Research paper

Networks of primary and secondary care services: how to organise services so as to promote efficiency and quality in access while reducing costs

Ricardo Farinha BSc MSc
Associate at Boston Consulting Group, Lisbon, Portugal

Mónica Duarte Oliveira BSc MSC PhD
Lecturer, Engineering and Management Department, Instituto Superior Técnico; Researcher, Centre of Management Studies, Instituto Superior Técnico, Technical University of Lisbon, Portugal; Visiting Research Fellow, Operational Research Group, London School of Economics and Political Science, UK

Armando Brito de Sá MSc PhD MD
Professor of General Practice at the Faculdade de Medicina, University of Lisbon, Portugal

ABSTRACT

Governments in countries with national health systems have been concerned with how to organise services so as to achieve improvements in efficiency and quality in healthcare delivery, as well as to control costs. In this study, a stochastic discrete event simulation model to study the organisation of primary and secondary care services is proposed. The model was built with reference to the context of the Portuguese NHS, was implemented in the Simul8 software program and was applied to the Portuguese Setúbal healthcare subregion (SHCR). For its application, a database with 2005 production, resource and cost indicators was built to calibrate and validate the applied model. After validation, three different policy scenarios were tested: the first one concerning a 10% increase in demand for primary care services; the second considering a shift between specialists and generalist physicians; and a third regarding restructuring of primary care services. Results show that although the current system is not prepared to cope with a rise in demand, the other scenarios indicate that there is room for primary care reforms to increase the system’s efficiency and accessibility, while lowering total costs.

Keywords: discrete event simulation, networks of services, organisation and planning, primary and secondary care interface, referral systems

How this fits in with quality in primary care

What do we know?

Faced with increasing costs, countries with a nationally funded system are under pressure to reduce them, while promoting the system’s equity, quality of access and efficiency in healthcare delivery. Also, there is a wide recognition of the need to shift resources from the secondary to the primary care sector, so that health systems increase their health-promotion role (in comparison to healthcare treatment). Nevertheless, few studies have quantified the impacts of policies shifting resources between the secondary and the primary care centres in a systematic way.
Introduction

Most health systems are under pressure to contain increases in their costs, while improving quality, efficiency and equity. To achieve these objectives, it is essential to develop tools capable of helping public planners in national health service (NHS)-based countries, by providing information about the system’s performance and the potential impact of future changes.

The Portuguese NHS is no exception and has as key objectives the pursuit of equity, efficiency, quality, accountability and the devolution of power.1 Yet, several problems have been referred to the system: the excessive use of emergency services, the long waiting times for surgery and some outpatient consultations, the concentration of resources in secondary care and urban areas, and the increase in the overall system’s cost.2 Thus, tools are needed to give information about the system’s current and future performance.

This study presents a model that informs planning of the two main levels of public planned primary and secondary healthcare services in NHS countries. Primary care represents the first level of contact, in the form of primary care centres (PCCs), acting as gatekeepers and regulating the entrance and referral of users. From the services offered, we consider that PCCs provide ambulatory care (provided for users by appointment), and emergency care services that do not need an appointment and can be provided by any physician. Secondary care, provided in hospitals, is more specialised and answers to situations not treated in PCCs. Portuguese hospitals are divided into four categories:3 district hospitals (DHs), able to provide basic services; central hospitals (CHs) that also provide other more specialised services to wider populations; (highly) specialised hospitals (SHs), not accessible for direct use; and level 1 hospitals (L1Hs), similar to DHs but more focused in recovery and extended care. Comparable hospital hierarchies and gatekeeping systems exist in other NHS-based countries, and the problem of organising and balancing these services is common in these health systems.4,5 Hospitals provide three main types of services: emergency care, inpatient care and outpatient consultations.

Primary and secondary services tend to be organised into hospital referral networks (HRNs). HRNs connect services and providers, exploring their complementarity while maximising the resources’ use. The referral process is of utmost importance, for it is the family physician at the PCC that makes the first assessment of the patient, and decides the best way to approach the problem(s). Thus, decisions at the PCC level impact in the overall NHS costs and in the secondary care workload.

Accordingly, despite the wide recognition of the importance of the interface between primary and hospital care services,6–10 few studies have quantified the impact of health policies on efficiency, quality in access and costs of running a network of primary and secondary providers.7,11,12 Studies generally acknowledge that there should be a shift of resources from the secondary care to the primary care sector (so as to give a higher emphasis to health promotion instead of focusing on treatment), but there is little evidence on the impact of shifting resources. This study proposes a stochastic discrete event simulation (DES) to study the organisation and interaction of primary and secondary care services within an HRN. DES is useful to model complex systems and to deal with the stochastic nature of some variables, such as demand. Although DES models have been scarcely used at a macro level, they are flexible in modelling the interactions between the different services and providers and in capturing randomness in demand, and they provide a wide set of outputs. These characteristics make them a useful choice to analyse the configuration of HRNs. Simulation models, besides aiding in the planning process, might be also used to support the policy process by promoting a better understanding of benefits of policies and by defining targets on production, access and financial indicators.
The proposed model was applied to the Setúbal healthcare subregion (SHCR) in Portugal as a case study. Figure 1 shows the geographic location of PCCs and hospitals within that subregion. This administrative area was chosen because it contains a well-defined area of health care and it comprises two subregions: the Setúbal Peninsula, urban, densely populated, with a growing young population; and the Litoral Alentejano region, rural, sparsely populated, with an aging population. Overall, 21 PCCs and five hospitals (1 CH, 2 DHs and 2 L1Hs) were considered.

This study is organised into the following sections: brief review of related literature; description of the conceptual model and of its application to the Portuguese NHS (with 2005 information); analysis of results and policy-related scenarios; and drawing of some conclusions.

Literature review

Considering the problem of planning and reorganising HRNs, there have been three main methodologies for approaching it:13 direct experimentation, mathematical programming, and simulation. Direct experimentation consists of testing HRN policies directly on the system, normally in a controlled way. Although it might be simple, it is costly and time consuming, and results might not be reproducible in other contexts. Mathematical modelling consists of representing the systems’ objectives through mathematical models that pursue optimal solutions for the structure of the system. Mathematical modelling is adequate for obtaining optimal solutions, but there might be some difficulties in modelling dynamic systems, it demands simplifications and very often also demands the use of heuristic methods for obtaining results, which sometimes leads to suboptimal results. Finally, simulation models are used for ‘the imitation of the operation of a real-world process or system over time ... to draw inferences concerning the operating characteristics of the real system’.14 Their main advantage is their flexibility, both in modelling the interaction between elements of the system and in the definition of associated parameters. This confers on simulation models a great ease of experimentation, allowing different policy scenarios and hypotheses to be tested in an easy, low-cost, risk-free and quick way.

Considering our objectives of analysing networks of primary and secondary services, a simulation model was selected as the most appropriate method to describe the users’ progress during several events in the healthcare system (e.g. patients) while taking into consideration the stochastic nature of demand. In other words, a stochastic model is needed, based on patient and care events that occur at a discrete time. This type of simulation is known as discrete event simulation (DES), and has been used in several medically related areas, e.g. epidemiology, health promotion and prevention,15,16 and the design17,18 and management19,20 of healthcare systems. The contribution of DES to the study of health-related problems has been generally recognised.21 Although simulation methods have been extensively used in some areas, they have been mostly used to analyse problems at the micro level (within organisations), while analysis at a macro level has mostly used mathematical programming methods.22,23 To our

Figure 1 PCC and hospital locations, numbers and their referral areas (black dashed arrows), with the emergency RRHs represented with grey arrows.
knowledge, DES models have not been used to analyse referral networks at a macro level, or to test the impact of policies changing the interface between the primary and secondary care sectors. Nevertheless, some studies on the Portuguese health system have somewhat analysed referral processes, either at the country or regional and unit level. This work proposes a DES methodology to study networks of primary and secondary care providers. It might be seen as a decision support tool to help planners to decide upon referral and resourcing policies and upon the balance between primary and secondary care.

Developed model and case study

The model had to take into account the resources available and used to provide healthcare services (ambulatory and emergency care in primary care, emergency, inpatient care and outpatient consultations in secondary care), the HRN linking healthcare units and services, the associated costs with provision of services, and the stochastic demand for health care. Our model considers only two hospital levels: district and central hospitals. The interaction between services and healthcare units is presented in Figure 2.

Figure 2 should be read as follows: the solid black arrows represent the direct entries of users in the system (computed afterwards in absolute demand values). These entries might be through PCCs (the system’s gatekeepers), or through emergency services (secondary care level). From this point onwards, patients’ movements are determined in accordance with the referral networks in use (and treated in a probabilistic way in accordance with past data). The model uses information on the probability of a patient, after using one of the entry points, being sent home (leaving the model, dashed black arrows), or being sent to another service. If the latter case occurs, we can make a distinction between a referral between levels (dashed grey arrows) or within secondary care (solid grey arrows). Within the referral system, the following referral options apply: referral from the PCC to an emergency service or to an outpatient consultation; and referral from inpatient care to an appointment with the user’s physician at the primary care level. There are both inter-hospital referrals (e.g. between inpatient care and emergency services) and intra-hospital referrals (inpatient admission directly after entrance into a hospital’s emergency service or making a new appointment for an outpatient consultation within the same hospital).

In analytical terms, the system is described as a set of 21 equations which establish the links between parameters and variables of the model – parameters make use of information provided by health authorities, and variables represent production and cost indicators computed within the system (these equations are available, on request, from the corresponding author).

The proposed model simultaneously considers the referral and catchment areas for populations (population served by each level of providers), whereas the smallest areas are the ones served by a PCC (see Figure 3). In terms of secondary care, in addition to offering services to the local population that are also available in the DH, the CH also provides specialised services to a wider population.

The model was implemented using Simul8 software (see Figure 4). To calculate the model’s parameters and validation, 2005 production and costs data were collected from several Portuguese health authorities.

Figure 2 Representation of the conceptual model with several healthcare services and providers.
We have faced some difficulties with this process due to the existence of sources with incomplete or contradictory information, or whose data-collection process was not clear. Validation of the model followed a black-box approach, with the model's output (number of consultations and referral values) being compared with the information collected from the real system and their respective confidence intervals. Given the lack of data on waiting lists, we had to assume that there were no initial queues to access the services. The model was run in an AMD 3800+ with three gigabytes of RAM, using Simul8 13.0 and Excel 2007 software. The warm-up period was 630 720 minutes (20% more than the longest event in the model, assumption made according to reference 27) and the data-collection interval was 525 600 minutes (one year, to coincide with the real data-collection period). The results were obtained after trials of five runs, and 95% confidence intervals for the average value were obtained. The results were found as expected, i.e. the model was found to be reproducing the current health network of services (values within the confidence intervals). (Specific information on the model and data in use may be requested from the corresponding author.)

Running the model produces results on several indicators, which include: queue mean sizes in the services, number of patients admitted and in the queues, average waiting times (and standard variation), number of consultations carried out, level of resources used (doctors in our model), costs, and other indicators built in on purpose, such as the weight of emergency care in total primary care costs.
Results and scenario analysis

Results replicating the system for 2005 have shown a high level of usage of medical resources for primary care, around 90% for ambulatory service and 20–30% lower for emergency care (as expected); secondary care resources were also highly used (90–100%). Total costs are presented in Table 1. With regard to costs, the Garcia de Orta’s weight (the only CH modelled), and the high weight of secondary care costs should be noted. This reinforces the idea of a system dominated by secondary care and highly specialised care (for which delivery is more costly). As a result of these facts, we have tested the impact of several policy-related scenarios on the model (see Table 2).

Scenario I

To test the system’s response to an increase in demand (for instance, as a result of population aging, or of tackling existing waiting lists or answering to current unmet need for care), we tested a 10% increase in demand for services, keeping all the remaining parameters of the model constant. Results show that the current system would reach a rupture point where most currently available resources are used at a level of 100%. This happens to 12 of the PCCs, and to the emergency and outpatient consultation services for the majority of hospitals, leading to the appearance of additional queues and an increase of waiting times. These results indicate that the current system is not prepared to cope with current and future increase in demand. Nevertheless, more information would be needed to run a more in-depth analysis of this scenario. In terms of costs (see Table 2), this scenario implies a substantial increase in the level of primary care costs, due to the direct pressure on the gatekeeping system.

Scenario II

A value around 50% has been suggested as the best proportion for both generalist and specialised physicians. Using Portuguese 2005 data, generalists only represent 37% of the total number of physicians. We tested a shift of physicians from secondary to primary care. Keeping the total number constant, the number of generalists was raised by 20%, and the number of specialised physicians diminished by 12%

<table>
<thead>
<tr>
<th>Table 1 Cost results obtained from the model for the year 2005</th>
</tr>
</thead>
<tbody>
<tr>
<td>Services/hospitals</td>
</tr>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td>2005 values reproducing the situation in 2005 for secondary care</td>
</tr>
<tr>
<td>Emergency (M€)</td>
</tr>
<tr>
<td>Inpatient care (M€)</td>
</tr>
<tr>
<td>External consultations (M€)</td>
</tr>
<tr>
<td>Total (M€)</td>
</tr>
<tr>
<td>Weight of each hospital in total hospital costs (%)</td>
</tr>
<tr>
<td>Total (M€)</td>
</tr>
<tr>
<td>2005 values reproducing the situation for primary care</td>
</tr>
<tr>
<td>174</td>
</tr>
</tbody>
</table>
Considering that a change in the resources distribution would also imply changes in utilisation, the demand for ambulatory care was raised by 5%, the referral rate from primary care to outpatient consultations was reduced by 20% (it is expected that a closer

<p>| Table 2 Cost results obtained from the model for the tested scenarios |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|</p>
<table>
<thead>
<tr>
<th>Services/hospitals</th>
<th>Garcia de Orta</th>
<th>Barreiro</th>
<th>Setúbal</th>
<th>Montijo</th>
<th>Litoral Alentejano</th>
<th>Total (Variation)</th>
<th>Weight variation of each service in total costs (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tested scenarios: results for secondary care</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Emergency care (% variation)</td>
<td>Scenario I</td>
<td>6.65</td>
<td>9.22</td>
<td>5.82</td>
<td>1.07</td>
<td>9.94</td>
<td>6.87</td>
</tr>
<tr>
<td></td>
<td>Scenario III</td>
<td>–0.02</td>
<td>0.06</td>
<td>0.04</td>
<td>0.00</td>
<td>0.17</td>
<td>0.03</td>
</tr>
<tr>
<td>Inpatient care (% variation)</td>
<td>Scenario I</td>
<td>2.68</td>
<td>6.52</td>
<td>5.18</td>
<td>0.74</td>
<td>6.90</td>
<td>4.21</td>
</tr>
<tr>
<td></td>
<td>Scenario II</td>
<td>–2.20</td>
<td>–2.31</td>
<td>–3.18</td>
<td>–1.15</td>
<td>–2.81</td>
<td>–2.45</td>
</tr>
<tr>
<td></td>
<td>Scenario III</td>
<td>0.00</td>
<td>0.07</td>
<td>0.08</td>
<td>0.00</td>
<td>0.10</td>
<td>0.04</td>
</tr>
<tr>
<td>External consultations (% variation)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scenario I</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>7.57</td>
<td>9.20</td>
<td>0.40</td>
<td>–0.79</td>
</tr>
<tr>
<td>Scenario III</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.32</td>
<td>0.01</td>
<td>–0.01</td>
</tr>
<tr>
<td>Total (% variation)</td>
<td>Scenario I</td>
<td>2.86</td>
<td>6.06</td>
<td>4.01</td>
<td>1.84</td>
<td>8.49</td>
<td>4.0</td>
</tr>
<tr>
<td></td>
<td>Scenario III</td>
<td>0.00</td>
<td>0.06</td>
<td>0.04</td>
<td>0.00</td>
<td>0.16</td>
<td>0.03</td>
</tr>
<tr>
<td>Weight variation of each hospital in total costs (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scenario I</td>
<td>–0.52</td>
<td>0.36</td>
<td>0.01</td>
<td>–0.05</td>
<td>0.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scenario II</td>
<td>0.04</td>
<td>0.13</td>
<td>–0.20</td>
<td>0.03</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scenario III</td>
<td>–0.01</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tested scenarios: results for primary care</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scenario I</td>
<td>8.03</td>
<td>9.33</td>
<td>0.10</td>
<td>5.43</td>
<td>0.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scenario II</td>
<td>4.65</td>
<td>–0.10</td>
<td>–0.36</td>
<td>–1.47</td>
<td>2.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scenario IIIa</td>
<td>20.77</td>
<td>–87.63</td>
<td>–6.56</td>
<td>–7.37</td>
<td>–5.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scenario IIIb</td>
<td>15.14</td>
<td>–87.63</td>
<td>–6.64</td>
<td>–5.37</td>
<td>–3.68</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Values presented indicate the percentage variation between the scenario results and values for 2005 (values in italic are the absolute difference between percentages)
follow-up of patients by a generalist physician leads to a slightly lower demand for secondary-level services, and the demand for emergency services was reduced by 5% (given the easier access to primary care services).

Results show that the reinforcement of the number of physicians at primary care level seems to allow for a better service, with the reduction of waiting times and resources used (for values around 80% in ambulatory care). For secondary care, the assumption of the demand reduction has shown that with a more-efficient gatekeeping process, the waiting times and the size of the queue for outpatient consultations significantly diminish (reduction of 71% in the time of access and 75% in the size of the queue for the three larger hospitals). Simultaneously, there was an increase in the weight of primary care in the cost's structure (an additional 2.26%), and a total saving of €7.19 million.

This policy scenario captures the potential impact of stimulating access to primary care and potentially reducing the demand for secondary care (possible by the gatekeeper role of general practitioners), while increasing the role of health promotion (versus treatment). The advantages of systems centred in primary care have been recognised in the literature. Analysis of this scenario indicates that it would allow for savings given the shift towards primary care that tends to be provided at lower cost. Nonetheless, this policy scenario should be seen as a long-term policy because it demands shifts in medical education numbers by specialty, in resources' planning and in population habits.

**Scenario III**

To simulate restructuring of primary care services, we tested the impact of closing 80% of the PCC’s emergency services (three PCCs located in areas with lower accessibility to hospital emergency services were kept open). Parallel to the closure of these services, we tried to simulate easier access to ambulatory care within PCCs. Some policy attempts towards this scenario have been tested recently in Portugal, with the creation of family health units (FHUs) together with experimental payment systems (EPS). Small emergency services in PCCs have been show to deliver low-quality and inefficient services given their lack of resources to answer to urgent demand.

This scenario attempts to catch the effect of closure of small-dimension emergency services and transferring the freed resources to ambulatory care, together with 95% of the respective demand (it is assumed that better access to general practitioners would lead to a lower use of services). The average consultation duration and cost after that change were computed as a weighted mean of the duration (or cost) of the ambulatory care and the emergency consultation. Finally, taking into account the latest report available regarding the impact of the FHUs/EPS in Portugal, the cost of each consultation is expected to decrease by 14.4% (scenario IIIa). In scenario IIIb we consider that there was no reduction on that cost. One of the consequences of this policy scenario has been the widening of the opening hours of ambulatory care services. We have also modelled this by extending the daily ambulatory care opening hours by 30 minutes for restructured PCCs.

Results show that while activity in secondary care effectively remains unchanged, there are improvements in primary care: an easier access to services (reduction of the average time of access and queues), and a reduction of around 30% in the resources used. Simultaneously, and from a conservative perspective (scenario IIIb), the costs of primary care decreased about 15%, which would allow a global saving of 5.37% to NHS costs. In scenario IIIa, the global savings represent a decrease of 7.37% in total costs.

**Conclusions**

This study has proposed a model to analyse networks of healthcare providers and services. This approach seems to be useful to analyse the impact of policy scenarios and planning options on the system’s efficiency and costs, and on the population access to services. The conceptual model was implemented using Simu8 software and applied to the Portuguese NHS. The underlying model and the implemented model could easily be adapted to test the effect of policies changing the balance between the primary and the secondary care sectors in other countries, such as the countries of the UK, and for several autonomous regions of the Spanish health system. For example, the proposed methodology could be used to test the impact of reorganising services within and across primary care trusts (which integrate hospitals and primary care centres) in England. Given the highly flexible nature of the DES methodology, the differences between the Portuguese NHS and other health systems might be accommodated with changes in the interactions between levels of care (primary and secondary), in the resources modelled, in the definition of the model’s outputs and naturally in the calibration of parameters (many of these modifications depend on the available data). The model presented is thus an exploratory and generic approach to analysing networks of providers.

Analysis has shown that the quality of results depends upon the quality and detail of data routinely collected from providers. The model was applied to a case study of the Portuguese NHS, and calibrated and
validated for the year 2005, and was shown to produce a wide set of useful information, including information on waiting times, queues, efficiency in the use of resources, and costs. The results of the applied model should be analysed with some caution given the weaknesses of data and the limitations of the modelling approach in use. In addition to the difficulty in obtaining real production data concerning the health system, and the problems of dealing with gaps and inconsistencies of available data, the model focused on the analysis of a limited number of services and resources. There were no data available on waiting times to validate the model, and demand did not account for current waiting lists. Several assumptions were thus needed, and one should take into account these assumptions when reading the results (e.g. no allowance for waiting lists for services and for unmet need means that we are underestimating the costs and not accounting for some access issues). Considering these factors, we suggest some key future developments for this work: improvement of the modelling of the interface mechanisms between primary and secondary care providers; running the model with more-detailed information on production, waiting lists and waiting times and decomposed financial information (that information should be used for a better calibration and validation of the model); development of the treatment of uncertainty and sensitivity analysis in the model; and the modelling of other services provided by primary and secondary care providers and by the tertiary care sector, as well as modelling of other resources used by providers.

For Portugal, the model was tested for three different scenarios, and results indicate that the current system is not expected to cope with a 10% increase in demand, while a shift of resources from secondary to primary care and a reorganisation of PCCs might improve efficiency and quality in the system (better utilisation of resources and lower waiting times) while decreasing costs. It seems that there are potential gains from strengthening the role of PCCs in Portugal.

REFERENCES

23. Oliveira MD and Bevan G. Modelling the redistribution of hospital supply to achieve equity taking account of

CONFLICTS OF INTEREST
None.

ADDRESS FOR CORRESPONDENCE
Mónica Duarte Oliveira, Centro de Estudos de Gestão do Instituto Superior Técnico, Technical University of Lisbon, Avenida Rovisco Pais, 1049-001 Lisbon, Portugal. Email: monica.oliveira@tagus.ist.utl.pt

Received 6 February 2008
Accepted 2 June 2008