A DEMAND AND CAPACITY MODEL FOR HOME-BASED INTERMEDIATE CARE: OPTIMIZING THE ‘STEP DOWN’ PATHWAY

Alison Harper
Martin Pitt
PenCHORD, PenARC
University of Exeter Medical School
St. Luke's Campus, Heavitree Rd
Exeter EX1 1TE, UK

Manon De Prez
Zehra Önen Dumlu
Christos Vasilakis
University of Bath
School of Management
Claverton Down
Bath BA2 7AY, UK

Paul Forte
Richard Wood
NHS Bristol, North Somerset and
South Gloucestershire CCG
South Plaza, Marlborough Street
Bristol BS1 3NX, UK

ABSTRACT

Intermediate care supports timely discharge from hospital for patients with complex healthcare needs. The purpose of ‘step-down’ care is to enable patients to leave hospital as soon as medically fit, avoiding costly discharge delays and consequent risks to patient health and wellbeing. Determining optimal intermediate care capacity requires balancing costs to both acute hospital and community care providers. Too much community capacity results in underutilized resources and poor economic efficiency, while too little risks excessive hospital discharge delays. Application of discrete-time simulation shows that total costs across the acute-community interface can be minimized by identifying optimal community capacity in terms of the maximum number of patients for which home visits can be provided by the service. To our knowledge, this is the first simulation study to model the patient pathway from hospital discharge through to community visits. Simulation modeling has supported short-term resource planning in a major English healthcare system.

1 INTRODUCTION

Many health and care systems are under unprecedented pressure as a result of long-term demographic change and the associated challenges that an increasing and ageing population can bring. Health and care organizations worldwide have responded to the growing proportion of elderly people in their populations with strategies that attempt to distribute resources more optimally between health and social care. The aim has often been to minimize acute hospital utilization and expenditure by supporting people to stay at their usual residence for longer (e.g. Shrank et al. 2018; de Bruin et al. 2018). The benefits of ‘ageing in place’ can be quantified in terms of cost to the system (Bauer et al. 2019), and to the wellbeing of patients who prefer to remain in a familiar environment (Pani-Harreman et al. 2020). Additionally, across the Organization for Economic Co-operation and Development (OECD) countries, national expenditure on
social care is associated with higher life expectancy, lower prevalence of chronic diseases, and lower all-cause mortality (Rubin et al. 2016).

The delivery of community health and social care in the UK represents a complex multi-organizational system centering on the requirement to achieve timely hospital discharges through prevention and rehabilitation. Community-based healthcare support is related to the treatment or control of a health condition outside of an acute (hospital-based) setting, while social care supports activities of daily living, such as washing and dressing. In the UK, relative constraints in public expenditure on health and social care over the last decade have been linked with negative outcomes (Watkins et al. 2017), with the growth in unmet need for adult social care resulting in legislative change. In 2021, all areas of England are moving toward Integrated Care Systems, which aim for more coordinated service delivery between local councils, the National Health Service (NHS) and other partners, including the independent and voluntary sectors. Specifically, there is an emphasis on enhanced use of technology, resources and data to support people to stay safe at home for longer (NHS 2020a).

Associated with demographic change, many parts of the world have seen a sharp rise in the number of emergency admissions over the last decade for those aged 85 years and older (e.g. van den Broek et al. 2020; The Health Foundation 2018). Elderly patients are at increased risk of delayed discharges, where despite being deemed medically fit to be discharged, arrangements for continuing care are not finalized. Adverse effects for patients as a result of these delays include increased risk of mortality, hospital-acquired infections, cognitive decline, and reduced physical and mental wellbeing (Everall et al. 2019). Additionally, care in an acute setting (i.e. a hospital bed) is considerably more expensive than providing care in community settings, and it has been estimated in England that approximately 500,000 acute bed days were lost due to discharge delays directly attributable to non-availability of social care in 2019 (NHS 2019), with an estimated annual cost to the NHS of treating older patients in hospital who no longer have an acute clinical care need of £820 million (National Audit Office 2016).

The provision of home-based health and social care as a bridging service between hospital and home can support patients to return to their own home more quickly following a hospital stay. In the UK, the Discharge to Assess (D2A) care model describes three time-limited intermediate care pathways which provide ‘step-down’ care following an acute hospital stay. Patients are funded for their ongoing health and social care needs for a period of up to six weeks, in either a bedded or home-based environment. Pathway 1 (P1) provides additional support at home or in the patient’s usual residence; Pathway 2 (P2) provides rehabilitation in a temporary, non-acute bedded setting; Pathway 3 (P3) provides temporary, ongoing healthcare support where complex/significant health and social care needs are present. Nationally, approximately 50% of patients leave hospital with no ongoing support, 45% enter P1, and the remaining 5% enter P2 and P3 (NHS 2020b). Note that the underpinning principle for effective D2A care is that patients should leave hospital when they no longer need the level of medical care provided at an acute hospital. In practice, this requires significant and timely coordination between healthcare organizations and adult social care services.

A D2A P1 pathway (see Figure 1) includes the following elements:

- A patient’s previous and ongoing health and care requirements are assessed during acute admission.
- A patient is assessed as medically fit for discharge, and a referral is made.
- A health and social care assessment is undertaken in the home environment. Needs may include sign-posting/advocacy, provision of equipment, rehabilitation, and social care intervention.
- Care is multi-disciplinary, and may include both community health and social care teams.
- P1 interventions typically last 1-2 weeks, up to a maximum of 6 weeks.
- The required number of daily visits by care workers are variable, and usually taper over the duration of service.
Following discharge from intermediate care, ongoing healthcare needs fall under the remit of primary healthcare teams, led in the UK by the patient’s General Practitioner (also known as a family doctor). Any ongoing social care needs are provided on an income-assessed basis by Local Authorities. Despite the emphasis on prompt discharge and assessment in the community, the variable nature of the provision of care and the high demand for P1 pathway services makes capacity planning highly challenging, with acute hospital discharge delays, or ‘delayed transfers of care’, a common occurrence.

![Figure 1: Flow from hospital-based care to the Discharge to Assess (D2A) P1 pathway (step-down home-based intermediate care)](image)

Ensuring effective resource management of the P1 pathway is essential to provide a cost-optimal service with high benefits to patients. However, given the interrelation of healthcare services along a pathway (Moore et al. 2017) it is necessary to consider not just the P1 service but other services along the pathway which can be affected by P1 resource configuration, specifically ‘upstream’ in the acute hospital. While excess P1 capacity results in idle resources and poor economic efficiency, too little capacity risks excessive delays in discharging patients from the more costly acute setting, with consequent impacts on patient wellbeing and hospital efficiency. Delays to discharging patients can propagate through the entire hospital system, resulting in cancelled elective surgery, delays to admission from the Emergency Department, and ambulance handover delays (Li et al. 2019).

The objective of this study is to identify optimal P1 capacity in terms of achieving the lowest aggregate costs across acute and community services. This is approached using a discrete-time simulation (DTS) model of the P1 visits over time, with application to a major healthcare system in south west England. The novelty of the approach taken is modelling service times and capacity in terms of number of visits by care workers which taper over time, in comparison to most health care simulation studies which consider service times in terms of length of stay in a bed or cubicle. This model forms part of a larger suite of applied models used for decision-support by commissioners and managers who are planning resource allocation at the interface between acute and community health and social care.

The remainder of this paper is structured as follows. Section 2 outlines the existing literature in community health and social care. Section 3 describes the methods, data and application, Section 4 presents the results and Section 5 discusses the implications and future work.

2 LITERATURE REVIEW

Academic interest in simulation models applied to health care services has seen increased interest over the last few decades (Fone et al. 2003, Brailsford et al. 2009, Sobolev et al. 2011, Mohiuddin et al. 2017). Most
of the literature concentrates on the improvement of hospital based acute services and patient flow such as outpatient appointment scheduling, surgery scheduling or intensive care unit management (Gupta and Denton 2008; Vasilakis et al. 2007; Chan et al. 2012). Few studies capture the outflow of patients following transfer or discharge from acute to community care, which comprises health and care facilities where patients rehabilitate or packages of care that support recently discharged patients at their usual place of residence. This paucity of published modelling studies contrasts with the importance of the interaction between acute and community care, given that lack of capacity in post-discharge facilities and poor coordination may have an impact on delayed transfers of care at hospitals (NHS England 2018; NHS England 2020; Jones et al. 2019) and hence, have a negative impact on patient waiting times and other health and system measures (Jasinarachchi et al. 2009).

A literature search was performed for modelling studies on capacity planning in community care and/or long-term care. In the rest of this section, we focus on some key papers which particularly explore this problem using mathematical modelling and simulation methods. Palmer et al. (2017) reviewed research focusing on modelling patient flow and patient outcomes for community care services using operational research methods. They underline that multiple services, patient mix and different health-states are rarely considered, especially when capacity, demand and timing of patient use varies. Cardoso et al. (2012) developed a comprehensive simulation model based on a Markov cycle tree structure to predict demand for long term care in Portugal. They incorporate different health states and socioeconomic characteristics of patient. Although they identify the demand in terms of individuals in need of each type of long-term care service, the resources including domiciliary visits and the associated costs, no interaction is assumed nor analysed between acute care and long-term care demand. There are other modelling studies focusing on capacity planning in the long-term care setting (Li et al. 2016; Lin et al. 2012; Patrick et al. 2014; Zhang et al. 2012) where capacity is captured as the number of available beds in the facilities. These studies simulate required capacity to reduce wait times for long term care. Zhang et al. (2012) optimised a discrete event simulation (DES) model and Patrick et al. (2014) developed a Markovian simulation model to find the required capacity for long-term care to achieve a certain service level. Li et al. (2016) and Lin et al. (2012) looked at both residential and community-based long-term care which, also known as home and community-based services, and provide capacity estimates for both types of long-term care. However, the care need for those patients was modelled as a constant insofar the number of visits over a certain length of stay was fixed over time. Demirbilek et al. (2018) aimed to maximize the number of patient visits in home healthcare by improving the decision about whether to accept or reject a new patient and if accepted, when to schedule that service. They express the length of stay of patients in weeks with a constant number of visits required in each week, and the simulation is limited to one nurse.

Some studies focus on capacity planning for specific types of disease. For example, Bayer et al. (2010), Monks et al. (2016) and Wood & Murch (2019) reported on the development of a DES model to identify capacity requirements in the acute and community for stroke care services. Even though these studies include community care in their models, it is modelled in a simplistic way as end destinations. Deo et al. (2013) developed a stochastic dynamic program for capacity allocation of community-based healthcare delivery for childhood asthma in a non-profit setting, where available capacity is allocated among patients of different health states while maximizing quality-adjusted life-years for the entire patient cohort. Diaz et al. (2015) evaluated patient flow between acute care and home care for patients with chronic disease. Congestion and capacity of resources were also considered. Through a scenario analysis they evaluate the impact of different patient routes and resource allocations on the level of demand for services and the cost of providing care. Ei-Darzi et al. (2000) developed a DES model to understand bed requirements and utilisation of resources for patients within a geriatric hospital and community homes. Through several analyses they show how the interaction between acute and community care can affect inflow and capacity requirements. These types of models can become a roadmap to decision makers, such as hospital planners and clinicians, to assess long term effects of changes in their system.
In this paper, we address the gap in the literature that models community care capacity across the acute-community interface. Additionally, community capacity is defined in terms of the number of visits required, which can change for any patient during the course of care. The next section details the modelling approach.

3 METHODS

3.1 D2A P1 Pathway Conceptual Model

The conceptual flow map in Figure 1 describes a simple, generic discharge process for patients who are discharged from hospital into the D2A P1 step-down intermediate care pathway. Figure 2 illustrates the scope of the simulation model described in this paper, comprising the process from referral into P1. The iterative modelling process forms part of a collaboration between modelers and health and social care staff in a major healthcare system in south west England, but is being developed to be more widely applicable. The healthcare system involved encompasses one Clinical Commissioning Group (CCG), two major acute hospital providers, three Local Authorities, and one community healthcare provider. It serves a population of approximately one million people.

All model parameters were estimated in collaboration with the healthcare service involved (‘the intermediate care service’). As data is required from acute hospitals, community healthcare, and community social care providers, calibration and validation is ongoing as an integral part of the process of embedding the simulation model into routine operational decision-support and planning.

Figure 2: Conceptual map of D2A P1 pathway for the developed simulation model

3.2 Model Development and Application

The discrete-time simulation (DTS) model presented in this paper has been developed using R (3.6.4), as an improvement to a deterministic Excel-based model used for capacity planning by the collaborating health and care organizations. Parameters were initially acquired from this spreadsheet model. This parameterized iteration of the model is open source, and available here: https://github.com/nhs-bnssg-analytics/winter-simulation-conf-2021-files. While referrals into P1 have been previously estimated based on a moving average for the time of year, with COVID-19 adjustments derived from an externally available SEIR (susceptible-exposed-infected-recovered) model (Booton et al, 2021), in this initial P1 DTS development phase a static mean referral rate was assumed. The referral rate was modelled as a Poisson process. Patients who are referred but unable to be discharged into P1 remain in hospital, and represent a delayed discharge. The duration of P1 service, the initial daily visit requirements, and the final daily visit requirements are sampled from normal distributions.

The duration of service has been estimated as a weighted average of the mean durations-of-service provided per Local Authority for P1 pathways. The mean number of visits required per day is not routinely
collected. The intermediate care service base their planning on an average of three care visits per day, however it is known that this tapers over the duration of service. It is also known that initial visits of three per day may require two care staff per visit, hence the initial number of visits required was set to a mean of 4 (truncated to upper limit = 6), and the final number of visits to a mean of 2. A sequence of visits from the sampled initial number to the sampled final number of visits across a duration of service was generated per simulated patient. For example, if the initial number of visits is sampled as 6, the final number of visits is sampled as 2, and the duration of service is 10 days, the visit sequence for the patient is [6,6,5,5,4,4,3,3,2,2].

Patients with a long sampled duration of service, and/or a high initial/end daily visit requirement subsequently block patients with lower service requirements from entering the P1 system. For this reason, if there are no available resources to start a service immediately, an arriving patient is scheduled to start on the following day. A patient with a greater care need is more likely to be delayed for ongoing care than a patient with a lower care need, whose visit sequence can be integrated into the remaining P1 capacity. 

Estimated acute and community healthcare costs are reference costs from financial year 2017/2018 (NHS Improvement 2018). Reference costs record the average (aggregated) unit cost to an NHS trust of providing defined services to NHS patients in a given financial year. Table 1 summarizes all parameters.

To determine the optimum P1 capacity which minimizes the total cost of acute delayed transfers of care (DTOCs) and the cost of providing surplus P1 capacity, a cost-capacity model simulated the overall cost across a range of feasible capacities. The capacity is the number of visits in the system, i.e. the number of patients admitted into P1 (‘slots’), multiplied by the average number of visits per day. The minimum number of visits required to reach a stable system was estimated by visual inspection of the outputs. The maximum number of visits in the system was calculated as the capacity required for 100% of patients to be discharged on the day they are referred, that is, zero delays. The cost of acute care was calculated as the cost of one hospital day multiplied by the mean delay per simulated patient and the number of patients discharged per week.

Having determined an optimal capacity which minimizes both the number of acute DTOCs and the underutilization of community capacity, two capacity scenarios were simulated:

(i) The current capacity of 164 slots (an average of 492 visits, given an average 3 visits/day)
(ii) The optimal capacity as determined by the cost-capacity output

Over a run of 40 days: (1) the number of patients in the system, (2) the number of delayed patients, and (3) the visits capacity utilization, were measured as outputs, comparing the two scenarios. The model has a warm-up duration of 1000 days (determined visually, and over-estimated for experimentation) and is replicated 200 times. Initial results have been shared with service stakeholders to support urgent planning during the COVID-19 pandemic, and to provide feedback toward the next iteration of the model.

Table 1: Parameters for D2A P1 simulation model

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimated values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arrival rate</td>
<td>Sampled from a Poisson distribution with a fixed mean over time</td>
</tr>
<tr>
<td>Initial/end visits</td>
<td>Truncated normal distributions (0-6 visits/day, tapering over the visit sequence)</td>
</tr>
<tr>
<td>Service duration</td>
<td>Normal distribution with fixed mean</td>
</tr>
<tr>
<td>Community cost</td>
<td>Reference cost for one intermediate care visit * given capacity</td>
</tr>
<tr>
<td>Acute cost</td>
<td>Reference cost for one hospital day * mean delay per patient * weekly number of patients discharged from hospital</td>
</tr>
</tbody>
</table>

4 RESULTS

Across a range of hypothetical P1 capacities, the combined cost of acute care delays and community capacity costs were estimated (Figure 3). The optimum capacity was determined to be 510 visits per day. Dividing this by the average number of visits across the duration of service suggests that 170 slots (patients)
are optimally accommodated by the system. Capacity below this level rapidly increases the costs of delays, while higher capacity steadily increases the cost of P1 underutilization. At the time of the study, the current capacity in the intermediate care system across the three Local Authorities was an average of 492 visits per day (corresponding to 164 slots), with a resultant higher cost associated with the expected higher number of DTOCs, given the input parameters. Sensitivity analysis was conducted for costs 10% above and below the reference costs presented in Table 1. Changes in acute care costs have more effect on aggregate costs at lower P1 capacities than at higher capacities, while P1 community costs influence the total aggregate costs, but not the cost-optimal capacity. The implications for acute and community care are when the P1 service is operating under-capacity, which both increases aggregate costs, and is more sensitive to cost uncertainty.

Plotting the same range of capacities against the number of delayed discharges illustrates that the cost-optimal capacity of 510 visits is a trade-off between DTOCs and P1 utilization (Figure 4). This was investigated further with DTS time-series output comparing capacity scenarios.

The (1) optimal and (2) current P1 capacity were fixed, while all other parameters remained unchanged. These scenarios were run for 40 days, starting at the current date. The optimal scenario (Figure 5) shows a very small number of delayed patients, with a mean of 1 patient, and a median of zero; 75% of the time, there are no DTOCS, which remain below 10 people at the 95th centile. These relatively few delayed
patients are those with a higher service need in terms of duration of service and/or visit sequence required. This compares with the current scenario (Figure 6), with a mean of 3 patients and a 95th centile reaching 19 delayed patients.

![Optimal capacity scenario: Number of acute patients delayed per day](image1)

Figure 5: Number of patients delayed (DTOCs) given the *optimal* number of visits per day --- mean; quantiles (shaded areas) = 75%, 90% and 95%. The median and the 75% quantile are 0.

![Current capacity scenario: Number of acute patients delayed per day](image2)

Figure 6: Number of patients delayed (DTOCs) given *current* number of visits per day --- mean; quantiles (shaded areas) = 75%, 90% and 95%. The median is 0.

The number of patients in the P1 service for the *optimal* and *current* capacity scenarios were compared. The average number of patients in the system is lower in the *optimal* capacity scenario compared with the *current* capacity scenario. This can be explained by the utilization of visits in the P1 system. In the *optimal* capacity scenario, throughput allows the majority of patients to enter, be processed, and leave. In the *current* capacity scenario, utilization is higher and patients are more likely to backlog in P1 and the acute setting. Figure 7 shows the percentage utilization of visits over time. The average utilization is around 90% and the system rarely reaches full utilization. Figure 8 shows the *current* scenario percentage utilization, which often averages over 95% and is at full utilization 75% of the time.
5 DISCUSSION

To the best of our knowledge, our simulation modelling study is novel in many ways. We developed a model to assess capacity in community care including the interaction between acute care and community care where we define capacity in terms of the number of visits required, and this number can change for any patient during the course of care. Also, using real data for parameter estimation, we identify through simulation the optimal number of visits needed while minimizing the aggregate costs associated with both community care resources and bed-blockages at the acute level.

As a stochastic system, a trade-off should be made between reducing acute DTOCs and uneconomical P1 capacity. An optimal level of capacity, given variation in the required number of daily visits at the start and end of the duration of service, can be determined by balancing the costs of DTOCs with the costs of additional P1 capacity. The findings of this study have shown that optimizing aggregate acute and community costs reduces the duration of delays as well as the number of patients delayed in the acute sector. This is important for both patients and for the wider health and social care system, reducing the risks and costs to both. The simulation results have shown that at a level of P1 capacity below the optimal level, the
system rapidly deteriorates in terms of higher utilization of P1 resources, higher number and duration of DTOCs, and higher overall costs.

One potential solution to further reduce delays is to reduce the variation in arrivals into the P1 pathway, the duration of service, or the sequence of visits required. However, the simplifications within the simulation model indicate the difficulty of implementing this potential solution in reality. The model aggregates three Local Authority processes, but in reality, there is very little sharing of resources except at geographical borders, hence future work will model each process separately. In addition, the duration of service contains both health care and social care elements. As these are funded separately and use different providers, these parallel processes also need to be modelled separately. The most significant barrier to this is the lack of social care data. This is a recognized challenge which is currently being addressed with social care stakeholders to understand what data can be made available. Third, there are behavioral components which are not captured in the model, for example movement between Pathways 1-3 to use available capacity, meaning a proportion of patients enter a suboptimal discharge pathway, risking readmission. When further calibrated and validated, this P1 model will be integrated with P2/P3 counterparts, and movement between pathways will be captured, enhancing model realism.

To date, discussion with stakeholders within the health and care system has been positive and encouraging, with staff having indicated the value of the model toward supporting operational planning. Datasets to further refine model parameters are under preparation. The use of predicted referrals as inputs, and setting initial conditions is being undertaken as a key part of resource planning, and will be used for modelling all pathways, ensuring the outputs of the simulation provide maximum value for resource planning for intermediate care commissioners and providers.

ACKNOWLEDGMENTS

This work was supported by Health Data Research UK, which is funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, National Institute for Health Research, Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (South Western Ireland), British Heart Foundation and Wellcome (award number CFC0129).

REFERENCES


Harper, Pitt, De Prez, Önen Dumlu, Vasilakis, Forte, and Wood


AUTHOR BIOGRAPHIES

**ALISON HARPER** is a research associate at the University of Exeter Medical School, in the *Peninsula Collaboration for Health Operational Research and Development*. She gained her PhD from University of Exeter Business School. Her research interests are applied health and social care research using data science and quantitative methods to model and improve services, with a particular focus on implementation and impact. Her email address is a.l.harper@exeter.ac.uk

**MARTIN PITT** is Director of *The Peninsula Collaboration for Health Operational Research and Development*, Exeter University: Medical School. He is particularly interested in using data science for healthcare improvement. He helped to establish and coordinates MASHnet - The UK Network for Modelling and Simulation in Healthcare. He was recently appointed as first President of the Association of Professional Healthcare Analysts (AphA). His email address is m.pitt@exeter.ac.uk

**MANON DE PREZ** achieved her Masters in Business Analytics at the University of Bath School of Management in the UK. She is currently working as a digital analyst for a telecommunications company in Canada. Her email address is deprezmanon@gmail.com.

**ZEHRA ÖNEN DUMLU** is a research associate at the University of Bath School of Management in the Information, Decisions and Operations division. She earned her Ph.D. in Industrial Engineering and Operations Management from Koç University. Her research interests focus healthcare operations management. She is mainly interested in developing and investigating models under uncertainty that will help improve healthcare operations and global health systems. Her email address is zod23@bath.ac.uk

**CHRISTOS VASILAKIS** is Chair and Subject Group Lead in Management Science in the Information Decision and Operations Division and Director, Bath Centre for Healthcare Innovation and Improvement (CHI²) both at the University of Bath School of Management in the UK. In 2020 he was a Visiting Scholar at Stanford University, Department of Management Science and Engineering and SURF Stanford Medicine research centre. His email address is c.vasilakis@bath.ac.uk

**PAUL FORTE** holds a part-time role with the London School of Hygiene & Tropical Medicine, and is an independent health planning and management consultant. He earned his Ph.D. in the School of Geography at the University of Leeds. His interests include whole system planning and management approaches in health and social care and working closely with stakeholders to support and educate the application of models in local decision making. His email address is paul.forte@nhs.net.

**RICHARD WOOD** Richard Wood is Head of Modelling and Analytics at Bristol, North Somerset and South Gloucestershire CCG and a Visiting Research Fellow at University of Bath School of Management. His interests are at the interface of academic theory and practical application. His background is in applied mathematics and through his career he has worked in diverse sectors such as mathematical biology, banking, and financial risk modelling. His email address is Richard.wood16@nhs.net