

This is a post-print (i.e. final draft post-refereeing) of a manuscript submitted to Journal of the Operational Research Society.

To cite this article: Santos, S.P., Belton, V. and S. Howick (2008). "Enhanced performance measurement using OR: A case study". *Journal of the Operational Research Society*, 59(6): 762-775.

To link to this article: <http://dx.doi.org/10.1057/palgrave.jors.2602397>

ENHANCED PERFORMANCE MEASUREMENT USING OR: A CASE STUDY

Sérgio P. Santos*
Faculdade de Economia
Universidade do Algarve
Campus de Gambelas
8005-139 Faro
Portugal
Phone: +351 289 800915
Fax: +351 289 800064
E-mail: ssantos@ualg.pt

Valerie Belton
University of Strathclyde
Department of Management Science
40 George Street
Glasgow G1 1QE
United Kingdom

Susan Howick
University of Strathclyde
Department of Management Science
40 George Street
Glasgow G1 1QE
United Kingdom

(*) Corresponding Author

Abstract

The case study described in this paper aims to illustrate how qualitative and quantitative system dynamics modelling and multicriteria analysis can be used in an integrated way to enhance the process of performance measurement and management in the radiotherapy department of a major UK cancer treatment centre. The complexity of the radiotherapy process and its significance for patients present particular challenges for performance measurement and management. The paper discusses the benefits arising and the practical difficulties faced in the study.

Keywords: performance measurement; health services; multiple criteria analysis; system dynamics.

Introduction

Formal or informal performance measurement is common practice in most organisations and it is well established that this plays a critical role in signalling the level of success in achieving objectives and identifying where improvement efforts are required. Several performance measurement frameworks have been developed to address these needs; among these the balanced scorecard (Kaplan and Norton, 1992), the performance pyramid (Lynch and Cross, 1991), the results and determinants framework (Fitzgerald et al., 1991), and more recently, the performance prism (Neely, Adams and Kennerley, 2002), are better known examples.

An evaluative review of the literature highlights two key issues that need to be properly addressed if performance measurement systems are to provide a stronger basis from which to manage effectively for improvement. Firstly, an effective design process is required to help identify an appropriate set of performance measures and develop a system which: promotes understanding about the dynamic relationships and trade-offs between performance measures; facilitates synthesis of the performance information into a meaningful and manageable form; and enables decision makers to understand the reasons why a particular level of performance is observed. Secondly, in order to promote continuous performance improvement through better decision-making, the different stages of the performance measurement and management process must become part of a continuous cycle. Unless performance assessment is viewed as an integral part of a process of organisational learning, enabling decision makers to change the way they think and act and to make a more effective use of the available information, its ultimate impact will be limited.

Although it has been recognised that there are a number of Operational Research/ Management Science (OR/MS) methods which can potentially enhance the processes of performance measurement and management (e.g., Dyson, 2000; Smith and Goddard, 2002; Ackermann et al., 2005), published accounts of their use are rare. For example, Shutler and Storbeck (2002) comment on the difficulty in filling a complete special issue of an OR journal on the topic of performance management due to the lack of papers.

In a previous paper (Santos, Belton and Howick, 2002) the authors proposed a theoretical framework incorporating the integrated use of multicriteria decision analysis (MCDA) and system dynamics (SD) to assist performance measurement and management. The testing of the proposed framework in an organisational context, with the longer-term aim of refining the methodology, was identified there as an objective for further research.

In this paper we describe a case study, conducted in the radiotherapy department of the Beatson Oncology Centre, one of the major UK cancer treatment centres, which was part of a programme of action research to meet the above objective. The key research objectives of the study were to explore the value of the integrated use of SD and MCDA in providing support for performance measurement and management, and the nature of the practical difficulties and challenges that arise in doing so. Further to this, the paper contributes to the growing literature on the use of OR/MS in healthcare management (see, for example, the recent special issue of JORS edited by Davies and Bensley, 2005 and the forthcoming issue of the European Journal of Operational Research edited by Brailsford and Harper, editorial available online, October 2006) and in particular to that on the use of SD (e.g., Dangerfield and Roberts, 1999; Wolstenholme, 1999; Lane, Monefeldt and Rosenhead, 2000; Brailsford et al., 2005) and MCDA (e.g., Min, Mitra and Oswald, 1997; Bohanec, Zupan and Rajkovic, 2000; Bots and Hulshof, 2000; Mosmans, Praet and Dumont, 2002; Dolan, 2005).

We begin by outlining the performance measurement and management process and the potential to enhance this through the use of SD and MCDA. A brief description of the context of the case study follows. This is expanded as we describe in detail the development of a performance measurement system in accordance with the proposed framework. We conclude with a discussion of the impacts and challenges of the intervention.

The performance measurement and management process

At the most general level, a robust framework for performance measurement and management should revolve around a four-stage process, as illustrated in Figure 1, which assists the:

- *design* of the measurement system. The design process enables decision makers to develop an integrated set of performance measures, to set targets and to define the procedures for periodic data gathering;
- *measurement* of key metrics. The measurement process enables organisations to learn how they have performed (i.e., *what* happened). However to fully understand performance results, they must be properly analysed;
- *analysis* of metric results. The analysis process enables decision makers to understand *why* the organisation is performing as it is and to identify the need for corrective actions. An informed understanding of performance is an essential basis for design of effective action (to improve poor performance or to maintain and reinforce good performance);
- and *improvement* of operations. Through further analysis, the improvement process enables decision makers to explore the likely effects of alternative corrective actions on the organisation's performance, leading to better-informed decisions.

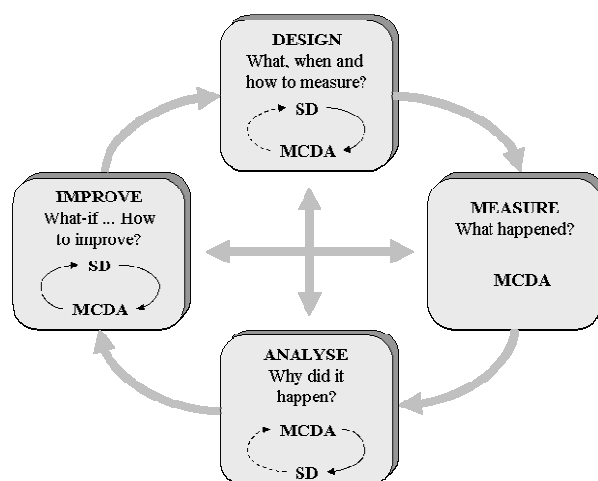


Figure 1 – The performance management process

Once measurement results have been collected, analysed, communicated (internally and externally), and used for the development of corrective action, effective performance management requires that the organisation's strategy and measurement system be updated accordingly to ensure that it continues to meet changing needs. Then the cycle begins again. Successful accomplishment of these four main tasks represents the foundation of good performance management. As described in Santos, Belton and Howick (2002), and indicated in Figure 1, we believe that SD and MCDA can make a significant contribution to the successful accomplishment of these four tasks and the closing of the performance measurement loop.

Both SD (see, for example, Forrester, 1961 and Sterman, 2000) and MCDA (see, Belton and Stewart, 2002, for an overview of the field and Keeney and Raiffa, 1976, or von Winterfeldt and Edwards, 1986, for more detailed descriptions of the multi-attribute value analysis approach used in this study) have been the focus of intensive research for more than three decades, and considerable developments have been seen in theory and practice. Each has individually proved their potential to inform and support decision-making. The strength of SD is in capturing the dynamics of a system, and that of MCDA is in providing a framework for exploring multiple, often conflicting, goals. Although the dynamic and multi-criteria nature of any system of performance measures is generally acknowledged (e.g., Mapes, New and Szwajczewski, 1997; Tatikonda and Tatikonda, 1998; Bititci, Turner and Begemann, 2000; Da Silveira and Slack, 2001), there is little published on the use of either SD or MCDA for performance measurement and management. Exceptions are: Rangone (1996), Bititci, Suwignjo and Carrie (2001), and Clinton, Webber and Hassell (2002), who use or mention the use of the Analytic Hierarchy Process approach to MCDA; and Davis and O'Donnell (1997), Boland and Fowler (2000), and Akkermans and van Oorschot (2005), who use or mention the use of SD. To the best of our knowledge, there is no published exploration of the synergy to be derived from the integrated use of MCDA and SD as part of a performance measurement system. Indeed, reported accounts of the formal integration of MCDA and SD are relatively few and they are in a different context (e.g., Hammond et al., 1977, 1978; Mumpower, Veirs and Hammond, 1979; Gardiner and Ford, 1980; Dennis et al., 1983; Gruver, Ford and Gardiner, 1984; Reagan-Cirincione, et al., 1991; Andersen and

Rohrbaugh, 1992; Brans et al., 1998; Kunsch, Springael and Brans, 1999; Rosas-Flunger, 2000; Kunsch, Chevalier and Brans, 2001; Schmidt and Gary, 2002; Springael, Kunsch and Brans, 2002; Chevalier, Kunsch and Brans, 2004; Olson, Dimitrova-Davidova and Stoykov, 2005; Rabelo et al., 2005).

Performance Management in a Radiotherapy Department

Radiotherapy along with surgery is a major modality in the management of cancer and it is used as a primary treatment for many types of cancer. Most commonly, radiotherapy is administered from an external source, usually a linear accelerator (also known as LinAc), with the beam being concentrated on the tumour. At the simplest level, radiotherapy may be part of a programme of treatments intended to cure the patient's tumour (radical radiotherapy), or it may be used to relieve symptoms of advanced disease (palliative radiotherapy). Whatever the intent, the use of radiation for the treatment of cancer is a complex, multi-step process, involving trained personnel who carry out a variety of interrelated activities.

Whilst measurement and improvement of performance are major concerns of most radiotherapy departments, these are particularly challenging managerial tasks due primarily to the complex nature of cancer treatment and to the multiple and sometimes conflicting needs and demands on the several stakeholders involved, including patients, health care professionals, administrative and managerial staff, as well as central Government. The wide range of views held by these stakeholders regarding what represents good performance demands a balanced set of measures to assess that performance and to ensure that trade-offs between objectives are explicitly considered.

The case study discussed here was conducted over a period of several months in the Beatson Oncology Centre in Glasgow, one of the United Kingdom's largest centres for non-surgical cancer treatment. It was organised in two main stages: the first stage was concerned with analysing the existing performance measurement practices within the department; and the second stage aimed to explore the use of SD and MCDA as means to add value to these practices. The study involved repeated visits to the

department, to interview clinical and managerial staff, including the superintendent radiographer (who has overall responsibility for radiographic staff and their standards of care), extensive study of relevant documentation, and collection and analysis of numerical data.

Findings from the first stage revealed that although tools and processes to measure organisational performance existed within the radiotherapy department, there was still considerable scope for improvement. In particular, the department could benefit from the development of a structured, systemic performance measurement framework based on a comprehensive and balanced set of performance measures.

The second stage of the study involved a series of meetings with key personnel responsible for the service under review, with the aim of: (1) developing a 'new' performance measurement system incorporating the integrated use of SD and MCDA, and (2) assessing the extent to which the 'new' system could overcome some of the shortfalls of the existing measurement practices as identified in the first stage of the case study research.

In the following three sections we discuss the process used to design the 'new' system to permit analysis of the levels of performance of the department, and exploration of possible improvement strategies. In the final section, we discuss the ability of this process to add value to the existing performance measurement practices within the department.

Designing a 'New' Performance Measurement System

Identifying the key performance areas

It is often suggested that the design of a performance measurement system should start with a clear identification of the organisation's mission and strategic objectives (e.g., Wisner and Fawcett, 1991) or with the identification of the stakeholders' wants and needs (e.g., Neely, Adams and Crowe, 2001). Assisted by the information gathered during the first stage of the case study, we began the second stage by trying to describe

what an ideal radiotherapy facility would look like from the point of view of the main stakeholders and what the main objectives of such a system would be.

A system of ideal radiotherapy cancer care would, among other things, ensure that: (1) all clinically eligible cancer patients within a facility's catchment area have timely access to radiotherapy services; (2) the care provided is clinically effective, in line with agreed standards, and is delivered with the minimum disruption to the patients' lives; (3) equipment is utilised at maximum efficiency and staff workload levels allow the delivery of a safe and high quality service; and (4) there are enough resources available to ensure the provision of an effective, efficient and accessible radiotherapy service. These four main themes – Fair Access to Services, Clinical Effectiveness and Outcomes, Efficiency, and Capacity and Capability – strongly reflect those found in the *Performance Assessment Framework* (NHS Executive, 1999), and were identified by the team tasked with developing the measurement system as key dimensions of performance on which to concentrate.

It is important to note that although the identification of these dimensions of performance constituted an important step in the overall assessment process, they did not immediately present a workable framework for detailed evaluation of the performance of the radiotherapy department. To thoroughly assess, for example, if a radiotherapy department is providing clinically effective health care or if it is ensuring that people's ability to obtain health care is related to their needs, these dimensions should be broken down further. In general, these dimensions should be decomposed to a level where they can be easily assessed. That is, they should be ultimately decomposed into performance measures.

Understanding the cause and effect relationships between areas of performance

Having identified the four main objectives and key dimensions of performance on which to concentrate, the system design process then focused on identifying the measures that best capture achievement of these objectives. At this stage, particular emphasis was placed on understanding and describing cause and effect relationships. A useful step in this direction and, ultimately, in identifying an appropriate set of measures was the use

of a simple causal loop diagram (CLD) to capture the key feedback structure of the radiotherapy treatment dynamics within the department. Causal loop diagrams are maps showing the causal links among a set of variables operating in a system, with arrows from a cause to an effect. CLDs are developed following well-established guidelines (e.g., Coyle, 1996, Chapter 2; Sterman, 2000, Chapter 5) and despite their limitations (e.g. Richardson 1986, 1997) they have proved highly effective in representing the feedback structure of systems. Using this tool, several feedback loops, and interlinkages between different loops, were identified, mapped, and subsequently integrated into a single diagram (Figure 2). This was developed in close interaction with the superintendent radiographer and was based on the information collected during several interviews. The purpose was to obtain a view of the components of the system and the dynamic interactions between them, leading to insights about the key performance drivers.

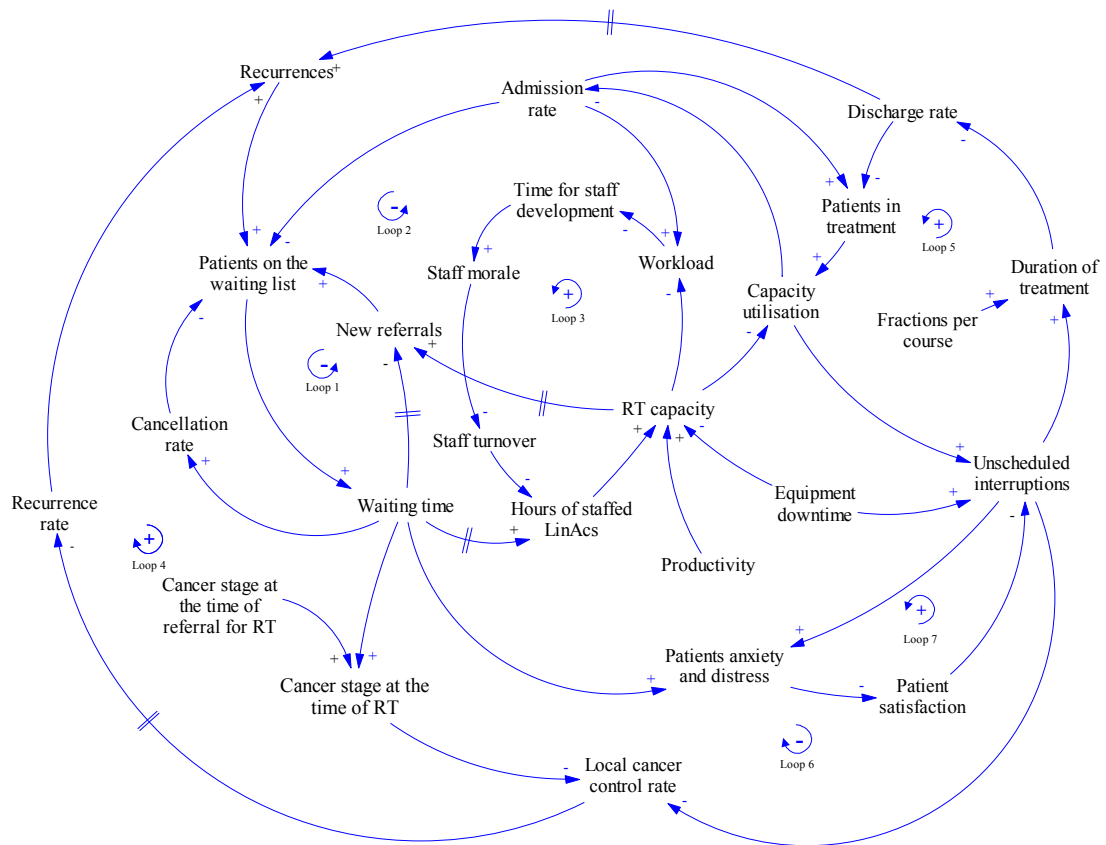


Figure 2 – A Simplified Causal Loop Diagram for the Radiotherapy Treatment Dynamics

Figure 2 is a simplified representation of the real system. Nevertheless, this simple model enables us to illustrate the insights that can be gained from its use. The CLD contains several feedback loops which reflect the interplay of the structural and behavioural forces in the system. Loop 1 and 2, for example, illustrate the key feedback structure of average waiting time for radiotherapy treatment. Loop 1 represents the demand side: other things being equal higher demand for radiotherapy leads to an increase in the number of patients on the waiting list, hence to longer average waiting times, but, in general, longer average waiting times lead to lower demand. As indicated in the diagram by the double bar symbol //, there is a delay between the moment when information about waiting times becomes available, and when it affects decisions about the use of radiotherapy. That is, it takes time for cancer patients and especially their surgeons and/or clinical oncologists to perceive changes in the radiotherapy waiting time and adjust their prescribing behaviour by looking for alternative forms of treatment (e.g., chemotherapy and/or surgery). It is important to note that in addition to new patients, the waiting list is also fed by patients who return for further treatment. As indicated, the recurrence rates are also, to some extent, influenced by the waiting times. Delays in treatment may allow progression of the tumour resulting in a worse outcome and, therefore, increasing the risk of recurrences.

Thus, the provision of timely treatment for cancer patients is critical as it increases the opportunity for optimal outcomes, including improved rates of cure, longer survival, and better quality of life. Because cancer cells continue to grow and divide much faster than normal cells while a patient waits for treatment, local control of the cancer may become compromised, with increased toxicity of treatment or lower rates of organ preservation if patients face prolonged waits.

The wait for treatment is also psychologically damaging for patients wishing to start therapy as quickly as possible and is therefore a major source of dissatisfaction.

Loop 2 is the managerial loop for regulating the department's capacity and it represents the supply side: increased waiting times result, after some delay, in more resources being allocated to radiotherapy treatment which, through an increase in admissions, leads ultimately to shorter average waiting times. As illustrated, the waiting list is reduced and the number of patients in treatment is increased, by an increase in

admission rate. Although high admission rates may lead to desirable consequences (e.g., lower waiting times), they may also lead to undesirable ones. The higher the admission rates and, consequently, the number of patients in treatment, the higher the staff workload. It is well known that unacceptably high workload levels limit the time available for staff development and, as a result, are damaging for morale (see Loop 3 in Figure 2).

The number of patients in treatment governs the extent to which the department's capacity is utilised. When that utilisation falls and there are vacancies, patients are admitted from the waiting list until all the available places are filled. It is also important to mention that utilisation levels must allow for maintenance downtime and for the treatment of urgent and, therefore, unscheduled patients. The patients in treatment are, in turn, reduced by the discharge rate. The number of patients discharged per month depends on the average duration of treatment. The overall radiation dose is delivered as a number of fractions, for example, a patient's course of treatment might consist of 30 fractions delivered on a daily basis (Monday to Friday) over a period of 6 weeks. The higher the number of fractions (i.e. the number of visits for radiotherapy) and the duration of gaps per course of treatment, the longer the treatments will last. On average, the discharge rate is simply a delayed version of the admission rate, the duration of the delay being the average duration of the radiotherapy treatment. Furthermore, and as previously noted, after a delay a proportion of those discharged will return for further treatment. Both the waiting times and the gaps in treatment cause the patient condition to deteriorate, adversely affecting the local control of the tumour, and therefore playing a critical role in the recurrence rates. Up to a point, a decrease in the local cancer control rate will increase the fraction of patients recurring and decrease the delay before they do so. Beyond that point, a decrease in the local cancer control rate will have an opposite effect, as patients become too ill to receive radiotherapy.

It is also important to note that overall cure rates are primarily determined by tumour stage at clinical presentation. As illustrated in the CLD, the local cancer control rates, which are an important measure of the clinical effectiveness of radiotherapy treatment, are affected by endogenous factors such as waiting times and gaps in treatment. However, they are also determined by exogenous factors such as the cancer

stage and grade or aggressiveness at the time the patient is referred to the radiotherapy department, factors over which the department has no control. If cancer is advanced at presentation or recurs after initial therapy, the outcome is often poor.

Selecting the key performance measures

Based on the CLD and on the increased understanding that emerged from its development, a small set of performance measures was derived and structured in a performance measures tree as shown in Figure 3. These measures, some of which were later broken down into more specific performance indicators, formed the basis of the performance measurement system. It is important to mention that this set was not intended to be comprehensive in covering all aspects of the department's activities but to throw light on some particularly important issues for its effective management. For example, while several staff groups are necessary to deliver radiotherapy, including consultant clinical oncologists, physicists, electronic technicians and therapy radiographers, it was decided to use the number of clinical oncologists and therapy radiographers as a proxy for the staffing levels. This decision was taken for simplification purposes and data unavailability for other types of staff. However, when carrying out an exhaustive performance assessment exercise, we recommend the use of data regarding the number of all types of professionals. It is also important to mention that the measures in Figure 3 were chosen because they support the goals and strategies of the department.

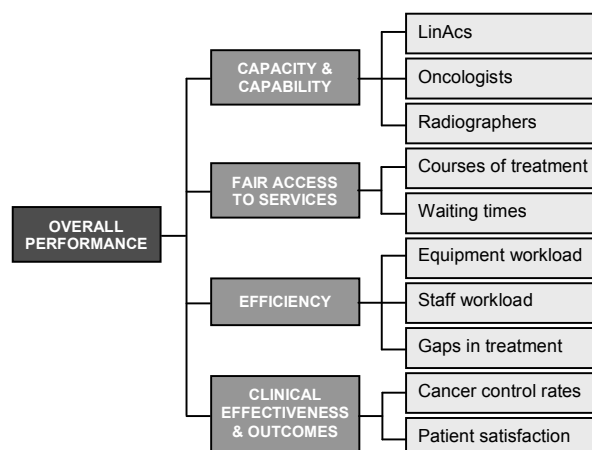


Figure 3 – Performance Measures Tree

The selection of this set of measures was significantly influenced by the key objective, which was to design a performance measurement system that would not only assess how the radiotherapy department was performing but also, and most importantly, would provide insights regarding the main causes for the performance observed and also insights about future performance.

The performance measure ‘unscheduled interruptions (gaps) in treatment’ is a good example of one that meets these objectives. Firstly, as revealed by the CLD and associated mapping process, it provides valuable information regarding the actual performance of the department. Monitoring gaps in treatment gives, for example, an indication of the efficiency in the use of resources. The causes for unscheduled treatment interruptions can be department-related, but may also be precipitated by deterioration in the patient’s condition. Nevertheless, a poor performance in this measure may indicate excessive equipment downtime, caused either by a deficient planning of machine servicing or by frequent breakdowns in machines.

Secondly, it helps to explain the level of performance achieved by other measures. Enforced and unexpected breaks in treatment are psychologically damaging for patients and a major source of dissatisfaction. Therefore, a poor performance in this measure may help in explaining patient satisfaction levels. Similarly, gaps in treatment may also be a major source of delays for patients awaiting treatment. When gaps in treatment occur, and no allowance has been made for compensation during the prescribed treatment period, the missed fractions are usually given at the end of the period, which means that the treatment will take longer than planned and, consequently, it will take longer for new vacancies to become available.

Finally, it has a predictive power as it provides useful insights regarding future performance and its likely implications for the department. It is well known that unscheduled and uncompensated interruptions in treatment have a detrimental effect on the local control of some cancers, significantly reducing the chances of achieving a cure. In particular, these interruptions impact significantly on the probability of cancer recurrence and on patients’ survival. Therefore, if the value of the measure ‘gaps in treatment’ is very poor at a particular point in time, it is very likely that the future performance of measures such as recurrence rates and survival rates will be poor.

A similar rationale was used to select the other performance measures. As mentioned previously, it is a critical objective of any radiotherapy facility to provide a clinically effective treatment to patients. Loco-regional control rates are a key outcome indicator of the underlying effectiveness of treatment and, therefore, they should be closely monitored. Furthermore, they are an important leading performance measure or performance driver of other outcome measures, including survival, quality of life, and recurrences. Having decided to measure the local cancer control rate, the next stage was to use the CLD to analyse what are the main factors under the control of the department that impact on this performance measure. As Figure 2 illustrates, both the waiting times and the unscheduled gaps in treatment have a detrimental influence on the local cancer control rate, and therefore, should be monitored. In addition, Figure 2 shows that both measures are a major source of patients' dissatisfaction, which is another key measure to consider. An important question then arose: how can the department control the waiting times and the gaps in treatments? As we discuss below, an effective planning of the machines' maintenance may help to reduce the gaps, and in this way, also reduce the waiting times. The problem of waiting time defies, however, a quick and simple answer because of the complexity of factors that might influence it. These may be short-term difficulties, such as machine breakdowns, temporary staff shortages, or long-term factors, such as an imbalance of patients and resources. Therefore, it is important that the measurement system includes measures that capture these factors, either directly or indirectly. In particular, measures should be selected to indicate if the department has enough resources available to provide an efficient, effective and accessible radiotherapy service. Through this process, a set of performance measures was identified (please refer to Figure 3).

As can be seen from Figure 3, the measures selected relate to structure, processes and outcomes, over which the radiotherapy department has some degree of control and which provide a sufficiently balanced assessment across the four key areas of performance considered.

Having identified the performance dimensions and measures considered to be relevant in evaluating the department's performance, the next step was to set targets for each of these measures and to find out how well the department was achieving them.

Measuring the Performance of the Department

The procedure we used to evaluate the radiotherapy department against each individual measure, and to arrive at an indicator of overall performance, made use of a hierarchical, weighted additive value function and was supported by the multicriteria decision support system V·I·S·A (Visual Interactive Sensitivity Analysis), as proposed elsewhere (Santos, Belton and Howick, 2002). Figure 4 illustrates elements of the model used to evaluate the annual performance of the department in the year 2002/03.

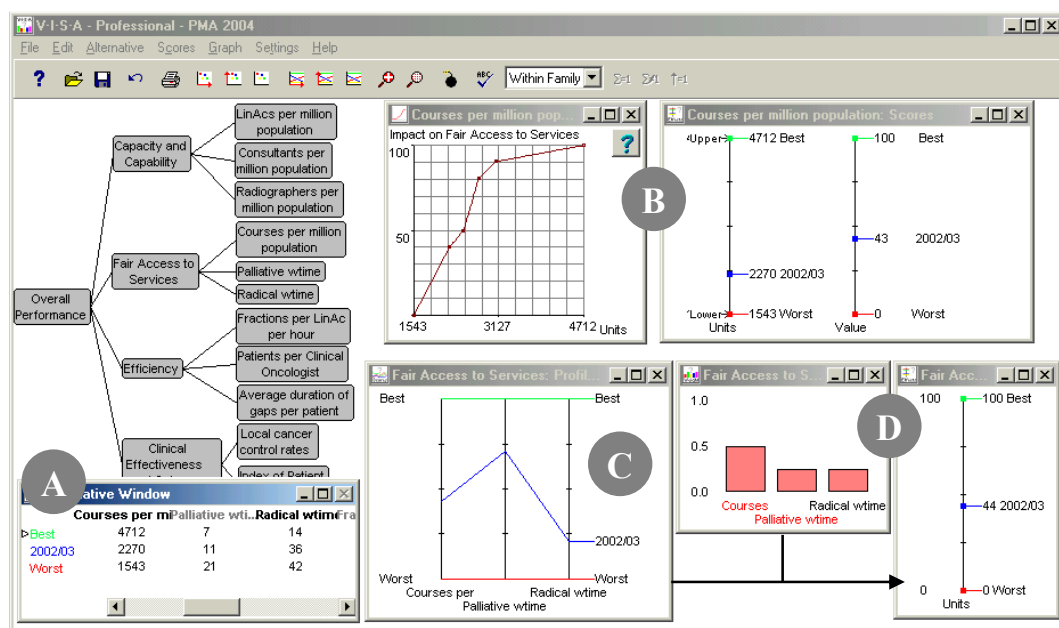


Figure 4 – A Section of the V·I·S·A Model

The measures discussed earlier are grouped under the four dimensions indicated to give the performance tree shown. The raw data in the *Alternative Window* (Figure 4-A) indicates the value of the performance measures on the natural scales for the year 2002/03. For example, we can see that from April 2002 to March 2003, the department delivered an average of 2,270 courses of radiotherapy per million population, and that the average waiting time from deciding to administer a course of radiotherapy to starting treatment was 36 days for patients awaiting radical treatment, and 11 days for patients awaiting palliative treatment.

Although these figures convey useful information, they are of limited value by themselves since they may be difficult to interpret in the absence of reference targets. For example, while a waiting time of 36 days for radical and, therefore, potentially curative treatment may be considered an acceptable performance in a cancer centre with limited radiotherapy resources, this is not likely to be the case in a cancer centre with no capacity shortages. It was therefore important to establish targets for each performance measure, and to analyse how well the department performed against these targets. Although it was desirable to define these targets based on a deep knowledge of the key stakeholders value systems, due to time restrictions on the part of some of these stakeholders (e.g., health care professionals and managerial staff), an alternative but also theoretically acceptable approach was adopted. The targets for each performance measure were defined by reference either to recommendations issued by prominent organisations in the area of radiotherapy treatment (e.g., Royal College of Radiologists) or to ‘industry benchmarks’ drawn from the results of surveys commissioned by the Department of Health regarding radiotherapy services in England and Wales (e.g., Department of Health, 1999; 2001; 2002). Based on this information, two reference points were defined for each performance measure, corresponding to the worst and best performance that could realistically occur. These levels were assigned values of 0 and 100 respectively. Once these reference points were determined, a value function was defined to assess intermediate scores using percentiles. Although this does not define a smooth value function, as would be expected, it was felt appropriate to fit the empirical data available. Figure 4-B shows a possible value function for the performance measure ‘courses of radiotherapy per million population’, and how the performances on the natural scale are converted, within the V-I-S-A model, into value scores on a 0 to 100 scale. In this case, a score above 50 means that the department performed better than half of its peers, while scores below 50 mean that the department performed worse than half of them.

Having defined the value functions for each performance measure, it was relatively straightforward to evaluate the levels of performance of the department against each of the bottom-level performance measures in the tree. Figure 4-C shows, for example, the performance of the department against the measures related to the

dimension ‘fair access to services’. Each performance measure is represented by a vertical bar and the performance profile of the radiotherapy department is shown by a coloured line crossing these bars at the appropriate point. It is easily seen if the department has a good (or poor) all-round performance, or combines significant strengths in some measures with weaknesses in others. In this particular example, it can be seen that the department is performing comparatively better on the measure ‘palliative waiting time’ than in the measures ‘courses of radiotherapy per million population’ and ‘radical waiting time’. However, an important question remains. Overall, is the radiotherapy department providing a fair access to its services? By themselves, the levels of performance achieved do not provide conclusive evidence of poor or good performance, unless they are analysed together with the department priorities and ‘values’.

Due to the limited resources available, it is important for health care organisations to define priorities. For example, is it a priority of a particular radiotherapy department to provide a relatively quicker service to palliative patients or to radical patients? Is it to broaden the access to radiotherapy even if this means that patients will have to wait longer, or is it to provide treatment to fewer patients but ensuring that the treatment will be provided quickly? Considering the performance levels represented in Figure 4, it is easy to conclude that the overall performance of the department regarding the fairness of access to its services will be higher, the higher the weight attributed to the performance measure ‘palliative waiting time’, and the lower the weight given to the performance measure ‘radical waiting time’.

When the scores of each performance measure within a performance dimension are combined with the respective weights, we obtain the overall performance score of the department in this dimension. Figure 4-D shows the aggregate score for the dimension ‘fair access to services’ if the relative importance of each of the measures contributing to this dimension is captured by the weights shown. The aggregate scores on each of the four key dimensions can be combined with appropriate weights to obtain an overall indicator of the performance of the department. This overall score can assist the management team to track performance over time and to see if particular changes made in the operations of the department have improved or hindered the overall

performance, whilst bearing in mind that the real learning comes from understanding why this has resulted from the actions taken and the stated priorities.

The weights reflect acceptable trade-offs among measures and should be elicited from stakeholders in accordance with their preferences and priorities for each performance measure, based on well-established axioms of decision-making. There are many questioning procedures which seek to elicit this information (see, for example, Kenney and Raiffa, 1976). However, in this study we did not attempt to elicit a definitive set of weight values; instead the emphasis was on using the model to illustrate the possibility to derive aggregate measures and explore the consequences of different priorities.

It is well known that achieving effective use of performance information calls for proper presentation to and discussion with users. To these ends we tested the use of a ‘traffic lights’ colour coding, which is familiar to some health care organisations in the UK (see Department of Health, 2000), as illustrated in Figure 5. This highlights areas of strength in the organisation (marked green) and areas for improvement (marked yellow or red), thus supporting planning for a process of continuous improvement. The colours assigned to each performance measure were based on the V·I·S·A scores.

It is important to emphasise that Figure 5 was used exclusively for illustrative purposes, as the lack of readily available data for some of the performance measures (e.g., local cancer control rates and patient satisfaction) prevented us from undertaking a definitive evaluation of the department at the time the case study was conducted.

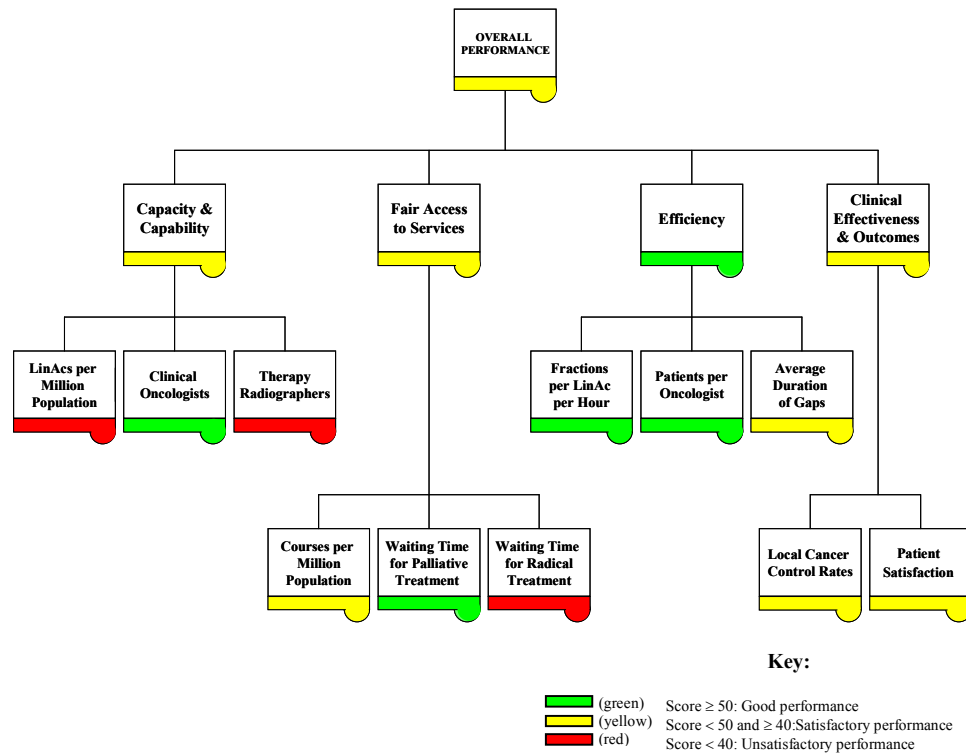


Figure 5 – Performance Results Using a ‘Traffic Lights’ System

Analysing the Results and Testing Improvement Alternatives

Having completed the organisational diagnosis, the next stage was to seek insights into the reasons for the levels of performance achieved and to explore possible responses. As discussed earlier, the Multi-Attribute Value Function (MAVF) approach, allowed us to look at individual performance measures and to synthesise that information to indicate how the radiotherapy department performed on each of the key dimensions and overall. However, whilst this information is important for monitoring the performance of the department and to indicate where improvement is required, it is of limited value if it is seen as an end in itself. As previously emphasised, to be effective, a performance measurement and management system has to support decision-making. It must inform decision makers about the reasons for poor (or good) performance, and provide insights about possible actions to achieve desired changes.

However, identifying the causes of problems and developing appropriate solutions are frequently difficult processes for the unaided decision maker. In order to try to understand the reasons for the levels of performance observed, we adopted a systems perspective. In particular, the analysis of the performance results focused on the underlying causal structure of the radiotherapy system discussed earlier (Figure 2). We use the performance measure ‘radical waiting time’ to illustrate this thinking.

As seen in Figure 4-C, the level of performance achieved in the measure ‘radical waiting time’ was unsatisfactory, resulting in a red traffic light indicating a substantially unmet target in Figure 5. With the objective of exploring possible causes of the observed level of performance, we started by analysing its pattern of behaviour over the previous four years. Considering that we adopted a systemic perspective in the development of the measurement system, we were interested in analysing patterns of change alongside information snapshots.

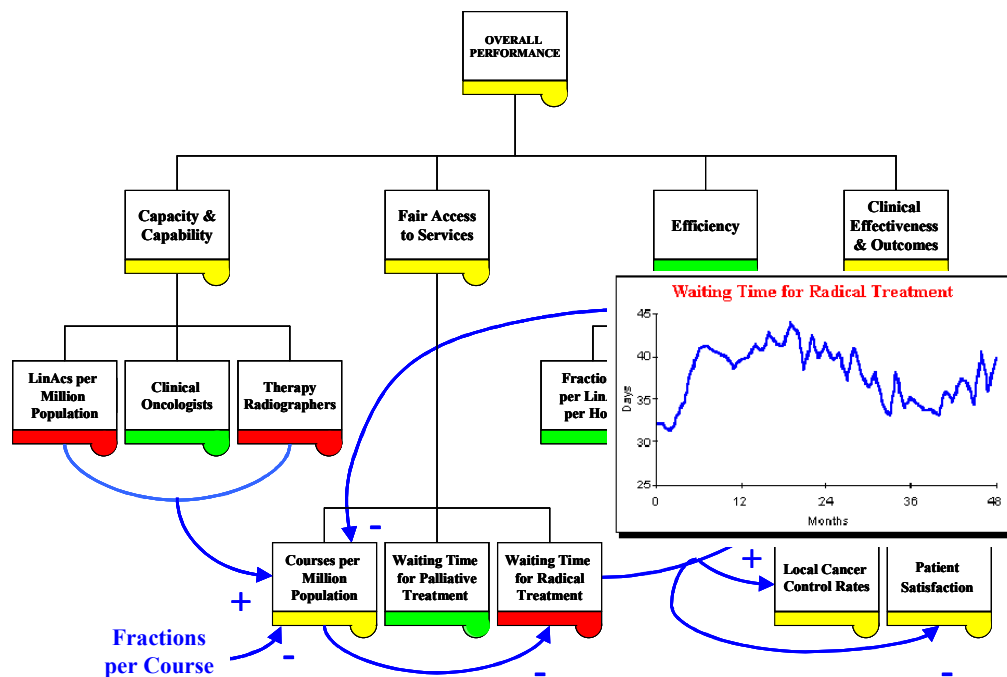


Figure 6 – Pattern of Behaviour of Radical Waiting Times

This revealed that the radical waiting times presented an oscillatory mode of behaviour over these previous four years (see Figure 6), suggesting that it has been generated by a negative feedback loop with delay, in accordance with loop 2 in the CLD. As previously mentioned, loop 2 is the managerial loop regulating the department’s

capacity. Long waiting times result, after some delay, in more resources being allocated to the radiotherapy department which leads ultimately to shorter average waiting times. Inspection of the capacity of the department over an identical time period (measured by the number of hours worked per month and the number of LinAcs available) revealed that the capacity may have played an important role in the behaviour of the waiting time performance for radical treatments.

It is indeed important to note that the MAVF analysis revealed that the levels of performance in two of the three capacity measures, is rather unsatisfactory. This is confirmed by the red indicators assigned to the measures ‘LinAcs per million population’ and ‘therapy radiographers per million population’ (Figures 5 and 6), which show that the capacity available in the department is below the minimum capacity judged to be necessary to provide a fair access to services. Coupled with the information about the links between measures (shown in Figure 2 and again in Figure 6) this suggested that capacity shortages may have played a key role in the level of the performance measure ‘radical waiting time’. This observation was in accordance with the belief of the management team regarding the fact that the unsatisfactory delays for radical treatments were mostly related with the capacity of the department, which was not believed to be enough to cope with the demand for the radiotherapy services. It was recognised that the department was understaffed, in part due to a national shortage of some staff groups (including radiographers and physicists) and ill-equipped due to underfunding. Nevertheless, there was a planned expansion of human resources and equipment.

However, and most importantly, the mapping of the CLD highlighted that many other factors in addition to available capacity influence the time that patients have to wait and should not be neglected in the analysis. Consideration of these factors, which was assisted by the CLD, led to valuable conclusions. Firstly, it highlighted the need to monitor, and if possible reduce, the equipment downtime. LinAc downtime, due to equipment breakdowns or regular servicing, reduces the capacity of the department and, therefore, negatively affects waiting times. Secondly, it highlighted that ‘changes’ in the average duration of treatment, due to increases in gaps in treatment and to changes in the average number of fractions prescribed per course of treatment, are likely to have played

a critical role in the behaviour of the waiting times. While the average duration of treatment for palliative patients remained broadly constant during the period of analysis, this was not the case for the average duration of treatment for radical patients. The pattern of behaviour of this variable mimicked, in some extent, the pattern of behaviour of the variable 'radical waiting times'. This was an interesting finding because it had not been anticipated by the management team. Furthermore, this issue is particularly interesting because fractionation is a controversial issue. The effect of different fractionation policies on the prognosis of patients, and on their experience of side effects, remains unclear in most cases, and more research is needed to clarify this issue (Commission for Health Improvement, 2001). Considering that local fractionation policies do vary considerably between radiotherapy facilities, and they are often developed based on clinical traditions or working practices rather than in solid empirical evidence, it could be very valuable to explore the effect that changes in these policies would have in the waiting times for treatment.

In order to analyse in greater detail the effect that these and other changes can have on the dynamics of the system, and to explore possible performance improvement actions, we developed a simple SD simulation model using PowerSim. Along with the explicit representation of the system's internal structure, a user-friendly interface was built to allow us to analyse different system configurations and scenarios (see Figure 7). The output of a PowerSim simulation/scenario was used to define a new alternative in the V-I-S-A model (with the two softwares being linked via a spreadsheet). By introducing changes in some parameters and running the SD model we could observe the changes that occur over time in the variables of interest, and the consequences of these changes as reflected in the multicriteria performance measurement framework. Figure 7 shows the investigation of a new scenario in which the average number of fractions per radical patient is reduced relative to the 2002/3 level. The impact on waiting times is seen in the displayed graph and the performance profile of the new scenario can be compared with the status quo in the V-I-S-A display. Considering that some of the performance measures are conflicting, it is likely that different actions will have different strengths and weaknesses. By cycling between the SD model and the MCDA

procedure, decision makers could enhance their understanding about these strengths and weaknesses, and make a better informed decision.

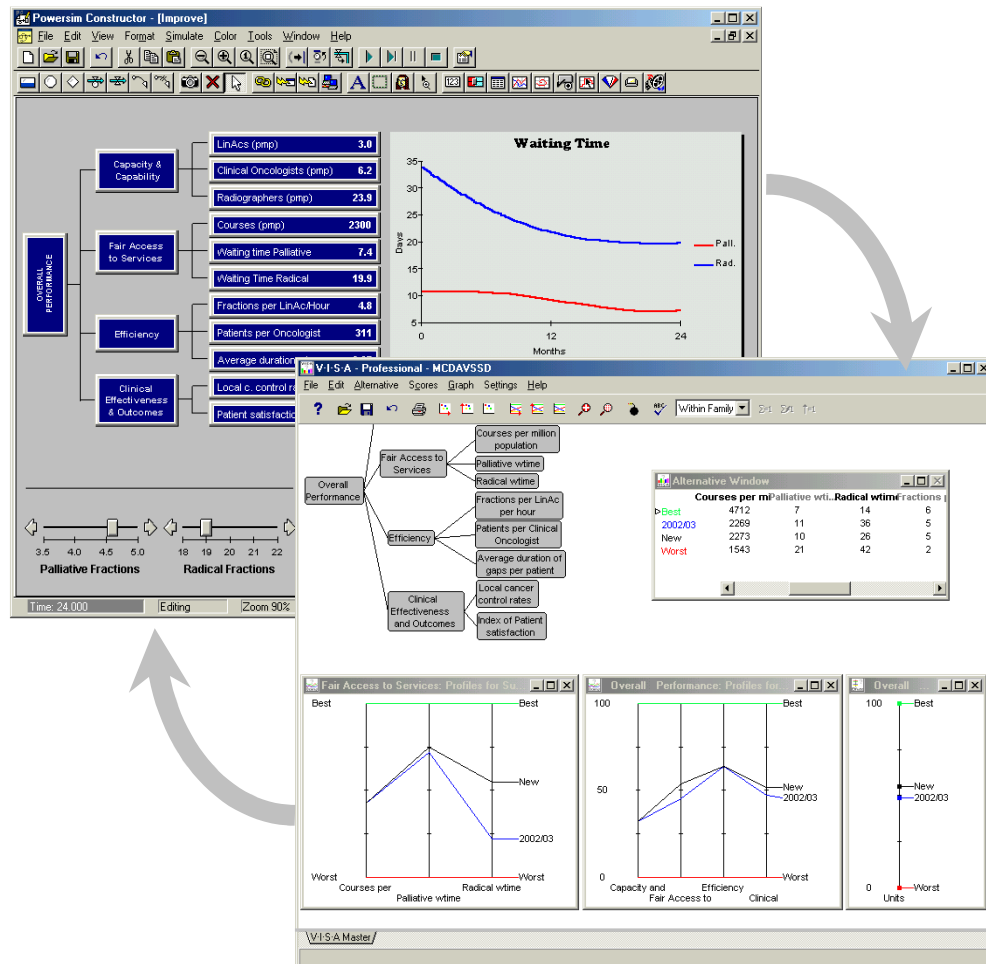


Figure 7 – Portion of the Simulation and MCDA Model

An important feature of the V-I-S-A software is that it allows interactive exploration of changes in inputs to the model, with the impacts being seen immediately in easy to interpret visual format. This sensitivity analysis can be used to explore the robustness of “conclusions” to priorities and values of different stakeholders, and when “disagreement” exists to highlight the rationale for this, thereby furthering discussion and understanding of others’ perspectives. The ability to “play” with the model is also a powerful means of building confidence in it and understanding of it. It is important to emphasise, however, that the possibility of running both the SD and MCDA approaches

in a single piece of software would make the implementation of the integrated approach easier as the user would have to manage only one interface.

Despite the restrictions on data availability and the limited involvement by the key stakeholders in the process of building the SD model, the model developed was still considered to be a valuable tool by those most involved in developing the measurement system. It allowed us to explore and assess the potential to analyse the effect of different actions on the system's performance measures over time. Furthermore, by linking the simulation model with the multicriteria decision support system the output could be easily synthesised and we could immediately evaluate the effect of these actions on each performance measure, aggregate dimension (such as "fair access to services") and overall performance. Had we had the opportunity to obtain reliable estimates for all the variables, we could have explored, for example, which of the factors discussed previously most influenced the radical waiting times. Furthermore, we could also have used the simulation model to test and compare alternative actions to improve the system's performance.

In spite of the limitations of the simulation model, the process adopted showed that the use of SD and MCDA in this integrated form, can offer a powerful frame to analyse why a particular level of performance is observed and the ways by which changes in system's performance may occur.

Discussion and Conclusions

This paper has discussed how SD and MCDA can enhance performance measurement and described the design of a performance measurement system for a radiotherapy department. This system aimed to address some of the limitations of the existing performance measurement practices, which included: absence of a structured measurement system able to evaluate the performance of the department on a regular basis; and an excessive focus of the performance assessment activities on a small number of individual measures at the expense of a more systemic and balanced view of performance.

The practical relevance of the use of SD and MCDA and their ability to enhance the practice of performance measurement and management within the radiotherapy department were assessed by using the information gained during the actual design of the performance measurement system, together with a follow-up workshop with the medical director, general manager and superintendent radiographer of the Beatson Oncology Centre. The case study has illustrated that the use of SD and MCDA, as part of an appropriate performance measurement system design process can bring new insights to inform and support the different stages of performance measurement and management within the department. In particular, the case study provided evidence that these approaches, used complementarily, permit:

- the development of a holistic view of the system being assessed, captured as a qualitative causal map, and improved understanding about the dynamic interdependencies and trade-offs among performance measures, leading to;
- the identification of an appropriate set of performance measures, which support;
- a meaningful synthesis, using MCDA, of the available performance information into a few key indicators of performance and/or into a single indicator of overall performance;
- increased understanding of the reasons why a particular level of performance is observed, and the ways by which changes in the organisation's performance arise from the combined use of these two models;
- rapid, but thorough, analysis and evaluation, in a risk free environment, of the effects of different actions on the department's performance can be carried out using the quantitative SD and MCDA models, fostering learning through experimentation;
- a proactive rather than reactive management of performance as a consequence of better understanding of the system and ability to predict the impact of changes.

However, the procedure used for designing and measuring the performance of the department is not without its limitations. Firstly, it is inherently subjective in the choice of measures, in their weighting and potentially also in the scoring process. However, this subjectivity is made explicit and is informed both by analysis and by objective data. Furthermore, thorough sensitivity analysis is recommended to explore the

robustness of conclusions to changed assumptions and to shed light on the significance of subjective judgements. Secondly, the use of MAVT to develop an aggregate performance measure, although theoretically well founded is based on strong assumptions about the decision makers' preferences (e.g. preference independence and a compensatory preference structure). Formal checks should always be made and if the assumptions are not observed, a restructured or alternative form of model should be considered. Thirdly, while the use of MCDA allows decision makers to see easily, via the aggregative model, how their beliefs and preferences turn into a single or a few key indicators of overall performance, this simplicity occurs at the expense of detail. This loss of information is, however, easily compensated by drilling down into the hierarchy as needed, to see the performance of lower level measures and further by reference to the SD model. Fourthly, the use of a MCDA approach requires judgements that may be difficult to elicit from stakeholders. Finally, quantitative SD models can be very demanding in terms of modelling skills, time, resources and availability of data. It is, therefore, fundamental to assess whether the quantification of the qualitative models is going to add enough value to the process, over and above what has already been gained, to make it worthwhile. Overall, however, while limitations and implementation challenges exist, the case study evidence suggests that the benefits of the use of SD and MCDA can compensate its limitations.

Finally, it is important to mention that the case study was not meant to be truly representative of all cancer treatment facilities, but instead was drawn to illustrate the applicability of and the insights emerging from the use of SD and MCDA to add value to the performance measurement and management process. However, although no two radiotherapy departments are exactly alike, the nature of the main problems they face is identical and, therefore, it is anticipated that similar conclusions would probably be obtained using other radiotherapy departments.

Acknowledgements - Financial support for the research carried out by Sérgio P. Santos was gratefully received from the Fundação para a Ciência e a Tecnologia, under grant SFRH/BD/823/2000. Acknowledgement is also due to the staff at the Cancer Centre that participated in this study. A very special thank you is due to Professor Alan Rodger, Dr. Adam Bryson, Ms Isobel Neil and Ms Franky Milne for their help, time and commitment, without which this study would not have been possible.

References

- Ackermann F, Belton V, Shafti S and Wisniewski M (2005). Building a Balanced Scorecard the OR Way. Paper presented at IFORS 2005, Hawaii, USA.
- Akkermans HA and van Oorschot KE (2005). Relevance assumed: a case study of balanced scorecard development using system dynamics. *J Opl Res Soc* **56**(8): 931-941.
- Andersen DF and Rohrbaugh J (1992). Some Conceptual and Technical Problems in Integrating Models of Judgement with Simulation Models. *IEEE T Syst Man Cyb* **22**(1): 21-34.
- Belton V and Stewart TJ (2002). *Multiple Criteria Decision Analysis: An Integrated Approach*. Kluwer Academic Publishers: London.
- Bititci US, Suwignjo P and Carrie AS (2001). Strategy management through quantitative modelling of performance measurement systems. *Int J Prod Econ* **69**:15-22.
- Bititci US, Turner T and Begemann C (2000). Dynamics of performance measurement systems. *Int J Opns Prod Mngt* **20**(6): 692-704.
- Bohanec M, Zupan B and Rajkovic V (2000). Applications of qualitative multi-attribute decision models in health care. *Int J Med Inform* **58**:191-205.
- Boland T and Fowler A (2000). A systems perspective of performance management in public sector organisations. *Int J P S Mngt* **13**(5):417-446.
- Bots PWG and Hulshof (2000). Designing multi-criteria decision analysis processes for priority setting in health policy. *J Multi-Criteria Decision Anal* **9**(1-3): 56-75.
- Brailsford SC and Harper P (Eds.) (2006). OR in Health. Special Issue of *Eur J Opl Res*. Forthcoming.
- Brailsford SC, Lattimer VA, Tarnaras P and Turnbull JC (2005). Emergency and on-demand health care: modeling a large complex system. *J Opl Res Soc* **55** (1): 34-42.

- Brans JP, Macharis C, Kunsch PL, Chevalier A and Schwaninger M (1998). Combining multicriteria decision aid and system dynamics for the control of socio-economic processes. An iterative real-time procedure. *Eur J Opl Res* **109**(2): 428-441.
- Chevalier A, Kunsch PL and Brans JP (2004). A contribution to the development of strategic control and planning instruments: An acquisition case study. *International Transactions in Operational Research* **11**(2): 155-168.
- Clinton BD, Webber SA and Hassell JM (2002). Implementing the Balanced Scorecard Using the AHP. *Mngt Acc Quarterly*. Spring 2002. **3**(3): 1-11
- Commission for Health Improvement (2001). *National Service Framework Assessments No. 1 – NHS Cancer Care in England and Wales*. Commission for Health Improvement: London.
- Coyle RG (1996). *System Dynamics Modelling: A Practical Approach*. Chapman & Hall, London.
- Da Silveira G and Slack N (2001). Exploring the trade-off concept. *Int J Opns Prod Mngt* **21**(7): 949-964.
- Dangerfield BC and Roberts CA (Eds.) (1999). Health and health care dynamics. Special Issue of the *Sys Dynam Rev* **15**(3).
- Davies R and Bensley D (Eds.) (2005). Meeting Health Challenges with OR. Special Issue of *J Opl Res Soc* **56**(2).
- Davis A and O'Donnell J (1997). Modelling complex problems: system dynamics and performance measurement. *Mngt Acc*. 18-20.
- Dennis RL, Stewart TR, Middleton P, Downton M, Ely D and Keeling MC (1983). Integration of technical and value issues in air quality policy formation: A Case Study, *Socio Econ Plan Sc* **17**(3): 95-108.
- Department of Health (1999). *A Survey of Radiotherapy Services in England (1999)*. Department of Health: London.
- Department of Health (2000). *The NHS Cancer Plan - A plan for investment, a plan for reform*. Department of Health: London.
- Department of Health (2001). *A Survey of Radiotherapy Services in England and Wales (2001)*. Department of Health: London.
- Department of Health (2002). *A Survey of Radiotherapy Services in England and Wales (2002)*. Department of Health: London.

- Dolan JG (2005). Patient priorities in colorectal cancer screening decisions. *Health Expect* **8**(4): 334-344.
- Dyson RG (2000). Strategy, performance and operational research. *J Opl Res Soc* **51**(1): 5-11.
- Fitzgerald L, Johnston R, Brignall S, Silvestro R and Voss C (1991). *Performance Measurement in Service Businesses*. CIMA Publishing: London.
- Forrester J W (1961). *Industrial Dynamics*. MIT Press, Cambridge, Massachusetts.
- Gardiner PC and Ford A (1980). Which Policy Run is Best, and Who Says So?. In Legasto AA, Forrester JW and Lyneis JM (Eds.). *System Dynamics: TIMS Studies in the Management Sciences* 14, North-Holland, Amsterdam, 241-257.
- Gruver WA, Ford A and Gardiner PC (1984). Public Policy Analysis Using Three Systems Science Techniques. *IEEE T Syst Man Cyb* **SMC-14**(2): 355-361.
- Hammond KR, Klitz JK and Cook RL (1978). How systems analysts can provide more effective assistance to the policy maker. *Journal of Applied Systems Analysis* **5**(2): 111-136.
- Hammond KR, Mumpower JL and Smith TH (1977). Linking Environmental Models with Models of Human Judgment: A Symmetrical Decision Aid. *IEEE T Syst Man Cyb* **SMC-7**(5): 358-367.
- Kaplan RS and Norton DP (1992). The Balanced Scorecard - Measures That Drive Performance. *Harvard Bus Rev* **70**(1): 71-79.
- Keeney RL and Raiffa H (1976). *Decisions with Multiple Objectives*. John Wiley & Sons: New York.
- Kunsch PL, Chevalier A and Brans JP (2001). Comparing the adaptive control methodology (ACM) to the financial planning practice of a large international group. *Eur J Opl Res* **132**(3): 479-489.
- Kunsch PL, Springael J, and Brans JP (1999). An adaptive multicriteria control methodology in sustainable development case study: a CO₂ ecotax. *Belgian Journal of Operations Research, Statistics and Computer Science* **39**: 109-143.
- Lane DC, Monefeldt C and Rosenhead JV (2000). Looking in the wrong place for health care improvements: a system dynamics study of an accident and emergency department. *J Opl Res Soc* **51**(5): 518-531.
- Lynch RL and Cross KF (1991). *Measure Up! The Essential Guide to Measuring Business Performance*. Mandarin: London.

- Mapes J, New C and Szwejczewski M (1997). Performance trade-offs in manufacturing plants. *Int J Opns Prod Mngt* **17**(10): 1020-1033.
- Min H, Mitra A and Oswald S (1997). Competitive benchmarking of health care quality using the analytic hierarchy process: An example from Korean cancer clinics. *Socio Econ Plan Sci* **31**(2):147-159.
- Mosmans A, Praet JC and Dumont C (2002). A decision support system for the budgeting of the Belgian health care system. *Eur J Opl Res* **139**(2): 449-460.
- Mumpower J, Veirs V and Hammond KR (1979). Scientific Information, Social Values, and Policy Formation: The Application of Simulation Models and Judgment Analysis to the Denver Regional Air Pollution Problem. *IEEE T Syst Man Cyb SMC-9*(9): 464-476.
- Neely A, Adams C and Crowe P (2001). The Performance Prism in Practice. *Measuring Business Excellence* **5**(2): 6-12.
- Neely A, Adams C and Kennerley M (2002). *The Performance Prism: The Scorecard for Measuring and Managing Business Success*. Financial Times Prentice Hall: London.
- NHS Executive (1999). *The NHS Performance Assessment Framework*. NHS Executive: London.
- Olson DL, Dimitrova-Davidova P and Stoykov I (2005). System Dynamics Model of a Transition Firm. *Managerial Finance* **31**(3): 67-80.
- Rabelo L, Eskandari H, Shalan T and Helal M (2005). Supporting simulation-based decision making with the use of AHP analysis. In Kuhl ME, Steiger NM, Armstrong FB and Joines JA (Eds.). *Proceedings of the 2005 Winter Simulation Conference*, Orlando, Florida, 2042-2051.
- Rangone A (1996). An analytical hierarchy process framework for comparing the overall performance of manufacturing departments. *Int J Opns Prod Mngt* **16**(8): 104-119.
- Reagan-Cirincione P, Schuman S, Richardson GP and Dorf SA (1991). Decision Modeling: Tools for Strategic Thinking. *Interfaces* **21**(6): 52-65.
- Richardson GP (1986). Problems with causal-loop diagrams. *Sys Dynam Rev* **2**(2): 158-170.
- Richardson GP (1997). Problems in causal loop diagrams revisited. *Sys Dynam Rev* **13**(3): 247-252.

- Rosas-Flunger R (2000). The system dynamics approach and the methodology for multicriteria decision aid as tools for organizational learning. In *Proceedings of the 18th International Conference of the System Dynamics Society*, Bergen, Norway, 176-177.
- Santos SP, Belton V and Howick S (2002). Adding Value to Performance Measurement by Using System Dynamics and Multicriteria Analysis. *Int J Opns Prod Mngt* **22**(11): 1246-1272.
- Schmidt MJ and Gary MS (2002). Combining system dynamics and conjoint analysis for strategic decision making with an automotive high-tech SME. *Sys Dynam Rev* **18**(3): 359-379.
- Shutler M and Storbeck J (2002). Performance management (Part Special Issue Editorial). *J Opl Res Soc* **53**(3): 245-246.
- Smith PC and Goddard M (2002). Performance management and Operational Research: a marriage made in heaven?. *J Opl Res Soc* **53**(3): 247-255.
- Springael J, Kunsch PL and Brans JP (2002). A Multicriteria Based System Dynamics Modelling of Traffic Congestion Caused by Urban Commuters. *Central European Journal of Operations Research* **10**(1): 81-97.
- Sterman JD (2000). *Business Dynamics Systems Thinking and Modeling for a Complex World*. McGraw-Hill, London.
- Tatikonda LU and Tatikonda RJ (1998). We Need Dynamic Performance Measures. *Mngt Acc* September: 49-53.
- von Winterfeldt D and Edwards W (1986). *Decision Analysis and Behavioural Research*. Cambridge University Press, Cambridge.
- Wisner JD and Fawcett SE (1991). Linking firm strategy to operating decisions through performance measurement. *Prod Invent Mngt* **32**(3): 5-11.
- Wolstenholme EF (1999). A patient flow perspective of U.K. Health Services: Exploring the case for new “intermediate care” initiatives. *Sys Dynam Rev* **15**(3): 253-271.