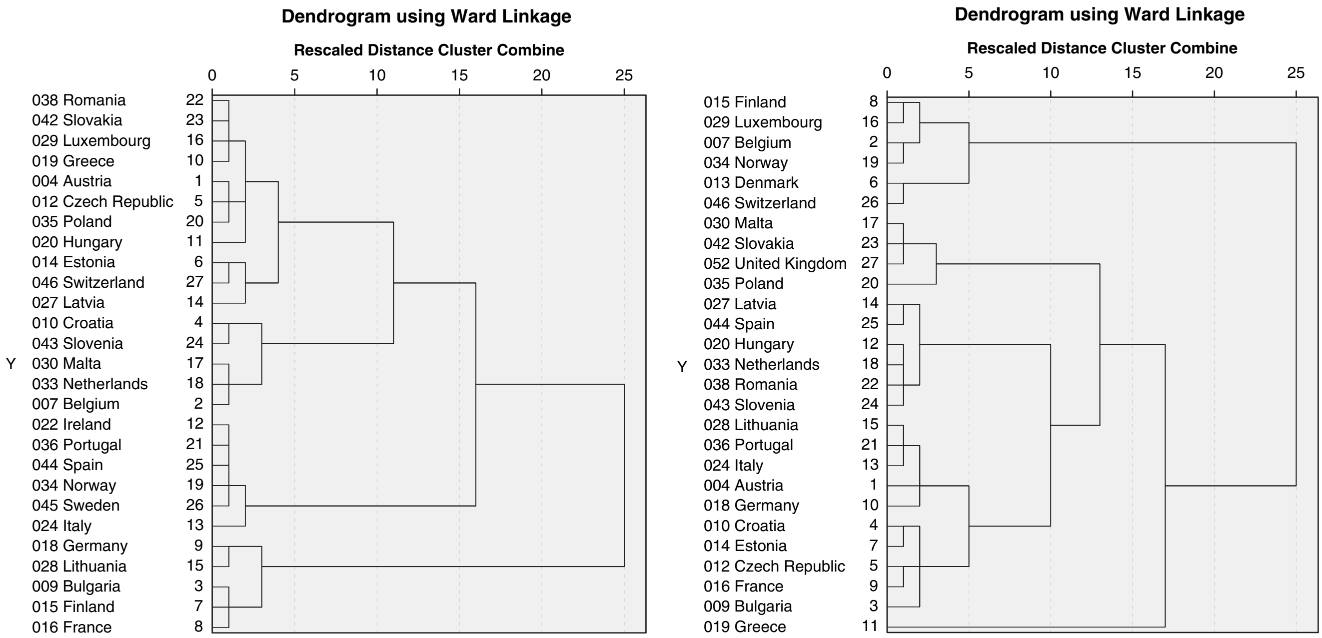


Figure A7 Dendrogram obtained from applying HCA with Ward's method to physical resources (left) and human resources (right) indicators.

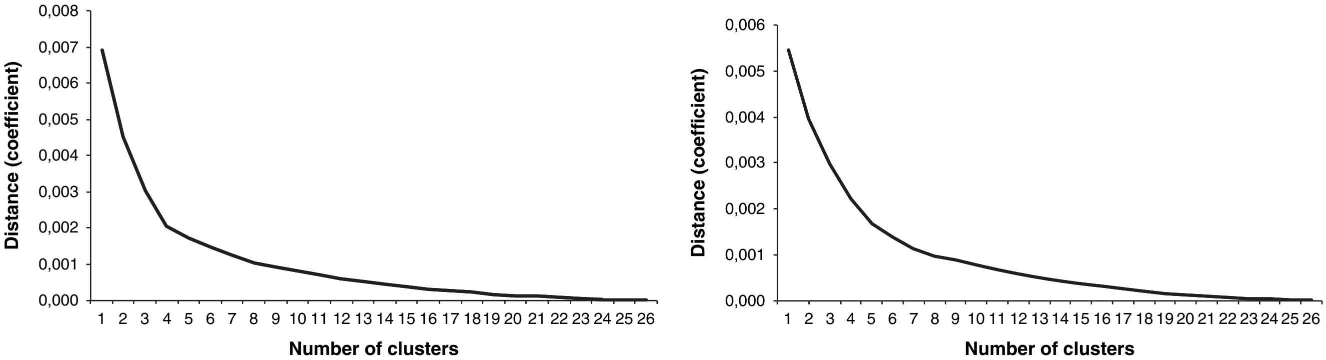


Figure A8 Scree plot for defining the elbow, that is, the cutting point in the number of clusters in healthcare physical (left) and human (right) resources.