



# The False Economy of Seeking to Eliminate Delayed Transfers of Care: Some Lessons from Queueing Theory

Richard M. Wood<sup>1,2,4</sup> · Alison L. Harper<sup>3,4</sup> · Zehra Onen-Dumlu<sup>2,4</sup> · Paul G. Forte<sup>1,4</sup> · Martin Pitt<sup>3,4</sup> · Christos Vasilakis<sup>2,4</sup>

Accepted: 22 November 2022  
© The Author(s) 2022

## Abstract

**Background** It is a stated ambition of many healthcare systems to eliminate delayed transfers of care (DTOCs) between acute and step-down community services.

**Objective** This study aims to demonstrate how, counter to intuition, pursual of such a policy is likely to be uneconomical, as it would require large amounts of community capacity to accommodate even the rarest of demand peaks, leaving much capacity unused for much of the time.

**Methods** Some standard results from queueing theory—a mathematical discipline for considering the dynamics of queues and queueing systems—are used to provide a model of patient flow from the acute to community setting. While queueing models have a track record of application in healthcare, they have not before been used to address this question.

**Results** Results show that ‘eliminating’ DTOCs is a false economy: the additional community costs required are greater than the possible acute cost saving. While a substantial proportion of DTOCs can be attributed to inefficient use of resources, the remainder can be considered economically essential to ensuring cost-efficient service operation. For England’s National Health Service (NHS), our modelling estimates annual cost savings of £117m if DTOCs are reduced to the 12% of current levels that can be regarded as economically essential.

**Conclusion** This study discourages the use of ‘zero DTOC’ targets and instead supports an assessment based on the specific characteristics of the healthcare system considered.

## Key Points for Decision Makers

It is a stated aim of many healthcare systems to ‘eliminate’ delays in transfer of care between acute and community services.

Mathematical modelling suggests that this requires uneconomically large amounts of downstream community capacity, the cost of which outweighs the acute cost saving.

For England’s National Health Service (NHS), substantial cost savings are possible if delays are reduced to the 12% that can be considered economically essential.

## 1 Introduction

In a nutshell, a delayed transfer of care (DTOC) occurs when a patient is ready for discharge from hospital but is still occupying a bed [1]. This can happen for a variety of reasons, such as disputes or patient choice, but is typically due to insufficient downstream capacity for patients requiring some form of step-down or continuing non-acute care [2–4].

In England’s National Health Service (NHS), 1.14m acute bed days were lost due to DTOCs in the 12-month period before the coronavirus disease 2019 (COVID-19) pandemic was declared in March 2020 [5]. Some of these (28%) were attributable to a lack of provision of local authority-funded social care, while approximately half (48%) were related to the unavailability of short-term step-down care (otherwise known as intermediate care) provided by NHS-funded community services [1].

Extended author information available on the last page of the article

At first glance, it would appear that DTOCs are wholly negative and should be reduced to zero within well-performing healthcare systems, and this is the impression given by many NHS organisations. Manchester and Surrey talk of plans to “eliminate DTOCs” [6, 7]. A major London healthcare system claims that hospitals have “accepted their challenge to eliminate DTOCs” [8] and, in light of COVID-19, the Isle of Wight system states “it has never been more important to eliminate DTOCs” [9]. Meanwhile, in Scotland, the Health Secretary has pledged to “eradicate bed blocking” [10]. Mention of ‘eliminate’ and ‘eradicate’ in relation to DTOCs also extends to the academic literature [11, 12]. However, should hospitals and healthcare systems really seek to eliminate DTOCs?

Eliminating DTOCs caused by insufficient downstream community capacity is entirely possible if the arrivals (patients becoming ready for discharge from the acutes) and lengths of stay (in the community service) are perfectly constant. For example, if we have exactly five acute patients becoming ready for discharge every day to bed-based community care (community hospitals or in ‘short-stay’ care home beds) and each patient has a community length of stay (LOS) of exactly 12 days, then 60 beds’ worth of community capacity would be sufficient. With this amount of capacity, there would never be any ‘queue’ at the acute hospitals (i.e. zero DTOCs) and all community capacity would be fully utilised all of the time.

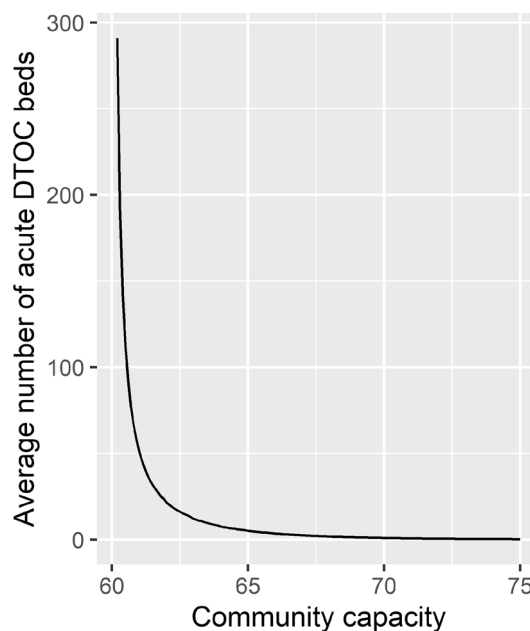
However, we know from Queueing Theory, a mathematical subdiscipline of Operational Research concerning the dynamics of queues, that when there is variability in the arrival rate and/or LOS then some amount of queueing is inevitable given the various peaks and troughs that will occur over time, such as having a number of arrivals in a short time period or having a group of long-staying patients admitted all at once [13, 14]. To prevent a significant queue forming then extra ‘slack’ is required in the system, and this can be achieved through increasing community bed capacity. The amount of additional capacity required depends on the amount of queueing considered tolerable (Fig. 1).

While able to reduce costly queueing due to DTOCs, this extra downstream capacity comes at a cost itself in terms of the additional community beds, domiciliary carers, and other support staff required. The question is, at what point do these additional costs start to outweigh the associated savings in reducing acute DTOCs? If this is known, and appropriate (non-zero) DTOC targets are set, then healthcare systems can work towards community discharge pathways of greater cost optimality.

The aim of this study was to illustrate the false economy of seeking to eliminate DTOCs and to show that some amount of ‘bed blockage’ is beneficial to reducing the total

costs across the continuum of care. This is approached through applying established Queueing Theory formulae to a major NHS healthcare system in South West England to calculate the optimal amount of acute DTOCs and community capacity required. Sensitivity to a range of parameters is considered, including the acute- community cost ratio and level of variability in community LOS, as well as the size of the healthcare system. Cost savings for England are also estimated if the optimal amount of DTOCs is to be achieved.

It should be noted that this study considers only the financial costs related to the part of the patient pathway of interest, i.e. the costs of caring for acute patients experiencing DTOCs and of operating the downstream community service. By not taking into account the wider costs potentially associated with patient ‘deconditioning’ (i.e. those related to additional care needs), the analysis presented here essentially assumes that patient care (beyond medical fitness for discharge) and outcomes are not impacted by the care setting. The possible effects of this assumption are considered further in the Discussion section. Note also that only ‘step-down’ care is considered in this study (the healthcare systems of some countries may also include ‘step-up’ care for admission avoidance purposes).



**Fig. 1** Effect of post-acute community capacity on the average number of acute beds subject to DTOCs, assuming a mean of five acute patients becoming ready for community discharge each day and a mean community LOS of 12 days. Modelling with some typical variability around arrivals (assumed Poisson) and LOS (assumed exponential) shows that the ‘averages-based’ community capacity ( $5 \times 12$ ) to treat 60 concurrent patients is insufficient, with large acute DTOCs developing. Results are produced using the method detailed in the Methods section. *DTOCs* delayed transfers of care, *LOS* length of stay

## 2 Methods

### 2.1 Queueing Theory

To address the aims of this study, it is necessary to introduce some terminology from Queueing Theory. First, we have *arrivals*. In our case, these are patients becoming ready for discharge from the acute hospital and requiring continuing care in the community setting. As is typically assumed for healthcare models, the arrival of one patient can be considered independent of the arrival of another [15]. Thus, the number of arrivals occurring in an interval of time can be described by a Poisson distribution, meaning the time between successive arrivals is exponentially distributed, or *Markovian*. Second, we have *service times*, or LOS. In our case, this represents the durations of time for which patients are under the care of the community service provider (starting from when they are transferred from the acute hospital), and is subject to some degree of variability. Third, we have *capacity*, which represents the maximum number of patients that can be cared for concurrently by the community service provider (either at home through care visits or in a bed-based unit). In Queueing Theory, Kendall's Notation is used to summarise these three properties [16]. In our case, we have an  $M|G|c$  queue, i.e. Markovian arrivals, some sort of general service time (LOS) distribution, and a capacity for  $c$  patients.

The formulae needed for the  $M|G|c$  queue is actually based on derivations for the  $M|M|c$  queue (i.e. the same queue but assuming that LOS is exponentially distributed, or *Markovian*). For this type of queue, Erlang's C Formula, an established result in Queueing Theory [17], gives the probability of an arriving patient finding no available community capacity, and thus experiencing a DTOC (Eq. 1):

$$\bar{C} = \left[ 1 + (1 - \rho) \left( \frac{c!}{(c\rho)^c} \right) \sum_{k=0}^{c-1} \frac{(c\rho)^k}{k!} \right]^{-1} \quad (1)$$

Here,  $\rho$  is the *traffic intensity*, which measures the ratio of demand to capacity. This is calculated through dividing the mean of the Poisson arrival rate,  $\lambda$ , by the mean service rate,  $c\mu$ , where  $\frac{1}{\mu}$  is the mean of the general service time distribution representing community LOS. The smaller  $\rho$  is, the less pressure the service will be under, and thus the less queueing there will be.

The average waiting time,  $\bar{W}$ , for an  $M|M|c$  queue is calculated [18] as (Eq. 2):

$$\bar{W} = \frac{\bar{C}}{c\mu - \lambda} \quad (2)$$

This is related to the  $M|G|c$  queue through the approximation (Eq. 3) [19]:

$$W \approx \frac{V^2 + 1}{2} \bar{W}. \quad (3)$$

Here,  $V$  measures the amount of dispersion in the LOS distribution (specifically, its coefficient of variation, calculated by its standard deviation over its mean).

Average queue length, here representing the number of acute DTOC beds (i.e. those occupied at any time due to unavailable downstream capacity), can be calculated through the average waiting time by Little's Law (Eq. 4) [20]:

$$L = \lambda W. \quad (4)$$

Therefore, if an acute bed costs  $\alpha$  per day, then the average daily cost of DTOCs borne in the acute setting is  $\alpha L$ . If the cost ratio of community to acute care is  $\beta$  (presumed less than one), such that the cost of community care is  $\beta\alpha$  per provided-for patient per day, then the corresponding daily cost borne in the community setting is  $\beta\alpha c$ . The total daily cost is thus calculated as the addition of the variable cost of acute discharge delays and the fixed cost of procured community capacity (Eq. 5):

$$T = \alpha L + \beta\alpha c. \quad (5)$$

It should be noted that the value of this equation does not depend on the exact values of  $\lambda$  and  $\mu$ , but their ratio as expressed in  $\rho$  (recalling  $\rho = \lambda/c\mu$ ), as well as the coefficient of variation of LOS,  $V$ . This can be shown by combining Eqs. 1–5 to yield (Eq. 6):

$$T = \alpha \frac{V^2 + 1}{2} \frac{\rho}{1 - \rho} \left[ 1 + (1 - \rho) \left( \frac{c!}{(c\rho)^c} \right) \sum_{k=0}^{c-1} \frac{(c\rho)^k}{k!} \right]^{-1} + \beta\alpha c \quad (6)$$

As can be seen, the value of  $T$  depends only on  $\alpha$ ,  $\beta$ ,  $V$ ,  $c$ , and  $\rho$ , and not  $\lambda$  and  $\mu$  directly. In other words, the total cost does not depend on the exact mean arrival and service rate values (only their ratio).

### 2.2 Data and Calibration

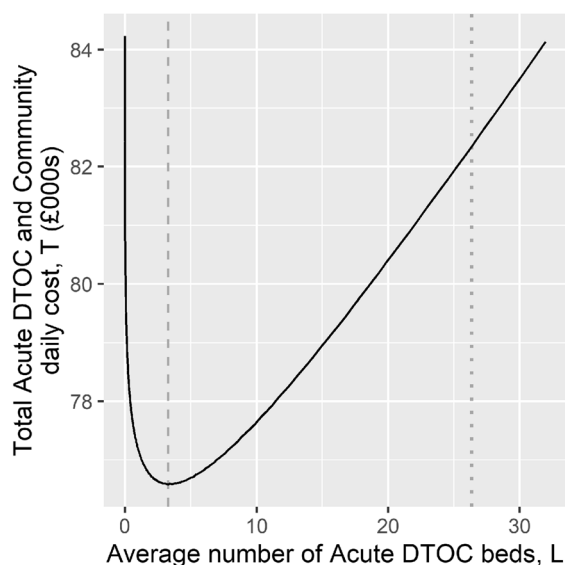
We apply the above formulae to a major healthcare system located in and around Bristol, in the south west of England. The Bristol, North Somerset and South Gloucestershire (BNSSG) healthcare system serves a 1 million resident population across a mixture of large metropolitan areas and rural and coastal locations [21]. The large metropolitan area of Bristol contains a higher proportion of younger individuals and has a more culturally and ethnically diverse demographic [22]. Rural and coastal areas contain a greater proportion of older individuals and pockets of severe deprivation [23]. Overall, the age profile of BNSSG is similar to that of England [21].

The cost of an acute bed day ( $\alpha$ ) is estimated at £346, based on the latest available national reference cost data that contains such a metric, from the financial year 2017/2018 [24]. From local data from the financial year 2020/2021, we estimated an average daily community cost across both NHS-funded home-based and bed-based care of £138 (i.e.  $\beta = 138/346 = 0.399$ ). Using these data, we also estimated the coefficient of variation of community service LOS ( $V$ ) of 0.965 (thus indicating slightly less dispersion than the commonly used exponential distribution, which has  $V = 1$ ). Community service capacity ( $c$ ) was estimated at 531, noting that this should be considered a relatively large service provider, given the size of the population covered. Finally, traffic intensity ( $\rho$ ) was estimated at 0.98, which suggests a very busy service (also supported by the data and the high levels of community service occupancy observed). For more details on how the parameter estimates were obtained, see the electronic supplementary material (ESM).

### 3 Results

#### 3.1 Optimal Amount of Acute Delayed Transfers of Care

With the Queueing Theory formulae calibrated to the local healthcare system, we can examine the main aim of this study: highlighting the economic benefit for some amount of DTOCs, or ‘bed blocking’ in acute care. In varying the hypothetical levels of NHS-funded, time-limited ‘step-down’ community care capacity,  $c$ , within Eqs. 4 and 6, we can



**Fig. 2** Modelled results showing the optimal amount of acute DTOCs (left-hand side) and community capacity (right-hand side) required to minimise total acute and community cost. Optimal levels are repre-

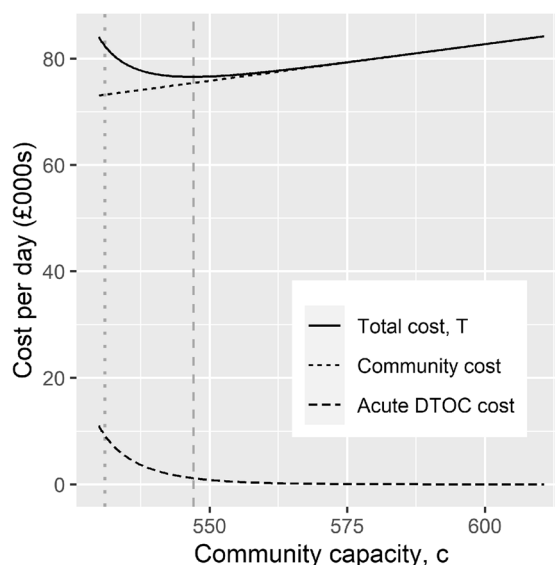
sent by the dashed grey vertical lines, with the dotted grey vertical lines representing the actual levels at the healthcare system studied. *DTOCs* delayed transfers of care

plot the relationship between acute DTOCs and total costs (Fig. 2, left-hand side). This clearly shows that the minimum total cost is achieved when operating at an above-zero amount of average acute DTOCs.

Moreover, it shows that costs increase sharply when average acute DTOCs are lessened to the point of elimination. This is being driven by the large amounts of downstream community capacity required to achieve such low levels of queueing. Conversely, if there is insufficient community capacity, then acute DTOC costs increase rapidly and non-linearly, making a significant contribution to the total cost (Fig. 2, right-hand side). Figure 2 also indicates that the studied healthcare system could reduce costs by increasing community capacity by 3%, from 531 to the theoretical optimal of 547. This would reduce the average number of acute DTOC beds from 26.3 to 3.3 and realise a 7% saving in daily operating costs.

#### 3.2 Effect of Acute-Community Cost Ratio and Variability in Length of Stay

Changing the ratio,  $\beta$ , of community to acute unit costs can have a profound effect on optimal resource allocation and utilisation (Fig. 3, left-hand side). As community unit costs increase, then a greater premium is put on community capacity, meaning that it becomes more cost effective to accept a larger amount of acute DTOCs. Essentially, as the cost gap narrows, community capacity is penalised in favour of more ‘bed blocking’. Here, if the community cost ratio is increased from 0.4 to half that of an acute bed, then achieving cost optimality means trading away capacity to care for



sented by the dashed grey vertical lines, with the dotted grey vertical lines representing the actual levels at the healthcare system studied. *DTOCs* delayed transfers of care

two patients in the community for an extra one acute bed required to support the greater level of discharge delays.

Another aspect that can affect resource allocation and utilisation is the amount of variability in community LOS, referred to previously by  $V$  (Fig. 3, right-hand side). When there is zero variability in LOS (i.e. every patient stays exactly the same length of time in community care) then less community capacity is required, and acute DTOCs and total costs also reduce slightly (note that some acute DTOCs still occur given variability in the arrival rate). When LOS variability increases, then so too does the optimal capacity requirement due to the greater volume of potential ‘shocks’ that need to be absorbed (i.e. longer patient stays). If these are not absorbed through additional community capacity, then they are manifest in the more costly acute setting in the form of DTOCs. While the total cost differences appear small on the plotted scale, they can provide considerable annual savings. For instance, if through more targeted efforts to reduce longer community stays, LOS variability could be reduced from 0.965 (dotted grey vertical line) to 0.5 (bottom-right panel), then this

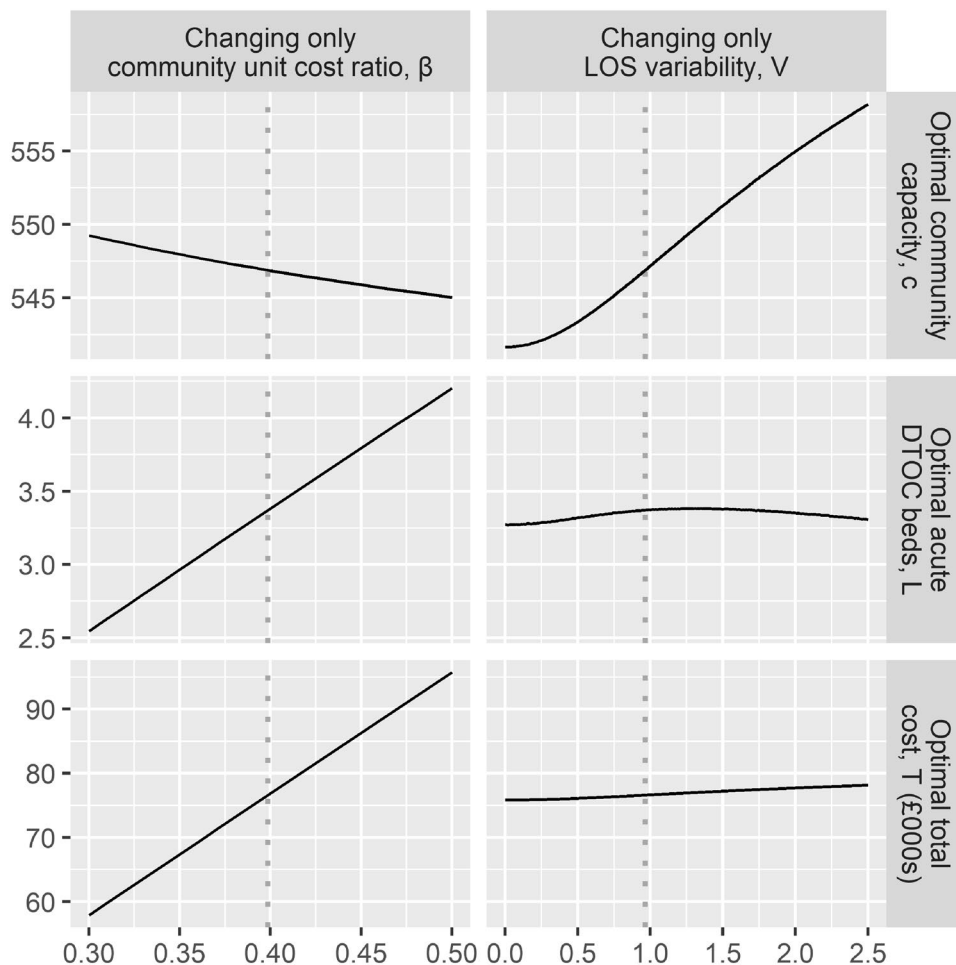
would account for a £510 daily saving, totalling £186k annually.

### 3.3 Effect of Healthcare System Size

Up to this point, all modelling has been based on a particular ‘arrival rate’ of acute patients becoming ready for discharge to the community setting; specifically, that arrival rate of the relatively large healthcare system considered here. Yet, what amounts of acute DTOCs should smaller or even larger healthcare systems work to? Smaller healthcare systems (essentially by definition) would have a lower throughput (arrival rate) and larger healthcare systems a higher throughput. The premise being tested is whether the optimal amount of acute DTOCs scales linearly with size, or whether there is, as would reasonably be expected, an element of queueing ‘economies of scale’.

Modelled results show that the relationship is non-linear (Fig. 4). For the first ten daily patient arrivals, an average 2.2 ‘blocked’ acute beds should be accepted. Yet, for the second ten daily arrivals, only an extra 0.9 acute beds should be tolerated, and fewer still (0.7) for the next set of

**Fig. 3** Modelled results showing the effect of changing only the community cost ratio (left-hand side) or the amount of LOS variability (right-hand side) on optimal levels of community capacity, acute DTOCs, and optimal cost. The dotted grey vertical lines represent the actual levels of community cost ratio and LOS variability at the healthcare system studied. *DTOCs* delayed transfers of care, *LOS* length of stay



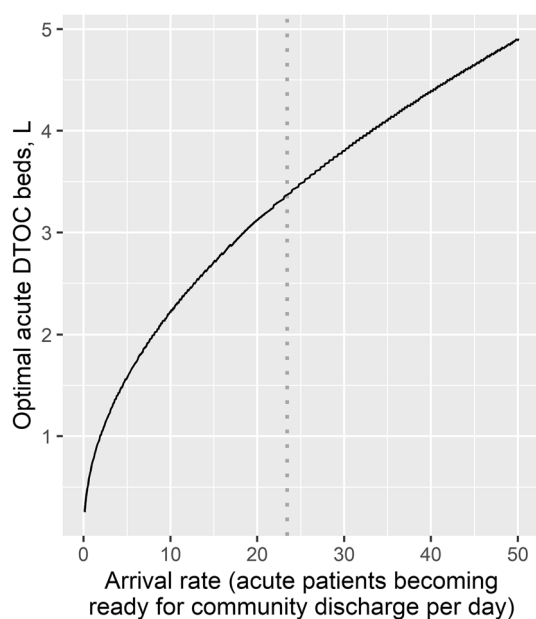


ten. Eventually, the advantages of this queueing ‘economy of scale’ settles down, and the relationship becomes linear for progressively larger systems (Fig. 4). Essentially, this reveals that smaller healthcare systems need to accept proportionately more acute DTOCs to be cost-optimal than larger healthcare systems.

### 3.4 Cost Saving Potential for England

It was earlier estimated that for the studied 1 million population healthcare system, a saving of 7% could be made on total costs were pathway resources appropriately rebalanced. The 7% saving would represent £5725 per day, totalling £2.1m annually. Key to the specific value of this saving is the typical number of acute beds being occupied by DTOC patients awaiting transfer to a community care setting. This is 26.3 for the healthcare system studied but, were it closer to the optimal 3.3 acute beds, then clearly there would be less potential cost saving as the pathway would already be closer to optimality (Fig. 2).

Proportionately upscaling the £2.1m saving to the national level therefore requires sufficient similarity to this ‘starting place’. Using the same centrally provided data as for the studied healthcare system, and applying the same definitions, gives an average 1489 acute beds occupied by community DTOC patients in all of England (see the ESM). With the approximate 56 million England population, this corresponds to an average 26.6 acute DTOC beds occupied



**Fig. 4** Modelled results showing the optimal amount of acute DTOC beds for different sized healthcare systems, as represented through the ‘arrival’ rate of acute patients requiring community care upon discharge. The dotted grey vertical line represents the arrival rate at the healthcare system studied. *DTOCs* delayed transfers of care

per 1 million population. This compares favourably to the above-mentioned 26.3 (for the 1 million BNSSG population), suggesting that local circumstances are not unrepresentative of those nationally.

Thus, multiplying through the £2.1m by 56 provides a £117m estimate of the potential annual cost saving in England. As calculated earlier, this would be achieved through a 3% expansion of post-acute community capacity, which would reduce acute DTOCs to the 12% of current levels that can be regarded as economically essential. This additional capacity would cost £45m, versus the £162m cost saving in the acute setting. Note that the net cost saving relates only to improvements in NHS-funded community care and does not represent the full cost saving potential that may be possible if improvements were also made to local authority social care (which, as mentioned in the Introduction section, is accountable for 28% of DTOCs).

## 4 Discussion

Although a stated aim for many NHS systems [6–10], this study finds that seeking to ‘eliminate’ DTOCs is likely a false economy: simply put, the additional community costs that would be required to achieve this aim are greater than the possible acute cost saving. We find that a large proportion (88%) of the average 26.6 acute DTOC beds per million population in England’s NHS can be attributed to inefficient use of resources. This indicates substantial scope for improving the efficiency of complex discharge pathways, through significantly reducing, but not eradicating, acute DTOCs. This study makes an important contribution to the literature: while others have linked community resources to acute DTOCs, there has been a deficit of work considering optimal resource allocation.

In terms of the £117m cost saving estimated here, the literature provides few examples for direct comparison. The National Audit Office has previously estimated a gross cost of £820m for acute DTOCs in England [25]. However, this figure, calculated in 2016, accounts for delays caused by local authority-funded social care as well as those attributable to NHS-funded community services (the scope of the current study). It also includes a 2.7-fold uplift (of fairly nebulous provenance) applied to the DTOC volumes “to account for patients no longer benefiting from acute care, excluded from the definition of delayed transfers of care”. While the non-acute costs required to achieve the removal of acute DTOCs are estimated, there is no consideration to the variability that realistically exists and which prohibits the kind of simplistic analysis based on DTOC ‘elimination’. In another study, Allan et al. [2] use regression modelling to consider the effect of local authority-funded social care spending on acute DTOCs in England. They find that “home

care supply significantly reduced DTOCs” and that increases to social care capacity over the period from 2011 to 2016 led to overall cost savings of between £73m and £274m.

With respect to practical implications, the findings of this study suggest that aiming for ‘zero DTOC’ targets is misguided. Instead, an assessment should be made based on the particular characteristics and conditions of the considered hospital or healthcare system. This would avoid the larger community capacity requirements associated with ‘eliminating’ DTOCs, which is particularly important given the current and longer-term workforce recruitment and retention issues in the healthcare sector [26]. Managers can also lessen required capacity through reducing inherent variability in the system. Here, we demonstrate the effect of this on community capacity through hypothetically adjusting length of stay variability. Efforts could also be made to reduce the variability in ‘arrivals’, e.g. through ‘smoothing’ the scheduling of elective care procedures for patients expected to require continuing post-acute care [27]. Another option, not explored in this current study, is to optimise the real-time procurement of short-term agency capacity, to coincide with community demand peaks as they occur. This would involve balancing the higher costs of agency resources against the value they bring in reducing acute DTOCs and increasing community capacity utilisation (such an approach has previously been considered in the critical care setting [28]).

In terms of technical limitations, it should be acknowledged that one mathematical formula presented in the Methods section (Eq. 3) is an approximation and not an exact result, although it has been found to be “usually an excellent approximation” [29]. It is also worth acknowledging that data from a variety of sources are used for model calibration, since no single source was available. While such data limitations are a recognised reality of working across various healthcare settings [30], and especially community care [31], they would not be expected to materially affect the overall conclusions of this study (although they may have some limited effect on the particular results; for instance, if more patients are routed to cheaper home-based rather than bed-based community care, then cost optimality would result from increasing community capacity by some amount, while further penalising acute DTOCs).

With respect to the scope of this study, it should be noted that our analysis covers only DTOCs attributable to NHS-funded community care, and not those related to disputes, patient choice, and insufficiency of local authority-funded social care capacity (the latter of which could be considered using a similar approach to that taken here). In terms of financial costs, the scope is restricted only to those directly associated with acute and community service provision, and not those related to other parts of the patient pathway. Upstream, DTOCs associate with

increased acute bed occupancy, which can lead to emergency department overcrowding and, in turn, ambulance offload delays, all of which require additional capacity and cost to address [32, 33]. DTOCs may also lead to a greater risk of hospital-acquired infection, again requiring additional resources to address. Downstream, the effect of DTOCs on functional and cognitive decline (so-called ‘deconditioning’) could mean a costlier long-term care placement when cheaper home-based care could otherwise have sufficed [34]. The effect of both these upstream and downstream effects would further penalise acute DTOCs, somewhat lowering the tolerated ‘optimal’ levels reported in the Results section; so too would consideration to the patient’s own preference for timely transfer upon discharge readiness, although further thought would be required as to how this ‘quality of life’ aspect can be accounted for in equations otherwise denominated by financial cost alone.

## 5 Conclusions

The additional community costs required to ‘eliminate’ DTOCs are greater than the possible acute cost saving. Some amount of DTOCs should therefore be regarded as economically essential, and the use of ‘zero DTOC’ targets should be discouraged. Secondly, this study finds that Operational Research methods such as Queueing Theory can be advantageous to investigating policy matters in this domain.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s40258-022-00777-2>.

**Acknowledgements** The authors are grateful to the reviewers for their helpful comments and suggestions, which have improved the quality and legibility of this article.

## Declarations

**Funding** 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).

**Conflicts of interest** Richard M. Wood, Alison L. Harper, Zehra Onen-Dumlu, Paul G. Forte, Martin Pitt, and Christos Vasilakis declare they have no conflicts of interest.

**Data availability** All data generated or analysed during this study are included in this published article.

**Code availability** The code and data used are available from the corresponding author upon reasonable request.

**Author contributions** RMW: Conceptualisation, methodology, data curation, formal analysis, investigation, funding acquisition, resources, visualisation, writing—original draft. ALH: Methodology, writing—review and editing. ZO-D: methodology, writing—review and editing. PGF: Methodology, writing—review and editing. MP: Funding acquisition, methodology, writing—review and editing. CV: funding acquisition, methodology, writing—review and editing.

**Ethics approval** Not applicable.

**Consent to participate** Not applicable.

**Consent for publication (from patients/participants)** Not applicable.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

## References

- National Health Service (NHS) England. Delayed Transfers of Care Data 2019-20. 2020. <https://www.england.nhs.uk/statistics/statistical-work-areas/delayed-transfers-of-care/delayed-transfers-of-care-data-2019-20/>. Accessed 1 Nov 2022.
- Allan S, Roland D, Malisaukaite G, Jones K, Baxter K, Gridley K, Birks Y. The influence of home care supply on delayed discharges from hospital in England. *BMC Health Serv Res*. 2021;21(1):1297. <https://doi.org/10.1186/s12913-021-07206-5>.
- Hendy P, Patel JH, Kordbacheh T, Laskar N, Harbord M. In-depth analysis of delays to patient discharge: a metropolitan teaching hospital experience. *Clin Med*. 2012;12(4):320. <https://doi.org/10.7861/clinmedicine.12-4-320>.
- Holmås TH, Kamrul Islam M, Kjerstad E. Between two beds: inappropriately delayed discharges from hospitals. *Int J Health Care Finance Econ*. 2013;13(3):201–17. <https://doi.org/10.1007/s10754-013-9135-4>.
- World Health Organisation. WHO Director-General's opening remarks at the media briefing on COVID-19 – 11 March 2020. <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020>. Accessed 1 Nov 2022.
- Greater Manchester Health and Social Care Partnership. Strategic Partnership Board. 2017. <https://democracy.greatermanchester-ca.gov.uk/documents/s6025/>. Accessed 1 Nov 2022.
- Surrey County Council. Vanguard New Models of Care – Review of Community Beds Model in North-East Hampshire and Farnham Clinical Commissioning Group. 2015. <https://mycouncil.surreycc.gov.uk/documents/s25857/151106%20NE%20Hants%20Farnham%20CCG%20Vanguard.pdf>. Accessed 1 Nov 2022.
- North East London Joint Committee. Outer North East London Joint Health Overview and Scrutiny Committee November 2009. <https://democracy.haverling.gov.uk/Data/Joint%20Health%20Overview%20&%20Scrutiny%20Committee/20091126/Agenda/26%20November%202009,%20Joint%20Health%20Overview%20and%20Scrutiny%20Committee,%20Agenda.pdf>. Accessed 1 Nov 2022.
- Tozer C. Keeping care going in a pandemic. 2020. <https://www.themj.co.uk/Keeping-care-going-in-a-pandemic/217159#>. Accessed 1 Nov 2022.
- Scottish Parliament. Meeting of the Parliament (Hybrid) 09 November 2021. <https://www.parliament.scot/chamber-and-committees/official-report/what-was-said-in-parliament/meeting-of-parliament-09-11-2021?meeting=13392&iob=121554>. Accessed 1 Nov 2022.
- El-Bialy R, Funk L, Thompson G, Smith M, St John P, Roger K, et al. Imperfect solutions to the neoliberal problem of public aging: A critical discourse analysis of public narratives of long-term residential care. *Canadian Journal on Aging/La Revue canadienne du vieillissement*. 2022;41(1):121–34. <https://doi.org/10.1017/S0714980821000325>.
- Harrop N. The four hour target is dead; long live the four hour target! *Clin Gov Int J*. 2013. <https://doi.org/10.1108/CGIJ-07-2013-0024>. Accessed 1 Nov 2022.
- Bittencourt O, Verter V, Yalovsky M. Hospital capacity management based on the queueing theory. *Int J Product Perform Manag*. 2018;67(2):224–38. <https://doi.org/10.1108/IJPPM-12-2015-0193>.
- Harrison GW, Shafer A, Mackay M. Modelling variability in hospital bed occupancy. *Health Care Manag Sci*. 2005;8(4):325–34. <https://doi.org/10.1007/s10729-005-4142-8>.
- Green L. Queueing analysis in healthcare. In: *Patient flow: reducing delay in healthcare delivery*. Boston: Springer; 2006. pp. 281–307. [https://doi.org/10.1007/978-0-387-33636-7\\_10](https://doi.org/10.1007/978-0-387-33636-7_10).
- Kendall DG. Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chain. *Ann Math Stat*. 1953;24(3):338–54.
- Erlang AK. Solution of some problems in the theory of probabilities of significance in automatic telephone exchanges. *Post Office Electric Eng J*. 1917;10:189–97.
- Sztrik J. Basic queueing theory. *Univ Debrecen Fac Inform*. 2012;193:60–7.
- Lee AM, Longton PA. Queueing processes associated with airline passenger check-in. *J Oper Res Soc*. 1959;10(1):56–71. <https://doi.org/10.1057/jors.1959.5>.
- Little JD. A proof for the queuing formula:  $L = \lambda W$ . *Oper Res*. 1961;9(3):383–7. <https://doi.org/10.1287/opre.9.3.383>.
- BNSSG Healthier Together. Primary Care Strategy 2019–2024. 2020. Available at: <https://bnssghealthiertogether.org.uk/wp-content/uploads/2021/07/Healthier-Together-BNSSG-primary-care-strategy.pdf>. Accessed 1 Nov 2022.
- Bristol City Council. JSNA Health and Wellbeing Profile 2021/22. 2021. <https://www.bristol.gov.uk/files/documents/1540-jsna-2021-population/file>. Accessed 1 Nov 2022.
- North Somerset Council. Joint Health and Wellbeing Strategy. 2021–2024. 2021. <https://www.n-somerset.gov.uk/sites/default/files/2022-02/Health%20and%20Wellbeing%20Strategy%202021-24.pdf>. Accessed 1 Nov 2022.
- National Health Service (NHS) Improvement. 2017/18 Reference Costs. 2019. <https://webarchive.nationalarchives.gov.uk/ukgwa/20200501111106/https://improvement.nhs.uk/resources/reference-costs/>. Accessed 1 Nov 2022.
- National Audit Office. Discharging older patients from hospital. 2016. <https://www.nao.org.uk/wp-content/uploads/2015/12/Discharging-older-patients-from-hospital.pdf>. Accessed 1 Nov 2022.
- Anderson M, O'Neill C, Clark JM, Street A, Woods M, Johnston-Webber C, et al. Securing a sustainable and fit-for-purpose UK health and care workforce. *The Lancet*. 2009.



- 2021;397(10288):1992–2011. [https://doi.org/10.1016/S0140-6736\(21\)00231-2](https://doi.org/10.1016/S0140-6736(21)00231-2).
27. Kc DS, Terwiesch C. Benefits of surgical smoothing and spare capacity: an econometric analysis of patient flow. *Prod Oper Manag.* 2017;26(9):1663–84. <https://doi.org/10.1111/poms.12714>.
28. Griffiths JD, Price-Lloyd N, Smithies M, Williams JE. Modelling the requirement for supplementary nurses in an intensive care unit. *J Oper Res Soc.* 2005;56(2):126–33. <https://doi.org/10.1057/palgr.ave.jors.2601882>.
29. Whitt W. Approximations for the GI/G/m queue. *Prod Oper Manag.* 1993;2(2):114–61. <https://doi.org/10.1111/j.1937-5956.1993.tb00094.x>.
30. The Health Foundation. Data Linkage: joining up the dots to improve patient care. 2018. <https://www.health.org.uk/newsletter-feature/data-linkage-joining-up-the-dots-to-improve-patient-care>. Accessed 1 Nov 2022.
31. Witham MD, Frost H, McMurdo M, Donnan PT, McGilchrist M. Construction of a linked health and social care database resource—lessons on process, content and culture. *Inform Health Soc Care.* 2015;40(3):229–39. <https://doi.org/10.3109/17538157.2014.892491>.
32. Affleck A, Parks P, Drummond A, Rowe BH, Ovens HJ. Emergency department overcrowding and access block. *Can J Emerg Med.* 2013;15(6):359–70. <https://doi.org/10.1017/S148180350002451>.
33. Li M, Vanberkel P, Carter AJ. A review on ambulance offload delay literature. *Health Care Manag Sci.* 2019;22(4):658–75. <https://doi.org/10.1007/s10729-018-9450-x>.
34. Everall AC, Guilcher SJ, Cadel L, Asif M, Li J, Kuluski K. Patient and caregiver experience with delayed discharge from a hospital setting: a scoping review. *Health Expect.* 2019;22(5):863–73. <https://doi.org/10.1111/hex.12916>.

## Authors and Affiliations

Richard M. Wood<sup>1,2,4</sup>  · Alison L. Harper<sup>3,4</sup> · Zehra Onen-Dumlu<sup>2,4</sup> · Paul G. Forte<sup>1,4</sup> · Martin Pitt<sup>3,4</sup> · Christos Vasilakis<sup>2,4</sup>

✉ Richard M. Wood  
richard.wood16@nhs.net

<sup>1</sup> UK National Health Service (BNSSG ICB), NHS Bristol, North Somerset and South Gloucestershire Integrated Care Board, 360 Bristol, Marlborough St, Bristol BS1 3NX, UK

<sup>2</sup> School of Management, University of Bath, Bath, UK

<sup>3</sup> Medical School, University of Exeter, Exeter, UK

<sup>4</sup> Health Data Research UK, South West Better Care Partnership, Bristol, UK