On Hospice Operations under Medicare Reimbursement Policies

Barış Ata*
Northwestern University

Bradley L. Killaly†
Emory University

Tava Lennon Olsen‡
The University of Auckland

Rodney P. Parker§
The University of Chicago

September, 2010; revised February, 2011, November 2011, May 2012

Abstract

This paper analyzes the United States Medicare hospice reimbursement policy. The existing policy consists of a daily payment for each patient under care with a global cap of revenues accrued during the Medicare year, which increases with each newly admitted patient. We investigate the hospice’s expected profit and provide reasons for a spate of recent provider bankruptcies related to the reimbursement policy; recommendations to alleviate these problems are given. We also analyze a hospice’s incentives for patient management, finding several unintended consequences of the Medicare reimbursement policy. Specifically, a hospice may seek short-lived patients (such as cancer patients) over patients with longer expected length-of-stay and the effort with which they seek-out, or recruit, such patients will vary during the year. Further, the effort they apply to actively discharge patients whose condition has stabilized may also depend on the time of year. These phenomena are unintended and undesirable but are a direct consequence of the Medicare reimbursement policy. We propose an alternative reimbursement policy which ameliorates these shortcomings.

*Email: b-ata@kellogg.northwestern.edu
†Email: bkillal@emory.edu
‡Email: t.olsen@auckland.ac.nz
§Email: rodney.parker@chicagobooth.edu
1 Introduction and Problem Motivation

Hospices are health care providers that cater to patients in the end phases of their lives who choose to undertake palliative care in lieu of further curative care. Since 1983 Medicare has reimbursed hospices for the care provided to eligible patients. This reimbursement consists of a daily payment for care, but the American federal government’s exposure is limited by an annual cap that depends on the number of patients admitted in a year. The Medicare hospice benefit is generally regarded as a success because it improves the quality of life while saving Medicare (compared with a patient continuing curative care) an average of $2309 per hospice user (Taylor et al., 2007). However, some recent issues have arisen to question the efficacy of this reimbursement policy. Specifically an increasing number of providers have entered bankruptcy, and the blame for this is attributed to the structure of the reimbursement policy (Sack, 2007). We investigate this claim analytically, study hospice provider incentives under the current policy, and explore alternative Medicare policy possibilities.

More specifically, we formulate a model for hospice profit and use it to examine the potential causes for hospices receiving payments exceeding the cap and the reasons behind potential bankruptcies. However, the payment scheme elapsing over a finite horizon raises further issues beyond the profitability of the provider. The annualized accounting involved may be leading to some undesirable traits in the rate of hospice admissions and discharges, such as patient recruiting and discharge rates which differ across diseases and change during the year. If this is the case, it is likely contrary to the United States Government’s equity objectives.

There is indeed evidence to suggest that some untoward patient management practices occur in the hospice industry, some of which are compelled by the reimbursement policy’s cap. Jenkins et al. (2010) report hospice providers confirm that they modify their practices when they are in danger of exceeding their cap. In a survey of 55% of the hospice providers in Alabama (a state frequently mentioned by MedPAC, 2010, as having providers exceeding their cap), 24.4% of respondees reported modifying their actions when faced with the prospect of a binding cap. Of such modifications, the two most cited behaviors were discharging patients (17% of all respondees) and marketing “to a certain type of patient (e.g., cancer patients)” (10.4% of all respondees) to alleviate cap problems; we seek to model these two behaviors in our work. Actions such as these are a sufficient concern that the Medicare Payment Advisory Commission has advised the Secretary of Health and Human Services to direct the Office of the Inspector General to investigate the financial
relationships between “hospices and long-term care facilities such as nursing facilities and assisted living facilities that may represent a conflict of interest and influence admissions to hospice,” and “the appropriateness of hospice marketing materials and other admissions practices and potential correlations between length of stay and deficiencies in marketing or admissions practices” (MedPAC, 2010).

This paper investigates the hospice manager’s optimal recruitment problem and finds that, indeed, the manager has an incentive to purposefully seek cancer and other short-lived patients when the hospice’s cap might be exceeded. We also investigate the incentives behind the practice of “live discharges” (MedPAC, 2010). A live-discharge is a living patient who is released from a hospice. Given the unpredictable trajectory of terminal diseases, some patients’ diseases simply do not follow expectations, a patient may recover, or a patient may simply elect to resume conventional medical treatment, which of course results in their leaving the hospice. However, a hospice caring for patients who have exceeded their contribution to the cap may feel the pressure to discharge such patients, however unethical (and potentially illegal) such a practice may be, a possibility we explicitly include in our model. Indeed, our model shows that the hospice manager may find it optimal to live-discharge patients particularly towards the end of the fiscal year if the current cap position is not desirable. Moreover, in those cases the rate of live-discharges increases toward the end of the fiscal year. To remedy these unintended consequences, we propose an alternative reimbursement policy and show that it indeed alleviates these nonstationary recruiting or live-discharge patterns.

The remainder of the paper is structured as follows. Section 2 briefly describes the existing Medicare reimbursement policy and presents the institutional details and motivation for our work, including Medicare and hospice-centered references. In §3 the hospice profitability model is formulated and analyzed, and remedies to the policy’s shortcomings are considered. In §4 we formulate and analyze a dynamic deterministic recruitment and discharge model, which highlights several disturbing unintended consequences of the existing policy. We also perform a simulation study to study the effects of stochasticity on our conclusions. Section 5 presents a policy to overcome the unintended consequences of the existing policy. Concluding remarks are provided in §6. Proofs are relegated to an appendix (Appendices A and B) throughout and Appendix C contains details of our simulation study; these appendices may be obtained from the authors.
2 Medicare Reimbursement Policy and Literature Review

First, we review the features of Medicare’s reimbursement policy relevant for our purposes; the reader is referred to MedPAC (2010) for a more comprehensive description of the current policy. We then review the relevant literature on the hospice system and Medicare reimbursement program.

The Medicare year runs from November 1 through October 31 of the following year. To be admitted to a hospice, a patient needs the signature of two physicians (typically, one will be the patient’s primary attending physician and the other will be employed by the hospice), certifying that the patient is not expected to live more than six months from admission,\(^1\) and that the patient agrees to forgo any curative care and undertake palliative care only.

During the Medicare year, a hospice receives a payment for a patient under hospice care for a part or entire day. This payment differs according to whether the patient is receiving routine home care ($142.91 per day), continuous home care ($834.10 per day), in-patient respite care ($147.83 per day), or general inpatient care ($635.74 per day).\(^2\) These payment rates can differ slightly by region in the United States depending on estimates of the cost of operation in these regions but do not depend upon the disease afflicting the patient. In 2002 and 2003, 93% of reimbursed hospice days were paid at the routine home care rate, 4.1% were continuous home care days, 2.7% were inpatient respite care days, and 0.2% were general inpatient care days (MedPAC, 2006). We focus on routine home care in our models, which are the vast majority of reimbursed days.

The second part of the Medicare hospice reimbursement policy is a payment cap, applied to the entire hospice (i.e., the cap is not patient specific),\(^3\) intended to limit the government’s exposure. At the beginning of the Medicare year, this cap is zero but it increases by $23,014.50\(^4\) for every newly admitted patient. The daily payment rates and cap increase quantities are adjusted from year to year, but unlike the daily rates, the cap does not vary by geography. At the end of the Medicare year, if the hospice’s cap is less than the total of the daily payments received then this

\(^{1}\)If the patient lives beyond the initial six month period, the patient can be recertified for two sequential 90 day periods, following by an unlimited number of sequential 60 day periods. Until 1990, there was a limit of 210 days over which a hospice could receive payments for a patient.


\(^{3}\)Medicare applies a second cap, which is rarely enacted. It limits the proportion of inpatient care days to 20% of all reimbursed days. This is to encourage hospice care in the patient’s home to be the primary method of delivery. Any days exceeding the 20% level will be reimbursed at the routine home care rate.

\(^{4}\)Again for the year 2009.
excess amount must be repaid to Medicare. If the cap is greater than the payments received, then
the hospice did not receive as many payments as they were eligible for but no adjustment is made.
The cap is reset to zero at the beginning of the new Medicare year.

Since the creation of the Medicare hospice benefit in 1983, there has been much research on
many aspects of hospices but, to the best of our knowledge, none have directly addressed the issues
focused on here. In fact, much of the literature does not address the reimbursement policy at all. An
exception is Fraser (1985) who describes the ethical and policy implications of Medicare’s hospice
reimbursement policy but does not identify the issues of provider bankruptcy or non-stationary
and disease-dependent recruitment and discharge, the foci of our study. GAO (2004) investigates
whether modifications to the reimbursement policy are warranted but limit their focus to the
comparison of the per diem rates and the costs of care. Huskamp et al. (2001) consider how the
Medicare rules affect care. There is consistent evidence in the literature that the hospice benefit
reduces Medicare costs (e.g., Campbell et al., 2004) while enhancing end of life care (Pyensen et
al., 2004; Taylor et al., 2007).

There has been previous empirical work that examines common characteristics of hospices that
exceed their caps. In particular, MedPAC (2010) reports the following attributes. They tend to: be
for-profit freestanding facilities;\(^5\) have a smaller patient census; treat a larger share of patients with
Alzheimer’s disease and other neurological conditions; exhibit significantly longer length-of-stay
(LOS), even when patient mix is taken into account; and have a proportion of patients with stays
exceeding 180 days (for particular diseases) substantially higher than those hospices below the cap.

Further, there is an increasing trend of hospices receiving payments greater than what they were
permitted (i.e., exceeding their caps) and having to repay this excess to the Federal government.
For example, MedPAC (2010) reports the percentage of all hospices exceeding their caps rose from
2.6% in 2002 to 10.4% in 2007 and points to two explanatory factors. Firstly, there has been an
increase in the proportion of longer stay patients and secondly, there has been an increase in the
LOS of the longer stay patients. In 1998 forty-seven percent of all hospice users had noncancer
diagnoses, whereas this has risen to 69 percent in 2008 (MedPAC, 2010). While MedPAC (2010)
reports the median LOS remained steady at 17 days between 2000 and 2008, the 90th percentile
grew from 141 days to 235 days. In brief, the short stays kept at a similar LOS but the long
stays grew longer. There appear to be no definitive explanations for these extended LOS, although
MedPAC is sufficiently concerned that they recommended Congress direct the Secretary of Health

---

\(^5\)Freestanding facilities are those not operated by a hospital, home health agency, or skilled nursing facility.
and Human Services to require a medical review of all stays exceeding 180 days in hospices where such stays make up 40% or more of all cases (MedPAC, 2010). If the trends of increasing proportions of noncancer patients and longer lifespans for the longest-living noncancer patients continue, we envision there will be more and more financially distressed hospices, particularly if they happen to be smaller providers. The results in our paper suggest this will lead to further and more extreme distortions of incentives. Given that MedPAC is aware of these trends, we suspect they will intervene to preserve the hospice benefit for truly terminal patients, either through stricter enforcement of the six-month LOS estimate or another mechanism. Modeling this is beyond the scope of our paper.

Concerning hospice costs, MedPAC (2010) finds that the average provider costs per day can vary by hospice type, finding that for-profit based hospices are less costly than non-profit hospices; rural hospices are less costly than urban hospices; and, curiously, hospices exceeding the payment cap are less costly than those below the cap. It is also found that the daily costs are higher at admission and discharge than regular care, so providers with longer average LOS have lower daily costs, which may explain why hospices exceeding the payment cap are less costly. Killaly et al. (2007) find the providers’ marginal and average costs are higher for cancer patients than non-cancer patients. MedPAC also examines hospice margins and generally all hospice types are profitable except hospital based providers (which consistently have negative margins from 2001 to 2007, presumably the result of greater overhead allocation) and providers with the lowest patient volumes. MedPAC (2010) finds hospice margins increase with patient volume in every year under study but projects the aggregate margin to drop from 5.9% in 2007 to 4.6% in 2010 across all hospices.

Regarding live-discharges, Taylor et al. (2008) suggests 15.5% of hospice users were discharged alive from years 1993 to 2000. MedPAC (2010) finds live discharges are far more prevalent amongst providers exceeding their caps (46% of all discharges) than those not exceeding their caps (16%) in 2007, and consequently recommended the Office of the Inspector General investigate the “appropriateness of enrollment practices among hospices with unusual utilization patterns.” Carlson et al. (2009) find that live-discharges are more prevalent for smaller hospices and that long-stay patients may be more susceptible to this practice, two factors which our findings suggest can lead to diminished profitability, although newer hospices were also commonly found to be live-dischargers, perhaps implying some inexperience in judging LOS may play a factor.
3 Model of the Current Policy

This section presents a static model of the existing Medicare policy for hospice reimbursement. We consider elements of the industry and market that may affect hospice profitability, including patient census, patient disease mix, and length-of-stay (LOS) uncertainty. The detrimental nature of poor mix realization, lack of scale, and uncertainty are well recognized in the operations management literature. For example, Eppen (1979) recognized the value of pooling inventory in the context of warehouses. Such lessons are instructive for analyzing hospice operations and we create a model to do so. To the best of our knowledge, there does not appear to be work in the operations management literature dealing with a cap akin to the Medicare hospice cap, although there do exist inventory papers with shared limited production capacity (e.g., Evans, 1967). Recently, there has been a substantial increase in the interest of applying operations management techniques to healthcare related topics (see, for example, Brandeau et al., 2004).

We express the general form for the hospice’s profit given the Medicare reimbursement policy and then simplify the demand model to make it tractable. The fiscal year is considered as a whole; specific dynamics caused by year-end effects will be discussed in Section 4. In particular, all patients are assumed to arrive, be cared for, and be reimbursed for in the same fiscal year. While in practice there will be patients that live from one Medicare year to the next, we feel this model captures the key effects regarding hospice profitability at a high level without losing tractability or insights.

A hospice provider faces uncertainty concerning the patients’ LOS. It is recognized that these uncertain LOSs differ by patient disease (Christakis et al., 1996), although the Medicare reimbursement policy is independent of disease-type. Our models consider only two disease types, although the results are robust to this characterization. Patients are classified as type 1 or type 2, characterized by the former suffering shorter mean LOS and the hospice incurring a lower marginal cost for the latter. In a broad sense, one could think of type 1 patients as those suffering from cancer and type 2 as non-cancer, although this categorization is far from perfect as there are several noncancerous diseases with shorter LOS than several cancer diseases (e.g., chronic kidney disease has a mean LOS of 28-30 days, CMMS, 2009). For the remainder of the paper, we refer to two patient types.

Let $N_1$ be the number of type 1 (short LOS) patients admitted in a year and $N_2$ be the number of type 2 patients admitted. Assume that $X_{1j}$ is the remaining life of type 1 patient $j$ ($j = 1, \ldots, N_1$), and $X_{2k}$ is the remaining life of type 2 patient $k$ ($k = 1, \ldots, N_2$). Further, assume type $i$ patients
cost \( c_i \) per day to treat, \( i = 1, 2 \), and let \( A \) be the fixed cost of operating the hospice. Let \( r \) be the daily pre-cap reimbursement rate for patients and \( K \) the cap adjustment per admitted patient. Then

\[
\text{Profit Rate (per year)} = r \left( N_1 \sum_{j=1}^{N_1} X_j^1 + N_2 \sum_{k=1}^{N_2} X_k^2 \right) \wedge K(N_1 + N_2) - c_1 \sum_{j=1}^{N_1} X_j^1 - c_2 \sum_{k=1}^{N_2} X_k^2 - A.
\]

As discussed above, this ignores beginning and end-of-horizon effects. We will refer to this model as our static model and relaxation of this assumption will occur with the dynamic model of Section 4.

For simplicity we assume \( N_1 \) and \( N_2 \) are not random: \( N_1 = \lambda_1 \) and \( N_2 = \lambda_2 \). We also assume \( \{X_j^1\} \) and \( \{X_k^2\} \) are independent and identically distributed (i.i.d.) sequences, and independent of each other. Then, assuming \( \lambda_1 \) and \( \lambda_2 \) are sufficiently large, we conclude by the central limit theorem that

\[
\text{uncapped revenues} = r \left( \sum_{j=1}^{N_1} X_j^1 + \sum_{k=1}^{N_2} X_k^2 \right) \sim rN(m, \sigma^2),
\]

where \( m = (\lambda_1 m_1 + \lambda_2 m_2) \), \( \sigma^2 = (\lambda_1 \sigma_1^2 + \lambda_2 \sigma_2^2) \), and \( m_i \), \( \sigma_i^2 \) denote the mean and the variance of \( X_i^1 \), respectively, for \( i = 1, 2 \). Let \( \phi(\cdot) \) and \( \Phi(\cdot) \) denote the pdf and cdf of the standard normal distribution, respectively. The following proposition characterizes the hospice’s annual profit.

**Proposition 1** The expected annual (static) profit is given by

\[
\pi = \lambda \left\{ K - (K - r\bar{m}) \Phi \left( \frac{K/r - \bar{m}}{\bar{\sigma}/\sqrt{\lambda}} \right) - r\bar{\sigma} \frac{\phi \left( \frac{K/r - \bar{m}}{\bar{\sigma}/\sqrt{\lambda}} \right)}{\sqrt{\lambda}} - \bar{c} - \frac{A}{\lambda} \right\}
\]

where \( \lambda = \lambda_1 + \lambda_2 \), \( \bar{m} = \frac{\lambda_1}{\lambda} m_1 + \frac{\lambda_2}{\lambda} m_2 \), \( \bar{\sigma} = \sqrt{\frac{\lambda_1}{\lambda} \sigma_1^2 + \frac{\lambda_2}{\lambda} \sigma_2^2} \) and \( \bar{c} = \frac{\lambda_1}{\lambda} (c_1 m_1) + \frac{\lambda_2}{\lambda} (c_2 m_2) \). Static profit \( \pi \) is concave increasing in \( K \), concave in \( r \), decreasing in \( \sigma_1 \) and \( \sigma_2 \), and linearly decreasing in \( c_1 \), \( c_2 \), and \( A \).

The following insights are immediate from Proposition 1. Firstly, high costs or low revenues obviously affect profits adversely. Secondly, so long as the per patient margin is positive, large volumes are better for the hospice manager. Thirdly, the hospice’s profit decreases as the variability (\( \bar{\sigma} \)) increases. Finally, the effects of patient mix and LOS are also important in determining profitability.\(^6\)

As mentioned in the introduction, hospices can go bankrupt due to several reasons. The most important among these are volume and mix. A hospice with insufficient volume cannot cover its

\(^6\)Even though Section 4 shows that the annualized cap provides incentives for undesirable behavior, pooling the cap across the patients helps the hospice manager. If the cap were simply per patient, then the presence of variability in LOS could be extremely detrimental to the hospice because of lack of any risk pooling.
fixed costs. Further, the negative impact of LOS variability diminishes as $\lambda$ increases. To be specific, it can be seen from (1) that a hospice’s profit loss due to LOS variability decreases as $\lambda$ increases, due to the term $\bar{\sigma}/\sqrt{\lambda}$, which captures effective LOS variability.

The second important reason for bankruptcy is improper mix. Intuitively, the extremes of the mix spectrum hurts the hospice profit. To see this, consider a hospice serving primarily cancer (short LOS) patients. Because its patients do not live very long the hospice’s per patient revenue is low, and in particular, insufficient to cover its fixed cost unless it serves a very large number of patients. At the other extreme, consider a hospice serving primarily non-cancer (long LOS) patients. Because its patients live a long time, the cap constraint will bind. Therefore, the hospice will accrue revenues during only a portion of a patient’s stay (because the cap binds) whereas it incurs caring costs throughout the patient’s stay, which may cause the hospice to go bankrupt. The ideal mix values are those which allow the hospice manager to leverage the benefits of both types of patients: cancer (short LOS) patients help build a large cap, but the hospice manager cannot take advantage of that cap with cancer patients only, whereas non-cancer (long LOS) patients help convert the cap to revenue.\footnote{To glean further insights from Proposition 1, consider the asymptotic regime where $\lambda$ gets large while the mix remains the same. Then, using the fact that $\lim_{z \to \infty} z(1 - \Phi(z))/\phi(z) = 1$, cf. Zipkin (2000), the hospice’s profit can be approximated for sufficiently large $\lambda$, as follows. If $K \neq r\bar{m}$, then $\pi \approx \lambda[K \wedge (r\bar{m}) - \bar{c} - A/\lambda]$. Similarly, if $K = r\bar{m}$, then $\pi \approx \lambda[K \wedge (r\bar{m}) - (r\bar{\sigma}\phi(0))/\sqrt{\lambda} - \bar{c} - A/\lambda]$. These equations highlight the importance of the mix and the cap through the term $K \wedge (r\bar{m}) - \bar{c}$, which shows that extreme values of mix may hurt profitability. These equations also validate that the hospice becomes more profitable as $\lambda$ increases so long as per patient margin is positive. The LOS variability has a first-order effect in profitability only when $K = r\bar{m}$, in which case increasing $\lambda$ helps increase profit through risk pooling as well because of the term $\bar{\sigma}/\sqrt{\lambda}$. Otherwise, i.e., $K \neq r\bar{m}$, the revenue loss due to the imbalance between the cap and per patient revenue dominate the revenue loss due to LOS variability.}

One remedy we propose to prevent possible bankruptcy is for appropriate hospices to merge. There are several potential benefits from this. Firstly, the merged hospice will have larger volumes and expected profits are increasing in scale. Secondly, the merged hospice enjoys the well-known benefits of pooling (see, e.g., Eppen, 1979) where the relative variability is reduced and Proposition 1 shows the hospice profit decreases in variability. Finally, and most importantly, if the constituent hospices are chosen well, the resulting patient mix in the merged provider could result in a more robust operating mix. For example, MedPAC (2010) highlights the issues of hospices in Alabama and Mississippi exceeding the cap, due primarily to the exceedingly long LOSs, and Sack (2007) highlights the dramatically shorter LOSs in South Dakota; such hospices could be merged.\footnote{Mergers to correct mix imbalance only make sense if patient mix is relatively stable across facilities.}
Despite merging hospices being attractive from a profit perspective, the practicalities of implementing such mergers must also be examined, which is beyond the scope of this paper. We will note that the government is in the prime position to act as a “match-maker” to identify candidates for and encourage such mergers and, perhaps more practically, to remove any regulatory hurdles to inhibit the mergers. For example, hospices in different states with a common owner are not currently permitted to operate with a common cap. However, by doing so they may also make the hospice benefit more expensive to Medicare (although still within their budgeted cap for admitted patients) and might need to reduce the hospice per diem to make such a recommendation revenue neutral, which is not without its own complications.\(^9\) Further study on the precise cost implications may be necessary before implementation.

As noted above, when carefully chosen, such mergers could potentially alleviate unfortunate systemic patient mix issues. Further, we feel that correcting unprofitable patient mix through hospice mergers is a more socially equitable (and ethical) solution than allowing hospices to adjust their mix by actively searching for patients of the “right” type (inducing non-terminally ill patients to join the hospice) or live-discharging patients (who are eligible for the benefit) to improve the hospice’s position relative to the cap; under this latter solution all patient types may no longer receive equal hospice access. In the following section, we show that hospices indeed have an incentive to actively search for the right patient type and live-discharge certain patients to improve their position relative to the cap, and that the optimal search and live-discharge strategies change over the course of a year, further exacerbating equal access concerns.

4 The Hospice Manager’s Problem

This section presents the patient management challenge a hospice manager faces in order to preserve profitability. We formulate and analyze a dynamic deterministic model where there are regular arrivals of hospice patients of two types during the Medicare year; in addition to these regular arrivals (at differing rates) of patients, hospices are allowed to seek, or “recruit,” additional admissions of these patient types during the year, and are allowed to “live-discharge” patients whose condition has stabilized. The recruiting rates of additional patients (of each type) and the discharge rates for the pool of stable patients are the decision variables of this dynamic model. Our model assumes that hospices do have the ability to recruit beyond their natural arrival rates but that such recruit-

\(^9\)Medicare has sometimes reduced the hospice benefit per diem for some types of care.
ing has a convex increasing cost, because diseconomies of scale appear natural in a limited and competitive market. Similarly, we also allow the hospice manager to live-discharge patients, should their condition stabilize, also with a convex increasing cost. Note that we do this in an effort to inform policy makers of the incentives inherent in the current system, and do not propose these rates as a prescription for hospice management.

4.1 A Fluid Model

The model advanced in this section treats arrivals of patients to each class as fluid arriving at the system at a constant rate. In particular, class \( i \) customers arrive at rate \( \lambda_i \) \((i = 1, 2)\) per time unit. Although all patients are diagnosed as terminally ill at admission to the hospice, the diagnosis may turn out to be false in some cases, and those patients may be live-discharged (should their condition stabilize). Let \( a \) and \( b \) index the patients who are truly terminally ill and those who may be live-discharged, respectively. For \( i = 1, 2 \), \( \gamma^a_i \) and \( \gamma^b_i \) denote their fraction \((\gamma^a_i + \gamma^b_i = 1)\). We take this approach because assuming all patients may be discharged would be problematic because discharging the truly terminal is both immoral and illegal. Our model is based on a single fiscal year, with time indexed by \( t \in [0, T] \).

Terminally ill patients and the misdiagnosed ones have different length-of-stay (LOS) distributions. The lifespan of a class \( i, j \) patient has mean \( m^j_i \) and density \( f^j_i(\cdot) \) for \( i = 1, 2 \) and \( j = a, b \). Although the hospice manager cannot tell apart the patients who are misdiagnosed at admission, she can do so during their stay at the hospice. As before, let \( r_j^i \) and \( c_j^i \) denote the daily revenue and cost of caring for a class \( i, j \) patient \((i = 1, 2, j = a, b)\), respectively. Consider a patient of class \( i, j \) admitted at time \( t \in [0, T] \), and let \( r^i_j(t) \) and \( c^i_j(t) \) denote potential revenues to be collected from Medicare and the cost of caring for that patient over the remainder of the fiscal year \([t, T]\), respectively.

At the end of the year, we use a terminal value \( v^j_i \) to denote the ongoing value of a class \( i, j \) patient \((i = 1, 2 \text{ and } j = a, b)\) who lives into the following fiscal year. Note that this value may be negative if the hospice’s revenue is severely constrained by its cap, so that the patient costs money to treat but brings in little to no revenue. Assuming a fixed terminal value for all patients (regardless of when they were admitted) is only exact if LOS distributions are exponential and must otherwise be considered an approximation based on the average LOS. Terminal values are explored further in Section 4.2. We let \( v^i_j(t) \) denote the terminal value attributed to a class \( i, j \) patient arriving at time \( t \). The following proposition characterizes \( r^i_j(t), c^i_j(t) \), and \( v^i_j(t) \).
Proposition 2  For \( t \in [0, T] \), \( i = 1, 2 \), and \( j = a, b \),

\[
\begin{align*}
    r_j^i(t) &= r \int_0^\infty [x \wedge (T - t)] f_j^i(x) \, dx, \\
    c_j^i(t) &= c_j^i \int_0^\infty [x \wedge (T - t)] f_j^i(x) \, dx, \quad \text{and} \\
    v_j^i(t) &= v_j^i \int_{T-t}^\infty f_j^i(x) \, dx.
\end{align*}
\]

For ease of notation we define

\[
\begin{align*}
    r_i(t) := \sum_{j=a}^{b} \gamma_j^i r_j^i(t), \quad c_i(t) := \sum_{j=a}^{b} \gamma_j^i c_j^i(t), \quad v_i(t) := \sum_{j=a}^{b} \gamma_j^i v_j^i(t).
\end{align*}
\]

The hospice manager is required to admit all arriving patients but also faces a decision as to whether or not to actively recruit more patients. Recall that, during admission, the hospice manager cannot tell apart the patients who may be live-discharged. Let \( \alpha_i(t) \) denote the rate at which the hospice manager recruits class \( i \) patients at time \( t \). There is a convex increasing cost \( s_i(\alpha) \) associated with recruiting (or searching for) class \( i \) patients at rate \( \alpha \). For concreteness, assume \( s_i(\alpha) = \frac{1}{2} \eta_i^2 \alpha^2 \) for \( i = 1, 2 \) and \( \alpha \geq 0 \), where \( \eta_1^2, \eta_2^2 > 0 \) are given parameters.

Moreover, the hospice manager may choose to live-discharge patients who are eligible. We assume for simplicity that the hospice manager learns which patients are eligible for live-discharge soon after their admission to the hospice. Let \( \theta_i(t) \) denote the rate at which the hospice manager live-discharges class \( i \) patients at time \( t \). By live-discharging a class \( i \) type \( b \) patient, the hospice manager forgoes potential revenues of \( \tilde{r}_i(t) \) and terminal value of \( \tilde{v}_i(t) \), but saves the caring cost of \( \tilde{c}_i(t) \) for \( i = 1, 2 \); these can be calculated as in Proposition 2 given the length-of-stay distributions of class \( i \) patients to be live-discharged.

The hospice manager is constrained while making the live-discharge decisions in two ways: Firstly, the number of eligible patients may be limiting. Secondly, live-discharging additional patients becomes increasingly more difficult. Our model does not keep track of the evolution of the number of patients in the hospice because doing so would require a complex history-dependent model, which is intractable analytically. This complexity stems from the “memory” of the length-of-stay distributions and prevents us from incorporating the constraint regarding the number of patients eligible for live-discharge.\(^{10}\) Instead, for tractability, we assume a convex increasing cost \( g_i(\theta) = \frac{1}{2} \eta_i^2 \theta^2 \) associated with the live-discharge rate of \( \theta \) (for \( i = 1, 2 \)) to capture the difficulties in

\(^{10}\)The Center for Medicare and Medicaid Studies (2010) states that any remaining portion of the cap the live-discharged patient contributed to the hospice can remain at the hospice unless that patient elects to transfer to another hospice immediately or later (the originating hospice would then relinquish that part of the cap to the other
the live-discharges indirectly. Furthermore, Taylor et al. (2008) reports that 15.5% of all patients are live-discharged, supporting that the number of patients eligible for live-discharges are not too few. Therefore, replacing hard constraints by a convex increasing cost for live-discharges may be a reasonable proxy.

Let $R(0)$ and $C(0)$ denote the potential revenue and the caring cost associated with patients in the hospice at the beginning of the fiscal year. Then given the hospice manager’s recruiting policy $\alpha(\cdot)$ and the live-discharge policy $\theta(\cdot)$, the cumulative (potential) revenues up to time $t \in [0, T]$, denoted by $R(t)$, is given by

$$R(t) = R(0) + \sum_{i=1}^{2} \int_{0}^{t} r_i(s)[\lambda_i + \alpha_i(s)]ds - \sum_{i=1}^{2} \int_{0}^{t} \tilde{r}_i(s)\theta_i(s)ds.$$ 

Similarly, the cumulative caring cost incurred by the hospice manager up to time $t$, is given by

$$C(t) = C(0) + \sum_{i=1}^{2} \int_{0}^{t} c_i(s)[\lambda_i + \alpha_i(s)]ds - \sum_{i=1}^{2} \int_{0}^{t} \tilde{c}_i(s)\theta_i(s)ds$$

and the cumulative terminal value associated with patients in the hospice at time $t$ is given by

$$V(t) = \sum_{i=1}^{2} \int_{0}^{t} v_i(s)[\lambda_i + \alpha_i(s)]ds - \sum_{i=1}^{2} \int_{0}^{t} \tilde{v}_i(s)\theta_i(s)ds.$$

The cumulative recruiting costs $S(t)$ and the live-discharge cost $D(t)$ up to time $t$ are given by

$$S(t) = \sum_{i=1}^{2} \int_{0}^{t} s_i(\alpha_i(x))dx \quad \text{and} \quad D(t) = \sum_{i=1}^{2} \int_{0}^{t} g_i(\theta_i(x))dx.$$  \hspace{1cm} (2)

As mentioned earlier, a crucial feature of the Medicare reimbursement policy is that the hospice’s revenue is constrained by a cap, which increases with the number of patients admitted during the fiscal year. To be specific, the cap at time $t$ is given by

$$K(t) = K \sum_{i=1}^{2} \int_{0}^{t} (\lambda_i + \alpha_i(x))dx.$$  \hspace{1cm} (3)

Therefore, the realized revenue at the end of the fiscal year is given by $\min \{K(T), R(T)\}$, and the hospice manager’s problem (P) can be written as follows: Choose search rates $\alpha_i(\cdot)$ and live-discharge rates $\theta_i(\cdot)$ for $i = 1, 2$ dynamically so as to

11 We assume that those patients who are in the hospice at the beginning of the fiscal year are not present at the end due to death or being live-discharged. Therefore, those patients do not contribute to the terminal value $V(T)$.
maximize \( \min(K(T), R(T)) + V(T) - C(T) - S(T) - D(T) \) 

subject to \( \alpha_i(t) \geq 0, \theta_i(t) \geq 0 \) for all \( i, t \).

Although the hospice manager’s problem (P) is an optimal control problem (and hence, infinite dimensional), its dual is much simpler. Indeed, in Appendix B, we show that the dual formulation can be reduced to a one-dimensional convex optimization problem, enabling an explicit solution to both the dual and the hospice manager’s original problem. The proof of this relies on the duality theory for optimal control problems developed by Rockafellar (1970), which is also introduced in Appendix B. For \( q \in [0, 1] \), define

\[
F(q) = -K(\lambda_1 + \lambda_2)T + R(0) + \sum_{i=1}^{2} \lambda_i \int_0^T r_i(s)ds + \sum_{i=1}^{2} \int_0^T \frac{-\tilde{r}_i(t)}{\eta_i} [-\tilde{r}_i(t)q + \tilde{c}_i(t) - \tilde{v}_i(t)]^+ dt \\
+ \sum_{i=1}^{2} \int_0^T \frac{r_i(t) - K}{\eta_i} [K + (r_i(t) - K)q - c_i(t) + v_i(t)]^+ dt.
\]

(4)

In what follows, we assume that the following holds:

\[
\min \{r_i(t), K\} + v_i(t) - c_i(t) > 0 \text{ for some } i, t \in (0, T),
\]

(5)

which ensures the strict monotonicity of \( F \) and the uniqueness of the dual optimal solution.

In Appendix B, \( F(\cdot) \) will be shown to be the derivative of the dual objective function. The following proposition shows that the inverse \( F^{-1} \) of \( F \) is well defined.

**Proposition 3** \( F \) is continuously differentiable and strictly increasing.

We are now ready to state our main result.

**Theorem 1** If \( F(0) > 0 \) then let \( q^* = 0 \), if \( F(1) < 0 \) then let \( q^* = 1 \), otherwise let \( q^* = F^{-1}(0) \).

Then the hospice manager’s optimal recruiting and live-discharge rates for \( i = 1, 2 \) and \( t \in [0, T] \) are given by

\[
\alpha_i^*(t) = \frac{[K(1-q^*) + q^*r_i(t) - c_i(t) + v_i(t)]^+}{\eta_i}, \quad \text{and}
\]

\[
\theta_i^*(t) = \frac{[-\tilde{r}_i(t)q^* + \tilde{c}_i(t) - \tilde{v}_i(t)]^+}{\eta_i}.
\]

We have thus explicitly characterized the optimal recruiting and live-discharge rates for a given hospice. For either type of patient, the hospice may choose to never actively recruit those patients, relying entirely on their natural arrival rates, they may recruit patients throughout the year, only at the beginning of the year, or they may only recruit towards the end of the Medicare year. Similar patterns apply to live-discharging. We now exercise these findings numerically.
4.2 Numerical Study of the Fluid Model

In the numerical experiments of this subsection the lifespans of patients will be modeled as gamma distributions. A randomized sample of the 1993 cohort of Medicare hospice beneficiaries (184,843 data points) was obtained, including the number of days between the date of admission and the date of death for 27 disease categories. The histograms describing the number of days survived since admission for each disease were monotone decreasing (in time ‘buckets’ of one day), each characterized by a very high frequencies at low numbers of days and decreasing into what could be described as “light tails.” Each individual disease’s histogram was curve-fitted and we also grouped diseases into type 1 (all cancer diseases and end-stage renal failure) and type 2 (all noncancer diseases other than renal failure) and fitted those histograms. There are a number of parametric distributions which tended to fit these histograms quite well and the gamma distributions tended to be consistent good performers for many diseases and across various measures of fit. The gamma distribution has the added benefit that the mixing of gamma distributions (for a common scale parameter) will also be a gamma distribution, which serves the purpose of separating those distributions for subtypes $a$ and $b$, resulting in $\Gamma^a_1(0.1567, 455.14)$, $\Gamma^a_2(0.8777, 455.14)$, $\Gamma^b_1(0.2198, 550.61)$, and $\Gamma^b_2(1.2308, 550.61)$ (these are the shape and scale parameters, respectively).

For the purposes of our examples, we doubled the scale parameter (to 1101.22) for the type 2 distributions to represent those hospices experiencing longer distributions for type 2 patients, where binding cap scenarios are more likely. There are a couple of reasons for why the fitted distributions may not result in binding cap scenarios. Firstly, these data are for a randomized sample of patients drawn from the entire country and will not reflect a “binding cap” scenario for a single hospice (the cap is designed not to bind for an average hospice). Secondly, these data are from the 1993 cohort, which was a time before the upper tails of the type 2 distributions began elongating, so such distributions are unlikely to contribute to a binding cap.

Taylor et al. (2008) contains data where the proportion of patients who enter hospice and leave prior to death is 11.3%.\footnote{Taylor et al. (2008) report that among the 1218 patients under study, 1029 died under continuous care, 131 were live-discharged and not readmitted, and 58 were live-discharged and then readmitted and died under hospice care (a few patients in this third group were discharged a second time but then died outside of hospice care). We excluded this third category since we wished to merely examine those patients who were live discharged a single time and then died. Thus, our proportion of potentially live-dischargable patients is $131/(131+1029)=11.3\%$.} In the absence of any other data, we use this data to set $\gamma^b_1 = 1 - \gamma^a_1 = \gamma^b_2 = 1 - \gamma^a_2 = 0.113$, reflecting the proportion of incorrectly diagnosed (or not terminally
ill) patients at admission. The Medicare parameters for the 2007 Medicare year are $K = \$21,410$ and $r = \$130.79$. The natural daily arrival rates of patient are chosen as $\lambda_1 = 1$ patient per day and $\lambda_2 = 9$ patients per day, intended to reflect the proportions of a hospice experiencing cap issues.\footnote{Note that total arrival rates of 10 patients per day would reflect a relatively large hospice (Wright and Katz, 2007) but the simulation model that will be presented shortly required arrival rates of this magnitude to avoid issues with integrality. However, the fluid model and intuition given below is robust to the choice of $\lambda$.}

We have taken the daily cost of caring as $c_1 = \$60$ per day and $c_2 = \$45$ per day, from Killaly et al. (2007). The recruiting and live-discharging parameters are $\theta_{ij}^l = 15,000$ for $i = 1, 2$ and $j = l, s$; varying these parameters results in greater or lesser recruiting and live-discharging rates but no additional insight. We set the terminal values as $v_j^l = (\psi r - c_i)m^l_j$ where the exogenous parameter $\psi$ is exercised between 0 and 1 below, as a robustness check. Note that, for all by type 2b, the average life of patients living into the next year is actually less than $m^l_j$ because the given gamma distribution has increasing failure rate so $-c_im^l_j$ is a lower bound on the terminal value and $(r - c_i)m^l_j$ is an upper bound on the terminal value. For ease of calculation, we use $\tilde{r}_i(t) = r^b_i(t)$, $\tilde{c}_i(t) = c^b_i(t)$, and $\tilde{v}_i(t) = v^b_i(t)$.

In the following examples, we examine the optimal recruiting and live-discharging policies of the hospice manager for a variety of $R(0)$ and $\psi$ values. Varying $\psi$ is necessary as it does not seem possible to determine precisely the ongoing value of those patients alive at the end of the Medicare year; we have however determined a base-case value of $\psi = 0.15$ by using a steady-state simulation as described in the following section. $R(0)$ is also varied to reflect that a hospice will commence a new year with living patients who reflect a stream of revenues in the new year. The hospice will adjust their optimal recruiting and live-discharging for varying values of $R(0)$ and $\psi$. Of course, $R(0)$, $C(0)$, and $\psi$ will be related as they reflect the amount of revenue and costs patients enrolled at the start of the year accrue, and the value of a patient at the end of the year, respectively. Finding $\psi = 0.15$ is our attempt to relate $R(0)$ and $\psi$, but this approach is imperfect so we exercise our model numerically across these parameters.

In Figure 1 we see the effect of varying $R(0)$ for $\psi = 0.15$. At lower values of $R(0)$, the cap is not binding and we observe the hospice will recruit type 2 patients at a higher rate than type 1 patients (the daily margin of patient type 1 is $130.79-60 = $70.79 and of patient type 2 is $130.79-13$...
Figure 1: The optimal recruiting and live-discharge rates for various values of $R(0)$ for $\psi = 0.15$.
(Key: $\alpha_1(\cdot)$ – solid line, $\alpha_2(\cdot)$ – dashed line, $\theta_1(\cdot)$ – dotted line, $\theta_2(\cdot)$ – dotted-dashed line)

Figure 2: The optimal recruiting and live-discharge rates for various values of $\psi$ for $R(0) = 55,000,000$. (Key: $\alpha_1(\cdot)$ – solid line, $\alpha_2(\cdot)$ – dashed line, $\theta_1(\cdot)$ – dotted line, $\theta_2(\cdot)$ – dotted-dashed line).
$45=\$85.79), although both patient types will be sought, which is unsurprising as each patient type is profitable. As the cap begins to bind (Figure 1(b)), we see that the hospice extends the duration of their recruiting and the recruiting of type 1 patients is sustained longer at a higher level and begins to dominate that of the type 2 patients during the latter parts of the year. This is due to the fact that the patients at the end of the year are costly because $\psi = 0.15$ ($v_i < 0$ for $i = 1, 2$) and type 1 patients can bump up the cap as much as type 2 patients but live shorter durations and thus will be less costly if they happen to remain living at the end of the year. (Clearly, many of them do not survive until then unless they are recruited very close to the end of the fiscal year.) This effect continues as $R(0)$ increases, as we see in Figure 1(c). It should be noted, however, that the fact that all patients who live into the following year are costly (regardless of their number) is an artifact of our specific model and choice of $\psi = 0.15$.

We also observe the hospice will live-discharge patients at a greater rate towards the end of the year to reduce the chance of possessing these costly patients at the end of the year. Moreover, by live-discharging patients the hospice manager also avoids the caring costs for those patients during the current fiscal year. Because the cost of live-discharging is assumed to be convex increasing, the hospice will conduct this over a period at the end of the year rather than all at once on the final day, for example. As $R(0)$ continues to increase, these trends will continue: live-discharging will continue over a longer period with type 2 discharges dominating type 1 later in the year (because $v_2 < v_1 < 0$) and discharge rates will increase. This is a means of reducing these starting patients who contribute nothing positive towards the cap but would consume it if they remained under care. As one would expect, for much lower values of $R(0)$ when the cap is not binding, little live-discharging occurs. This is simply because the hospice has little need to reduce its patient census and can accommodate the revenues from those patients under the cap (to which they do not contribute at all). For $\psi = 0.15$ and $R(0) = 55,000,000$ we find that 9.1% of all patients are live discharged, which is comparable with the numbers found by Kutner et al. (2004) and other references. Clearly $R(0)$, which reflects the potential revenues of patients left over from the previous year, has a considerable effect upon the optimal actions of the hospice. Compensating actions such as additional recruiting (to increase the cap) or additional live-discharging (to reduce the revenues) are clearly necessary as $R(0)$ increases.

In Figure 2 we observe the effect of increasing $\psi$ for a fixed value of $R(0)$.$^{15}$ When $\psi = 0,$

\[\begin{align*}
\theta_2(t) & \text{ in } 17
\end{align*}\]

\[\begin{align*}
\text{The terminal parameters are as follows:} & \text{ (a) } v_1 = -6503.3, v_2 = -16548.9, \text{ (b) } v_1 = -4376.9, v_2 = -9334.2, \text{ (c) } v_1 = -2250.5, v_2 = -2119.4, \text{ (d) } v_1 = -124, v_2 = 5095.4, \text{ and (e) } v_1 = 2002.4, v_2 = 12310.2. \text{ The line for } \theta_2(t) \text{ in}
\end{align*}\]
\[ \sum_{i=1}^{2} \int_{0}^{t} (K - r_i(x)) \lambda_i dx - R(0). \]

\[ K(t) - R(t). \]

Figure 3: The gap between the cap and potential revenues \((R(0) = \$55M, \psi = 0.15)\).

\(v_1 = -6503.28\) and \(v_2 = -16548.9\) so the recruiting of type 1 is higher than that of type 2 and both will decrease towards the end of the year. Likewise, the terminal costs are so significant that sizable live-discharging will take place, with more type 2 discharged than type 1. As \(\psi\) increases, the gap between \(v_1\) and \(v_2\) narrows. As this happens, the rate of type 2 recruiting will increase, eventually dominating the rate of type 1 recruiting, both eventually increasing at the end of the year. Likewise, as \(\psi\) increases and the terminal values become more positive (that is, there is no negative effect of carrying patients into the following year), the need for live-discharging reduces significantly and ultimately disappears altogether.

Ultimately, the manager’s objective is to ensure the hospice is not exposed to a large deficit in the difference between the cap and revenues received. This is verified by examining Figure 3, which is for \(R(0) = \$55M\) and \(\psi = 0.15\). In Figure 3(a) the difference between the cap and the revenues for the natural arrivals (i.e., absent any recruiting) is displayed, demonstrating a shortfall of \$13.8 million at time \(T\). Once the hospice’s recruiting and live-discharging is also included, the gap between the cap and the revenues, displayed in Figure 3(b), will be closed at time \(T\). The last result we wish to highlight is that when we increase the coefficient of variation of the LOS distribution, the recruiting and live-discharge rates for both disease types increases.

Thus far in this section, we have assumed a deterministic model. A natural question to ask is what are the effects of random arrivals and LOS on the intuition provided. The following section provides a simulation study to address this question.
4.3 A Simulation Study

To test the effects of stochasticity on the system we created a discrete event simulation in C++. While using this simulation to find optimal recruiting and discharge policies is not tractable, we can use it to test some reasonable hypotheses for the system. In particular, we hypothesize that uncertainty should amplify the effects of the cap due to random shocks, resulting in more type 1 patients being recruited, especially towards the end of the horizon, and perhaps decreasing type 2 recruiting.

The simulation was created to mimic the fluid model except with the addition of stochastic arrivals and LOS. In particular, multiple replications of a single year (i.e., terminating) model were performed, where the terminal values were set equal to those of the numerical study of the fluid model and an initial number of customers were introduced at the start of each year to generate $R(0)$ and $C(0)$. However, because terminal values and initial customers are actually endogenous to the system, in addition to this “terminating simulation,” we also created a steady-state version. Admissions were assumed to take place at the start of each day, with a random number of patients arriving at a rate equal to the sum of the natural arrivals and the recruiting rate. In order to test the effects of variability, the number of arrivals each day was sampled from a discretized gamma distribution with coefficient of variation set to 1 for the base case (in which case the number of arrivals each day follows a discretized exponential distribution). LOS distributions were also taken from a gamma distribution with parameters as given previously for the fluid model.

In order to test the above hypotheses, a heuristic was created based on state-dependent control (recall that finding the optimal control in the stochastic setting is not tractable). At each day, the hospice manager wishes to decide recruiting and discharge rates in order to maximize expected revenue minus costs over what is left of the horizon. As in the fluid model, this is implemented by solving an optimization problem on the revenue, cost, and terminal value functions, where the decision variables are the recruiting and discharge rates for the remaining days in the year. Only the calculated recruiting and discharge rates for the current day are used and those for the rest of the year are discarded (and recalculated the following day). Further details on the heuristic may be found in Appendix C. We ran the heuristic on all parameter choices in Figures 1 and 2 and the average improvement over the fluid control was 6.29%, illustrating that there is typically a benefit to state-dependent control in the presence of stochasticity. As mentioned previously, we also created a steady-state simulation using the heuristic that imputes terminal values.\textsuperscript{16}

\textsuperscript{16}We ran the steady-state simulation with the heuristic policy, under the base-case parameters from the fluid
In Figure 4 we observe the difference between the terminating-simulation heuristic recruiting and the recruiting in the fluid policy and the difference between the terminating-simulation heuristic discharging and the discharging in the fluid policy for $\psi = 0.15$ and 2968 initial patients split appropriately between the different classes, which corresponds to an $R(0)$ of approximately $55\text{ million}$.

Notice that the rates shown in Figure 4 are averages over all replications. In general, the confidence intervals for these rates increased in width throughout the year; this is intuitive because the effects of stochasticity would tend to accumulate throughout the year and in some years no recruiting would be desirable by the end and in others much recruiting may be needed.

Figure 4(a) shows that in the simulation we see more recruiting of both types of patients towards the end of the horizon. This supports our hypothesis that in a stochastic system cancer patients may be more heavily sought, particularly at the end of the year when revenues may have exceeded the cap. However, there is also more recruiting of non-cancer patients towards the end of the year, contradicting our initial hypothesis that there should be less recruiting of these patients. On the numerics section, in order to create a base-case estimate for both the terminal values and the number of customers present at the start of the year. At the end of each year, the estimates for the terminal values are updated using an algorithm described in Appendix C. Note that terminal values are used within the heuristic optimization in order to determine the appropriate levels of recruiting and/or discharging and therefore an iterative approach is taken to their computation.

We also produced these graphs for 2171 initial patients, which are the values obtained from the steady-state simulation and correspond to an $R(0)$ of approximately $40.2\text{ million}$, but found in that case the heuristic did almost no discharging and the figure was less informative.
other hand, the difference between the heuristic and fluid recruiting of type 1 patients is greater
than that of type 2 patients at the end of the year, again supporting our intuition.

In Figure 4(b), we see that live-discharging under the heuristic is at first less and then more than
the fluid model, with an overall average level that is lower. One possible reason is that the stochastic
model may delay discharging until uncertainty has been further resolved. As noted above, with
fewer patients at the start of the year ($R(0) =$40.2 million) the heuristic does far less discharging
than the fluid model, although the decreased discharging of patients was paired with increased
recruiting by the heuristic. Given the possibility of “favorable” sample paths under stochasticity,
where no discharging is needed, this appears to be consistent with the delayed discharging decisions
shown in Figure 4(b).

In order to further test the effects of stochasticity we ran the terminating simulation with
$\psi = 0.15$ under a variety of values for the squared coefficient of variability of both the number of
arrivals per day and also the LOS distributions. Specifically, we scaled all squared coefficients of
variability up by ten times and down by ten times while keeping the means constant. We found
that the improvement over the fluid control policy increased with increased variability, verifying our
intuition that a state-dependent policy, such as the heuristic, will have more value as the system
gets more variable. Further, there was a clear trend of profit decreasing and both recruiting and
discharging levels of both types of patients increasing with increasing variability of arrivals. When
the squared coefficients of variability of the LOS distributions were altered both recruiting and
discharge levels increased significantly under increased variability; for this case, similar trends can
also be generated using the fluid model as it allows for two-parameter LOS distributions.

We believe that nonstationary recruiting and live-discharge rates are undesirable from a policy
perspective because they do not align with the equal access objective of a social planner. The
immediate implication of nonstationary recruiting is that a prospective patient’s desirability to a
hospice depends on the calendar. A cancer patient would not be as highly sought in December
than they might be in October (the Medicare year runs from November 1 to October 31). In a
similar vein, nonstationary live-discharging is highly undesirable. The government’s motivation in
establishing the Medicare hospice benefit was to facilitate comfortable palliative care for patients
during their end-of-life by third party providers. The nonstationary active recruiting and live-
discharges which can arise as the optimal policy and is observed in practice, are clearly unintended
consequences of the government’s Medicare reimbursement policy. The following section presents
a new policy for Medicare to consider, which is intended to alleviate this unintended behavior.
5 The Legacy Policy and its Analysis

This section introduces and analyzes a policy to overcome the nonstationary behavior observed under the current Medicare reimbursement policy seen in Section 4. We label this policy the legacy policy as it explicitly addresses the possibility of beneficiaries living into the next Medicare year. The implementation of this policy requires the hospice to segregate the tracking of the patients admitted in each Medicare year until they expire, but is otherwise no more burdensome in terms of administration than the current Medicare policy. Indeed, our goal is a policy that maintains the same fundamental framework of the Medicare policy but does not have an inherent incentive for non-stationary recruiting and discharge. That is, we assume that Medicare wishes to maintain a cap (to limit its exposure, especially as the declaration of a patient being terminally ill can be difficult), to keep the cap pooled (to mitigate risk to the hospice), and to keep payment rates constant across disease types and time frame (for ease of implementation).

The policy consists of allowing the hospice to continue receiving revenues for all the patients living at the end of the year until any remaining cap is exhausted or all these patients expire, whichever occurs first. If the cap is exhausted before these patients expire, then the hospice receives no further revenues for these patients but continues to incur the costs of caring for them. So the key difference between the legacy policy and the current Medicare policy is that once a new Medicare year arrives, current patients (those receiving hospice services currently) will count against the previously accumulated cap (if it is not exhausted) as opposed to the current practice of counting them against the cap accumulated in the new year. In practice, this policy is likely most simply implemented by allowing exactly one extra legacy year and assuming that the probability of having patients live into a third Medicare year with the cap still not exhausted is negligible.\(^{18}\)

Under the modified policy, because the accounting is over the entire patient life rather than just over the Medicare year, \(r_i(t) = rm_i, c_i(t) = c_i m_i\) and \(\tilde{r}_i(t) = r \tilde{m}_i, \tilde{c}_i(t) = c_i \tilde{m}_i\) for all \(i, t\), where \(\tilde{m}_i\) denotes the mean of the residual lifespan of patients who are live-discharged. Similarly, \(R(0) = C(0) = 0, v_i = \tilde{v}_i = 0,\) and \(v_i(t) = \tilde{v}_i(t) = 0\) for all \(i, t\). Then the fluid model of the hospice manager’s problem can be adapted to choosing recruiting rates \(\dot{\zeta}(\cdot)\) and live-discharge rates \(\dot{\Theta}(\cdot)\) in order to

\[
\max_{\dot{\zeta}} \min_{\dot{\Theta}} \left\{ K \sum_{i=1}^{2} (\lambda_i T + z_i(T)), r \sum_{i=1}^{2} (m_i(\lambda_i T + z_i(T)) - \tilde{m}_i \Theta_i(T)) \right\}
\]

\(^{18}\)Notice that this policy would mean that a start-up hospice would have no cap adjustment to pay-back until the end of its second year of operation.
Lemma 1

subject to

\[ z_i(t) = \int_0^t \dot{z}_i(s) \, ds \quad \text{with} \quad \dot{z}_i(t) \geq 0 \quad \text{for all } i, t, \]  
\[ \Theta_i(t) = \int_0^t \dot{\Theta}_i(s) \, ds \quad \text{with} \quad \dot{\Theta}_i(t) \geq 0 \quad \text{for all } i, t. \]  

The following lemma will be useful in proving our main result.

Proposition 4

Any optimal solution of the formulation (6)-(8) is a stationary recruiting policy, i.e., \( \alpha_i(t) = \dot{z}_i(t) = \ddot{\alpha}_i \) and \( \theta_i(t) = \dot{\Theta}_i(t) = \ddot{\theta}_i \) where \( \ddot{\alpha}_i \geq 0, \ddot{\theta}_i \geq 0 \) for all \( i, t. \)

Therefore, the hospice’s manager’s problem can be stated without loss of generality (setting \( T = 1 \)) as follows: Choose the stationary recruiting rates \( \alpha_1, \alpha_2 \) and the stationary live-discharge rates \( \theta_1, \theta_2 \) so as to

\[
\max \left\{ \min \left[ K \sum_{i=1}^{2} (\lambda_i + \alpha_i), r \sum_{i=1}^{2} (\lambda_i + \alpha_i)m_i - \tilde{m}_i\theta_i \right] - \sum_{i=1}^{2} c_i[m_i(\lambda_i + \alpha_i) - \tilde{m}_i\theta_i] \right. \\
- \sum_{i=1}^{2} s_i(\alpha_i) - \sum_{i=1}^{2} g_i(\theta_i) \right\} \tag{9}
\]

subject to \( \alpha_i \geq 0, \theta_i \geq 0, i = 1, 2. \)

To facilitate the statement of this section’s main result, define

\[
\pi_1 = \frac{\sum_{i=1}^{2} \frac{(rm_i - K)^2}{q_i^2} + \sum_{i=1}^{2} \frac{r(\tilde{m}_i)^2c_i}{q_i^2} - \sum_{i=1}^{2} (rm_i - K)[\lambda_i + \frac{(r-c_i)m_i}{\eta_i}]}{\sum_{i=1}^{2} \frac{(rm_i - K)^2}{q_i^2} + \sum_{i=1}^{2} \frac{(\tilde{m}_i)^2}{q_i^2}},
\]

\[
\pi_2 = \frac{\sum_{i=1}^{2} \frac{(rm_i - K)^2}{q_i^2} + \frac{r(\tilde{m}_i)^2c_i}{q_i^2} - \sum_{i=1}^{2} (rm_i - K)[\lambda_i + \frac{(r-c_i)m_i}{\eta_i}]}{\sum_{i=1}^{2} \frac{(rm_i - K)^2}{q_i^2} + \frac{(\tilde{m}_i)^2}{q_i^2}},
\]

\[
\pi_3 = \frac{\sum_{i=1}^{2} \frac{(rm_i - K)^2}{q_i^2} - \sum_{i=1}^{2} (rm_i - K)[\lambda_i + \frac{(r-c_i)m_i}{\eta_i}]}{\sum_{i=1}^{2} \frac{(rm_i - K)^2}{q_i^2}}.
\]

The following lemma will be useful in proving our main result.

Lemma 1

Assume \( c_1 > c_2 \). Then the following hold:

(i) \( \pi_1 < c_2/r \) if and only if \( \pi_2 < c_2/r \);

(ii) if \( \pi_1 < c_2/r \), then \( \pi_3 < c_1/r \);

(iii) \( \pi_2 < c_1/r \) if and only if \( \pi_3 < c_1/r \).
We are now ready to state the main result of this section.

**Proposition 5** Let \( \alpha_i^*, \theta_i^* \) for \( i = 1, 2 \) be the optimal solution to (9). Then

1. If \( \sum_{i=1}^{2}(rm_i - K)\left[\frac{(r-c_i)m_i}{\eta_i} + \lambda_i\right] < 0 \), then
   \[ \alpha_i^* = \frac{m_i}{\eta_i}(r-c_i) \] and \( \theta_i^* = 0 \) for \( i = 1, 2 \).

2. If \( \sum_{i=1}^{2}(rm_i - K)\left[\frac{(r-c_i)m_i}{\eta_i} + \lambda_i\right] > \sum_{i=1}^{2}\left(\frac{rm_i - K}{\eta_i}\right)^2 + \sum_{i=1}^{2}\frac{r(m_i)^2c_i}{\eta_i} \), then
   \[ \alpha_i^* = \frac{K-c_im_i}{\eta_i} \] and \( \theta_i^* = \frac{c_im_i}{\eta_i} \) for \( i = 1, 2 \).

3. If \( 0 \leq \sum_{i=1}^{2}(rm_i - K)\left[\frac{(r-c_i)m_i}{\eta_i} + \lambda_i\right] \leq \sum_{i=1}^{2}\left(\frac{K-rm_i}{\eta_i}\right)^2 + \sum_{i=1}^{2}\frac{r(m_i)^2c_i}{\eta_i} \), then we have the following cases:
   
   (a) If \( \pi_1 < c_2/r \), then
   \[ \alpha_i^* = \frac{K(1-\pi_1) + rm_i\pi_1 - c_im_i}{\eta_i} \] and \( \theta_i^* = \frac{m_i(c_i - r\pi_1)}{\eta_i} \) for \( i = 1, 2 \).

   (b) If \( \pi_1 \geq c_2/r \), then we have the following two sub-cases:
   
   i. If \( \pi_2 < c_1/r \), then \( \theta_1^* = m_1(c_1 - r\pi_2)/\eta_1, \theta_2^* = 0 \) and
   \[ \alpha_i^* = \frac{K(1-\pi_2) + rm_i\pi_2 - c_im_i}{\eta_i} \] for \( i = 1, 2 \).

   ii. Otherwise, i.e., \( \pi_2 \geq c_1/r \), then
   \[ \alpha_i^* = \frac{K(1-\pi_3) + rm_i\pi_3 - c_im_i}{\eta_i} \] and \( \theta_i^* = 0 \) for \( i = 1, 2 \).

Depending on the parameters, Proposition 5 has three cases for the solutions of the optimal recruiting rate under the legacy policy.

- In case 1, the reimbursements determine the revenue rate (i.e., the cap does not bind).
- In case 2, the cap binds in the optimal solution (i.e., the cap determines the revenue rate).
- In case 3, the cap equals the reimbursement rate (perfectly balanced).
In case 1 of Proposition 5 the potential revenues of the hospice will not use the available cap and the recommended recruiting rate will then simply solve the resulting first order condition of (9). This results in rates which simply balance out the contribution over the remaining life of the patient against the cost of recruiting them; and there is no need to live-discharge anyone. These rates are increasing in $r$ and $m_i$, and decreasing in $\eta_i$ and $c_i$. Note that if the cap also does not bind under the original policy then we would expect to see similar levels of recruiting (and zero discharges). Thus, the legacy policy is designed for hospices where the cap is of relevance to the revenues and will make little difference to hospices that typically do not meet their caps.

In case 2 the cap is binding and the resulting optimal recruiting rate simply takes the cap as the total revenue, subtracts the total costs of caring, and balances this quantity against the cost of recruiting these patients. The rate of live-discharging balances the remaining lifetime costs of those patients against the cost of discharging them. Thus, the recruiting rates are increasing in $K$ but decreasing in $\eta_i$, $m_i$, and $c_i$. Note that depending on the cost structure, the live-discharge rate can be quite high. In case 3 the recruiting and live-discharging rates of cases 1 and 2 are blended in such a way to create a perfectly balanced situation where the weighted total revenues equals the available cap.

An observation is that although Proposition 5 states that the legacy policy will deliver a stationary policy, it does not imply the recruiting rates will be identical across diseases or even equal in proportion to volumes. Indeed, Proposition 5 confirms the recruiting rates will differ, according to the underlying characteristics and economics for each disease. So, while the legacy policy addresses the problem of nonstationarity of the existing reimbursement policy, it does not address the issue of the relative profitability across diseases (accounting for the cost of care, the cost of recruiting, and the lifespans of patients) which causes different rates of recruiting and live-discharging between diseases. This is consistent with the existing policy and could only be addressed by Medicare instigating a disease-specific reimbursement policy. Indeed, Killaly et al. (2007) recommends revisiting Medicare’s disease-invariant per diem reimbursement, but we believe a disease-dependent payment rate would need to be very carefully chosen; otherwise the system will be prone to gaming such as hospices classifying admissions based on the higher margin disease in the case of co-morbid patients. Thus, more careful consideration is merited.

Table 1 presents the legacy policy’s optimal recruiting ($\alpha_i^*$) and live-discharge ($\theta_i^*$) rates for $\psi = 0.15$ and $R(0) = $55M and various recruiting ($\eta_i^*$) and live-discharge ($\eta_i^*$) costs (for $i = 1, 2$). For comparison purposes, the time-averages of the fluid model’s optimal recruiting and live-discharge
\[
\begin{aligned}
\eta_i^j & \quad \alpha_1^i & \quad \alpha_2^i & \quad \theta_1^i & \quad \theta_2^i & \quad \frac{1}{T} \int_0^T \alpha_1(t) \, dt & \quad \frac{1}{T} \int_0^T \alpha_2(t) \, dt & \quad \frac{1}{T} \int_0^T \theta_1(t) \, dt & \quad \frac{1}{T} \int_0^T \theta_2(t) \, dt \\
5000 & 2.530 & 2.635 & 1.539 & 1.154 & 0.4131 & 0.5563 & 0.7115 & 3.306 \\
10000 & 1.289 & 1.228 & 0.9436 & 1.168 & 0.2917 & 0.3157 & 0.3926 & 1.808 \\
15000 & 0.8757 & 0.7596 & 0.7452 & 1.173 & 0.3649 & 0.2934 & 0.3544 & 1.463 \\
20000 & 0.6689 & 0.5252 & 0.6461 & 1.176 & 0.3813 & 0.2906 & 0.3724 & 1.256 \\
\end{aligned}
\]

Table 1: The legacy policy optimal stationary recruiting \((\alpha_1^i)\) and live-discharge \((\theta_1^i)\) rates and time-average of the optimal recruiting and live-discharge rates for various levels of the recruiting and live-discharge costs, \(\eta_i^j\) for \(i = 1, 2\) and \(j = s, l\) \((\psi = 0.15, R(0) = 55,000,000)\).

There are several observations. Firstly, as might be expected, as the costs of recruiting increase the legacy policy recruiting rates decrease, as it is more expensive to seek patients. However, as the costs of live-discharging increase, the live-discharging of cancer patients falls while the live-discharging of noncancer patients rises. It is important to note that all these costs are increasing concurrently, suggesting there is a degree of substitution between the live-discharging. Moreover, for low values of \(\eta_i^s\), the legacy recruiting rates start as \(\alpha_1^i < \alpha_2^i\) but as \(\eta_i^s\) increase the fluid model migrates to a cap constrained scenario and we likewise see the legacy policy migrates to \(\alpha_1^i > \alpha_2^i\) (in Table 1’s example, we see the gap between the fluid model’s average recruiting rates is initially large and then narrows). The reverse happens with the legacy policy’s live-discharge rates: at low values of \(\eta_i^s\), \(\theta_1^i > \theta_2^i\) but this reversed to \(\theta_1^i < \theta_2^i\) at higher levels. This suggests another substitution effect is going on: the legacy policy is substituting recruiting for live-discharging of cancer patients as the costs increase but the reverse (live-discharging for recruiting) for noncancer patients.

We also notice that the legacy policy recommends higher recruiting and type 1 live-discharge rates and lower type 2 live-discharge rates than the average numbers from the fluid model for these negative terminal values. The first suggestion is that while the legacy policy overcomes the undesirable nonstationarity of the fluid model, it appears to encourage other behaviors which Medicare may not find desirable, although this insight is tempered by the knowledge that the fluid model averages may be somewhat depressed by their negative terminal values, quantities that the legacy policy is not subject to. Both recruiting and live-discharging are actions which Medicare finds concerning.

It is notable that the revenues accrued under the legacy policy are greater than those of the fluid
policy and hence the cost to Medicare has increased under this policy. For example, for $\psi = 0.15$, $R(0) = -55M$, and $\eta_i = 15000$ the legacy policy accrues revenues of $90,926,088$, a 9.2% increase over the fluid policy’s revenues of $83,290,870$; again, this must be tempered by the observation that the comparison is not completely analogous due to the terminal values of the fluid model. Obtaining an accurate estimate of both the revenue effects and the effects on average recruiting and discharge rates of implementing the legacy policy would likely be best done through a pilot study. Reimbursement rates could then also be adjusted by Medicare to make the policy revenue neutral.

Although the legacy policy eliminates the first order nonstationary recruiting/discharge behavior (as captured by the fluid model), there may still be a second order nonstationary behavior due to the uncertainty in patient arrivals and LOS realizations. More specifically, the hospice manager may still have an incentive to recruit cancer patients towards the end of the Medicare year if she is cap constrained due to inherent uncertainty in the environment. In order to test the magnitude of such effects we again used simulation with a heuristic, analogous to that in Section 4.3, designed to maximize expected revenue minus costs under the proposed new accounting system (see Appendix C for details).

Figure 5 shows the various possibilities for recruiting of type 1 patients under the legacy and original heuristics. It can be seen that the recruiting levels under the legacy heuristic are indeed remarkably stationary until the last month or so of the year, while the recruiting levels under the
original heuristic vary widely throughout the year. We found similar results for recruitment of type 2 and discharges of type 1 patients, although discharges of type 2 were only slightly more stationary under the legacy policy, possibly due to its decreasing failure rate distribution being approximated by its mean in our heuristic. This can be seen in Table 2, which shows the time-average squared coefficient of variation of the recruiting and discharge rates across the simulated sample paths. It appears likely that the counteraction to stochastic variability at the end of year is less concerning to Medicare than a calculated non-stationarity resulting from the current policy’s incentives.

While more complicated policy structures could be designed, there is benefit derived from the relative simplicity of the existing and legacy policies from an implementation perspective. Even for these simple policies, an opportunity has grown for Regional Home Health & Hospice Intermediary (RHHI) firms to act as Medicare billing agencies on behalf of the hospices. If the policies became more complicated, then the likelihood is that such intermediaries will consume even more channel profits. A benefit of the legacy policy is that it strongly resembles the existing policy and is unlikely to involve much more difficulty in implementation (the difference is that the hospice will need to segregate any revenues received on behalf of those beneficiaries living into the following year from those newly admitted in that following year).

6 Concluding Remarks and Discussion

We have examined Medicare’s hospice reimbursement policy, both from profitability and patient recruitment/live-discharge perspectives. Our primary goal has been to inform policy makers and stir debate. From the profitability perspective, we are able to discern the aspects of the policy or market conditions which lead to potential losses. Specifically, we find that if a provider has a lack of scale and/or an imbalance in the mix of patients, they run a risk of not receiving sufficient revenues to reach profitability. This could be because the patient census lived too long and the cap limited the revenues gained from Medicare while the provider continued to incur the costs of caring for the
patients, or because the patient census lived too short for the provider to gain sufficient revenues to cover the fixed overhead of operation. We suggest a potential remedy is for the government to encourage the merging of appropriate providers, or at a minimum remove regulatory hurdles that deter such mergers.

Another aspect of the current Medicare hospice reimbursement policy we investigated was the manager’s optimal recruitment and live-discharge policies. Using a dynamic fluid model, we demonstrate that the manager has an incentive to recruit patients of differing diseases at rates which differ across disease types and change during the Medicare year. For example, the manager might seek to recruit type 1 (short LOS) patients at a rate which dominates the recruitment rate for type 2 (long LOS) patients towards the end of the year. The basic reason for this is that type 1 patients can increase the cap by the same amount as type 2 patients but are expected to live for shorter durations, which is an important consideration when patients living into the next Medicare year are taken into consideration. We also show that the manager has differing incentives for live-discharging patients whose conditions have stabilized throughout the year. These are clearly an unintended and disturbing consequences of the current policy and there is strong anecdotal evidence to suggest that such behavior indeed occurs in practice (e.g., Jenkins et al., 2010).

Medicare (and Medicaid) funds a variety of programs aside from hospices. A natural instinct of the policy maker when constructing a set of rules that distributes taxpayer money is to protect the reserves from people trying to take advantage of the system; this is the motivation for the cap. Other public programs have similar caps to limit the government’s exposure. While the specifics of each program may differ somewhat, we believe our lessons on the unintended consequences of such caps are likely to be transferable and our models can act as prototypes for anyone wishing to model the specifics of those programs.

We design and analyze an alternative policy, called the legacy policy, which allows the hospice to continue receiving revenues for these remnant patients providing any positive remnant cap exists at the end of the year, until the last patient expires or the remnant cap is exhausted, whichever occurs first. Importantly, the remnant cap and the remnant patients are tracked separately from both the new cap launched at the start of the new year and any newly admitted patients. We show this alternative policy restores stationarity into the manager’s problem (at least in the deterministic model studied), which is compatible with an objective of equal access to hospice care. An attractive attribute of the alternative policy is that it closely resembles the existing policy so that its implementation is not expected to be disruptive to Medicare or hospice providers.
Even with the legacy policy in place however, the optimal recruitment and discharge rates for different disease types will differ. There are a number of possible remedies for this issue. In particular, Medicare could reimburse at different rates for different diseases, they could adjust the cap increment for different disease types, or they could move to a fixed plus variable reimbursement for differing patients. All of these suggestions raise significant new issues, such as the classifying of patients with co-morbidities into a single class, the incentives for patient churn inherent in a fixed payment system, and the difficulty involved in calibrating payment rates to actual costs. Our focus has been on highlighting and alleviating the calendar-based recruitment incentives inherent in the current policy. We leave a broader policy study of all incentives under the current scheme as the subject of future research.

To sum up, under any government policy there will always be unintended consequences; this paper sheds light on some of these under the Medicare hospice reimbursement program. In particular, we studied both the efficacy of the program with respect to hospice profitability and the hospice providers’ incentives for patient recruitment and live-discharges under the program. The primary remedies we suggest, namely the merging of appropriate providers and the new legacy policy, seek to mitigate the consequences of undesirable patient mix at a hospice. With respect to implementing the legacy policy, Congress may wish to run a pilot program with the new policy or task the Center for Medicare and Medicaid Studies to use historical data to estimate its financial impact.

References


Appendix: Not Intended for Print Publication

A Proofs

Proof of Proposition 1. Expected patient revenue equals $E[R \wedge \kappa]$, where $R \sim rN(m, \sigma^2)$ and $\kappa = (\lambda_1 + \lambda_2)K$. Then, $E[R \wedge \kappa] = r\sigma E[Z \wedge z] + rm$ where $Z$ is the standard normal r.v. and $z = (\kappa - rm)/(r\sigma)$. Note that $E[Z \wedge z] = -\phi(z) + z[1 - \Phi(z)]$ which follows directly from the properties of the normal distribution (e.g., Zipkin, 2000, p.459). Then, substituting $\bar{m} = m/\lambda$, $\bar{\sigma}^2 = \sigma^2/\lambda$, and $\kappa = \lambda K$ and rearranging terms yields total expected revenue as

$$E[R \wedge \kappa] = \lambda \left[ K - (K - \bar{m}) \Phi \left( \frac{K/r - \bar{m}}{\bar{\sigma}/\sqrt{\lambda}} \right) - \frac{\bar{\sigma}}{\sqrt{\lambda}} \phi \left( \frac{K/r - \bar{m}}{\bar{\sigma}/\sqrt{\lambda}} \right) \right].$$

The expected cost is $c_1 m_1 \lambda_1 + c_2 m_2 \lambda_2 + A = \bar{c} \lambda + A$ and the result follows.

For the comparative static results, $\frac{\partial \pi}{\partial K} = -\frac{1}{2} \lambda \left( \frac{2}{\sqrt{\pi}} \int_0^\infty e^{-t^2} dt - 1 \right)$, where $\bar{\lambda} = \frac{\lambda(K - r\bar{m})}{\sqrt{2r\bar{\sigma}}}$.

Because $\frac{2}{\sqrt{\pi}} \int_0^\infty e^{-t^2} dt$ is the Gaussian error function, bounded above by 1, $\frac{\partial \pi}{\partial K} > 0$. Moreover, some additional algebra shows $\frac{\partial^2 \pi}{\partial K^2} < 0$. So, $\pi$ is concave increasing in $K$. Further,

$$\frac{\partial^2 \pi}{\partial r^2} = -\frac{e^{-\frac{(K-r\bar{m})^2}{2\bar{\sigma}^2}}}{\sqrt{2\pi \bar{\sigma}^3}} < 0, \quad \frac{\partial \pi}{\partial \sigma_1} = -\frac{\lambda_1 e^{-\frac{(K-r\bar{m})^2}{2\bar{\sigma}^2}}}{\sqrt{2\pi \bar{\lambda} \bar{\sigma}}} r\sigma_1 < 0, \quad \frac{\partial \pi}{\partial \sigma_2} = -\frac{\lambda_2 e^{-\frac{(K-r\bar{m})^2}{2\bar{\sigma}^2}}}{\sqrt{2\pi \bar{\lambda} \bar{\sigma}}} r\sigma_2 < 0,$$

$$\frac{\partial \pi}{\partial c_1} = -\lambda_1 m_1 < 0, \quad \frac{\partial \pi}{\partial c_2} = -\lambda_2 m_2 < 0, \quad \frac{\partial \pi}{\partial A} = -1 < 0.$$

Thus, $\pi$ is concave in $r$, decreasing in $\sigma_1$ and $\sigma_2$, and linearly decreasing in $c_1, c_2,$ and $A$.  

Proof of Proposition 2. Fix $t$, $i$, and $j$, and consider one unit of class $i,j$ fluid arriving at time $t$, the fraction remaining in the system at time $s \geq t$ equals $Pr(Y_i^j > s - t)$, where $Y_i^j$ is a generic random variable with density $f_i^j(\cdot)$. Note that lifetimes are not random in this model so $Pr(Y_i^j > s - t)$ is the exact fraction that remains, not the expected fraction. Then the terminal value $v_i^j(t)$ associated with one unit of class $i,j$ fluid arriving at time $t$ is given by

$$v_i^j(t) = v_i^j \Pr(Y_i^j > T - t) = v_i^j \int_{T-t}^\infty f_i^j(x)dx.$$

The cumulative (potential) revenue $r_i^j(t)$ over $[0, T]$ from one unit of class $i,j$ fluid arriving at time $t$ is given by

$$r_i^j(t) = r \int_t^{T} \Pr(Y_i^j > s - t) ds = r \int_0^{T-t} \Pr(Y_i^j > x) dx = r \int_0^{\infty} (x \wedge (T - t)) f_i^j(x) dx,$$
where the final equality is achieved with an interchange of integrals. The derivation of $c^j_i(\cdot)$ follows similarly.

\section*{Proof of Proposition 3.}

To simplify the analysis, we define for all $i, t, q$,
\[
\delta_i(t; q) = K + (r_i(t) - K)q - c_i(t) + v_i(t),
\]
\[
\tilde{\delta}_i(t; q) = -\tilde{r}_i(t)q + \tilde{c}_i(t) - \tilde{v}_i(t),
\]
and make the following technical assumption.

\textbf{Technical Assumption 1:} The number of times the functions $\delta_i(\cdot; q)$ and $\tilde{\delta}_i(\cdot; q)$ change sign on $[0, T]$ is uniformly bounded for all $q \in [0, 1]$.

This is a reasonable assumption given that $(r_i(\cdot), c_i(\cdot), v_i(\cdot))$ and $(\tilde{r}_i(\cdot), \tilde{c}_i(\cdot), \tilde{v}_i(\cdot))$ are smooth functions. Moreover, it is satisfied when the length of stay distribution is exponential or gamma.

By the technical assumption, there exists $N$, and for $i = 1, 2$, continuous trigger functions $\{\underline{\tau}_{ij}(\cdot), \bar{\tau}_{ij}(\cdot)\}_{j=1}^N$ and $\{\underline{\tau}_{ij}(\cdot), \bar{\tau}_{ij}(\cdot)\}_{j=1}^N$ corresponding to the points at which $\delta_i(\cdot; q)$ and $\tilde{\delta}_i(\cdot; q)$ change sign, respectively, such that for all $i, q$
\[
0 \leq \underline{\tau}_{11}(q) \leq \bar{\tau}_{11}(q) \leq \ldots \leq \underline{\tau}_{1N}(q) \leq \bar{\tau}_{1N}(q) \leq T,
\]
\[
0 \leq \underline{\tau}_{21}(q) \leq \bar{\tau}_{21}(q) \leq \ldots \leq \underline{\tau}_{2N}(q) \leq \bar{\tau}_{2N}(q) \leq T,
\]
which are defined as follows. We will only construct $\{\underline{\tau}_{ij}(\cdot), \bar{\tau}_{ij}(\cdot)\}_{j=1}^N$; and the construction of $\{\underline{\tau}_{ij}(\cdot), \bar{\tau}_{ij}(\cdot)\}$ follows similarly.

To this end, fix $i = 1, 2$ and $q \in [0, 1]$. If $\tilde{\delta}_i(\cdot; q)$ is always positive, then let $\underline{\tau}_{1j}(q) = 0$, $\bar{\tau}_{1j}(q) = T$ and $\underline{\tau}_{2j}(q) = \bar{\tau}_{2j}(q) = T$ for $j = 2, \ldots, N$. If $\tilde{\delta}_i(\cdot; q)$ is always non-positive, then let $\underline{\tau}_{ij}(q) = \bar{\tau}_{ij}(q) = T$ for all $j$. Otherwise, $\tilde{\delta}_i(\cdot; q)$ changes sign at least once. Let $0 < \gamma^1_i < \ldots < \gamma^M_i < T$ denote the times at which $\tilde{\delta}_i(\cdot; q)$ changes sign, where $1 \leq M \leq 2N$. For notational convenience, let $\gamma^{0}_i = 0$ and $\gamma^{M+1}_i = T$. There are two cases to consider:

\textbf{Case i:} The sign switches to positive at $\gamma^1_i$. Then
\[
(\underline{\tau}_{ij}(q), \bar{\tau}_{ij}(q)) = \begin{cases} (\gamma^j_i, T) & \text{for } 1 \leq j \leq \lfloor \frac{M}{2} \rfloor, \\ (T, T) & \text{otherwise.} \end{cases}
\]

\textbf{Case ii:} The sign switches to negative at $\gamma^1_i$. Then
\[
(\underline{\tau}_{ij}(q), \bar{\tau}_{ij}(q)) = \begin{cases} (\gamma^j_i, \gamma^{2j-1}_i) & \text{for } 1 \leq j \leq \lfloor \frac{M}{2} \rfloor, \\ (T, T) & \text{otherwise.} \end{cases}
\]
The trigger functions $\bar{t}_{ij}(\cdot)$ and $\bar{t}_{ij}(\cdot)$ are defined similarly. Moreover, assumption (5) ensures that for every $q$ and some $i$, $0 \leq \bar{t}_{ii}(\cdot) < \bar{t}_{ij}(\cdot) < T$. Given the trigger functions, we write

$$F(q) = -K(\lambda_1 + \lambda_2)T + R(0) + \sum_{i=1}^{2} \lambda_i \int_{0}^{T} r_i(s)ds$$

$$+ \sum_{i=1}^{2} \sum_{j=1}^{N} \int_{\underline{t}_{ij}(q)}^{\bar{t}_{ij}(q)} \frac{r_i(t) - K}{\eta_i} (K + (r_i(t) - K)q - c_i(t) + v_i(t))dt$$

$$+ \sum_{i=1}^{2} \sum_{j=1}^{N} \int_{\underline{t}_{ij}(q)}^{\bar{t}_{ij}(q)} \frac{-\bar{r}_i(t)}{\eta_i} (-\bar{r}_i(t)q + \bar{c}_i(t) - \bar{v}_i(t))dt.$$ 

Note that the trigger functions may not be differentiable at all points. However, it is straightforward to argue that their right and left derivatives exist. Therefore, a viable proof strategy to show differentiability is to establish that the right and left-derivatives of $F$ are equal. In the interest of brevity, we will proceed as if the trigger functions are differentiable, but it will be clear in calculating $F'$ that whether we use the left or the right derivatives makes no difference in calculating $F'$ because the terms involving $\bar{t}'_{ij}, \bar{r}'_{ij}, \bar{\lambda}'_{ij}, \bar{\tau}'_{ij}$ will vanish.

To be more specific, note by Leibnitz differentiation rule that

$$F'(q) = \sum_{i=1}^{2} \sum_{j=1}^{N} \left[ \int_{\underline{t}_{ij}(q)}^{\bar{t}_{ij}(q)} \frac{r_i(t) - K}{\eta_i} dt + \int_{\underline{t}_{ij}(q)}^{\bar{t}_{ij}(q)} \frac{(\bar{r}_i(t))^2}{\eta_i} dt \right]$$

$$+ \sum_{i=1}^{2} \sum_{j=1}^{N} \left[ \frac{\bar{r}_i(\bar{t}_{ij}(q) - K)}{\eta_i} \delta_i(\bar{t}_{ij}(q); q) \bar{\lambda}_{ij}(q) - \frac{r_i(\bar{t}_{ij}(q)) - K}{\eta_i} \delta_i(\bar{t}_{ij}(q); q) \bar{\lambda}'_{ij}(q) \right]$$

$$- \sum_{i=1}^{2} \sum_{j=1}^{N} \left[ \frac{\bar{r}_i(\bar{t}_{ij}(q))}{\eta_i} \delta_i(\bar{t}_{ij}(q); q) \bar{\lambda}_{ij}(q) - \frac{\bar{r}_i(\bar{t}_{ij}(q))}{\eta_i} \delta_i(\bar{t}_{ij}(q); q) \bar{\lambda}'_{ij}(q) \right],$$

where the last two summations vanish because each of the summands is zero (by definition of the trigger functions and $\delta_i, \tilde{\delta}_i$) regardless of whether one uses the left or the right derivatives. Then because the trigger functions are continuous, $F$ is continuously differentiable with

$$F'(q) = \sum_{i=1}^{2} \sum_{j=1}^{N} \int_{\underline{t}_{ij}(q)}^{\bar{t}_{ij}(q)} \frac{(r_i(t) - K)^2}{\eta_i} dt + \sum_{i=1}^{2} \sum_{j=1}^{N} \int_{\underline{t}_{ij}(q)}^{\bar{t}_{ij}(q)} \frac{\bar{r}_i(t)^2}{\eta_i} dt.$$ 

Moreover, $F'(q) > 0$ because $0 \leq \bar{t}_{ii}(\cdot) < \bar{t}_{ij}(\cdot) < T$ for some $i$ by assumption (5). Therefore, $F(q)$ is strictly increasing on $(0,1)$.

**Proof of Theorem 1.** Note that the hospice manager’s problem (P) is equivalent to the optimal control problem (49)-(55) presented in Appendix B, where $\alpha_i(t) = \bar{z}_i(t)$ and $\theta_i(t) = \bar{\Theta}_i(t)$ for all $i, t$. Similarly, the dual optimal control problem (69)-(74) is equivalent to the dual problem (D),
introduced in Appendix B, with $\dot{p}(t) = 0$ and $p(t) = (K(1 - q), q)$ for all $t$. Rockafellar (1970) provides a duality relationship between (49)-(55) and (69)-(74), whereby the two formulations have the same optimal objective. Moreover, by Theorem 5 of Rockafellar (1970), the optimal solutions to (49)-(55) and (69)-(74), must satisfy

$$ (\dot{p}(t), p(t)) \in \partial L(t, (z(t), \zeta(t)), (\dot{z}(t), \dot{\zeta}(t))) \quad (10) $$

for $t \in [0, T]$. Also note by Proposition 8.12 of Rockafellar and Wets (1997) that for any proper convex function $f$, its subgradient set $\partial f(\bar{x})$ at $\bar{x}$ is given by

$$ \partial f(\bar{x}) = \{u : f(x) \geq f(\bar{x}) + \langle u, x - \bar{x} \rangle \text{ for all } x\}. $$

Namely, for $v \in \partial f(\bar{x})$, we must have that $f(x) \geq f(\bar{x}) + \langle v, x \rangle - \langle v, \bar{x} \rangle \text{ for all } x$. Rearranging the terms gives

$$ \langle v, \bar{x} \rangle - f(\bar{x}) \geq \langle v, x \rangle - f(x) \text{ for all } x, $$

which holds with equality for $x = \bar{x}$. Therefore,

$$ \langle v, \bar{x} \rangle - f(\bar{x}) = \sup_{x} \{\langle v, x \rangle - f(x)\} = f^{*}(v). \quad (11) $$

Hence, we conclude that $v \in \partial f(\bar{x})$ if and only if $\bar{x}$ is an element of the set $\arg \max_{x} \{\langle v, x \rangle - f(x)\}$ in defining $f^{*}(v)$, c.f., (11). Using this observation, (10) holds if and only if, for $t \in [0, T]$,

$$ (z(t), \Theta(t), \zeta(t), \dot{z}(t), \dot{\Theta}(t), \dot{\zeta}(t)) \in \arg \max_{x,y} \{\langle \dot{p}(t), p(t) \rangle \cdot (x,y) - L(t, x, y)\}. \quad (12) $$

By Proposition 6, (12) is equivalent to having

$$ \dot{z}_{i}(t) = \frac{[p_{i}^{z} + r_{i}(t)p_{i}^{\zeta} + v_{i}(t) - c_{i}(t)]^{+}}{\eta_{i}^{z}} \quad \text{for } i = 1, 2. \quad (13) $$

$$ \dot{\Theta}_{i}(t) = \frac{[p_{i}^{\Theta} - \bar{r}_{i}(t)p_{i}^{\Theta} + \bar{c}_{i}(t) - \bar{v}_{i}(t)]^{+}}{\eta_{i}^{\Theta}} \quad \text{for } i = 1, 2. \quad (14) $$

Also, note by equivalence of the formulations (69)-(74) and (D) that

$$ p_{i}^{z} = K(1 - q), p_{i}^{\zeta} = q \quad \text{and} \quad p_{i}^{\Theta} = 0 \text{ for } i = 1, 2. $$

Similarly, by the equivalence of (49)-(55) and (P) we have that

$$ \alpha(t) = \dot{z}(t) \text{ and } \theta(t) = \dot{\Theta}(t) \text{ for } t \in [0, T]. $$

35
Therefore, (13)-(14) give the hospice manager’s optimal recruiting and live-discharge rates as stated in the theorem, where \( q^* \) is the optimal solution to the dual problem, characterized in Proposition 7.

**Proof of Proposition 4.** Let \( \alpha(\cdot) \) and \( \theta(\cdot) \) denote a feasible nonstationary policy. Then let

\[
\bar{\alpha}_i = \frac{1}{T} \int_0^T \alpha_i(s)ds \quad \text{and} \quad \bar{\theta}_i = \frac{1}{T} \int_0^T \theta_i(s)ds,
\]

and observe that for \( i = 1, 2, \)

\[
\int_0^T \frac{1}{T} s_i(\alpha_i(t))dt > s_i \left( \int_0^T \frac{1}{T} \alpha_i(t)dt \right) = s_i(\bar{\alpha}_i)
\]

(15)

and

\[
\int_0^T \frac{1}{T} g_i(\theta_i(t))dt > g_i \left( \int_0^T \frac{1}{T} \theta_i(t)dt \right) = g_i(\bar{\theta}_i)
\]

(16)

by Jensen’s inequality and strict convexity of the quadratic recruiting and live-discharging cost functions \( s_i(\cdot) \) and \( g_i(\cdot) \). Multiplying both sides of (15)-(16) by \( T \) gives

\[
\int_0^T s_i(\alpha_i(t))dt > Ts_i(\bar{\alpha}_i) \quad \text{and} \quad \int_0^T g_i(\theta_i(t))dt > Tg_i(\bar{\theta}_i)
\]

so that the recruiting and live-discharge costs are strictly larger for the nonstationary policy \( \alpha(\cdot) \) and \( \theta(\cdot) \) than its stationary counterpart \( \bar{\alpha} \) and \( \bar{\theta} \), while all other costs and the revenues are the same for the two policies. Thus, switching over to the stationary policy \( (\bar{\alpha}, \bar{\theta}) \) strictly improves the hospice’s profit.

**Proof of Lemma 1.** For notational simplicity, let

\[
a = \sum_{i=1}^{2} (rm_i - K) \left[ \lambda_i + \frac{(r - c_i)m_i}{\eta_i^2} \right] \quad \text{and} \quad b = \sum_{i=1}^{2} \frac{(rm_i - K)^2}{\eta_i^2}.
\]

(i) Note that \( \pi_1 < c_2/r \) is equivalent to

\[
br + \sum_{i=1}^{2} \frac{(rm_i)^2}{\eta_i^2} c_i - ra < c_2b + c_2 \sum_{i=1}^{2} \frac{(rm_i)^2}{\eta_i^2}.
\]

(17)

Subtracting \( c_2(r\tilde{m}_2)^2/\eta_2^2 \) from both sides of (17) shows that (17) is equivalent to

\[
br + \frac{(r\tilde{m}_1)^2}{\eta_1^2} c_1 - ra < c_2b + c_2 \frac{(r\tilde{m}_1)^2}{\eta_1^2},
\]

(18)

which, in turn, is equivalent to \( \pi_2 < c_2/r \).
(ii) Because $\pi_1 < c_2/r$ is equivalent to (18), it follows from $\pi_1 < c_2/r$ and $c_1 > c_2$ that
\[
br + \frac{(r \tilde{m}_1)^2}{\eta_1^i} - c_1 - ra < c_2 b + \frac{c_2 (r \tilde{m}_1)^2}{\eta_1^i} < c_1 b + c_1 \frac{(r \tilde{m}_1)^2}{\eta_1^i},
\]
which implies $\pi_3 < c_1/r$.

(iii) Note that $\pi_2 < c_1/r$ is equivalent to
\[
br + \frac{(r \tilde{m}_1)^2}{\eta_1^i} - c_1 - ra < bc_1 + \frac{c_1 (r \tilde{m}_1)^2}{\eta_1^i},
\]
which clearly is equivalent to $\pi_3 < c_1/r$.

\[\blacksquare\]

**Proof of Proposition 5.** We can make the hospice manager’s optimization problem convex and put it in the form of Boyd and Vandenberghe (2004) as follows:

\[
\begin{align*}
\min & -\xi + \sum_{i=1}^{2} c_i m_i \alpha_i - \sum_{i=1}^{2} c_i \tilde{m}_i \theta_i + \sum_{i=1}^{2} \frac{1}{2} \eta_i^s \alpha_i^2 + \sum_{i=1}^{2} \frac{1}{2} \eta_i^l \theta_i^2 \\
\text{subject to} & \xi - K \sum_{i=1}^{2} (\lambda_i + \alpha_i) \leq 0, \\
\xi - r \sum_{i=1}^{2} [m_i (\lambda_i + \alpha_i) - \tilde{m}_i \theta_i] & \leq 0, \\
-\alpha_i & \leq 0, i = 1, 2. \\
-\theta_i & \leq 0, i = 1, 2.
\end{align*}
\]

Letting $\mu_i, \nu_i^s, \nu_i^l$ for $i = 1, 2$ denote the Lagrange multipliers, we write the KKT conditions as follows (see p. 243 of Boyd and Vandenberghe, 2004). First consider setting the gradient of the Lagrangian to zero, which gives

\[
\begin{align*}
-1 + \mu_1 + \mu_2 & = 0, \\
c_i m_i + \eta_i^s \alpha_i - K \mu_1 - r m_i \mu_2 - \nu_i^s = 0, & i = 1, 2 \\
-c_i \tilde{m}_i + \eta_i^l \theta_i + r \tilde{m}_i \mu_2 - \nu_i^l = 0, & i = 1, 2.
\end{align*}
\]

That is,

\[
\begin{align*}
\alpha_i & = \frac{K \mu_1 + r m_i \mu_2 + \nu_i^s - c_i m_i}{\eta_i^s}, \\
\theta_i & = \frac{-r \tilde{m}_i \mu_2 + \nu_i^l + c_i \tilde{m}_i}{\eta_i^l}, \\
\mu_1 + \mu_2 & = 1.
\end{align*}
\]

37
Also, the feasibility and complementary slackness conditions give the following:

\[ \xi - K \sum_{i=1}^{2} (\lambda_i + \alpha_i) \leq 0, \]  
\[ \xi - r \sum_{i=1}^{2} [m_i(\lambda_i + \alpha_i) - \tilde{m}_i \theta_i] \leq 0, \]  
\[ \alpha_i \geq 0, \]  
\[ \theta_i \geq 0, \]  
\[ \mu_1 \left[ \xi - K \sum_{i=1}^{2} (\alpha_i + \lambda_i) \right] = 0, \]  
\[ \mu_2 \left[ \xi - r \sum_{i=1}^{2} [m_i(\alpha_i + \lambda_i) - \tilde{m}_i \theta_i] \right] = 0, \]  
\[ \nu_i^s \alpha_i = 0, \quad i = 1, 2 \]  
\[ \nu_i^l \theta_i = 0, \quad i = 1, 2 \]  
\[ \mu_i \geq 0, \quad i = 1, 2 \]  
\[ \nu_i^l \geq 0 \quad i = 1, 2, j = s, l. \]  

Then equations (19)-(31) will pin down the optimal solution.

Recall that \( K > c_i m_i \) for \( i = 1, 2 \) by assumption. Then it is easy to see that \( \alpha_i > 0 \) for \( i = 1, 2 \). Otherwise, by (19), \( \alpha_i = 0 \) implies

\[ \nu_i^s = m_i [c_i - (\mu_2 r + (1 - \mu_2)K/m_1)] < 0, \]

which contradicts (31). Therefore, in what follows we assume \( \alpha_i > 0 \) for \( i = 1, 2 \), which, in turn, implies by (28) that

\[ \nu_i^s = 0 \quad \text{for} \quad i = 1, 2. \]  

Despite this simplification, we still need to consider several cases. First, note that at least one of the inequalities (22) and (23) will bind. Therefore, we have the following three main cases to consider:

- **Case 1:** Constraint (22) does not bind.
- **Case 2:** Constraint (23) does not bind.
- **Case 3:** Both (22) and (23) bind.

Consider Case 1: Because the cap constraint, c.f. (22), does not bind, one would expect that the hospice manager recruits patients and does not live-discharge any, i.e. \( \theta_i = 0 \) for \( i = 1, 2 \). Indeed,
to see this note that $\mu_1 = 0$ and $\mu_2 = 1$ by (26) and (21). Also, if $\theta_i > 0$, then $\nu_i^l = 0$ by (29). Then by (20), $\theta_i = \tilde{m}_i(c_i - r)/\eta_i^l < 0$, which contradicts (25). Therefore, in case 1, we must have $\theta_i = 0$ for $i = 1, 2$. More specifically, because $\mu_1 = 0$ and $\mu_2 = 1$, we conclude from (19) and (32) that

$$\alpha_i = \frac{m_i(r - c_i)}{\eta_i^l} \quad \text{and} \quad \theta_i = 0 \text{ for } i = 1, 2.$$  \tag{33}

To conclude the part 1 of the proof, we note that case 1 arises if and only if (22) does not bind, but (23) does:

$$\xi < K \sum_{i=1}^2 (\lambda_i + \alpha_i) \quad \text{and} \quad \xi = r \sum_{i=1}^2 [m_i(\lambda_i + \alpha_i) - \tilde{m}_i \theta_i].$$

That is,

$$r \sum_{i=1}^2 [m_i(\lambda_i + \alpha_i) - \tilde{m}_i \theta_i] < K \sum_{i=1}^2 (\lambda_i + \alpha_i).$$

Substituting (33) and rearranging terms gives the following equivalent condition:

$$\sum_{i=1}^2 (rm_i - K) \left[ \lambda_i + \frac{m_i(r - c_i)}{\eta_i^l} \right] < 0 \tag{34}$$

which along with (33) prove part 1 of Proposition 5.

Next, consider case 2: constraint (23) does not bind. That is, the cap constraint binds and the potential revenue constraint does not. In this case, one would expect the hospice manager to both recruit patients (to relax the cap constraint) and to live-discharge patients to reduce costs. In other words, we expect to see $\theta_i > 0$ for $i = 1, 2$. Indeed, to see this note that by (27) and (21), $\mu_1 = 1, \mu_2 = 0$. Then it follows from (20) that $\theta_i \geq c_i \tilde{m}_i/\eta_i^l > 0$ for $i = 1, 2$. Therefore, we must have $\theta_i > 0$ for $i = 1, 2$ in case 2.

To be more specific, $\theta_i > 0$ implies $\nu_i^l = 0$ by (29). Then, it follows from (19)-(20) and (32) that

$$\alpha_i = \frac{K - c_i m_i}{\eta_i^l} \quad \text{and} \quad \theta_i = \frac{c_i \tilde{m}_i}{\eta_i^l} \text{ for } i = 1, 2.$$  \tag{35}

Moreover, observe that case 2 arises if and only if

$$\xi = K \sum_{i=1}^2 (\lambda_i + \alpha_i) \quad \text{and} \quad \xi < r \sum_{i=1}^2 [m_i(\lambda_i + \alpha_i) - \tilde{m}_i \theta_i].$$

That is,

$$K \sum_{i=1}^2 (\lambda_i + \alpha_i) < r \sum_{i=1}^2 [m_i(\lambda_i + \alpha_i) - \tilde{m}_i \theta_i].$$

Substituting (35) into this and rearranging terms shows that case 2 arises if and only if

$$\sum_{i=1}^2 \frac{r(\tilde{m}_i)^2 c_i}{\eta_i^l} + \sum_{i=1}^2 \frac{(rm_i - K)}{\eta_i^*} \sum_{i=1}^2 \frac{2}{(rm_i - K)} \left[ \lambda_i + \frac{m_i(r - c_i)}{\eta_i^*} \right], \tag{36}$$

39
which along with (35) proves part 2 of Proposition 5.

Finally, consider case 3, where both (22) and (23) bind. As a first step, we rule out the case \( \theta_1 = 0 \) and \( \theta_2 > 0 \). Suppose that this is possible. Then by (29), \( \theta_2 > 0 \) implies \( \nu_2' = 0 \), and by (20) we write \( \theta_2 = \tilde{m}_2(c_2 - \mu_2 r)/\eta_2' \), which in turn implies \( \mu_2 < c_2/r \) because \( \theta_2 > 0 \). At the same time, \( \theta_1 = 0 \) implies by (20) that \( \mu_2 = \nu_2'/(r \tilde{m}_1) + c_1/r \geq c_1/r \). Thus, we conclude both \( \mu_2 \geq c_1/r \) and \( \mu_2 < c_1/r \), which is a contradiction because \( c_1 > c_2 \). Next, we consider the following three subcases, each of which may arise:

**Case 3a:** \( \theta_i > 0 \) for \( i = 1, 2 \).

**Case 3b:** \( \theta_1 > 0 \) and \( \theta_2 = 0 \).

**Case 3c:** \( \theta_i = 0 \) for \( i = 1, 2 \).

Consider Case 3a: \( \theta_i > 0 \) implies \( \nu_i' = 0 \) for \( i = 1, 2 \) by (29). Then it follows from (19)-(20) and (32) that
\[
\alpha_i = \frac{K(1 - \mu_2) + r m_i \mu_2 - c_i m_i}{\eta_i'} \quad \text{and} \quad \theta_i = \frac{\tilde{m}_i (c_i - r \mu_2)}{\eta_i'} \quad \text{for} \quad i = 1, 2. \quad (37)
\]
Because we must have \( \theta_i > 0 \), (37) requires that
\[
\mu_2 < \frac{c_2}{r} < \frac{c_1}{r}. \quad (38)
\]
Moreover, because both (22) and (23) bind, we must have that
\[
K \sum_{i=1}^{2} (\alpha_i + \lambda_i) = r \sum_{i=1}^{2} [m_i (\lambda_i + \alpha_i) - \tilde{m}_i \theta_i]. \quad (39)
\]
Substituting \( \mu_1 = 1 - \mu_2 \) and (36) into (39) and rearranging terms give
\[
\mu_2 = \pi_1. \quad (40)
\]

Consider Case 3b: \( \theta_1 > 0 \) implies that \( \nu_1' = 0 \). Moreover, \( \theta_2 = 0 \) implies by (20) that \( \nu_2' = \tilde{m}_2 (r \mu_2 - c_2) \geq 0 \), which requires \( \mu_2 \geq c_2/r \). Also, \( \theta_1 > 0 \) requires \( \mu_2 < c_1/r \) by (19). So we must have
\[
\frac{c_2}{r} \leq \mu_2 < \frac{c_1}{r}. \quad (41)
\]
Also, it follows from (19)-(20) and (32) that
\[
\alpha_i = \frac{K(1 - \mu_2) + r m_i \mu_2 - c_i m_i}{\eta_i'} \quad \text{for} \quad i = 1, 2 \quad \text{and} \quad \theta_1 = \frac{\tilde{m}_1 (c_1 - r \mu_2)}{\eta_1'}. \quad (42)
\]
Because both (22) and (23) bind, (39) must hold. Then using (42) and the fact that \( \theta_2 = 0 \), we find that
\[
\mu_2 = \pi_2. \tag{43}
\]

Finally, consider Case 3c: \( \theta_2 = 0 \) for \( i = 1, 2 \). Note that \( \theta_i = 0 \) implies by (20) that \( \nu_i = \tilde{m}_i[r\mu_2 - c_i] \), which, in turn, implies \( \mu_2 \geq c_i/r \). That is, we must have
\[
\mu_2 \geq \frac{c_1}{r} > \frac{c_2}{r}. \tag{44}
\]
Moreover, it follows from (19) and (32) that
\[
\alpha_i = \frac{K(1-\mu_2) + rm_i\mu_2 - c_i m_i}{\eta_i^*} \quad \text{for } i = 1, 2. \tag{45}
\]
Because both (22) and (23) bind, using (44) and the fact that \( \theta_i = 0 \) for \( i = 1, 2 \), one can show that
\[
\mu_2 = \pi_3. \tag{46}
\]

To conclude the proof, note by Lemma 1 that the conditions (38) and (40), (41) and (43), and (44) and (46) partition the remaining subcases of case 3 above (over the parameter space of the problem) as in Part 3 of Proposition 5. Combining this partition and characterizations of recruiting and live-discharge rates in (37), (42), and (45) proves Part 3 of Proposition 5. \( \blacksquare \)

**B Duality Analysis**

In this appendix, we derive the dual formulation and some auxiliary results. To facilitate the statement of the dual formulation, let
\[
\begin{align*}
\beta_1 &= K(\lambda_1 + \lambda_2)T \\
\beta_2 &= \sum_{i=1}^{2} \int_{0}^{T} \lambda_i r_i(s)ds + R(0)
\end{align*} \tag{47, 48}
\]
It will be shown below that the dual formulation \((D)\) of \((P)\) can be stated as follows: Choose \( q \) so as to
\[
\begin{align*}
\min & \quad \beta_1(1-q) + \beta_2 q + \sum_{i=1}^{2} \int_{0}^{T} \frac{1}{2\eta_i^*} \left( [K(1-q) + r_i(t)q + v_i(t) - c_i(t)]^+ \right)^2 dt \\
& \quad + \sum_{i=1}^{2} \int_{0}^{T} \frac{1}{2\eta_i^*} \left( [-\tilde{r}_i(t)q - \tilde{v}_i(t) + \tilde{c}_i(t)]^+ \right)^2 dt \\
\text{subject to} & \quad 0 \leq q \leq 1.
\end{align*} \tag{D}
\]
This appendix proves this statement and (in Proposition 7) provides the optimal solution to this dual formulation.
For \( i = 1, 2 \) and \( t \in [0, T] \), let
\[
\dot{z}_i(t) = \int_0^t \alpha_i(s) ds \quad \text{and} \quad \dot{\Theta}_i(t) = \int_0^t \theta_i(s) ds
\]
(in particular, \( \dot{z}_i(s) = \alpha_i(s) \) and \( \dot{\Theta}_i(s) = \theta_i(s) \)). That is, \( z_i(t) \) and \( \Theta_i(t) \) denote the cumulative number of recruited and live-discharged, respectively, patients by time \( t \). Observe that the hospice manager’s problem (P) can be written as follows: Choose \( \dot{z}(\cdot), \dot{\Theta}(\cdot) \) and \( \dot{\zeta}(\cdot) \) so as to
\[
\min\{ -\min(\beta_1 + K z_1(T) + K z_2(T), \beta_2 + \zeta_1(T) + \zeta_2(T)) \} + \int_0^T \sum_{i=1}^2 [(c_i(t) - v_i(t))\dot{z}_i(t) + s_i(\dot{z}_i(t))] dt
\]
subject to
\[
\begin{align*}
z_i(t) &= z_i(0) + \int_0^t \dot{z}_i(s) ds, \; z_i(0) = 0, \\
\dot{z}_i(t) &= \dot{z}_i(0) + \int_0^t \dot{\zeta}_i(s) ds, \; \dot{z}_i(0) = 0, \\
\dot{\Theta}_i(t) &= \dot{\Theta}_i(0) + \int_0^t \dot{\Theta}_i(s) ds, \; \dot{\Theta}_i(0) = 0, \\
\dot{\zeta}_i(t) &\geq 0, \\
\dot{\Theta}_i(s) &\geq 0,
\end{align*}
\]
\[
\dot{\zeta}_i(s) = r_i(s)\dot{z}_i(s) - \tilde{r}_i(s)\dot{\Theta}_i(s).
\]

The cumulative revenue accrued until time \( t \) from patients of type \( i \) is \( \zeta_i(t) \) and \( \dot{\zeta}_i(t) \) is the revenue rate at time \( t \) from patients of type \( i \). To put this in the framework of Rockafellar (1970), define
\[
L(t, x, y) = \sum_{i=1}^2 [(c_i(t) - v_i(t))y_i^z + s_i(y_i^z) + \chi_{\{y_i^z \geq 0\}}] + \sum_{i=1}^2 [-(\tilde{c}_i(t) - \tilde{v}_i(t))y_i^\Theta + g_i(y_i^\Theta)] + \sum_{i=1}^2 [\chi_{\{y_i^z = r_i(t)\tilde{y}_i(t) - \tilde{r}_i(t)\tilde{y}_i(t)\geq 0\}}],
\]
\[
l(x, y) = \chi_{\{x^z = 0, x^\Theta = 0, x^\zeta = 0\}} - \min\{\beta_1 + Ky_1^z + Ky_2^z, \beta_2 + y_1^z + y_2^z\}
\]
for \( x = (x_1^z, x_1^\Theta, x_1^\zeta) \in \mathbb{R}^3 \), \( x = (x_1, x_2) \in \mathbb{R}^6 \), \( y_i = (y_i^z, y_i^\Theta, y_i^\zeta) \in \mathbb{R}^3 \), \( y = (y_1, y_2) \in \mathbb{R}^6 \), and \( x_j = (x_1^j, x_2^j) \), \( y_j = (y_1^j, x_2^j) \in \mathbb{R}^2 \) for \( j = z, \Theta, \zeta \), and \( \chi_{\leq F} \) is an “indicator” function taking values zero or infinity. Namely,
\[
\chi_{\leq F}(a) = \begin{cases} \infty & \text{if } a \notin F, \\ 0 & \text{otherwise.} \end{cases}
\]
Notice $L(t,x,y)$ is independent of $x$ because the profit function is independent of the cumulative number of recruited patients and cumulative revenue within the year; the cumulative number of patients and cumulative revenue is relevant only at time $T$ which are represented by $y$ in $l(x,y)$. Then the hospice manager’s problem can be written as follows: Choose the functions $\dot{z}(\cdot), \dot{\Theta}(\cdot), \dot{\zeta}(\cdot)$ so as to

$$\min l((z(0), \Theta(0), \zeta(0)), (z(T), \Theta(T), \zeta(T))) + \int_0^T L(t, (z(t), \Theta(t), \zeta(t)), (\dot{z}(t), \dot{\Theta}(t), \dot{\zeta}(t)))dt. \tag{59}$$

Following Rockafellar (1970) to derive the dual problem,\(^{19}\) define

$$m(d(0), d(T)) = l^*(d(0), -d(T)), \tag{60}$$

$$M(t, p, s) = L^*(t, s, p), \tag{61}$$

where $l^*$ and $L^*$ are convex conjugates of $l$ and $L$, respectively, $m$ is the terminal function dual to $l$, $M$ is the Lagrangian function dual to $L$, and $d(t) = (d^x(t), d^z(t)) \in \mathbb{R}^6$. The dual problem can then be stated as follows: Choose $p(\cdot)$ and $\dot{p}(\cdot)$ so as to

$$\min m(p(0), p(T)) + \int_0^T M(t, p(t), \dot{p}(t))dt. \tag{62}$$

The next step is to characterize $m, M$ which is done in the next proposition.

**Proposition 6** We have that

$$m(d^x(0), d^\Theta(0), d^\zeta(0), d^z(T), d^\Theta(T), d^\zeta(T)) = \beta_1 d^x_1(T)/K + \beta_2 d^\zeta_1(T) + \chi_{\{(d^x_1(T) - d^z_1(T))/K = \zeta, \zeta \}}$$

$$+ \chi_{\{d^\zeta_1(T) \geq 0, i = 1, 2, j = x, z, \}} + \chi_{\{d^\Theta(T) = 0\}}, \tag{63}$$

$$M(t, p, s) = \sum_{i=1}^2 \frac{1}{2\eta_i^2} \left( [p_i^* + r_i(t)p_i^\zeta + v_i(t) - c_i(t)]^2 + \chi_{\{s = 0\}} \right)

+ \sum_{i=1}^2 \frac{1}{2\eta_i^2} \left( [p_i^\Theta - \tilde{r}_i(t)p_i^\zeta - \tilde{v}_i(t) + \tilde{c}_i(t)]^2 \right). \tag{64}$$

Moreover,

$$\arg \max_{s, p} \{(s, p) \cdot (x, y) - L(t, x, y)\} = \left\{ (x, y) : \begin{array}{c}
y_i^* = \frac{p_i^* + r_i(t)p_i^\zeta + v_i(t) - c_i(t)}{\eta_i^*},
y_i^\Theta = \frac{p_i^\Theta - \tilde{r}_i(t)p_i^\zeta - \tilde{v}_i(t) + \tilde{c}_i(t)}{\eta_i^*}, \text{ and } \\
y_i^\zeta = r_i(t)y_i^* - \tilde{r}_i(t)y_i^\Theta \end{array} \right\}. \tag{65}$$

\(^{19}\)Here we are following Rockafellar’s (1970) notation as closely as possible to facilitate the use of his results. For example, the swapping of the order of arguments in $M$ and $L^*$ in (61) matches his equation (5.5) on p. 190.
**Proof.** Note that

\[
m(d^2(0), d^3(0), d^z(0), d^z(T), d^\zeta(T), d^z(T)) = l^*(d^2(0), d^3(0), d^z(0), -d^z(T), -d^\zeta(T), -d^z(T))
\]

\[
= \sup_c \{c(0)d(0) - c(T)d(T) - \chi_{\{c(0)=0\}}
+ \min_\{\beta_1 + Kc^z_1(T) + Kc^z_2(T), \beta_2 + c^z_1(T) + c^z_2(T)\}\}
\]

\[
= \sup_{c(T)} \{-c^z(T)d^z(T) - c^\zeta(T)d^\zeta(T) - c^\zeta(T)d^\zeta(T)
+ \min_{c(T)} \{\beta_1 + Kc^\zeta_1(T) + Kc^\zeta_2(T), \beta_2 + c^\zeta_1(T) + c^\zeta_2(T)\}\}
\]

\[
= \chi_{\{d^z(T)=0\}} + \sup_{c(T)} \{-c^z(T)d^z(T) - c^\zeta(T)d^\zeta(T)
+ \min_{c(T)} \{\beta_1 + Kc^\zeta_1(T) + Kc^\zeta_2(T), \beta_2 + c^\zeta_1(T) + c^\zeta_2(T)\}\}.
\]

The optimization problem on the right-hand side is equivalent to the following linear program:

\[
\max_{c(T)} \{\xi - c^z_1(T)d^z_1(T) - c^z_2(T)d^z_2(T) - c^\zeta_1(T)d^\zeta_1(T) - c^\zeta_2(T)d^\zeta_2(T)\}
\]

subject to

\[
\xi \leq \beta_1 + Kc^z_1(T) + Kc^z_2(T),
\]

\[
\xi \leq \beta_2 + c^\zeta_1(T) + c^\zeta_2(T).
\]

The dual linear program is given by

\[
\min \beta_1 y_1 + \beta_2 y_2
\]

subject to

\[
\begin{bmatrix}
-K & 0 \\
-K & 0 \\
0 & -1 \\
0 & -1 \\
1 & 1
\end{bmatrix}
\begin{bmatrix}
y_1 \\
y_2
\end{bmatrix}
= \begin{bmatrix}
-d^z_1(T) \\
-d^z_2(T) \\
-d^\zeta_1(T) \\
-d^\zeta_2(T) \\
1
\end{bmatrix},
\]

\[
y \geq 0,
\]

whose constraints are equivalent to the following:

\[
d^z_1(T) = d^z_2(T) \text{ for } j = z, \zeta,
\]

\[
d^z_1(T)/K + d^z_1(T) = 1,
\]

\[
d^\zeta_i(T) \geq 0 \text{ for } i = 1, 2, \text{ } j = z, \zeta.
\]
Moreover, its objective is given by $\beta_1 d_1^2(T)/K + \beta_2 d_1^2(T)$, from which (63) follows.

Similarly,

$$M(t, p, s) = L^*(t, s, p),$$

where

$$L^*(t, s, p) = \sup_{x, y} \{ (s, p) \cdot (x, y) - L(t, x, y) \}$$

$$= \sup_{x, t, s, y} \left\{ x^2 s_1^2 + x_2^2 s_2^2 + x_1^2 s_1^2 + x_2^2 s_2^2 + y^2 \text{t}^2 + y^2 \eta^2 + y^2 \right\}$$

$$- \sum_{i=1}^2 [(c_i - v_i(y))y_i^2 + \frac{1}{2} \eta_i^2(y_i^2)]$$

$$+ \sum_{i=1}^2 [(\bar{c}_i - \bar{v}_i(y))y_i^2 - \frac{1}{2} \eta_i^2(y_i^2) : y_i^2 \geq 0, y_i^2 \geq 0],$$

It is clear that we must have $s \equiv 0$ (and that $x^z, x^\zeta$ can take any value). Then

$$L^*(t, s, p) = \chi_{\{s=0\}} + \sup_{y^z, y^\zeta} \left\{ y^z_1 p_1^z + y^z_2 p_2^z + y^\zeta_1 p_1^\zeta + y^\zeta_2 p_2^\zeta + (r_1(y^z_1 - \tilde{r}_1(y^z_1))p_1^z$$

$$+ (r_2(y^z_2 - \tilde{r}_2(y^z_2))p_2^z - \sum_{i=1}^2 [(c_i(y) - v_i(y))y_i^2 + \frac{1}{2} \eta_i^2(y_i^2)]$$

$$+ \sum_{i=1}^2 [(\bar{c}_i(y) - \bar{v}_i(y))y_i^2 - \frac{1}{2} \eta_i^2(y_i^2) : y_i^2, y_i^2 \geq 0] \right\},$$

where we substituted $y^\zeta_i = r_i(y^z_1 - \tilde{r}_i(y^z_1))$. Notice that the optimization problem on the right-hand side decomposes so that

$$L^*(t, s, p) = \chi_{\{s=0\}} + \sum_{i=1}^2 \sup_{y^z_i} \left\{ y^z_i (p^z_i + r_i(y^z_1 - \tilde{r}_i(y^z_1)) - c_i(y) + v_i(y)) - \frac{1}{2} \eta_i^2(y_i^2) : y_i^2 \geq 0 \right\}$$

$$+ \sum_{i=1}^2 \sup_{y^\zeta_i} \left\{ y^\zeta_i (p^\zeta_i - \tilde{r}_i(y^z_1 - \tilde{r}_i(y^z_1)) + \bar{c}_i(y) - \tilde{v}_i(y)) - \frac{1}{2} \eta_i^2(y_i^2) : y_i^2 \geq 0 \right\}. \quad (66)$$

Then the first order conditions give

$$y_i^z = \frac{[p_i^z + r_i(y^z_1 - \tilde{r}_i(y^z_1)]^+}{\eta_i^z}, \quad (67)$$

$$y_i^\zeta = \frac{[p_i^\zeta - \tilde{r}_i(y^z_1 - \tilde{r}_i(y^z_1)]^+}{\eta_i^\zeta}. \quad (68)$$

Substituting these into (66) and using the definition $M(t, p, s) = L^*(t, s, p)$ gives (64). Moreover, (65) follows from (67)-(68), and the fact that $y_i^\zeta = r_i(y^z_1 - \tilde{r}_i(y^z_1))y_i^\zeta$ and that $x$ can take any value because $s \equiv 0$. 

•
Then combining Proposition 6 with (62) gives the dual formulation: Choose \( p(\cdot), \dot{p}(\cdot) \) so as to
\[
\min \beta_1 p_1^z(T)/K + \beta_2 p_1^\zeta(T) + \sum_{i=1}^{2} \int_{0}^{T} \frac{1}{2\eta_i^z} \left( \left[ p_i^z(t) + r_i(t) p_i^z(t) + v_i(t) - c_i(t) \right]^+ \right)^2 dt \\
+ \sum_{i=1}^{2} \int_{0}^{T} \frac{1}{2\eta_i^\zeta} \left( \left[ p_i^\zeta(t) - \tilde{r}_i(t) p_i^\zeta(t) - \tilde{v}_i(t) + \tilde{c}_i(t) \right]^+ \right)^2 dt
\]
(69)
subject to
\[
p_j^1(T) = p_j^2(T) \text{ for } j = z, \zeta, \]
(70)
\[
p_i^\zeta(T) + K p_i^1(T) = K, \]
(71)
\[
p_i^1(T) \geq 0 \text{ for } i = 1, 2, \ j = z, \zeta, \]
(72)
\[
p_i^\zeta(T) = 0 \text{ for } i = 1, 2. \]
(73)
\[
\dot{p}_i(t) = 0 \text{ for } i = 1, 2 \text{ and } t \geq 0. \]
(74)

Note that \( p(t) = p(T) \) for all \( t \in [0, T] \) because \( \dot{p}(t) = 0 \). Then letting \( q = p_1^z(T) \), using the constraint that \( p_1^z(T) + K p_1^\zeta(T) = K \), the hospice manager’s problem reduces to
\[
\min \beta_1 (1 - q) + \beta_2 q + \sum_{i=1}^{2} \int_{0}^{T} \frac{1}{2\eta_i^z} \left( \left[ K(1 - q) + r_i(t) q + v_i(t) - c_i(t) \right]^+ \right)^2 dt \\
+ \sum_{i=1}^{2} \int_{0}^{T} \frac{1}{2\eta_i^\zeta} \left( \left[ -\tilde{r}_i(t) q - \tilde{v}_i(t) + \tilde{c}_i(t) \right]^+ \right)^2 dt
\]
subject to \( 0 \leq q \leq 1 \).

**Proposition 7** The optimal solution \( q^* \) of the dual formulation is given as follows:
\[
q^* = \begin{cases} 
0 & \text{if } F(0) \geq 0, \\
F^{-1}(0) & \text{if } F(0) < 0 < F(1), \\
1 & \text{if } F(1) \leq 0, 
\end{cases}
\]
(75)

**Proof.** Letting
\[
G(q) = \beta_1 (1 - q) + \beta_2 q + \sum_{i=1}^{2} \int_{0}^{T} \frac{1}{2\eta_i^z} \left( \left[ K(1 - q) + r_i(t) q + v_i(t) - c_i(t) \right]^+ \right)^2 dt \\
+ \sum_{i=1}^{2} \int_{0}^{T} \frac{1}{2\eta_i^\zeta} \left( \left[ -\tilde{r}_i(t) q - \tilde{v}_i(t) + \tilde{c}_i(t) \right]^+ \right)^2 dt,
\]
the dual formulation can be written as follows:
\[
\min G(q) \text{ subject to } 0 \leq q \leq 1. \]
(76)
Adopting the trigger functions constructed for the proof of Proposition 3, we can write

\[ G(q) = \beta_1(1 - q) + \beta_2 q + \sum_{i=1}^{2} \sum_{j=1}^{N} \int_{L_i(q)}^{r_i(q)} \frac{1}{2\eta_i} (K(1 - q) + r_i(t)q + v_i(t) - c_i(t))^2 dt \]

\[ + \sum_{i=1}^{2} \sum_{j=1}^{N} \int_{L_i(q)}^{r_i(q)} \frac{1}{2\eta_i} (-r_i(t)q - \tilde{v}_i(t) + \tilde{c}_i(t))^2 dt, \]

from which it is straightforward to show that \( G'(q) = F \). This in turn yields \( G''(q) = F''(q) > 0 \) for all \( q \). Hence, \( G \) is strictly convex and the first order conditions are sufficient to pin down the unique optimal solution. Moreover, because \( F \) is strictly increasing, it follows that if \( F(0) \geq 0 \), then \( q^* = 0 \); and if \( F(1) \leq 0 \), then \( q^* = 1 \). Otherwise, we have an interior solution characterized by \( F(q) = 0 \), which yields \( q^* = F^{-1}(0) \).

\[ \blacksquare \]

C  Heuristic for the Simulation

This appendix describes the heuristic implemented in the simulation study. We wish to determine the total recruiting for type \( i \) (which will divided proportionally between types \( a \) and \( b \)) and the live-discharge rate for type \( i, b \), for \( i = 1, 2 \). From the fluid model we can calculate \( r_i^m(j) \), \( c_i^m(j) \), and \( v_i^m(j) \) for \( i = 1, 2 \), \( m = a, b \), and \( 1 \leq j \leq 365 \). Further, we define average class revenues and terminal values as \( \hat{r}_i(j) = \gamma^a_i r_i^a(j) + \gamma^b_i r_i^b(j) \) and \( \hat{v}_i(j) = \gamma^a_i v_i^a(j) + \gamma^b_i v_i^b(j) \) (recall that \( \gamma^a_i \) is the proportion of type \( a \) arrivals to class \( i \) and \( \gamma^b_i = 1 - \gamma^a_i \)).

Suppose that the current day is the start of day \( j \), \( 1 \leq j \leq 365 \). Define \( R(j) \) as the potential revenue earned thus far up to day \( j \) and let \( n_i(j) \) be the patients admitted thus far of type \( i = 1, 2 \). For \( k = j, \ldots, 365 \), let \( x_i^k \) be the recruiting rate for class \( i \) in period \( k \) and \( y_i^k \) the live-discharge rate for class \( i, b \), for \( i = 1, 2 \). Then we estimate cumulative potential revenue for the year as

\[ R = R(j) + \sum_{i=1}^{2} \sum_{k=j}^{365} [(\lambda_i + x_i^k)\hat{r}_i(k)] - \sum_{i=1}^{2} \sum_{k=j}^{365} [\hat{v}_i(k)y_i^k]. \]

We estimate the cumulative cap for the year as

\[ \text{CAP} = K \sum_{i=1}^{2} \left( n_i(j) + \sum_{k=j}^{365} (\lambda_i + x_i^k) \right). \]

Our mathematical program is to:

\[ \max \left\{ \min(R, \text{CAP}) + \sum_{i=1}^{2} \sum_{k=j}^{365} \left( (\hat{v}_i(k) - \hat{c}_i(k))(\lambda_i + x_i^k) - (v_i^b(k) - c_i^b(k))y_i^k \right) \right. \]

\[ \left. - \sum_{i=1}^{2} s_i(x_i^k) - \sum_{i=1}^{2} g_i(y_i^k) \right\} \]

(77)
subject to $x_i^k \geq 0, y_i^k \geq 0, i = 1, 2$.

This optimization can be solved analogously to that in Proposition 5 and pseudo-code for the heuristic is given below. First define:

- $\text{DayNumber}$ as the index of the current day.
- $\text{NatArrival[DayNumber]}$ as the expected revenue from natural arrivals from day $\text{DayNumber}$ to the end of the year.
- $\text{patientlist[i]->length}$ as the number of patients of type $i$ currently in the system.
- $w^m(j) = \alpha^m_i (j) - \alpha^m_i(j) + \hat{w}_i(j) = \gamma^m_i w^a_i(j) + \gamma^m_i w^b_i(j)$
- $\text{Item[i].Eta} = \eta_i^s$ and $\text{Item[i].EtaDis} = \eta_i^l$
- $\text{HeurDenomPart1[j]} = \sum_{k=j}^{356} (\hat{r}(j) - \text{CAP})^2 / \eta_i^s$
- $\text{HeurNumerPart1[j]} = \sum_{k=j}^{356} (\hat{r}(j) - \text{CAP}) / \eta_i^s$
- $\text{HeurDenomPart2a[j]} = \sum_{k=j}^{356} (\hat{r}(j))^2 / \eta_i^l$
- $\text{HeurNumerPart2a[j]} = \sum_{k=j}^{356} (\hat{r}(j) \hat{r}(j)) / \eta_i^l$
- $\text{HeurDenomPart2b[j]} = \sum_{k=j}^{356} \hat{w}(j) (\hat{r}(j) - \text{CAP})^2 / \eta_i^s$
- $\text{HeurNumerPart2b[j]} = \sum_{k=j}^{356} (\hat{r}(j) - \text{CAP}) / \eta_i^s + (\hat{w}(j) \hat{r}(j)) / \eta_i^s$

Then the discharge and recruiting rates are calculated using the following pseudo-code:

```plaintext
EstRev = YearsRevenue+NatArrival[DayNumber];
for (i=0; i<4; ++i) EstRev += r[i][DayNumber]*(patientlist[i]->length);
CapMiss = EstRev-CAP*NumAdmitTotal;
lagrange1 = -(CapMiss + HeurNumerPart1[DayNumber]+HeurNumerPart2a[DayNumber])/
           (HeurDenomPart1[DayNumber]+HeurDenomPart2a[DayNumber]);
lagrange2 = -(CapMiss + HeurNumerPart1[DayNumber]+HeurNumerPart2a[DayNumber])/
           (HeurDenomPart1[DayNumber]+HeurDenomPart2b[DayNumber]);
for (class=0;class<2;++class)
  if ((CapMiss+HeurNumerPart1[DayNumber] > 0)&&(lagrange1>-1))
    if (lagrange1<((c[1]-v[1])/(r-1))
      recruit = (lagrange1*(\hat{r}[class][DayNumber]-\text{CAP}) + \hat{w}[class][DayNumber])/(Item[class].Eta); discharge = (-lagrange1*r[class+2][DayNumber]- w[class+2][DayNumber])/(Item[class].EtaDis);
    else if (lagrange2<((c[0]-v[0])/(r-1))
      recruit = (lagrange2*(\hat{r}[class][DayNumber]-\text{CAP}) + \hat{w}[class][DayNumber])/(Item[class].Eta); discharge = (-lagrange2*r[0][DayNumber]- w[0][DayNumber])/(Item[0].EtaDis);
  if (class==0)
```

48
\[
\text{discharge} = \frac{(-\text{lagrange2}\times r[\text{class+2}][\text{DayNumber}] - w[\text{class+2}][\text{DayNumber}])}{\text{Item[\text{class}].EtaDis}};
\]
else discharge = 0;
else
\[
\text{lagrange3} = \frac{-\text{CapMiss} + \text{HeurNumerPart1[DayNumber]}}{\text{HeurDenomPart1[DayNumber]}},
\]
\[
\text{recruit} = \frac{(\text{lagrange3}\times (\hat{r}[\text{class}][\text{DayNumber}] - \text{CAP}) + \hat{w}[\text{class}][\text{DayNumber}])}{\text{Item[\text{class}].Eta}};
\]
discharge = 0;
else if ((\text{CapMiss} + \text{HeurNumerPart1[DayNumber]} > 0) && (\text{lagrange1} \leq -1))
\[
\text{recruit} = \frac{(\text{CAP} + \hat{w}[\text{class}][\text{DayNumber}] - \hat{r}[\text{class}][\text{DayNumber}])}{\text{Item[\text{class}].Eta}};
\]
discharge = \frac{\text{r[\text{class+2}][\text{DayNumber}] - w[\text{class+2}][\text{DayNumber}])}{\text{Item[\text{class}].EtaDis}};
else
\[
\text{recruit} = \frac{\hat{w}[\text{class}][\text{DayNumber}]}{\text{Item[\text{class}].Eta}};
\]
discharge = 0;

\text{RecruitingRate[i] = recruit;}
\text{DischargeRate[i] = discharge;}

We also create an equivalent heuristic for the legacy policy. This is done by setting the values in (77) and the associated code as follows: \( r^m_i(j) = rm^2_i, \) \( c^m_i(j) = cm^2_i, \) and \( v^m_i(j) = 0 \) for \( i = 1, 2, \) \( m = a, b, \)

Recall that \( \psi \) yields terminal values \( v^m_i = (\psi r - c_i)m^2_i, i = 1, 2 \) and \( j = a, b. \) We estimate a base-case for \( \psi \) by using a steady-state version of the simulation. We assume that \( \psi \) is uniform across all patient classes and is calculated as the estimated fraction of revenue for patients that end one year that will be of use the following year. In order to calculate this, at the end of each year we calculate if there was potential reimbursement left unused at the end of the year, and if so let \( \psi = 1. \) Otherwise, we calculate the extra revenue received this year (including from patients who were here at the start) that went unreimbursed and let \( \psi \) equal the ratio of that unreimbursed amount to total revenue.

To be specific, terminal values are estimated as follows (note that this calculation only makes sense in a steady-state world). At the end of the year we calculate:

- \text{CapMiss = YearsRevenue - CAP\times NumAdmitTotal} (this is the extra revenue received THIS year (including from patients who were here at the start) that went to waste or, if negative, the amount under the CAP that we ended the year with);
• $\text{EstRev}$ is the projected revenue to be received by the patients left at the end of the year (using their actual death times, although that won’t be available to a planner), with each class contributing $\text{QueueRev}[i]$;

• $\text{QueueCost}[i]$ is the projected cost to be incurred by the patients left at the end of the year from class $i$.

Finally, $\text{fracallocate}$ is calculated as the estimated fraction of revenue for patients that end one year that will be of use the following year (using this year’s value of $\text{CapMiss}$) so that

\[
\text{If } \text{CapMiss} \leq 0 \text{ then } \text{fracallocate}=1; \text{ else } \\
\text{if } \text{CapMiss}-\text{EstRev} \geq 0 \text{ then } \text{fracallocate}=0; \text{ else } \\
\text{fracallocate} = \frac{\text{CapMiss}}{\text{EstRev}}.
\]

The estimated terminal value for class $i$ is then

\[
(\text{fracallocate} \times \text{QueueRev}[i] - \text{QueueCost}[i]) / \text{NumPatients}[i]
\]

where $\text{NumPatients}[i]$ is the number of class $i$ left at the end of the year and $\text{fracallocate}$ becomes our estimate for $\psi$. Using the base case parameter values, this yields an estimated $\psi$ of 0.15.

Because the number of customers left at the end of the year, and hence the calculated terminal values, is likely to be quite variable we update the terminal values using an exponential smoothing type algorithm. In particular, if $\text{ALPHA}$ is the exponential smoothing parameter (set equal to 0.1 in our tests) then we calculate the new terminal value as

\[
\text{ALPHA} \times (\text{fracallocate} \times \text{QueueRev}[i] - \text{QueueCost}[i]) / \text{NumPatients}[i] + (1-\text{ALPHA}) \times \text{Terminal}[i]
\]

where $\text{Terminal}[i]$ is the previous terminal value for class $i$ patients.

D References


