Bowers J, Lyons B & Mould G (2012) Developing a resource allocation model for the Scottish Patient Transport Service, *Operations Research for Health Care*, 1 (4), pp. 84-94.

This is the peer reviewed version of this article

Developing a resource allocation model for the Scottish Patient Transport Service

John Bowers, Gillian Mould and Bob Lyons
University of Stirling, Stirling, United Kingdom

Abstract

The Patient Transport Service is a vital component of many healthcare systems. However, increasing demand and constrained resources impose great challenges, especially in Scotland where there is a substantial remote and rural population. This case study describes the development of a decision support tool for strategic resource allocation decisions. The tool had to be relevant to management's practical needs including transparency to a range of stakeholders and a flexible speedy response to help management explore various operational and policy options. However, the tool also had to demonstrate rigour and identify an efficient allocation of resources. In response to these requirements, the decision support tool was constructed from simple models, verified in comparisons with more rigorous and sophisticated approaches, notably a Dial-a-Ride genetic algorithm and an open vehicle routing simulation. Using this tool, management were able to: identify a more rational, strategic allocation of resources; quantify the remote and rural effect; examine trade-offs between service level and resource requirements within various scenarios for future demand.

Key words: health service; vehicle routeing; allocation; decision support systems; methodology.

1. Introduction

The Patient Transport Service (PTS) provides the routine, non-emergency transport essential for many patients attending hospital clinics. In an environment of highly constrained budgets, the Scottish PTS struggles to meet demand and has to continually review its use of resources. The distribution of resources had tended to reflect historic patterns of demand and localised policies so a national review was instigated to ensure a fairer allocation of PTS resources across Scotland. A decision support tool was needed to provide a rational, transparent basis for this strategic review of PTS resources and to help develop an understanding of the possible trade-offs between resource provision and the level of service. Allocating resources simply in proportion to the number of patient requests in each area ignores the very different transport problems across Scotland: a model was required that explicitly incorporated these critical geographical differences.

Although the decision support tool focussed on strategic issues in PTS resource allocation, it had to reflect efficient tactical management practice. This implied detailed analyses of local historic data describing the detailed geographical distribution of demand and travel times to determine the resource requirements in each region, incorporating the stochastic geographical and temporal variabilities. The model also had to provide a rapid, flexible response enabling management to explore alternative scenarios. In addition it had to be explicable to a wide range of stakeholders who might be affected by any consequent reallocation of PTS resources. The strategic modelling requirements suggested that simple models would be preferred, providing the transparency, flexibility and speed necessary for a successful implementation. But the need to reflect efficient tactical practice implied the use of more rigorous routing models. This methodological tension was fundamental to the case study. The model development involved a parsimonious selection of scheduling and routing heuristics, striving to identify the simplest acceptable models and then comparing their performance with more sophisticated approaches. In particular the Dial-a-Ride Problem (DARP) and open vehicle routing problem provided key comparative models. This case study illustrates how it is possible to achieve a compromise between simplicity and rigour, with the simpler models being built upon the foundations of more rigorous approaches.

The PTS decision support tool was developed in a number of stages, with a degree of iteration to revise elements of the tool where necessary:

- analysing historic data to identify the distributions of PTS demand by time and by postcode across Scotland;
- analysing historic travel time data to estimate the times to travel between postcodes, including patients' pick up or drop off;
- constructing a geographical module with simple routing heuristics to determine the vehicle routing and hence the resource requirements to satisfy any specified mean demand for a single 30 minute time window for each hospital; this was based on a series of Monte Carlo simulation experiments reflecting the geographical distribution of demand, summarised in a metamodel;
- testing the simple routing heuristics in a series of comparisons with a more rigorous DARP genetic algorithm;
- constructing a temporal module which used the metamodel to develop an aggregate estimate of the complete resource requirements, given the distribution of demand over time at each hospital;
- testing the assumptions in the aggregations of the temporal module in comparisons with an open vehicle routing simulation.

2. Problem definition

2.1 Studies of ambulance services

Studies of ambulance services have usually focused on emergency transport [1,2,3]. The PTS receives relatively little attention and while there are some examples of useful studies of patient transport within hospitals [4,5], their scale is significantly different from the PTS problem examining strategic resource requirements across Scotland. In comparison to the emergency ambulance service, there are several critical features of the PTS:

- while the daily PTS demand is variable, it is known 24 hours in advance;
- patients usually share PTS vehicles, providing the opportunity for efficient routing;
- response time is a critical service measure for the emergency service; the PTS service levels were defined by the time patients might spend waiting for transport and also their journey time, including diversions to accommodate fellow passengers.

2.2 The Patient Transport Service in Scotland

The PTS is delivered by the Scottish Ambulance Service with a number of regional centres and a headquarters in Edinburgh responsible for resource allocation across the regions [6]. While the PTS has to respond to many different transport needs, the routine journeys are those between patients' homes and hospitals. Many of the journeys involve patients travelling to outpatient clinics at hospitals but there are a variety of possible destinations including local daycentres: in this study the term "hospital" is used to describe all such destinations. This transportation task typically involves 8000 patient-journeys per day and 800 PTS vehicles. The PTS strives to ensure that transport is provided only to those patients with a real need; without such management, demand would readily exceed the capacity of the service. The vehicles are stationed at a variety of depots across Scotland, some near hospitals, others at population centres. The vehicles' daily schedules are produced at regional centres where the individual patient requests for transport are co-ordinated 24 hours in advance, using a mixture of a computerised vehicle routing system and local knowledge. The prime objectives are to synchronise the transport with the hospital appointments and to avoid excessive patient-journey times, whilst ensuring an efficient use of resources. Ideally patients should not arrive more than 30 minutes before their appointment time, defining a time window for the delivery of patients at hospital; a similar target applies to the return journey when patients should not wait any longer than 30 minutes before starting their journey home. While it is equally important to have sufficient resources for the homeward journeys, the process is essentially symmetrical to that of transporting patients to hospital and it will not be described in any detail.

A key performance measure is associated with patients' sharing PTS vehicles: the first patient collected will be taken on a journey that involves detours to collect additional patients, with further time at each stop to allow boarding. The diversions, from a simple journey transporting the first patient directly to the hospital, could entail a substantial "excess ride time". The diversion time was particularly important in the debate about equity for remote and rural areas where large detours might sometimes be necessary to make more efficient use of resources. The resource requirements in urban and rural areas can vary substantially, as noted in other studies of public transport [7]: seat capacity might be critical in cities while patient-journey time constraints may be more important in rural areas. This paper focuses on the PTS vehicles; as a first approximation staffing is proportional to the number of vehicles but a more accurate estimate of the staff requirement requires a detailed consideration of the shift patterns and the needs of different categories of patients.

2.3 Delivering a services in remote and rural areas

The PTS in Scotland faces particular challenges delivering a service across great geographical diversity while demonstrating equity for all users. The service must cope with city centre congestion but also provide efficient transport for remote and rural communities. Although the large majority of the Scottish population live in urban areas, there is a significant proportion in remote or rural areas defined as:

- rural, with the population living in settlements of fewer than 3000 people
- remote, as living more than a 30 minute drive from a population centre of 10000

This definition implies that 19% of the Scottish population live in rural areas, with 6.4% being classified as both remote and rural [8]. Many population centres of 10000-50000 do not have a district hospital providing comprehensive services and some patients may have to travel significant distances to obtain specialist healthcare. However, a large proportion of PTS requests are for transport to smaller local health and day centres. A specific measure of remoteness relevant to this study was the travel time to hospital assuming that the vehicle makes a direct journey with no diversions to collect other patients. Figure 1 records the mean direct journey times illustrating the variation in remoteness across Scotland, though this portrayal obscures the fact that some patients do have to travel great distances to receive specialist care. However, many of the more remote areas have small populations and the consequences for the national PTS resource requirements are not immediately apparent. A major challenge for the model was to capture the effect of remoteness and rurality in an objective and equitable manner such that any suggested allocation of PTS resources could be justified to a wide audience.

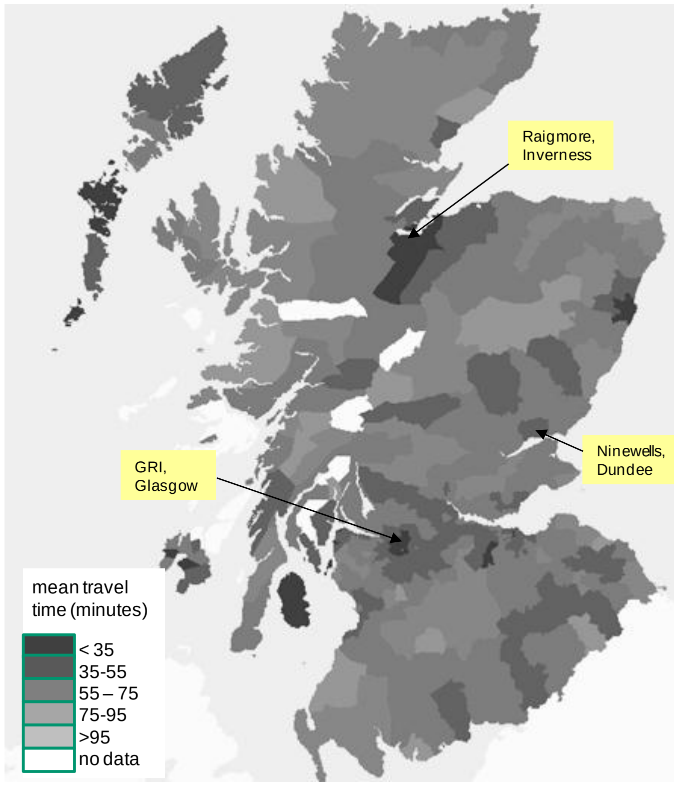

Figure 1 Mean PTS direct patient-journey time (minutes) to Scottish hospitals

2.4 The demand for the Patient Transport Service

The geographical distribution of demand was a key element of the study. Each patient-journey record contained the precise postcode of the patient's home address. However, such detail was not necessary and given the need to maintain confidentiality the analysis considered the postcode district, e.g. IV27 rather than IV27 4QA. This provided a reasonable degree of granularity for estimating travel times: 435 postcode districts span Scotland with a mean population of 12000 in each, though the more rural districts have considerably fewer people. In addition to the travel times, the time required to collect or set down a patient were incorporated in the estimate of the total transport time; later refinement of the analysis considered the effects of the varying traffic conditions during the day. Ideally, hospital appointments might be scheduled to maximise the opportunities for patients from similar geographical areas to share vehicles, minimising the PTS resource requirement. However, in practice clinical constraints normally dominate and transport needs are not considered when agreeing appointment times, thus the geographical and temporal distributions of the PTS demand are largely independent.

The eligibility of patients for the PTS is determined by their clinical need and mobility. One of the roles of the PTS model was to explore the consequences of changes in eligibility guidelines on demand and PTS resource requirements. A variety of vehicles are used in the PTS but the typical PTS vehicle has a capacity of 7 spaces. The vehicles can be readily adapted to meet patients' needs for additional space to accommodate escorts, wheelchairs or stretchers. Given the mean wheelchair, escort and stretcher requirement a vehicle with a notional 7 spaces has a typical capacity of 5.5 patients.

The temporal distribution of demand is illustrated by profile of arrivals at Ninewells hospital in Figure 2. Patients' appointment times dictate both the arrival and departure profiles; the arrival and departure times are specified in advance, incorporating a substantial allowance to accommodate the uncertainties within the clinic, or delays in arrival. This can imply that occasionally some patients have a long wait for their return transport. A more dynamic system might try to provide transport as soon as a patient has completed their appointment but this is likely to require additional resources and administration: further study would be needed to assess the costs or providing such a service. The PTS day usually begins with patients arriving for minor surgery and dialysis at 08:00. PTS traffic then grows with patients arriving for outpatient clinics, peaking at 10:00. A second batch of dialysis patients arrives at 13:00, followed by outpatients attending the afternoon clinics starting at 14:00 and then a final cohort of dialysis patients at 18:00. Figure 3 illustrates the associated profile of departures. The pattern is generally similar Monday-Thursday but hospitals may experience lower levels of PTS demand on Friday when there tend to be fewer clinics. These profiles are typical but the detail depends on the range of services and local practice in each hospital.

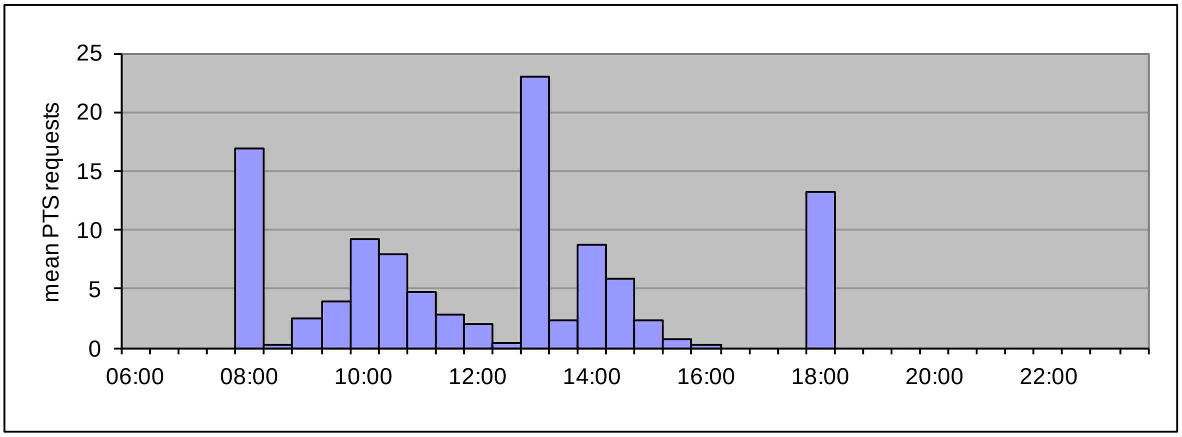

Figure 2 Mean PTS arrivals at Ninewells (30 minute intervals)

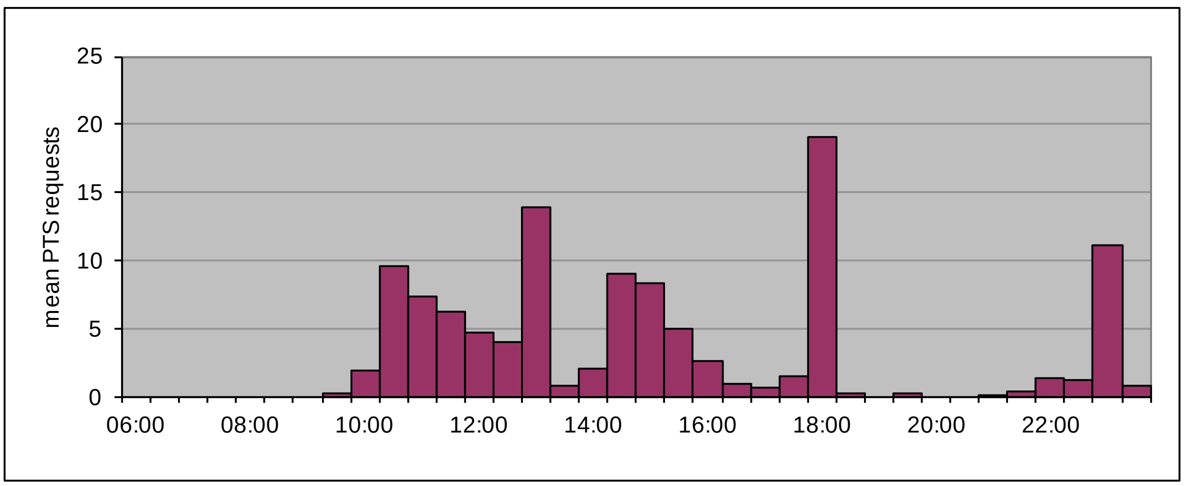

Figure 3 Mean PTS departures from Ninewells (30 minute intervals)

2.5 Modelling requirements

A major requirement was that the model's logic should be sufficiently transparent that it might be explained to a variety of stakeholders. Transparency encourages the greater engagement that is often critical in effective implementation [9] and is especially important when exploring a potentially controversial reallocation of resources [10]. Wherever possible simplicity was desirable in order to achieve a successful implementation with:

- transparency, enabling the logic to be explained to the various stakeholders who might be affected by the resource allocation decisions;
- flexibility, such that the model could accommodate additional features as the understanding of the working practices developed;
- speed of response with a decision support tool that could rapidly analyse the whole of Scotland's PTS activity for a range of policy options;

- timely delivery of the tool to support the scheduled review of the PTS.

Identifying the simplest requisite model is a key “craft skill” [11]: simplicity has to be combined with an appropriate degree of rigour. The model had to demonstrate good management practice implying an efficient use of resources, consistent with the most rigorous methods. But it also had to be relevant to all the needs of the Scottish Ambulance Service. Some research is predominantly concerned with rigour, to the detriment of relevance while other research is criticised as being too concerned with short term relevance; this study explicitly strives to bridge the potential gap between rigour and relevance [12].

2.6 Conceptual modelling

Conceptual modelling [13] emphasises the importance of explicitly considering the model scope and level of detail within a project life cycle reflecting the evolving understanding of the problem and modelling requirements. The conceptual modelling may well indicate that a single model will not suffice: it may be desirable to have two or more models providing different levels of detail [14]. Some studies have developed a suite of models, selecting the most appropriate for the particular task with a trade-off between solution quality and response time [5]. A dual approach was critical to the PTS model development, using rigorous tactical models to validate the simpler component models that were eventually included in the strategic decision support tool. The PTS model exhibited many of the features of the Dial-a-Ride Problem (DARP). In both problems there is a requirement to identify efficient routes to transport a set of passengers with specific pickup and delivery requests, and a potential to share vehicles [15]. There were also similarities with the open vehicle routing simulation [16] since the vehicles have considerable flexibility, finishing one journey and then starting to collect passengers for the next without returning to a central depot. The DARP and the open vehicle routing simulation could not offer the necessary flexibility, transparency, responsiveness and timely delivery. However, they did provide the essential foundation for the PTS decision support tool as a basis for comparison in a series of test cases, validating the use of the simpler models.

Developing a set of simple models with limited scope rather than attempting to construct a single generalised model is often recommended [17]. It was agreed that a parsimonious approach should be adopted, striving to identify the simplest component models that could still provide a sufficiently accurate estimate of the resource requirements. Each of these models had to be justified to the stakeholders before assembling the complete model and undertaking the full analysis. This gradual development via a set of simpler models enhanced all parties’ systemic understanding of resource allocation in the PTS, in addition to providing a practical approach to validation.

3. The theoretical basis of the model

3.1 An example of a Dial-a-Ride Problem

The PTS schedules vehicles in response to individual patients' needs rather than following fixed routes and schedules. Hence the PTS routing problem has many features in common with various DARP formulations [15] notably: time windows; static with a specific set of requests at each day; an objective to minimise patient dissatisfaction with constraints on diversion and early/late-ness. The PTS problem is multi-vehicle with the number of vehicles being modelled as part of the objective function, rather than a constraint: all transport requests are satisfied but the routes are selected in order to minimise the vehicle requirement subject to service level constraints.

Patients' diversion from the shortest routes to hospital was an important issue in the PTS problem, especially when examining the resource requirements in remote and rural areas. The diversion time or "excess ride time" has been recognised in DARP studies as a key measure of the passenger service level [18,19] and is particularly important in healthcare applications [20]. The PTS routing problem is relatively simple in many respects: the vehicles tend to deliver (or collect on the reverse journey) patients from a single hospital, or at least very similar locations; in comparison the standard DARP vehicle often transports passengers with a great variety of destinations. However, the PTS routing problem exhibits "open vehicle" features with the vehicles having the freedom to travel direct to the start of the next journey rather than always returning to a specified depot [20]. A critical characteristic is the scale: reported DARP schedules typically consider 100-200 passengers [15] while the PTS routing problem involves 8000 patient journeys on a typical day, although this may be considered as 30 independent sub-problems defined by the destination hospital.

3.2 DARP and stochastic demand

The standard DARP analysis considers a static problem, identifying an efficient solution to the transport of a specific set of passengers defined by their home and destination locations. Some studies have examined dynamic variants of DARP using insertion heuristics to accommodate new requests [18,21]. Other studies explore the stochastic nature of demand, typically using simulation [22,23,24,25]. The simulations, often incorporating many local operational issues [25], are valuable in assessing specific DARP systems: the vehicle availabilities are specified and the output of the simulation includes statistics describing the cost and service for the passengers. However, the objective of the PTS routing exercise was to provide a basis for strategic resource allocation, identifying the distribution of the resource requirement implied by the stochastic PTS demand. A Monte Carlo simulation framework provides a mechanism for estimating this distribution, iteratively applying a vehicle routing algorithm to determine the resource implications of multiple sets of daily PTS requests. Incorporating a DARP algorithm within a simulation always imposes practical modelling challenges with the multiple iterations increasing the scale of the routing problem. There is an emphasis on speed and robustness, accepting some trade-off with optimality [26]. This emphasis is even greater in the PTS model where additional iterations are needed to estimate the resource requirement.

3.3 DARP algorithms

A further challenge was the requirement that the model should provide responsive decision support, allowing management to explore future resource allocations in the context of changing demand and budgetary constraints. This implied a need for a responsive model that incorporated a fast, reasonable approximation for vehicle routing, rather than a slower, optimal approach. A variety of DARP algorithms are available to support tactical decisions, providing vehicle allocations and routes for specific sets of passenger requests. These employ various approaches, e.g. branch and bound, tabu search and genetic algorithms [15]. When analysing a single set of transport requests, the computer processing time required for such approaches is not usually a significant barrier to implementation, even when operating in real time. However, considerable computer processing time is needed when these methods are applied to substantial problems [18,19], as in the PTS model. The more rigorous DARP algorithms provided a valuable standard for comparison but they proved to be impractical for the strategic PTS model and a quicker, simpler approach was necessary.

Simple nearest neighbour heuristics have been adopted as components of vehicle routing algorithms in other DARP studies [19,27,28]. The algorithm reviews the options for selecting passenger i+1, given that a partial route including passenger i, has already been determined. It then calculates the extra time to divert and pick up each of the candidate additional passengers, compared to a direct journey to the destination with just i passengers. The consequences for the diversion time of the first passenger are then compared to the specified limit; this maximum diversion time is often the major constraint in DARP applications where the density of demand is relatively low, as in the PTS application. A refinement of this basic nearest neighbour heuristic is to extend the set of options, considering the possible next pairs of passengers i+1 and i+2. Experiments suggested that such refinements were not necessary in the current application and that the basic heuristic was sufficient, as confirmed in later validation of the model. Multiple vehicle routing usually takes either a sequential approach constructing the route for each vehicle in turn, as in this study, or a parallel approach developing the routes for each vehicle in the fleet simultaneously [29]. The parallel approach produces routes with a more equitable workload but it assumes that the fleet size is known; in the PTS routing problem, the objective is to minimise the number of vehicles required, thus a sequential approach is more appropriate.

4. Developing the decision support tool

4.1 Overview of the tool

The decision support tool assimilates the geographical and temporal patterns of demand with good vehicle routing practice to deduce resource requirements, given specified service levels. The interface provides the PTS management with a number of decision variables enabling staff to explore issues such as assumptions about future demand in different patient-categories or particular regions. The tool, known as PTARMIGAN (Patient Transport Allocation of Resources Model Incorporating Geographical Analysis), is currently in use at the Scottish Ambulance Service headquarters to routinely review policies and project future resource requirements. It consists of three modules encoded in Access, Excel and Visual Basic:

- An analysis of the historic PTS database, cleaning and importing relevant data from the database to provide the basis for the geographical and temporal modules. These data are collected routinely by the PTS during four separate weeks each year chosen to reflect the seasonality of demand, with typically 30000 records for each of the selected weeks. The data are provided by the PTS staff as part of the Scottish Ambulance Service routine quality management processes. The analysis is not described in this paper but it did require considerable effort, devising heuristics to clean the data and producing summary management reports as well as the crucial modelling input.
- A geographical module, described in 4.2-4.3, which uses the distribution of the patients' home locations and the travel time data to construct a metamodel $V(h,n,a)$. This metamodel describes the number of vehicles in transit a minutes before the specified arrival time, pursuing their various routes to satisfy a specified number of

PTS requests n for a single time window to a hospital h. The complete vehicle requirement can be summarised by the peak activity $V(h,n,0)$, when all the vehicles converge on the hospital, and the total vehicle-hours reflecting the different routes of varying durations.

– A temporal module, described in 4.6-4.7, using the output from the historic analysis to establish the distribution of PTS requests over time for each of the hospitals in Scotland; these patient demand profiles are then transformed into vehicle requirements using the metamodel derived in the geographical module.

The division of the geographical and temporal module was critical to the design of the PTS model. In many analyses this design separated much of the intensive calculation, including the Monte Carlo iterations, from the modelling of specific scenarios, ensuring that the model provided responsive decision support.

4.2 The geographical module

The geographical module answers the hypothetical question, "If n patients require transport to hospital h, arriving at the same time from diverse home locations, how many vehicles V are required?" A number of heuristics were examined based on the nearest-neighbour concept [27]. The classic multiple travelling salesmen "greedy algorithm", using a serial allocation routine [29] was adopted as the basis for the geographical module. The choice of heuristics was dictated by a parsimonious principle, selecting the simplest heuristics that could still provide reasonably efficient routing. An important requirement of the geographical module was that it should produce an equitable allocation of resources, with particular reference to the remoteness and rurality of a significant proportion of the Scottish population. A simple definition of equity was agreed with the stakeholders: the vehicle routing should strive to ensure that no patients' journey time to hospital exceeds their direct journey time by more d minutes; any diversion and collection of additional patients should be completed within the specified d minutes, whether the area is urban or remote and rural. Providing an equitable service inevitably implies that vehicles serving remote rural districts with scattered communities will tend to have fewer passengers. The parameter d offers a mechanism for exploring the trade-off between service level and resource utilisation: increasing the maximum permissible diversion time d may reduce the vehicle requirement but at the expense of a poorer service for all patients, whether urban or remote and rural. The specification of the value of d will depend on management's judgement about the needs of the service users; in this study, the typical values of d were 40-80 minutes.

4.3 The routing algorithm

The simple algorithm included in Figure 4 serially constructs routes for vehicles, selecting transport requests from those as yet unallocated. The maximum permissible diversion time is a significant constraint in this application implying that the construction of the routes begins with the most remote patient and then identifies acceptable diversions from a direct route to hospital to collect additional patients.

s_i	= seat requirement for patient i
c	= vehicle seat capacity
l_i	= location of patient i
l_h	= location of hospital h
$m(l_1, l_2)$	= travel time between locations, e.g. the homes of patients 1 and 2
p	= pick up time for a patient to board the vehicle
d	= maximum diversion time (from the direct route to hospital)

$M(l_1, l_2 \ldots l_j, l_h)$ = total travel time for the most remote patient i=1 to be transported to the hospital h along a route defined by the home locations of the selected j patients: $l_1, l_2 \ldots l_j, l_h$

$$M(l_1, l_2 \ldots l_j, l_h) = \sum_{k=1}^{j} \left(m(l_k, l_{k+1}) + p \right) + m(l_j, l_h) \tag{1}$$

Having identified and satisfied the PTS requests 1, 2...j the remaining unallocated patients are considered in order of their travel time from patient j. The additional travelling time implied by introducing an extra patient into the route is compared to the direct route for the first patient collected. The nearest unallocated patient to j satisfying the maximum diversion constraint, if such a patient exists, is adopted as patient j+1 such that:

$$M(l_1, l_2 \ldots l_j, l_{j+1}) + p - m(l_1, l_h) \leq d \tag{2}$$

and the vehicle's seat capacity is sufficient:

$$\sum_{k=1}^{j+1} s_k \leq c \tag{3}$$

When no more patients can be added to the vehicle's route, an extra vehicle is deployed, repeating the routing process until all patients' transport requests have been satisfied. The algorithm is described in terms of the journey from patients' home locations to the hospital but a similar approach was employed when analysing the return journeys.

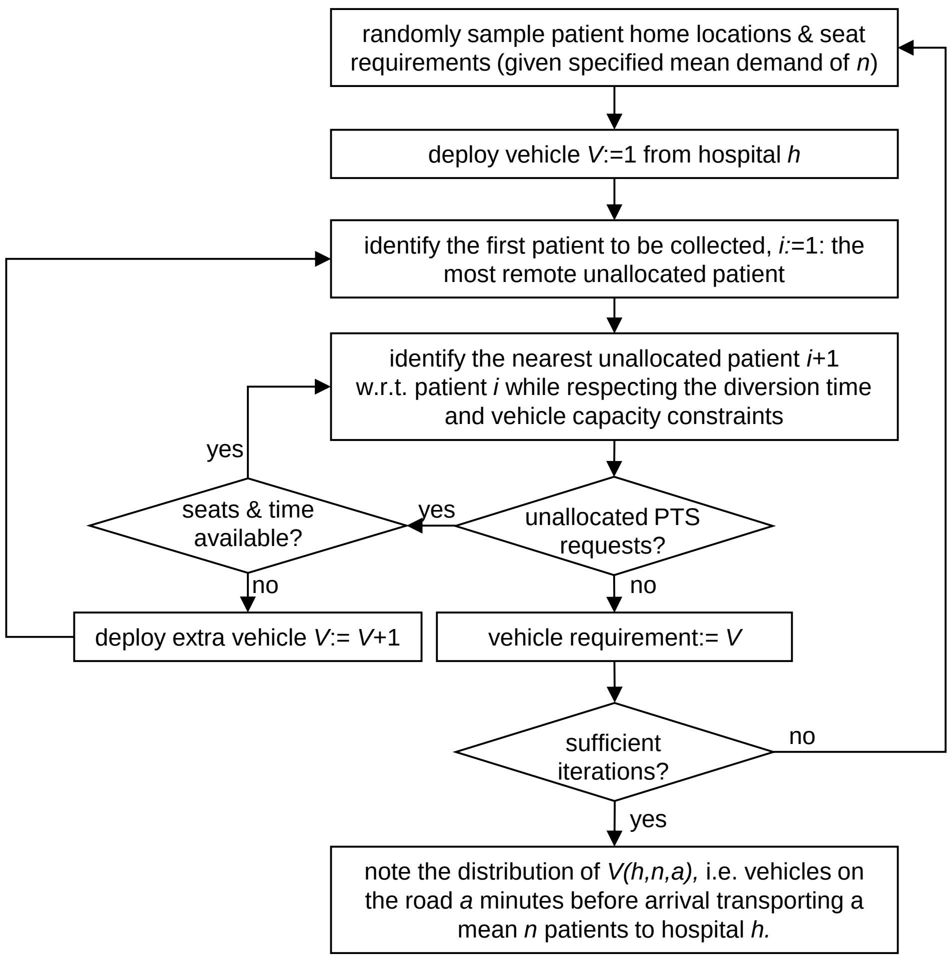

Figure 4 Estimating the distribution of the vehicle requirement $V(h,n,a)$ for delivering n patients a minutes before arrival at a specified hospital h

4.4 Comparison with a genetic algorithm for the DARP

While the refined routing algorithm had the attraction of simplicity, it was essential that it could provide sufficiently efficient routes. The output from the simple heuristics had to be comparable to that offered by more rigorous models. Flexibility in formulation was essential: the model had to be adapted as the shared understanding of the problem developed. A DARP genetic algorithm (DARP GA) was chosen as the basis for comparison because of its ability to readily incorporate multiple constraints and refinements in the objective function. A series of test cases were examined, comparing the performance of the routing algorithm with that of a DARP GA derived from similar applications [18]. Details of the comparison are provided in Appendix A. The two methods' mean performances were very similar with the journey time for the DARP GA's solutions being 99.7% of that offered by the geographical module. However, the geographical module's routing algorithm may not be suitable for other studies: each case requires a careful comparison with a well proven rigorous standard approach, such as the DARP GA.

4.5 Applying the routing algorithm

The routing heuristics were embedded within a Monte Carlo simulation framework as illustrated in Figure 4. This was used to estimate the vehicle requirements for repeated sets of PTS requests with a specified mean demand of n patients. At each iteration of the simulation the peak vehicle requirement $V(h,n,0)$ was recorded together with the vehicles in transit a minutes before arrival to develop the distribution of the complete vehicle requirement $V(h,n,a)$. This distribution reflected three stochastic elements:

- the number of patients;
- the seat requirements of each patient; patients may need more than one seat to accommodate an escort and a wheelchair or stretcher;
- the home locations of the patients.

The simulation described in Figure 4 was repeated considering different values of n to construct a response function summarising the relationship between demand and the vehicle requirement [30]. This process was replicated for each significant hospital destination h to produce a complete metamodel $V(h,n,a)$ which was then incorporated in the decision support tool. The more routinely debated management decision parameters, such as the volume of demand at different hospitals, could be adjusted and the resource requirements determined rapidly. However, when more fundamental assumptions, such as the value of the maximum diversion time d, were questioned, the full detailed stochastic routing analysis had to be repeated.

Figure 5 summarises the estimates of the mean (± 95% confidence interval) vehicle requirement for three hospitals for a range of mean demand within a single 30 minute time window (n=1, 2, 5, 10…30). All three hospitals exhibit some economy of scale as the number of PTS requests increases. Providing transport for just a few patients, n = 1-5, offers limited opportunities for sharing a vehicle without imposing an excessive diversion for the patients. With more requests there is a greater chance that more patients can share the same vehicle without exceeding the diversion time constraint.

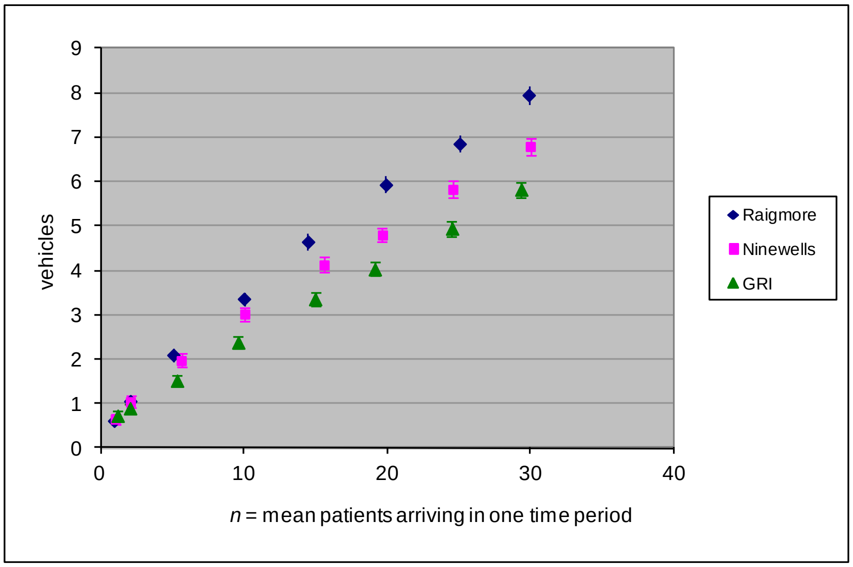

Figure 5 $V(h,n,0)$ = Vehicles required to deliver n patients to a specific hospital h (d = 60 minutes)

4.6 Modelling the complete journey

$V(h,n,a)$ describes the instantaneous requirement associated with the arrival of vehicles transporting a cohort of patients to hospital h all sharing a common appointment time window. However, this only captures part of the whole vehicle requirement to deliver n patients: estimating the full resource implication requires an analysis of the time to complete each vehicle-tour. Some of the vehicle-tours may be local with a short time on the road but others may be longer. The complete function $V(h,n,a)$ describes the mean number of vehicles in transit a minutes before their arrival in the specified time window: some hours before the specified arrival time, only a few vehicles may be in transit travelling from distant locations but more become active until all the necessary vehicles converge on the hospital at a = 0 such that $V(h,n,0)$ is the peak vehicle requirement.

Examples of these response functions are provided in Figure 6 depicting the mean vehicle requirement for the delivery of 15 patients to each of three representative hospitals h = Raigmore, Ninewells, Glasgow Royal Infirmary. The mean number of vehicles required to transport 15 patients to Raigmore, V(Raigmore,15,0) = 4.6. However, some of the vehicles are in transit for many hours, e.g. V(Raigmore,15,180) = 1.2 implying that, on average, there are 1.2 vehicles in transit 3 hours before the scheduled arrival time. The total resource requirement is dependent on both the peak vehicle demand, at a=0, and also the vehicle-hours in transit, $\sum V(h,n,a)$.

The function $V(h,n,a)$ is dependent on assumptions about the starting locations of each vehicle relative to the collection of the first patient. Usually vehicles begin their day at one of a number of stations close to population centres but the model allowed these station locations to be specified, producing modified functions $V(h,n,a)$. Hence the consequences of alternative station configurations could be explored: fewer, larger stations may have various practical managerial advantages but the additional vehicle mileage could negate these potential benefits.

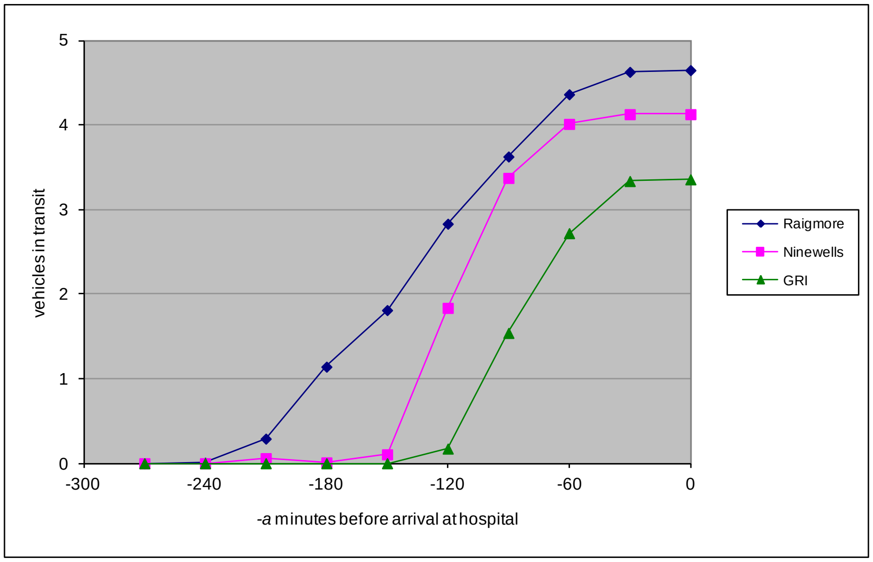

Figure 6 Mean vehicles in transit a minutes prior to the arrival of 15 patients at hospital h: $V(h,15,a)$

4.7 The temporal module

The profile of vehicles in transit, $V(h,n,a)$ illustrated in Figure 6, just describes the activity related to delivering a mean n patients with a common time window such that they can share vehicles. However, there are many appointment time windows (x) over 24 hours (x = 1..48, assuming 30 minute time windows) with a pattern of patient arrivals $R(h,x)$ illustrated by Figure 2. In some cases, there was a significant degree of seasonality in demand and sensitivity analyses were used to explore the consequent variations in resource requirements. The temporal module aggregates the profiles of vehicle in transit associated with the delivery of patients for all appointment time windows x constructing a profile $T_{arrivals}(h,t)$. At time t (t = 1..48, such that $a = 30(x\text{-}t)$) there may be several vehicles in transit collecting and transporting patients scheduled for arrival at later times x:

$$T_{arrivals}(h,t) = \sum_{x=t}^{48} V(h, R(h,x), x-t) \tag{4}$$

Most vehicles follow a rhythm of travelling into and away from the hospital with patients on board in both directions, avoiding journeys with empty vehicles. In practice vehicles can not deposit one set of patients and depart immediately with the next and a turn-round time y_1 is introduced, with a transformation of the functions $V(h,n,a)$ such that at time t the vehicle activity V' can be described by:

$$a < 0 \qquad V'(h,n,a) = 0 \tag{5}$$

$$0 \le a < y_1 \qquad V'(h,n,a) = V(h,n,0) \tag{6}$$

$$a \ge y_1 \qquad V'(h,n,a) = V(h,n,a\text{-}\, y_1) \tag{7}$$

$$T_{arrivals}(h,t) = \sum_{x=t}^{y_1+t-1} V(h, R(h,x), 0) + \sum_{x=y_1+t}^{48} V(h, R(h,x), x-(t+y_1)) \tag{8}$$

A similar process was used to produce an aggregate profile of vehicle activity $T_{departures}(h,t)$ associated with patient's departures $D(h,x)$ from hospital. The process is symmetrical to that of aggregating the vehicle activity due to arrivals with a turnround time y_2. At any time t there may be several vehicles involved in returning patients home $a = 30(t\text{-}x)$ minutes after their appointments at some earlier time x:

$a < 0 \qquad V'(h,n,a) = 0$ (9)

$0 \leq a < y_2 \qquad V'(h,n,a) = V(h,n,0)$ (10)

$a > y_2 \qquad V'(h,n,a) = V(h,n,a- y_2)$ (11)

$$T_{departures}(h,t) \sum_{x=t+1-y_2}^{t} V(h, D(h,x),0) + \sum_{x=0}^{t-y_2} V(h, D(h,x),(t-x)+y_2) \quad (12)$$

And the complete profile of vehicle activity, illustrated in Figure 7 is:

$$T(h,t) = T_{arrivals}(h,t) + T_{departures}(h,t) \quad (13)$$

Figure 7 illustrates the mean number of vehicles required to deliver and return patients to and from a single hospital. There is some flexibility in staff and vehicle deployment between hospitals within reasonable travelling time, so a further aggregation of demand was undertaken considering each region's hospitals and healthcare centres. The estimate of the complete regional profile of PTS vehicle activity had to be refined further, including some "out of area" demand requiring patients to be transported to other regions. The regional profiles provided the basis for reviewing the allocation of vehicles across Scotland and identifying any need for a redistribution of resources.

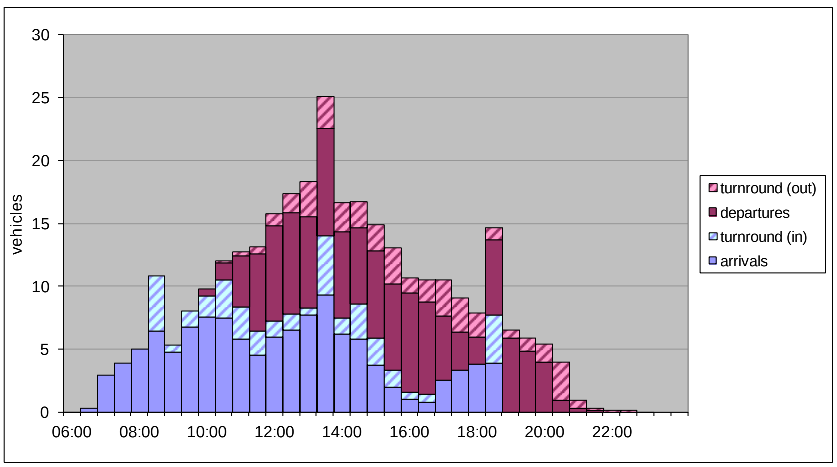

Figure 7 Vehicle activity in transporting patients to and from hospital *h*: $T(h,t)$

4.8 Comparison with an open vehicle routing simulation

The temporal module contained three critical assumptions:

- a simple 30 minute allowance for turn-round enabling a vehicle to travel from the last drop-off of one journey to the first collection of the next;
- the demand for vehicles can be considered a continuous variable, rather than being composed of a number of discrete journeys;
- the daily peak vehicle demand (either the mean or a percentile) based on the aggregation of the expected vehicle transit profiles provides a reliable measure of the vehicle requirement.

Examining a few specific examples, the client staff judged these assumptions to be reasonable, given the flexibility in shift patterns worked by the vehicle crews. However, more substantial verification was needed. This was undertaken using a Monte Carlo simulation of the open vehicle routing problem [16], as described in Appendix B. The experiments with this simulation indicated that the peak mean vehicle requirement is a robust basis for vehicle allocation, even in geographically diverse areas.

4.9 Model validation

The model was developed in collaboration with Scottish Ambulance Services staff and the regular reviews provided a degree of “white box” validation with the model development proceeding only when staff agreed its logic. Comparisons of the model’s utilisation and service statistics for the various regions revealed many differences. Some variation had been anticipated since the actual allocation of vehicles and staff was partly historic; indeed one of the roles of the model was to provide a more rational basis for distributing resources. The comparisons provoked valuable debate: some of the differences were attributed to the imperfect distribution of PTS resources but others prompted re-examination of the model. After a number of iterations sufficient agreement was obtained that it was agreed that the model provided a reasonable basis for management decision making.

5. Empirical results and modelling experience

5.1 Resource requirements in remote and rural communities

PTARMIGAN provided specific resource allocation recommendations for each region reflecting local demand and geography; these specific results are not included due to a need to maintain some confidentiality for the client organisation. A fundamental challenge posed at the start of this study was to quantify the remote and rural effect on PTS resource requirements across Scotland. In addition to the specific recommendations, the study offered some general insights, as illustrated in the analyses of three representative hospitals PTS needs. Figure 5 summarised the results of using the geographical module to estimate the resources required to deliver patients to three hospitals serving very different geographical areas: Raigmore provides care for patients from Inverness and also many remote and rural highland communities; Ninewells has many patients from urban Dundee but also a significant proportion from more remote towns; Glasgow Royal Infirmary, serves a predominantly urban population. The hospitals’ locations are noted on Figure 1. Table 1 summarises the resource requirement as the mean number of vehicles necessary and the vehicle-hours in transit, reflecting the profiles of vehicle activity as in Figure 8. Whatever the volume of demand (n = 5, 10..20), Raigmore typically requires 40% more vehicles to transport the same number of patients as Glasgow Royal Infirmary. But the vehicles transporting patients to Raigmore often spend longer in transit. Using vehicle-hours as a measure, the results of Table 1 suggest that transporting patients to Raigmore requires approximately 130% more resources than in Glasgow. This does not imply that the typical Raigmore patient experience is of long hours travelling to the hospital: for much of the time the vehicles may carry only 1 or 2 patients, collecting their remaining passengers closer to the destination.

n = patients delivered	mean vehicles			mean vehicle-hours		
	Raigmore	Ninewells	Glasgow	Raigmore	Ninewells	Glasgow
5	2.10	1.97	1.52	4.14	3.30	1.88
10	3.36	3.02	2.38	7.28	5.38	3.21
15	4.65	4.13	3.36	10.46	7.70	4.67
20	5.93	4.80	4.03	13.58	8.98	5.66

Table 1 The remote and rural effect on vehicle requirement (d = 60 minutes)

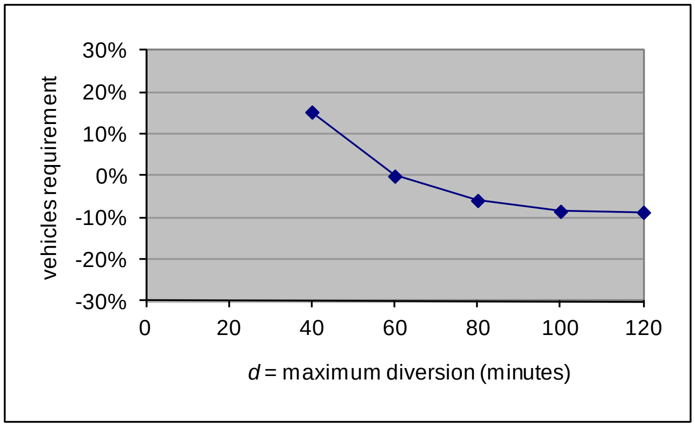

Figure 8 More diversions to collect additional patients can reduce the vehicle requirement

5.2 Resource allocation within a constrained budget

PTARMIGAN is used regularly to review the allocation of PTS vehicles across Scotland. In addition to the more routine reviews, the model is also used to explore issues of policy. For example, the initial runs of the model revealed a discrepancy between the aspirations of the PTS and its real capability: fulfilling all of the current PTS requests while meeting the service levels in terms of the delivering patients in their 30 minute windows and avoiding diversions of more than 60 minutes implied a vehicle and staff requirement that could not be met within the current budget. One option might be to relax the maximum permissible diversion time constraint, enabling some vehicles to collect and transport more patients. A series of experiments with the model suggested that this could reduce the vehicle requirement by 5-10%, as illustrated in Figure 8.

Rather than compromising service levels, an alternative approach was to manage demand. While it is essential that the PTS is available to all who have a real need, there may be some patients who might use other forms of transport. A transport network such as the PTS should have some economies of scale: a marginal reduction in demand could just result in a lower occupancy of vehicles, with little change in the total resource requirement. Experiments with the model captured the sensitivity of the national vehicle requirement to overall changes in the mean demand, as noted in Figure 9. Within a range of ± 30% the response is linear, though with some economy of scale such that a change in demand of 10% results in a change in vehicle requirement of 8%.

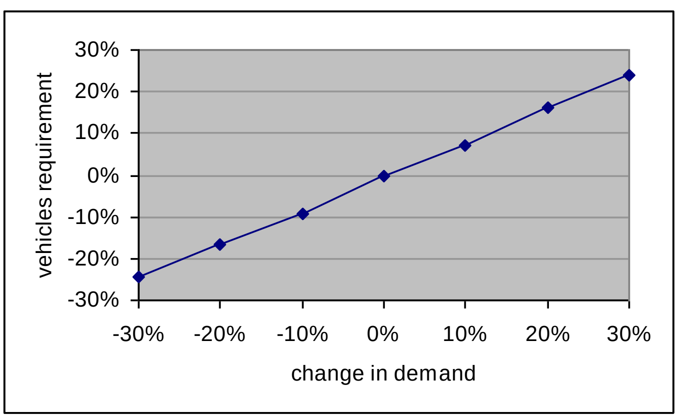

Figure 9 Sensitivity of the vehicle requirement to demand

5.3 Reflections on the modelling experience

PTARMIGAN integrated a number of models of patient demand, vehicle routing and aggregation over time to deduce the vehicle requirement. This approach enabled the construction of a comprehensive, practical decision support tool applicable across the whole of Scotland; the model was implemented successfully and is used routinely by Scottish Ambulance Staff. Some of the attractions of the simpler models were pragmatic, given the time available for the study and the relatively novel requirements which precluded the use of a single standard model. The use of simpler models also enhanced transparency, making the model more accessible to the various stakeholders and developing their understanding and confidence. The importance of stakeholder engagement has been identified as critical to successful implementation in other healthcare modelling studies [9]; it was particularly important in this study since numerous parties were competing for limited resources and the results had to be justified to management, staff and patients. It might also be argued that real-life vehicle routing is unlikely to correspond to the theoretical optimum: typically the data are imperfect and there are additional, sometimes tacit, constraints that need to be considered. A model based on simpler heuristics may well be a better basis for practical resource allocation.

It was essential that PTARMIGAN's outputs were based on efficient practice, demonstrating a responsible use of constrained resources. This was eventually achieved through the comparisons of the simpler models with more rigorous sophisticated approaches, such as the DARP routing genetic algorithms and the open vehicle simulation. However, the schedule for study required the delivery of an interim model and results to inform the client's preliminary review of PTS resource allocation. This timescale imposed a more pragmatic approach: initially more basic tests were employed, such as manual inspections of representative outputs as part of the regular project reviews with the client; it was only later that it was possible to return and complete the more rigorous verification.

6. Conclusions

The PTS model provided a practical basis for determining a fair allocation of resources across Scotland and assessing alternative policies. In particular, the constraint of the maximum diversion from a patient's direct route to hospital, provided a transparent and equitable mechanism for exploring resource requirements across geographically diverse regions, spanning both urban and remote and rural areas. Various trade-offs are possible, managing demand or increasing the maximum diversion can both reduce the vehicle requirement substantially.

PTS resource allocation exhibits features of a number of vehicle routing problems. Given the scale of the problem and the need to adopt a Monte Carlo simulation framework, a model fully incorporating the more sophisticated algorithms was impractical. Instead a series of simpler sub-models was used to determine the resource requirements. This approach helped develop all parties' understanding and confidence, essential in resource allocation decisions with multiple stakeholders. However, it is still important to verify the approximations inherent in such an approach. Techniques such as the DARP genetic algorithm formulation and simulations of the open vehicle routing problem were valuable in this role: experiments with the PTS resource requirements for three representative hospitals demonstrated that the simpler sub-models were sufficiently accurate for the current application. This case study illustrates how both simple and sophisticated models offer vital complementary qualities: it is possible to achieve a balance with the simpler models being built upon the foundations of the more rigorous approaches.

Funding sources

This study was funded by the Scottish Ambulance Service.

References

[1] O.I. Alsalloum, G.K. Rand, Extensions to emergency vehicle location models, Comput. Opns. Res. 33(9) (2006)2725-2743.

[2] L. Brotcorne, G. Laporte, F. Semet Ambulance location and relocation models, Eur. J. Opl. Res. 147(3) (2003) 451-463.

[3] G. Erdogan, E. Erkut, A. Ingolfsson, G. Laporte, Scheduling ambulance crews for maximum coverage, J. Opl. Res. Soc. 61(4) (2010) 543-550.

[4] A. Beaudry, G. Laporte, T. Melo, S. Nickel, Dynamic transportation of patients in hospitals, OR Spectr. 32(1) (2010) 77-107.

[5] T. Hanne, T. Melo, S. Nickel, Bringing Robustness to Patient Flow Management Through Optimized Patient Transports in Hospitals, Interfaces 39(3) (2009) 241-255.

[6] NHS Scotland (2011) Working for you: Scottish Ambulance Service Annual Report and Accounts 2010/11, available at: http://www.scottishambulance.com/TheService/annualreport.aspx, last accessed April 2012.

[7] T. Bektas, S. Elmastas, Solving school bus routing problems through integer programming. J. Opl. Res. Soc. 58(12) (2007) 1599-1604.

[8] Scottish Government (2008) Rural Scotland key facts, available at http://www.scotland.gov.uk/Publications/2008/08/27154843/0, last accessed April 2012.

[9] S. Brailsford, T. Bolt, C. Connell, J. Klein, B. Patel, Stakeholder engagement in health care simulation. In Rossetti MD, Hill RR, Johansson B, Dunkin A and Ingalls RG (eds). Proceedings of the 2009 Winter Simulation Conference: Austin, TX. IEEE: Piscataway, NY, (2009) 1840-1849.

[10] R. O'Loughlin, A. Kelly, Equity in resource allocation in the Irish health service - A policy Delphi study, Health Policy 67(3) (2004) 271-280.

[11] P. Bennett, J. Crosbie, P. Dick, Use of OR by government to inform health policy in England: Examples and reflections, Opns. Res. Health Care: 1(1) (2012) 1-5.

[12] R. Gulati, Tent poles, tribalism, and boundary spinning: the rigor-relevance debate in management research, Acad. Man. J. 50(4) (2007) 775–782.

[13] S. Robinson, Conceptual modelling for simulation Part I: definition and requirements, J. Opl. Res. Soc. 59(3) (2008) 278-290.

[14] S. Robinson, Conceptual modelling for simulation Part II: a framework for conceptual modelling, J. Opl. Res. Soc. 59(3) (2008) 291-304.

[15] J.F. Cordeau, G. Laporte, The dial-a-ride problem: models and algorithms, Ann. Opns. Res. 153(1) (2007) 29-46.

[16] U. Derigs, K. Reuter, A simple and efficient tabu search heuristic for solving the open vehicle routing problem, J. Opl. Res. Soc. 60(12) (2009) 1658-1669.

[17] M. Pidd, Just modeling through: A rough guide to modelling, Interfaces 29(2) (1999) 118-132.

[18] L. Coslovich, R. Pesenti, W. Ukovich, A two-phase insertion technique of unexpected customers for a dynamic dial-a-ride problem, Eur. J. Opl. Res. 175(3) (2006) 1605-1615.

[19] R.M. Jorgensen, J. Larsen, K.B. Bergvinsdottir, Solving the Dial-a-Ride problem using genetic algorithms, J. Opl. Res. Soc. 58(10) (2007) 1321-1331.

[20] E. Melachrinoudis, A.B. Ilhan, H. Min, A dial-a-ride problem for client transportation in a health-care organization. Comput. Opns. Res. 34(3) (2007) 742-759.

[21] O.B.G. Madsen, H.F. Ravn, J.M. Rygaard, A heuristic algorithm for a dial-a-ride problem with time windows, multiple capacities, and multiple objectives, Ann. Opns. Res. 60 (1995) 193-208.

[22] L. Fu, A simulation model for evaluating advanced dial-a-ride paratransit systems, Transp. Res. Part A - Policy Pract. 36(4) (2004) 291-307.

[23] S.N. Yang, M. Hamedi, A. Haghani, Integrated approach for emergency medical service location and assignment problem, Transp. Netw. Model. 1882 (2004) 184-192.

[24] P.R. Harper, A.K. Shahani, J.E. Gallagher, C. Bowie Planning health services with explicit geographical considerations: a stochastic location-allocation approach. Omega-Int. J. Mngt. Sci. 33(2) (2005) 141-152.

[25] C.H. Häll, M. Âgberg, J. Lundgren, A modeling system for simulation of dial-a-ride services. Public Transp, 4(1) (2012): 17-37.

[26] Z.H. Xiang, C.B. Chu, H.X. Chen, The study of a dynamic dial-a-ride problem under time-dependent and stochastic environments, Eur. J. Opl. Res. 185(2) (2008) 534-551.

[27] J.W. Baugh, G.K.R. Kakivaya and J.R. Stone, Intractability of the dial-a-ride problem and a multiobjective solution using simulated annealing, Eng. Opt. 30(2) (1998) 91-123.

[28] C.F.Daganzo. An approximate analytic model of many-to-many demand responsive transportation systems. Transp. Res, 12 (1978) 325-333.

[29] G. Clarke, J.W. Wright, Scheduling of vehicles from a central depot to a number of delivery points, Opns. Res. 12(4) (1964) 568-581.

[30] G.E. Box New York N.R. Draper (1987). Empirical Model-Building and Response Surfaces, Wiley, New York

Appendix A: Comparison with a genetic algorithm for the DARP

While the refined routing algorithm had the attraction of simplicity, it was essential that it could provide sufficiently efficient routes, as noted in 4.4. This required the analysis of a series of test cases and comparison with the performance of the chosen DARP genetic algorithm. A similar formulation to that employed in other DARP studies [19] was adopted, as illustrated with the binary chromosome representation of Figure A-1 describing one possible set of routes using five vehicles to transport 20 patients. The one significant difference in this formulation is that the number of vehicles is not fixed but the algorithm attempts to minimise the number deployed.

	depot	patient 1	patient 2	patient 3	patient 4	patient 5	patient 6	patient 7	patient 8	patient 9	patient 10	patient 11	patient 12	patient 13	patient 14	patient 15	patient 16	patient 17	patient 18	patient 19	patient 20
postcode district	DD1	PH1	PH1	PH1	DD8	DD1	DD4	DD4	DD3	PH1	PH1	DD8	PH2	PH2	DD5	DD7	DD1	PH2	DD3	DD4	DD3
vehicle 1	1	0	0	0	0	0	1	0	1	0	0	0	0	0	0	1	0	0	0	0	1
vehicle 2	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1	0	1	1	0
vehicle 3	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
vehicle 4	1	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0
vehicle 5	1	1	1	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0
vehicle 6	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure A-1 Binary chromosome representation (adapted from Jorgensen et al [19])

The genetic algorithm strives to find routes that avoid excessive diversions while satisfying constraints such as vehicle capacities. The objective function incorporates the key measures of success when routing any set of PTS requests, notably the total travel time and the number of vehicles required, given the vehicle capacities and maximum permissible diversion *d*. A relatively challenging convergence limit was specified for the genetic algorithm which typically resulted in approximately 5000 function calls for the larger problems. The intention was that the DARP genetic algorithm should provide a standard of comparison corresponding to very good but achievable practice. The performance of simple routing heuristics of the geographical module and the DARP genetic algorithm were compared in analyses of 40 sample sets of PTS requests. The requests were generated randomly, reflecting the geographical distribution at representative hospitals, with the mean number of requests n=5, 10, 15, 20. The two methods' mean performances were very similar with the journey time for the genetic algorithm's solutions being 99.7% of that offered by the geographical module. There was much agreement with the two methods producing similar solutions for the sets of PTS requests, as illustrated by the total travel times in Figure A-2. Given the restrictions on convergence for the genetic algorithm, there were a few occasions when the geographical module's routing algorithm offered better solutions with shorter travel times. These experiments suggest that while the algorithm of Figure 4 might be bettered by other methods, its performance is sufficient; the algorithm offers the combination of processing time, transparency and routing performance required in this particular study. However, the geographical module's simple heuristics may not be suitable for other studies: each case requires a careful comparison with a well proven rigorous standard approach, such as the DARP genetic algorithm.

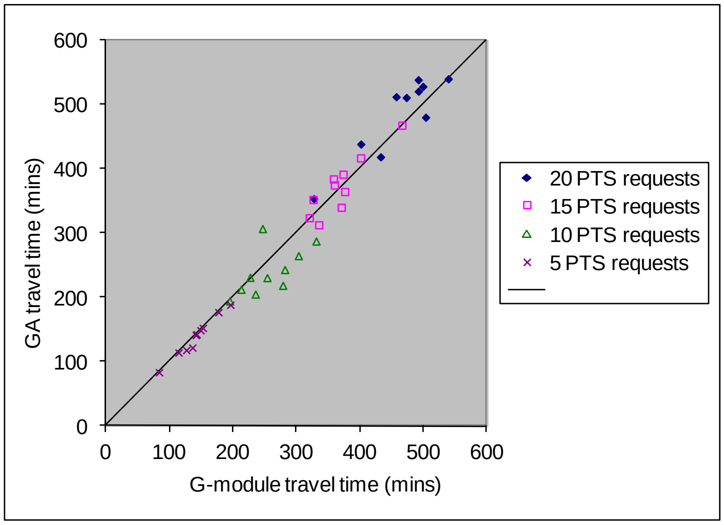

Figure A-2 Comparing the geographical module heuristics' performance with a genetic algorithm

Appendix B: Comparison with an open vehicle routing simulation

A series of simulation experiments was performed using an open vehicle routing simulation to examine the three representative hospitals' vehicle requirements in detail, testing the assumptions of the temporal module, see 4.8. This simulation combines the vehicle routing with a detailed tracking of the movement of the vehicles between journeys, avoiding the need for these assumptions. Such detailed modelling was practical for a few selected test cases but not when examining PTS activities across the whole of Scotland. A key variable was the number of vehicles available: if there were insufficient then some requests for transport would not be fulfilled. Table B-1 records the results of the simulation experiments, noting the ability to fulfil PTS requests (± 95% confidence interval). For example at Ninewells, the peak mean vehicle requirement was estimated to be 25.1; if 25 vehicles were available, the simulation suggests that 96.8% of journeys could be fulfilled. On 3.2% of occasions a vehicle was required but none was available. Unlike the emergency ambulance service, the availability of the PTS is not critical to patient safety. Failing to provide a vehicle is not desirable but in practice there is some flexibility and patients with less critical needs may have their requests for transport denied. Similar results were obtained for the other hospitals suggesting that the peak mean requirement provides a reasonable guide when allocating vehicles. Despite serving a significant remote and rural community, the approximations of the temporal module also appear reasonable in the analysis of the vehicle requirement at Raigmore: an allocation of 21 vehicles to the Raigmore area enabled 96.4% of journeys to be fulfilled.

hospital	peak mean vehicle requirement	vehicles allocated (1)	journeys fulfilled	vehicles allocated (2)	journeys fulfilled	vehicles allocated (3)	journeys fulfilled
Raigmore	21.4	19	94.0 ± 0.7 %	21	96.4 ± 0.5%	23	98.1 ± 0.4 %
Ninewells	25.1	22	94.1 ± 0.3 %	25	96.8 ± 0.3 %	27	98.2 ± 0.2 %
GRI	14.4	12	95.7 ± 0.6 %	14	98.4 ± 0.4 %	16	99.5 ± 0.2 %

Table B-1 Vehicle allocation and the ability to fulfill PTS requests

The relative robustness of the peak mean requirement as a basis for vehicle allocation is partly due to the similarity of the patterns of activity across most hospitals. Furthermore, the peak mean vehicle requirement is not as an extreme statistic as might first appear. The peak patient requests may be highly variable, as illustrated in Figure 2 and Figure 3 but the resultant vehicle requirement reflects the vehicles-in-transit profile of Figure 8. The vehicle requirement at time *t* is a function of patient arrivals over several hours after time *t*, and departures for several hours prior to time *t*. The reliability of the peak mean vehicle requirement as a measure for resource allocation could be explored further but these experiments provided sufficient evidence to justify its use in this application. Table B-1 also records the consequences of employing fewer vehicles: the ability to respond to PTS requests is impaired but this may be a necessary compromise if budgets are severely constrained. Increasing the allocation of vehicles, as in the third vehicle allocation of Table B-1, improves the service but some requests still cannot be met.