Channel Management in Outpatient Care: Implications of Telemedicine and Transportation Support

Nan Liu[1], Shan Wang[2], Noa Zychlinski[3]
[1] Carroll School of Management, Boston College, Chestnut Hill, MA 02467, USA
[2] School of Business, Sun Yat-sen University, Guangzhou, 510275, China
[3] Faculty of Data and Decision Sciences, Technion – Israel Institute of Technology, Haifa 3200003, Israel
nan.liu@bc.edu, wangsh337@mail.sysu.edu.cn, noazy@technion.ac.il

The COVID-19 pandemic accelerated telemedicine adoption, offering a convenient alternative to in-person care. However, televisits may not fully address health concerns and sometimes require supplementary in-person visits, consuming resources that could have been saved if the initial visit had been in-person. As the pandemic subsides, in-person visits are regaining popularity, prompting providers to reorient resources toward in-person care. Transportation support (or subsidies) for patients, funded by providers or the government, plays a critical role in facilitating in-person visits. In this evolving landscape of telemedicine, we study how an outpatient care provider can optimally balance virtual and in-person services and whether, and how, to engage with transportation subsidies. We develop a stylized queueing-game model to represent the operations of a revenue-maximizing provider serving patients who strategically choose between service channels. We find that provider size, measured by total capacity relative to demand, is key. Small and large providers perform best by focusing on one channel without offering subsidies, whereas medium-sized providers benefit from carefully balancing both channels alongside subsidies. Paradoxically, transportation subsidies, which make in-person care more accessible, may reduce overall patient access to care, even when fully funded by the government. This occurs because providers may shift capacity toward a higher-reimbursement channel, ultimately serving fewer patients. Differentiating payment rates between in-person and virtual visits can potentially prevent such reductions. Our study highlights the importance of capacity coordination between channels for providers and cautions policymakers that transportation support may unintentionally harm patient access. Properly designed financial incentives can help prevent such negative outcomes.

Key words: healthcare operations management, telemedicine, queueing game, capacity management

---

1. Introduction

The COVID-19 pandemic accelerated the adoption of virtual services in general and telemedicine in the context of healthcare services in particular (Bokolo 2020). Telemedicine enables the provision of remote clinical services through real-time communication between patients and healthcare providers, utilizing tools like video conferencing and remote monitoring (Monaghesh and Hajizadeh 2020). Virtual services, which reduce travel costs and are associated with lower waiting costs for patients, can improve the efficiency of care delivery and expand access to care (Wong et al. 2021). However, they also pose a risk of delivering low-value care (O'Reilly-Jacob et al. 2021). Research has shown that a virtual visit may lead to a supplementary in-person visit for the same medical concern within a short timeframe, potentially due to misdiagnosis or inadequate treatment (Shi et al. 2018). Consequently, telemedicine may increase follow-up care and hurt provider productivity (Bavafa et al. 2018, Li et al. 2021). The evidences, however, are not always consistent. Other studies indicate that the increase in supplementary in-person visits may be minimal (Reed et al. 2021).

Despite these mixed results, telemedicine is expected to continue growing beyond the COVID-19 era, although it cannot fully replace in-person visits (Rosenthal 2021). In-person care provides benefits that virtual care cannot replicate, such as better communication and a stronger doctor-patient relationship. Moreover, certain diagnosis and tests can only be conducted in person.

During the pandemic, payment parity was introduced to encourage the adoption of telemedicine to expand care access, by reimbursing televisits at the same rate as in-person visits. With the pandemic now over, the downsides of telemedicine have drawn increased attention. A prevailing view is that reimbursing virtual visits—typically involving fewer diagnostic services—at the same rate as in-person visits represents over-payment. This has shifted policy discussions toward payment equity rather than parity in the post-pandemic era (Shachar et al. 2020). Meanwhile, in-person visits are regaining popularity among both patients and providers.

In response to these shifts, providers have started re-orienting their resources toward in-person visits. A 2022 McKinsey report highlights that as the pandemic abates, more physicians are gravitating away from virtual care and would prefer a return to in-person care (Cordine et al. 2022). There has been a 13% increase in physicians recommending in-person visits over telehealth and a 10% rise in physicians offering in-person care only since July 2020. Providers are making various efforts to encourage and support in-person visits, such as improving the in-clinic encounter experience, investing in a better physical service environment, and offering transportation support.

Transportation support, in particular, plays a critical role in facilitating in-person visits, especially for rural populations and those facing transportation challenges. It can be funded by the government. For example, federal Medicaid regulations (42 CFR 440.170) require that states ensure transportation for beneficiaries to and from providers (Medicaid and CHIP Payment and Access Commission 2019), a benefit known as non-emergency medical transportation (NEMT). Covered transportation modes include taxis, buses, vans, and personal vehicles. States have considerable flexibility in administering NEMT and may deliver the benefit via fee-for-service, managed care contracts, or transportation brokers.

Transportation may also be subsidized by providers through means such as travel reimbursement, shuttle services, coordinated shared ride programs, parking vouchers, or transit passes. For example, Family Health Services in Ohio and Kheir Clinic in Los Angeles both provide their own no-cost transportation services for patients; the Upper Allegheny Health System in southwestern New York and Bradford Regional Medical Center in northwestern Pennsylvania partner with the Area Transportation Authority of North Central Pennsylvania to provide free transportation services (Graham 2022). In this paper, we use the term "transportation subsidy" broadly to refer to any support provided to help patients attend in-person visits.

Our research is motivated by this changing landscape of telemedicine, in which all stakeholders are rethinking how best to integrate virtual care into the broader care delivery system. In particular, we seek to understand the interplay between telemedicine and transportation support, and its implications for care delivery. We focus on non-emergency outpatient care settings, and investigate how providers can optimally balance virtual and in-person services, and whether and how they should engage with transportation subsidies. Our study also sheds light on how provider operational choices impact access to care.

When seeking outpatient care, patients are typically presented with two options—coming in for an in-person visit or having a virtual visit. Patients weigh the utilities and costs associated with these alternatives before deciding. In-person visits require waiting in the clinic, incurring both time spent and potential travel costs (e.g., transportation and parking). Conversely, virtual visits offer the convenience of engaging from home or another preferred location by patients, thus reducing their waiting and travel costs. For some health conditions, a virtual visit can fully address patients' concerns. However, in other cases, a supplementary in-person visit (or returning visit) may be required if the initial virtual treatment is inadequate or symptoms do not improve. In such scenarios, patients need to subsequently visit the clinic, incurring the costs and disutilities initially avoided.

Facing strategic patient choice, an outpatient care provider needs to determine the optimal system design and capacity allocation between in-person and virtual channels. Whereas the virtual channel is appealing for its convenience to both providers and patients, it may lead to returning in-person visits, which come with additional cost and capacity utilization. To support in-person visits, mitigate the need for returning visits, and improve overall operational efficiency, transportation subsidies may be offered, as discussed above. Though offering transportation subsidies can be useful to improve a provider's operational efficiency and revenue, its impact on patient access to care is unclear. Intuitively, transportation subsidies should expand patient access to care because they make in-person visits more accessible, but does this always hold true?

To answer our research questions, we develop a stylized queueing-game model to represent the operations of an outpatient care provider, who has a fixed daily capacity to allocate between two channels. The provider faces an exogenous stream of strategic patient demand. Each patient evaluates the expected waiting in each channel and the likelihood of needing a returning in-person visit, with three options available: choose an in-person visit, wait for a virtual visit, or balk. Balking encapsulates other care alternatives the patient may choose, such as visiting other care providers or taking some at-home treatments.

If an in-person visit is chosen, the patient's expected utility is the service reward net the expected waiting cost and travel cost associated with the in-person channel. If a virtual visit is chosen, the patient's expected utility incorporates the expected waiting cost in the virtual channel and the expected utility of a returning visit, which will incur additional waiting and travel costs to attend the in-person visit. Lastly, if the patient balks, the utility is normalized to zero. Importantly, patient choice is endogenous to the provider's capacity allocation because waiting cost depends on service capacity allocation between the two channels. Factoring into patient strategic choice, the provider aims to maximize the total revenue across both service channels by carefully allocating her daily capacity.

We first show that for any given capacity allocation, there exists a unique mixed-strategy equilibrium in patient choice. With three alternatives available (virtual visit, in-person visit, or balking), there are seven equilibrium regions based on how patients mix their choices. We then fully characterize the optimal capacity allocation strategy for the provider to maximize her revenue. The optimization incorporates patient equilibrium, which is endogenous to the capacity allocation. We find that the optimal system configuration (i.e., how much capacity to allocate to each channel) depends on the size of the system, measured by the total available service capacity relative to the overall patient demand. In a small system, the provider should pool her capacity to focus on one service channel to maximize efficiency. For a medium-sized system, she has enough capacity to prevent patient balking, but needs to carefully balance capacity between the two channels. For a large system, because no patients would balk regardless of the capacity allocation, a revenue-maximizing provider should prioritize the channel with the highest revenue margin.

Next, we consider the setting with transportation subsidies. We differentiate two scenarios: subsidies paid by the provider and subsidies funded by the government. For the first case, we fully characterize the joint optimal decision for capacity allocation and subsidy levels. For the second case, because the subsidies incur no cost to the provider, the decision focuses solely on capacity allocation. It is natural to expect that the use of transportation subsidy increases total revenue, regardless of who pays for it. What is more interesting is that the optimal system design can fundamentally change with the subsidies offered. For example, a system optimally designed to focus on the virtual channel without subsidies may be optimized by changing to the in-person channel with subsidies, or vice versa. Intuitively, transportation subsidies should increase the total rate at which patients access the service across both channels because subsidies can attract more in-person visits, which do not require returning visits and thus conserve service resources. However, our results show this is not always the case: offering subsidies can paradoxically reduce the total access rate, even when funded entirely by the government at no cost to the provider!

Our numerical analysis, calibrated with real-world data and empirical studies, confirms this "backfire" phenomenon. The underlying cause is the potential shift of the provider's optimal strategy: she may choose to serve fewer patients at a higher payment rate, leading to a decrease in total access rate. To mitigate such an adverse impact on social welfare, one solution we identify is in line with the idea of "payment equity" discussed above. Specifically, we show that differentiating payment rates between in-person and virtual visits can prevent reductions in access rate, regardless of who funds transportation subsidies. Our results, at a high level, also support the current Centers for Medicare & Medicaid Services (CMS) policy that grants states flexibility in administering NEMT, as this flexibility allows states to tailor transportation support in a way that aligns with provider incentives and care delivery goals to help prevent unintended reductions in access to care.

The rest of the paper is organized as follows. Section 2 reviews the relevant literature. Section 3 presents the base model, analyzes patient equilibrium strategy, and introduces the provider's problem. Section 4 discusses the the optimal capacity allocation. Section 5 considers the option of providing transportation subsidies. Section 6 conducts a numerical analysis of the impact of transportation subsidies on patient access to care. Section 7 draws conclusions and suggests a few directions for future research. All proofs of the technical results are shown in the E-Companion.

2. Literature Review

This research draws upon the literature on healthcare operations management (OM) and queueing studies with strategic customers and multiple service access channels. We review each stream below.

2.1. Healthcare OM

Within the healthcare OM literature, our work is related to the research that (1) employs queueing models to investigate system design questions in outpatient care, and (2) addresses the multi-channel care setting, in particular those with one telemedicine channel. Two representative studies from the first stream of literature include Green and Savin (2008) and Zacharias and Armony (2016). They develop stylized queueing models to investigate system design questions such as panel size selection and capacity decisions. An important phenomenon addressed by these two works is that patients may revisit the provider after (no-shows from) initial visits. Though this patient "returning" feature also presents in our model, our patient revisits require a different source of capacity compared to their initial visits. Moreover, patient demand in Green and Savin (2008) and Zacharias and Armony (2016) are treated as exogenous variables, whereas we model patients as strategic customers, rendering patient demand endogenous to the provider's decisions.

Depending on the clinical setting, patients may access care through different channels, e.g., in-person appointment visits, walk-ins, virtual visits, or visits to different providers. These channels can be managed by a single provider or coordinated among/independently offered by multiple providers. The literature addressing OM issues in such multi-channel care settings is growing. Liu et al. (2023) examine capacity management between appointment-based and walk-in services within a single provider setting. In their model, only strategic walk-in patients may balk, and patients are assumed to fully recover regardless of the chosen channel. In multi-provider settings, Shumsky and Pinker (2003) introduce a classic gatekeeper model to study gatekeeper-specialist dynamics. A more recent work by Sharma et al. (2020) focuses on the interaction between emergency rooms and general providers. Both of these studies explore incentive design, aiming to align different providers who offer in-person service towards societal or firm-centric objectives. On a separate topic, Huang et al. (2022) study the doctor-shopping behavior (i.e., patients seek opinions from multiple doctors without referrals) and its impact on social welfare. In contrast to the latter three studies on multi-provider settings, our research concentrates on the operations of a single provider who offers both in-person and virtual services.

With the rise of telemedicine, the body of literature assessing its operational impact is expanding. Rajan et al. (2019) and Bavafa et al. (2021) study the use of telehealth visits in chronic care. The former finds that telemedicine can improve access to care and increase provider revenue, whereas the latter shows that e-visits may negatively impact panel size and patient health. Tele-triage is another form of telemedicine, aimed at remotely assessing patient conditions to determine the appropriate care they need. Çakıcı and Mills (2021) and Guan et al. (2025) evaluate the use of tele-triage, and discover that it may increase cost and exacerbate inefficiencies in a healthcare system. In contrast to these studies, we focus on how an outpatient care provider can effectively integrate virtual visits into her operations.

A few recent studies have explored a hybrid care delivery model similar to ours, involving in-person and virtual channels with returning visits (or, more generally, imperfect treatment). Focusing on the Primary Care First (PCF) initiative, Adida and Bravo (2023) analyze blends capitation and fee-for-service payments with performance-based adjustments to incentivize redesigned primary care delivery, including remote care. They find that PCF can achieve socially optimal outcomes by adjusting payment structures based on state-specific health heterogeneity.

The most relevant work to ours is Çakıcı and Mills (2024), which investigates the repercussions of telehealth reimbursement policies on the accessibility of acute care. They show that returning visits resulting from virtual care create an incentive alignment problem, and thus pay parity for telehealth may hurt patient access. Our study departs from theirs in at least two significant ways. First, our base model without transportation subsidies *fully* characterizes the provider's optimal capacity allocation decisions under *all* possible demand-capacity scenarios, whereas Çakıcı and Mills (2024) focuses only on a high-workload setting. Second, in contrast to their focus on incentive structures designed during the pandemic to promote telemedicine adoption, we examine issues that are increasingly salient in the post-pandemic era: specifically, whether a hybrid care delivery model combining in-person and virtual services improves patient access to care, particularly in the presence of transportation subsidies.

In addition, Zychlinski (2024) tackles scheduling and capacity allocation decisions across three service channels: in-person, virtual, and supplementary service for returning virtual patients. Zang et al. (2024) explore how improving patient understanding of their telemedicine suitability through online triage affects the performance of a hybrid system. Wang et al. (2019) study dual-channel care (in-person and virtual) where some patients receiving virtual care require returning in-person visits. Their model concerns two independent providers who set their capacity separately in a capacity-abundant environment, whereas we study how a single provider should operate under any capacity situation. Complementing these recent studies, our work contributes to a more comprehensive understanding of how to effectively manage hybrid healthcare systems that incorporate telemedicine.

2.2. Queueing Studies with Strategic Customers

In terms of methodology, our research leverages the analysis of queuing systems with strategic customers. This broad research area, initiated by Naor (1969), explores customer join/balk decisions based on how sensitive customers are to waiting and aims to optimize system efficiency/social welfare by adjusting the service capacity, pricing, or priority schemes. Hassin and Haviv (2003) provide a comprehensive review of this area. Below we draw attention to recent studies most relevant to our work.

Hassin and Roet-Green (2020) study the impact of queue-length information on the performance of a service system where customers must travel to join. In their model, customers can observe the queue length before deciding whether to travel. Another recent relevant paper that considers travel costs is by Baron et al. (2022). They examine an omni-channel service system that offers both walk-in and online channels. To receive the final service, customers incur travel costs regardless of the channel chosen. They show that although the online channel increases revenue, in equilibrium it reduces customer utility and social welfare. They propose prioritizing walk-in customers to mitigate these effects. In contrast to these two works, we consider a healthcare setting that offers a virtual service channel with no travel costs and an in-person channel that requires travel for both initial and returning visits. We focus on capacity rather than (scheduling) priority decisions.

2.3. Our Contribution

We conclude this section by briefly summarizing our contributions. We develop a modeling framework for an outpatient care provider who offers both in-person and virtual services. Our model captures the key trade-offs that strategic patients face when choosing between these two channels. A notable feature of our model is that it captures the difference in care quality between channels: patients who choose the virtual channel may need an in-person return visit. In the post-pandemic era, in-person visits are regaining popularity, and transportation support is one of the most effective ways to facilitate these visits, either offered by the provider or subsidized by the government. We also study how transportation support can be incorporated into provider operations and its broader impact.

In our analysis, we first prove the existence and uniqueness of a mixed strategy equilibrium for patient choice, based on which we characterize the optimal capacity management strategy for a revenue maximizing provider. We find that the size of the system, measured by the total available service capacity relative to the overall patient demand, plays a critical role in determining the optimal capacity allocation of the provider. Small and large systems are better off focusing on one channel and do not need to offer transportation support, whereas medium-sized systems can benefit from a careful balance between the two channels alongside transportation support. However, we caution that providing transportation support can backfire and reduce overall patient access to care, even when fully funded by the government at no cost to the provider. To prevent such outcomes, one solution we identify is to differentiate payment rates between in-person and virtual visits.

3. The Model

We consider an outpatient care provider offering two service channels: in person and virtual, with a fixed daily capacity of μ. The provider decides how to allocate the in-person capacity (μ_f) and the virtual capacity (μ_v) such that $\mu_f + \mu_v = \mu$. The actual service time of patients may have some variability, but with an allocation (μ_f, μ_v), the provider is expected to be able to serve μ_f in-person patients and μ_v virtual patients per day. In practice, outpatient care providers often reserve specific slots in their daily appointment book for in-person and virtual visits, respectively. Our capacity allocation model reflects this common approach.

Homogeneous strategic patients arrive following a Poisson process with daily rate Λ. Each patient, upon his[1] arrival, strategically chooses between in-person and virtual service channels, if both are offered by the provider. Some patients may need a supplementary in-person visit with the same provider after the virtual visit. This could be due to the challenges of virtual diagnosis (as with

[1] For convenience, we shall refer to the provider as "she" and a patient as "he" in the rest of this article.

some skin conditions), the need for additional tests before prescribing medication (e.g., urinary tract infection), or the lack of improvement in patient symptoms with at-home treatments (such as in cases of persistent sore throats). We assume that for each patient, the *ex ante* probability of requiring a supplementary visit is δ, and a smaller δ indicates greater effectiveness of virtual care. We also assume that in-person visits always resolve patient health concerns, so patients do not need to revisit the provider after an in-person visit for the same issue. Note that supplementary visits resulting from the ineffectiveness of virtual care are fundamentally different from follow-up in-person visits triggered by other health concerns, e.g., chronic conditions. We use the term "supplementary" to emphasize that this visit addresses the same episode of health issue that was not fully resolved by the initial virtual visit. In contrast, follow-up visits are not the result of an ineffective initial visit but are often required by clinical protocols. We assume that the demand for such follow-up visits depends solely on population characteristics and can therefore be incorporated into the daily demand rate Λ.

Moreover, patients may have a range of other care options, such as seeking care at emergency departments, considering urgent care service, or resorting to self-care at home (Khairat et al. 2021). To encapsulate all these options, we include "balking" as a third option within the patient's choice set. This approach allows us to capture the full range of patient decisions in response to the offered services. Patients make their choices based on the expected utilities of these options, which depend on the expected wait time to get service, the associated travel cost, if any, as well as the need for a supplementary visit.

Given that patients are *ex ante* homogeneous, we identify the equilibrium in the class of mixed and symmetric strategies. Denote any mixed strategy by (p_f, p_v, p_b), where p_f, p_v, and p_b, respectively, represent the probabilities that a random patient, upon his arrival, chooses the in-person channel, the virtual channel, and balking. Note that $p_f + p_v + p_b = 1$ because these three options are exhaustive. Hence, we can simplify the representation of the mixed strategy as (p_f, p_v). Let $\lambda_f = p_f\Lambda$ and $\lambda_v = p_v\Lambda$ denote the respective arrival rate to the in-person channel and virtual channel. Then, the pair (λ_f, λ_v) represents the effective arrival rate to each channel when a mixed strategy (p_f, p_v) is adopted. Given that Λ is fixed, we also use (λ_f, λ_v) to denote the patient strategy for notational simplicity.

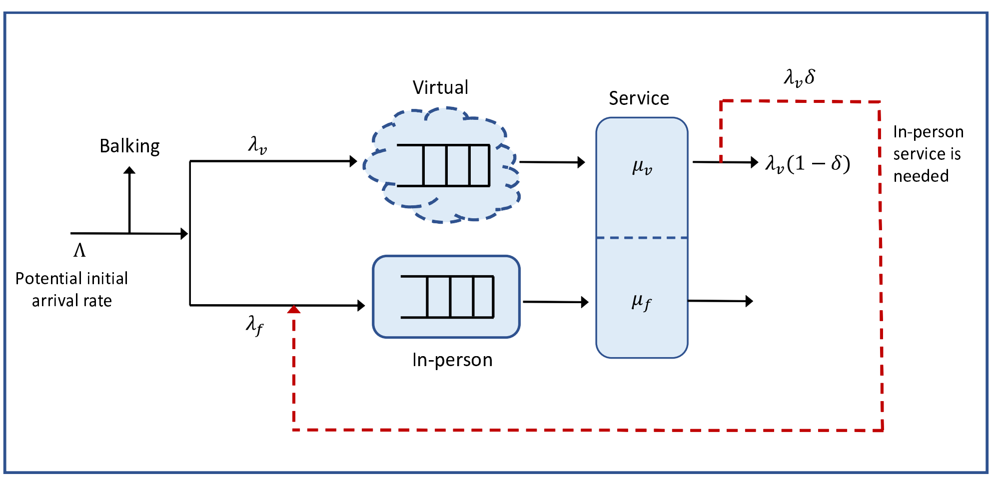
Our model can be extended to incorporate patient heterogeneity in travel cost and health needs (see Remark 1 and Remark 3, respectively). However, since patient heterogeneity does not alter the key insights, we focus on the current model with homogeneous patients for clarity and ease of discussion.

3.1. Patient's Utility

To analyze patient strategy, we start by examining the utility of each choice. We begin with the arrival–service process under any given strategy (λ_f, λ_v). Recall that the provider divides her daily capacity into in-person service slots and virtual slots, so that she has the capacity to serve μ_f in-person patients and μ_v virtual patients per day. In-person (virtual) slots are dedicated to in-person (virtual) patients only. Thus, one can view the in-person and virtual service channels as two independent "servers".

Inspired by prior literature employing stylized single-server queueing models to address strategic-level questions in outpatient care services (Green and Savin 2008, Liu and Ziya 2014, Zacharias and Armony 2016, Liu et al. 2023), we adopt a network of two stylized single-server queues to capture the evolution of system dynamics. When a patient requests a virtual visit, he is scheduled to the end of the virtual care queue (i.e., added to the virtual visit appointment backlog). Upon completion of virtual service, if a supplementary in-person visit is required, the patient is scheduled to the end of the in-person care queue. Similarly, a patient requesting an in-person visit is directly added to the end of the in-person care queue (i.e., to the in-person visit appointment backlog). In this framework, queue waiting encompasses both appointment delay and waiting in the actual or virtual service line for care. For tractability, and without losing key insights, we model both the arrival and service processes as Markovian, following Green and Savin (2008), Liu and Ziya (2014) and Liu (2016). Hence, this queueing network is a Jackson network (see Figure 1). In steady state, it functions as two M/M/1 queues, one representing the virtual channel and the other the in-person channel. The arrival and service rates are λ_v and μ_v, respectively, in the virtual queue, and $\lambda_f + \delta\lambda_v$ and μ_f, respectively, in the in-person queue.

Figure 1 **Illustration of the model featuring patients choosing among virtual visit, in-person visit or balking.**



Next, we quantify the expected utility associated with each option.

Cost of each in-person visit. If a patient chooses the in-person channel, he incurs a cost T, which represents the disutility associated with each in-person visit, such as inconvenience and cost due to transportation. For ease of discussion, We call T the travel cost. Since each queue in the Jackson network can be regarded as an M/M/1 queue, the expected waiting time in the system for each in-person visit is

$$W_f(\lambda_f, \lambda_v) = \frac{1}{\mu_f - (\lambda_f + \delta\lambda_v)}. \tag{1}$$

Let θ_f denote the waiting cost per unit of time in the in-person channel. Then, the expected total cost associated with each in-person visit is

$$C_f(\lambda_f, \lambda_v) = T + \theta_f W_f(\lambda_f, \lambda_v). \tag{2}$$

Cost of each virtual visit. The expected waiting time in the system for the virtual channel is

$$W_v(\lambda_v) = \frac{1}{\mu_v - \lambda_v}. \tag{3}$$

Let θ_v denote the waiting cost per unit of time for patients in this channel. Since patients choosing the in-person channel must commute to and stay in the clinic, whereas patients on the virtual channel can see providers in a more comfortable environment such as their homes, it is natural to impose that $\theta_f \geq \theta_v$.

Though the virtual channel incurs no travel cost and has a lower waiting cost rate, patients starting with the virtual channel may need a supplementary in-person visit, with all the additional disutility associated with it. The expected total cost of each virtual visit (followed by the necessary in-person visit) is, therefore,

$$C_v(\lambda_f, \lambda_v) = \theta_v W_v(\lambda_v) + \delta C_f(\lambda_f, \lambda_v). \tag{4}$$

Utility of each option. Patients receive a service reward R once their health concerns are addressed. Without loss of generality, we normalize the utility of balking to be 0. Let $U_f(\cdot)$, $U_v(\cdot)$, and $U_b(\cdot)$ represent the expected utilities of choosing in-person visit, virtual visit, and balking, respectively. Then, by (1)–(4), the utility for each alternative is:

$$U_f(\lambda_f, \lambda_v) = R - T - \frac{\theta_f}{\mu_f - (\lambda_f + \delta\lambda_v)}, \tag{5}$$

$$U_v(\lambda_f, \lambda_v) = R - \frac{\theta_v}{\mu_v - \lambda_v} - \delta T - \frac{\delta\theta_f}{\mu_f - (\lambda_f + \delta\lambda_v)}, \tag{6}$$

$$U_b(\lambda_f, \lambda_v) = 0.$$

3.2. Patient Equilibrium Strategy

Given (μ_f, μ_v), a strategy (λ_f, λ_v) is a symmetric equilibrium strategy if it constitutes the best response to itself. In other words, no patient can unilaterally increase his own utility by deviating from this strategy. More formally, the equilibrium effective arrival rates (λ_f, λ_v) must satisfy the following condition:

CONDITION 1. For $x = f, v$ or b, if $\lambda_x > 0$, then we have

$$U_x(\lambda_f, \lambda_v) = \max\{U_f(\lambda_f, \lambda_v), U_v(\lambda_f, \lambda_v), U_b(\lambda_f, \lambda_v)\}.$$

According to Condition 1, Table 1 provides a summary of the different strategies along with the conditions on their corresponding effective arrival rates and utilities. To illustrate various strategies, we use the following symbols: B for the pure strategy of balking, V for the pure strategy of choosing the virtual channel, and F for the pure strategy of selecting the face-to-face channel (i.e., the in-person channel). The mixed strategies, combining these choices, are denoted by the letter combinations of B, V, and F. Altogether, there are seven distinct types of strategies: B, V, F, BVF, BF, BV and VF.

Table 1 Equilibrium Strategies

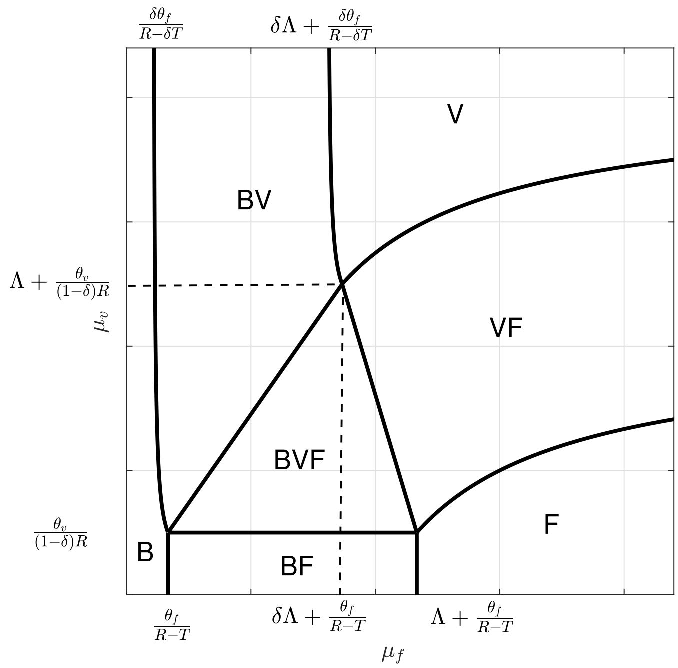
Strategy	Effective Arrival Rates	Utilities
B	$\lambda_f = 0, \quad \lambda_v = 0$	$U_f(0,0) \leq 0, \quad U_v(0,0) \leq 0$
V	$\lambda_f = 0, \quad \lambda_v = \Lambda$	$U_f(0,\Lambda) \leq U_v(0,\Lambda), \quad U_v(0,\Lambda) \geq 0$
F	$\lambda_f = \Lambda, \quad \lambda_v = 0$	$U_f(\Lambda,0) \geq U_v(\Lambda,0), \quad U_f(\Lambda,0) \geq 0$
BVF	$\lambda_f + \lambda_v \leq \Lambda$	$U_f(\lambda_f,\lambda_v) = U_v(\lambda_f,\lambda_v) = 0$
BF	$\lambda_f \leq \Lambda, \quad \lambda_v = 0$	$U_f(\lambda_f,0) = 0, \quad U_v(\lambda_f,0) \leq 0$
BV	$\lambda_f = 0, \quad \lambda_v \leq \Lambda$	$U_f(0,\lambda_v) \leq 0, \quad U_v(0,\lambda_v) = 0$
VF	$\lambda_f + \lambda_v = \Lambda$	$U_f(\lambda_f,\lambda_v) = U_v(\lambda_f,\lambda_v) \geq 0$

Following Table 1, we derive the explicit conditions on (μ_f, μ_v) that determine which strategy is adopted. These conditions define seven regions for service capacity allocation, namely Region B, V, F, BVF, BF, BV and VF. For instance, Region BVF encompasses capacity allocation (μ_f, μ_v) where patients choose the BVF strategy, and is described by the following inequalities:

$$\begin{aligned}
&\mu_v \geq \frac{\theta_v}{(1-\delta)R}, \\
&\mu_f - \delta\mu_v \geq \frac{\theta_f}{R-T} - \frac{\delta\theta_v}{(1-\delta)R}, \\
&\mu_f + (1-\delta)\mu_v \leq \Lambda + \frac{\theta_f}{R-T} + \frac{\theta_v}{R}.
\end{aligned}$$

The specific closed-form definitions for all regions can be found in Definition EC.1 of E-Companion EC.1.1. With the x-axis representing μ_f and the y-axis representing μ_v, Figure 2 depicts these seven regions.

Figure 2 **Illustration of the seven equilibrium regions.**



Along with Definition EC.1, Theorem 1 formally establishes the existence and uniqueness of the equilibrium strategy for patients, given any service capacity allocation (μ_f, μ_v).

Theorem 1 **(The Existence and Uniqueness of Patient Equilibrium)**. *The seven regions of (μ_f, μ_v) defined in Definition EC.1 are mutually exclusive and collectively exhaustive. And in each region, there exists a unique equilibrium strategy.*

1. *When $(\mu_f, \mu_v) \in$ Region B, there exists a unique equilibrium strategy – strategy B – such that $\lambda_f = \lambda_v = 0$.*
2. *When $(\mu_f, \mu_v) \in$ Region V, there exists a unique equilibrium strategy – strategy V – such that $\lambda_f = 0$ and $\lambda_v = \Lambda$.*
3. *When $(\mu_f, \mu_v) \in$ Region F, there exists a unique equilibrium strategy – strategy F – such that $\lambda_f = \Lambda$ and $\lambda_v = 0$.*
4. *When $(\mu_f, \mu_v) \in$ Region BVF, there exists a unique equilibrium strategy – strategy BVF – such that $\lambda_f = \mu_f - \delta\mu_v - \theta_f[R-T]^{-1} + \delta\theta_v[(1-\delta)R]^{-1}$ and $\lambda_v = \mu_v - \theta_v[(1-\delta)R]^{-1}$.*
5. *When $(\mu_f, \mu_v) \in$ Region BF, there exists a unique equilibrium strategy – strategy BF – such that $\lambda_f = \mu_f - \theta_f[R-T]^{-1}$ and $\lambda_v = 0$.*
6. *When $(\mu_f, \mu_v) \in$ Region BV, there exists a unique equilibrium strategy – strategy BV – such that $\lambda_f = 0$ and $\lambda_v = (\mathcal{B} - \sqrt{\mathcal{D}^2 + 4\delta^2\theta_f\theta_v})[2\mathcal{A}]^{-1}$, where $\mathcal{A} = (R-\delta T)\delta$, $\mathcal{B} = (R-\delta T)(\mu_f + \delta\mu_v) - \delta(\theta_f + \theta_v)$, and $\mathcal{D} = (R-\delta T)(\mu_f - \delta\mu_v) - \delta(\theta_f - \theta_v)$.*

7. *When* $(\mu_f, \mu_v) \in$ *Region* VF*, there exists a unique equilibrium strategy – strategy VF – such that* $\lambda_f = \Lambda - (-\tilde{\mathcal{B}} + \sqrt{\tilde{\mathcal{D}}^2 + 4\theta_f\theta_v})[2\tilde{\mathcal{A}}]^{-1}$ *and* $\lambda_v = (-\tilde{\mathcal{B}} + \sqrt{\tilde{\mathcal{D}}^2 + 4\theta_f\theta_v})[2\tilde{\mathcal{A}}]^{-1}$*, where* $\tilde{\mathcal{A}} = (1-\delta)T$*,* $\tilde{\mathcal{B}} = (\theta_f + \theta_v) + T(\mu_f - \Lambda - (1-\delta)\mu_v)$*, and* $\tilde{\mathcal{D}} = (\theta_f - \theta_v) + T(\mu_f - \Lambda + (1-\delta)\mu_v)$*.*

Remark 1. In the current model, all patients have the same travel cost T. The model can be extended to incorporate heterogeneous (uniformly distributed) travel costs. The resulting patient equilibrium has a similar structure to that described in Theorem 1; see E-Companion EC.2 for details.

In equilibrium, when μ_f is small, no patient will choose the in-person channel; when μ_v is small, no patient will choose the virtual channel. When both μ_f and μ_v are small, all patients will balk. It is worth noting that μ_f impacts the utilities of both channels—in-person and virtual. In particular, when μ_f is small, those who choose virtual visits can still suffer because some of them need a supplementary in-person visit for which they will have to wait a long time and, accordingly, have low utility. Hence, when μ_f is small, regardless of how large μ_v is, some patients will balk. When μ_f is large, however, regardless of how small μ_v is, no patient will balk. To the best of our knowledge, Theorem 1 is the first in the literature to *fully* characterize the patient equilibrium strategy across *all* possible provider capacity allocations in this type of model. This characterization is essential for deriving the global optimal solution to the subsequent provider's problem.

3.3. The Provider's Problem

Considering patient strategic behavior, the provider aims to maximize her expected daily rewards by choosing a capacity allocation (μ_f, μ_v) such that $\mu_f + \mu_v = \mu$. The provider receives payment r_f and r_v for each patient who chooses the in-person channel and the virtual channel, respectively. The provider's problem can be formally defined as follows.

$$\begin{aligned} \max_{\mu_f,\ \mu_v \geq 0} \quad & r_f\lambda_f + r_v\lambda_v \\ s.t. \quad & \mu_f + \mu_v = \mu, \\ & (\lambda_f, \lambda_v) \text{ is defined in Theorem 1.} \end{aligned} \tag{P}$$

For any given capacity allocation (μ_f, μ_v) in the feasible solution space, Theorem 1 confirms that there exists a unique patient equilibrium, and hence Problem (P) is well defined.

Remark 2. The provider's revenue structure in our model is quite flexible and can capture commonly-adopted reimbursement regimes in healthcare. The pair (r_f, r_v) naturally represents a bundle payment mechanism because r_v, covers the whole episode of care, which may involve a supplementary in-person visit. Under the fee-for-service payment mechanism, where r_v^{FFS} and r_f^{FFS} represent the payments for each virtual and in-person visit respectively, the model can still be applied by setting the provider's rewards (r_f, r_v) as $(r_f^{\text{FFS}}, r_v^{\text{FFS}} + \delta r_f^{\text{FFS}})$.

4. Optimal Capacity Allocation

In this section, we investigate the optimal capacity allocation in problem (P). Recall that Definition EC.1 and Theorem 1 partition the solution space for capacity allocation (μ_f, μ_v) into seven regions. We start by exploring the optimal capacity allocation in each region. Specifically, for any Region X defined in Definition EC.1, we incorporate the constraints of (μ_f, μ_v) into Problem (P), where X=B, V, F, BVF, BF, BV or VF. We denote the optimal solution and the corresponding effective arrival rates subject to Region X, if they exist, as $\left(\mu_f^{X*}, \mu_v^{X*}\right)$ and $\left(\lambda_f^{X*}, \lambda_v^{X*}\right)$. Then, we compare the optimal objective values of (P) across all regions to find the global optimal capacity allocation.

Region B, V and F are trivial cases where any feasible capacity allocation results in the same equilibrium. In Region BF, the equilibrium arrival rates are $(\lambda_f^{BF}, \lambda_v^{BF}) = (\mu_f - \theta_f[R-T]^{-1}, 0)$ according to Theorem 1. It is thus clear that the optimal choice is to maximize the value of μ_f. The following Proposition summarizes the optimal capacity allocation and the resulting patient equilibrium arrival rates in these relatively straightforward regions.

Proposition 1 **(Optimal Capacity Allocation in Regions B, V, F and BF)**. *In Regions B, V and F, any feasible allocation is optimal. In addition,* $(\lambda_f^{B*}, \lambda_v^{B*}) = (0,0)$, $(\lambda_f^{V*}, \lambda_v^{V*}) = (0,\Lambda)$ *and* $(\lambda_f^{F*}, \lambda_v^{F*}) = (\Lambda, 0)$. *For Region BF,* $(\mu_f^{BF*}, \mu_v^{BF*}) = (\overline{\mu}_f^{BF}, \mu - \overline{\mu}_f^{BF})$ *where* $\overline{\mu}_f^{BF}$ *is the largest* μ_f *such that* $(\mu_f, \mu - \mu_f) \in$ *Region BF, and* $(\lambda_f^{BF*}, \lambda_v^{BF*}) = (\overline{\mu}_f^{BF} - \theta_f \left[R-T\right]^{-1}, 0)$.

For Region BVF, since the equilibrium λ_v^{BVF} (λ_f^{BVF}) linearly increases in μ_v (μ_f) with everything else being fixed, one would expect the optimal capacity allocation to be a "bang-bang" type of control. The following proposition formalizes this intuition.

Proposition 2 **(Optimal Capacity Allocation subject to Region BVF)**. *Let* $[\underline{\mu}_v^{BVF}, \overline{\mu}_v^{BVF}]$ *denote the range for* μ_v *such that* $(\mu - \mu_v, \mu_v) \in$ *Region BVF. The optimal capacity allocation depends on the payment rates such that if* $r_v \geq (1+\delta) r_f$, *then* $(\mu_f^{BVF*}, \mu_v^{BVF*}) = (\mu - \overline{\mu}_v^{BVF}, \overline{\mu}_v^{BVF})$; *otherwise,* $(\mu_f^{BVF*}, \mu_v^{BVF*}) = (\mu - \underline{\mu}_v^{BVF}, \underline{\mu}_v^{BVF})$. *The corresponding equilibrium arrival rates* $(\lambda_f^{BVF*}, \lambda_v^{BVF*}) = (\mu_f^{BVF*} - \theta_f \left[R-T\right]^{-1} - \delta\mu_v^{BVF*} + \delta\theta_v \left[(1-\delta)R\right]^{-1}, \mu_v^{BVF*} - \theta_v \left[(1-\delta)R\right]^{-1})$.

In light of the equilibrium outcomes stipulated in Theorem 1, the process of deriving the optimal capacity allocation, which maximizes λ_v^{BV}, subject to Region BV, is quite involved. Three possibilities exist: Region BV prevails over Region V; Region V is infeasible and Region BV dominates Region BVF; or Region V is infeasible and Region BV is subordinate to Region BVF. Proposition 3 provides detailed descriptions of these scenarios.

Proposition 3 **(Optimal Capacity Allocation subject to Region BV)**. *Define* $\tilde{\mu}_v^{BV} = \left(\mu - (\delta(\theta_f - \theta_v) + (1-\delta)\sqrt{\delta\theta_f\theta_v})\left[R - \delta T\right]^{-1}\right)\left[1+\delta\right]^{-1}$. *Then, we have:*

– *If* $\mu \geq \Lambda(1+\delta) + \min\{\theta_f\,[R-T]^{-1} + \theta_v\,[(1-\delta)R]^{-1}, (\sqrt{\delta\theta_f} + \sqrt{\theta_v})^2\,[R-\delta T]^{-1}\}$*, any feasible allocation lying on the boundary between Region BV and Region V is optimal, leading to* $(\lambda_f^{BV*}, \lambda_v^{BV*}) = (0, \Lambda)$*.*

– *If* $\mu < \Lambda(1+\delta) + \min\{\theta_f\,[R-T]^{-1} + \theta_v\,[(1-\delta)R]^{-1}, (\sqrt{\delta\theta_f} + \sqrt{\theta_v})^2\,[R-\delta T]^{-1}\}$ *and* $(\mu - \tilde{\mu}_v^{BV}, \tilde{\mu}_v^{BV}) \in$ *Region BV, then* $(\mu_f^{BV*}, \mu_v^{BV*}) = (\mu - \tilde{\mu}_v^{BV}, \tilde{\mu}_v^{BV})$ *and* $\lambda_v^{BV*} = (\mu - (\sqrt{\delta\theta_f} + \sqrt{\theta_v})^2\,[R-\delta T]^{-1})\,[1+\delta]^{-1}$*.*

– *Otherwise,* $(\mu_f^{BV*}, \mu_v^{BV*}) = (\mu - \underline{\mu}_v^{BV}, \underline{\mu}_v^{BV})$*, where* $\underline{\mu}_v^{BV}$ *is the lower bound for* μ_v *such that* $(\mu - \mu_v, \mu_v) \in$ *Region BV, leading to* $\lambda_v^{BV*} = (\mu - \theta_f[R-T]^{-1} - \theta_v[(1-\delta)R]^{-1})\,[1+\delta]^{-1}$*.*

Finally, the optimal allocation and the effective arrival rates subject to Region VF are characterized by Proposition 4.

Proposition 4 **(Optimal Capacity Allocation subject to Region VF)**. *Let* $[\underline{\mu}_v^{VF}, \overline{\mu}_v^{VF}]$ *denote the range for* μ_v *such that* $(\mu - \mu_v, \mu_v) \in$ *Region VF.*

– *If* $r_v > r_f$*,* $(\mu_f^{VF*}, \mu_v^{VF*}) = (\mu - \overline{\mu}_v^{VF}, \overline{\mu}_v^{VF})$*, then* $\lambda_v^{VF*} = \min\{\Lambda, \delta^{-1}(\mu - \Lambda - \theta_f\,[R-T]^{-1} - \theta_v\,[(1-\delta)R]^{-1})\}$ *and* $\lambda_f^{VF*} = \Lambda - \lambda_v^{VF*}$*.*

– *If* $r_v = r_f$*, any feasible allocation is optimal and* $r_f\lambda_f^{VF*} + r_v\lambda_v^{VF*} = r_f\Lambda = r_v\Lambda$*.*

– *If* $r_v < r_f$*,* $(\mu_f^{VF*}, \mu_v^{VF*}) = (\mu - \underline{\mu}_v^{VF}, \underline{\mu}_v^{VF})$ *and* $(\lambda_f^{VF*}, \lambda_v^{VF*}) = (\Lambda, 0)$*.*

This proposition indicates that the optimal capacity allocation within Region VF depends on the relationship between r_f and r_v. If r_f is larger, the capacity should be allocated to the in-person channel as much as possible; if r_v is larger, however, the capacity should be allocated to the virtual channel as much as possible.

Following the above results on optimal capacity allocation subject to each region, we proceed to identify the global optimal capacity allocation. We start by defining several system-size categories to facilitate the subsequent discussions.

Definition 1. Given the system's primitives, we consider the following system-size categories with respect to the total service capacity μ.

1. **Small System:**

$$\max\left\{\frac{\theta_f}{R-T}, \underline{\mu}^{BV}\right\} \leq \mu < \Lambda + \frac{\theta_f}{R-T} + \frac{\theta_v}{(1-\delta)R}.$$

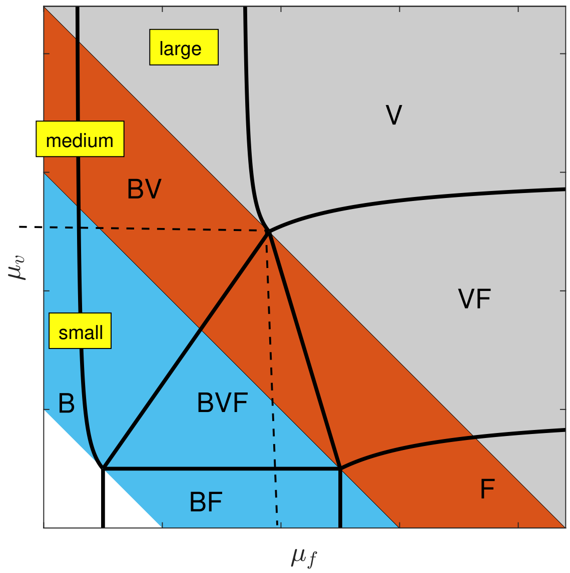
2. **Medium-sized System:**

$$\Lambda + \frac{\theta_f}{R-T} + \frac{\theta_v}{(1-\delta)R} \leq \mu < (1+\delta)\Lambda + \frac{\theta_f}{R-T} + \frac{\theta_v}{(1-\delta)R}.$$

3. **Large System:**

$$\mu \geq (1+\delta)\Lambda + \frac{\theta_f}{R-T} + \frac{\theta_v}{(1-\delta)R}.$$

Figure 3 Illustration of three system-size categories.



In the definition for Small System, the parameter $\underline{\mu}^{BV}$ denotes the minimum value of μ necessary to achieve Region BV. The detailed formulation of this parameter is provided in E-Companion EC.1.2. Figure 3 depicts the three distinct categories of system size in different colors. Note that we do not consider the region where μ is less than $\max\left\{\theta_f [R-T]^{-1}, \underline{\mu}^{BV}\right\}$, which corresponds to the uncolored, bottom-left corner in Figure 3. This is a less interesting region with very small capacity such that patients either balk entirely or mix between balking and one service channel.

Theorem 2 **(Optimal Capacity Allocation)**. *The optimal capacity allocation* (μ_f^*, μ_v^*) *for each system-size category can be described as follows.*

1. ***Small System:*** *If* $r_f \min\{\Lambda, \mu - \theta_f [R-T]^{-1}\} \geq r_v \lambda_v^{BV*}$, (μ_f^*, μ_v^*) *is in Region BF (with Region F considered a special case of Region BF); otherwise,* (μ_f^*, μ_v^*) *is in Region BV.*
2. ***Medium-sized Systems:*** *If* $r_v \lambda_v^{BV*} \geq r_f \Lambda + \delta^{-1}(r_v - r_f)^+(\mu - \Lambda - \theta_v [R(1-\delta)]^{-1} - \theta_f [R-T]^{-1})$, (μ_f^*, μ_v^*) *is in Region BV; otherwise,* (μ_f^*, μ_v^*) *is in Region VF (with Region F considered a special case of Region VF and dominating Region VF when* $r_f \geq r_v$*).*
3. ***Large System:*** *If* $r_v \geq r_f$, (μ_f^*, μ_v^*)*is in Region V; otherwise,* (μ_f^*, μ_v^*) *is in Region F.*

Remark 3. The current model assumes that all patients can initially be served by either the in-person or virtual channel. The model can be extended to incorporate heterogeneous health needs as follows: some patients, arriving at an exogenous rate of λ_E, must access the in-person channel. The provider "carves out" μ_E service capacity for them and then allocates the remaining $\mu - \mu_E$ capacity between the in-person and virtual channels for strategic patients who choose between the two. All results in Sections 3 and 4 directly apply to assist the provider in making capacity allocation decisions in this case.

In a small system, Region BVF is dominated by Region BV or Region BF, because the optimal solution lies on the region boundary, as discussed in Proposition 2. When the in-person channel offers greater benefits to the provider, she would attract the maximum number of patients to this particular channel by dedicating all capacity to the in-person channel. Conversely, if the virtual channel brings higher payments, the provider should strive to draw in as many patients as possible to the virtual channel. This can be achieved by setting a relatively large virtual service capacity and reserving an appropriate level of in-person service capacity for virtual patients requiring supplementary in-person services. In short, the *pooling effect* drives the optimal choice in a small system: the provider should use her limited capacity to focus on one type of patients.

For a medium-sized system, Region BVF proves to be sub-optimal. When the in-person payment rate r_f is sufficiently large, it is more beneficial to draw all patients towards the in-person service channel. Conversely, if the virtual service payment rate r_v is adequately large, directing all patients to the virtual channel (even with some balking) yields higher revenues. When the payment gap between in-person and virtual channels is not overly pronounced, attracting patients to both channels and not to lose any is the favorable approach. In a medium-sized system, the provider has enough capacity to avoid patient balking, but she needs to consider if she should attract all patients who intend to balk, and if so, how to balance capacity allocation between two channels.

Lastly, for a large system, no patients would balk regardless of how the service provider allocates capacity. Naturally, the provider would aim to channel patients to the more "profitable" option. In this case, the key decision driver is the payment or revenue margin per patient.

Theorem 2 reveals that the provider's optimal capacity strategy depends on the capacity-demand relationship in her own practice as well as the external payment environment. In practice, providers are rarely large systems, as healthcare resources are often limited. Most providers likely function as medium-sized systems unless capacity is extremely constrained. As discussed later, as the payment rates for in-person and virtual channels are relatively similar, balancing these two channels is crucial for managing operational efficiency.

5. Model with Transportation Subsidy

Transportation service is an important support mechanism for in-person visits. It can be financed directly by the provider or subsidized by the government, thus being cost-free to the provider. Transportation support takes a variety of forms, including reimbursement of transportation fees, shuttle services, coordinated shared ride programs, parking vouchers, and daily subway passes. In this section, we examine the impact of providing transportation support on patient behavior and provider operations. For ease of discussion, we will use "subsidy" as an umbrella term for these different forms of transportation support. Let $b \geq 0$ denote the transportation subsidy for each in-person visit. Regardless of who pays for the subsidy, the patient utility functions are modified as follows.

$$U_f^b(\lambda_f, \lambda_v) = R - T + b - \frac{\theta_f}{\mu_f - \lambda_f - \delta\lambda_v}; \tag{7}$$
$$U_v^b(\lambda_f, \lambda_v) = R - \frac{\theta_v}{\mu_v - \lambda_v} - \delta T + \delta b - \frac{\delta\theta_f}{\mu_f - \lambda_f - \delta\lambda_v}.$$

Generally, we allow b to exceed the travel cost T. However, when the subsidy is paid by the provider, we impose that it does not exceed the received payment, i.e.,

$$b \leq r_f \quad \text{and} \quad \delta b \leq r_v. \tag{8}$$

These conditions guarantee a positive expected net revenue from each visit for the provider.

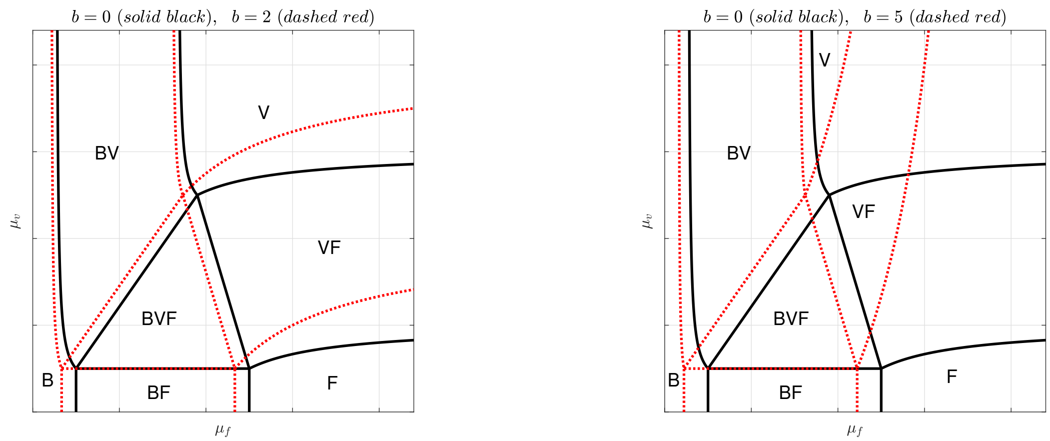
The impact of transportation subsidies on patient behavior is multifaceted. On the one hand, such subsidies may prompt patients to opt for the in-person service channel. On the other hand, they could lead patients to choose the virtual channel, potentially allowing them to take advantage of the subsidy in the event that a subsequent in-person visit becomes necessary. As the impact of transportation subsidies on patient equilibrium behavior does not depend on who pays for the subsidy, we start by analyzing patient equilibrium behavior for a given b. We then explore two scenarios: one where the subsidy is paid by the provider and another covered by the government. In each scenario, we investigate the impact of transportation subsidies on the overall access rate, i.e., the proportion of patients who do not balk. This total access rate can be regarded as an indicator of social welfare (Çakıcı and Mills 2024).

5.1. Patients' Equilibrium Strategy

Similar to the base model without a transportation subsidy (as detailed in Section 3.2), when a fixed transportation subsidy b is provided, the capacity allocation space can be partitioned into seven mutually exclusive and collectively exhaustive regions: namely, Region B, V, F, BVF, BF, BV, and VF. Their detailed definitions are presented in Definition EC.2 of E-Companion EC.1.3. Corollary EC.1 in E-Companion EC.1.3 outlines the equilibrium strategies of patients for any given service capacity allocation (μ_f, μ_v) and a non-negative transportation subsidy b. We omit details here as Corollary EC.1 is similar to Theorem 1.

With a transportation subsidy, patient equilibrium strategies share similar structures to those in the case without it. The only difference is that T is replaced by $T - b$, while other factors remain unchanged. This difference slightly alters the divisions of these regions as the utility functions now

Figure 4 **Illustrations of equilibrium regions under different values of b.**



Note. $\Lambda = 1$, $T = 4$, $R = 8$, $\delta = 0.7$, $\theta_f = 1$, $\theta_v = 0.6$.

incorporate b. It should be noted that the sign of $(T - b)$ influences the boundary shapes between Region V and Region VF, as well as between Region F and Region VF. Figure 4 depicts the changes in the regional structures for different values of b. We observe that with a positive transportation subsidy b, the boundaries shift to the left, leading to a smaller Region B. When $T - b$ is positive (as shown in the left plot), the boundary shapes remain the same. In contrast, when $T - b$ is negative (as shown in the right plot), the shapes change: Region V, where all patients opt for the virtual channel, shrinks significantly, while Region F, where all patients choose the in-person channel, expands. Overall, offering a transportation subsidy for in-person visits reduces balking and has the potential to increase service utilization, particularly for the in-person channel.

5.2. When the Provider Offers a Transportation Subsidy

5.2.1. The Provider's Optimal Joint Decision on Capacity and Transportation Subsidy In this subsection, we study the provider's joint decision regarding the transportation subsidy b and the capacity allocation (μ_f, μ_v). Specifically, the provider seeks to solve the following optimization problem.

$$\begin{aligned} \max_{\mu_f,\ \mu_v \geq 0} \quad & (r_f - b)\lambda_f + (r_v - b)\lambda_v \\ s.t. \quad & \mu_f + \mu_v = \mu, \\ & (\lambda_f, \lambda_v) \text{ is defined in Corollary EC.1.} \end{aligned} \tag{P1}$$

The complexity of (P1) arises from the fact that varying values of b result in distinct feasible equilibrium regions, leading to different optimal capacity allocation patterns.

We start our analysis by examining the optimal capacity allocation for an arbitrary value of b. Since the patient equilibrium exhibits a similar structure to that described in Section 3.2, all the optimization outcomes can be directly derived from the prior analysis in Section 4. To streamline the analysis, in Corollary 1, we consolidate all potential scenarios regarding the optimal capacity allocation for any given b.

Corollary 1 **(Equilibrium under Optimal Capacity Allocation with a Fixed b).** *Given a subsidy b, the optimal capacity allocation $(\mu_f^*(b), \mu_v^*(b))$ should lead to one of the following equilibrium scenarios.*

1. ***Scenario B***: *All patients balk.*
2. ***Scenario NV***: *No patient chooses the virtual channel. Therefore, $\lambda_v^{NV} = 0$ and $\lambda_f^{NV} = \min\{\Lambda, \mu - \theta_f [R - T + b]^{-1}\}$.*
3. ***Scenario NF***: *No patient chooses the in-person channel. Therefore, $\lambda_f^{NF} = 0$ and*

$$\lambda_v^{NF} = \begin{cases} \min\left\{\Lambda, \frac{\mu}{1+\delta} - \frac{(\sqrt{\delta\theta_f} + \sqrt{\theta_v})^2}{(1+\delta)(R-\delta T+\delta b)}\right\}, & \text{if } \frac{\delta(\theta_f - \theta_v) + (1-\delta)\sqrt{\delta\theta_f\theta_v}}{(R-\delta T+\delta b)} \le \frac{\theta_f}{R-T+b} - \frac{\delta\theta_v}{(1-\delta)R}, \\ \min\left\{\Lambda, \frac{\mu}{1+\delta} - \frac{\theta_f}{(1+\delta)(R-T+b)} - \frac{\theta_v}{(1+\delta)(1-\delta)R}\right\}, & \text{otherwise.} \end{cases} \tag{9}$$

4. ***Scenario VF***: *No patient balks. Therefore, $\lambda_f^{VF} = \Lambda - \lambda_v^{VF}$ and*

$$\lambda_v^{VF} = \frac{1}{\delta}\left(\mu - \Lambda - \frac{\theta_v}{(1-\delta)R} - \frac{\theta_f}{R-T+b}\right). \tag{10}$$

These four scenarios constitute a refined version of the regions discussed in Theorem 1. For instance, Scenario NV describes the equilibrium arrival rates assuming that the provider's optimal joint decision leads to Regions BF or F.

Next, we identify the optimal b within each specific scenario, and then, we derive the globally optimal joint decision of b^* and (μ_f^*, μ_v^*) through a comparison of these four scenarios. For Scenario B, the analysis is straightforward: there is no need to consider a transportation subsidy and hence the optimal b is zero. The analyses of other scenarios are more challenging because the value of b influences which scenario is possible. To facilitate our analysis, we first establish the range of b that renders a particular scenario feasible. We begin by establishing a common upper bound for b^* in the following lemma.

Lemma 1. *The optimal subsidy b^* must satisfy $b^* \le \bar{b}$ where*

$$\bar{b} = \begin{cases} \infty, & \text{if } \mu \le (1+\delta)\Lambda + \frac{\theta_v}{(1-\delta)R}, \\ \frac{\theta_f}{\mu - (1+\delta)\Lambda - \frac{\theta_v}{(1-\delta)R}} - R + T, & \text{otherwise.} \end{cases} \tag{11}$$

When b reaches $\bar{b}$, the provider can attract all patients under the optimal capacity allocation, and there is no need to consider the case of $b > \bar{b}$. In other words, $\bar{b}$ is the smallest b that allows the system to function as a "large" system, as per Definition 1. Following Lemma 1, we refine the upper bound for b^* within each specific scenario.

Let $\bar{b}_{NV}$, $\bar{b}_{NF}$ and $\bar{b}_{VF}$ denote the refined, scenario-specific upper bound of b^* when examining the optimal subsidy subject to scenarios NV, NF, and VF, respectively. Each of these refined upper bounds is the minimum of three valid upper bounds. The first is the global upper bound $\bar{b}$ presented in Lemma 1. The second is $\min\{r_f, \delta^{-1} r_v\}$ such that inequality (8) holds to ensure that the payment to the provider exceeds the subsidy. A larger subsidy b leads to a higher total arrival rate in each scenario. But the maximum total arrival rate is Λ, which defines the third valid upper bound for b. The detailed formulations for $\bar{b}_{NV}$, $\bar{b}_{NF}$ and $\bar{b}_{VF}$ can be found in E-Companion EC.1.3.2.

Now, we determine the lower bound of b^* when studying the optimal subsidy subject to a specific scenario. Let $\underline{b}_{NV}$, $\underline{b}_{NF}$ and $\underline{b}_{VF}$ represent the smallest b that respectively makes Scenarios NV, NF, and VF possibly feasible. For instance, $\underline{b}_{NV} = \theta_f \mu^{-1} - R + T$, which ensures that λ_f^{NV} is non-negative. The specific formulations for $\underline{b}_{NV}$, $\underline{b}_{NF}$ and $\underline{b}_{VF}$ can be found in E-Companion EC.1.3.2.

After characterizing these bounds of b^*, we can analyze b^* itself in each scenario described in Corollary 1. For Scenario B, the optimal $b^{B*} = 0$. For Scenario $X \in \{NV, NF, VF\}$, the provider solves

$$\max_{b \in [\underline{b}_X,\ \bar{b}_X]} (r_f - b)\lambda_f^X + (r_v - \delta b)\lambda_v^X, \tag{12}$$

where λ_f^X and λ_v^X are defined in Corollary 1, and $\underline{b}_X$ and $\bar{b}_X$ are defined in E-Companion EC.1.3.2. We use a superscript (e.g., B in b^{B*}) to denote the optimal b in each scenario.

We discuss the optimal subsidy by scenario. Proposition 5 presents the optimal subsidy subject to Scenario NV.

Proposition 5 **(Optimal b Subject to Scenario NV)**.

$$b^{NV*} = \begin{cases} \underline{b}_{NV}, & \textit{if } \tilde{b}^{NV} < \underline{b}_{NV}, \\ \tilde{b}^{NV}, & \textit{if } \underline{b}_{NV} \le \tilde{b}^{NV} \le \bar{b}_{NV}, \\ \bar{b}_{NV}, & \textit{if } \tilde{b}^{NV} > \bar{b}_{NV}, \end{cases}$$

where

$$\tilde{b}^{NV} = \sqrt{\frac{\theta_f (R - T + r_f)}{\mu}} - R + T.$$

Solving b^{NF*} for Scenario NF is challenging due to the complexity of the patient equilibrium involved. By imposing an additional technical assumption, we derive a closed-form expression for b^{NF*}.

Proposition 6 **(Optimal b Subject to Scenario NF)**. *Define*

$$f(b)=\frac{\delta(\theta_f-\theta_v)+(1-\delta)\sqrt{\delta\theta_f\theta_v}}{R-\delta T+\delta b}-\frac{\theta_f}{R-T+b}+\frac{\delta\theta_v}{(1-\delta)R}.$$

Let

$$\tilde{b}^{NF}=\begin{cases}\frac{(\sqrt{\delta\theta_f}+\sqrt{\theta_v})\sqrt{R-\delta T+r_v}}{\delta\sqrt{\mu}}-\frac{R-\delta T}{\delta}, & \text{if } \frac{(\sqrt{\delta\theta_f}+\sqrt{\theta_v})\sqrt{R-\delta T+r_v}}{\delta\sqrt{\mu}}-\frac{R-\delta T}{\delta}\leq b^0\\ \sqrt{\frac{\theta_f(\delta R-\delta T+r_v)}{\delta\mu-\frac{\delta\theta_v}{(1-\delta)R}}}-R+T, & \text{otherwise,}\end{cases}$$

where b^0 solves $f(b^0)=0$. Suppose that $f(0)<0$, then

$$b^{NF*}=\begin{cases}\underline{b}_{NF}, & \text{if } \tilde{b}^{NF}<\underline{b}_{NF},\\ \tilde{b}^{NF}, & \text{if } \underline{b}_{NF}\leq\tilde{b}^{NF}\leq\bar{b}_{NF},\\ \bar{b}_{NF}, & \text{if } \tilde{b}^{NF}>\bar{b}_{NF}.\end{cases}$$

The additional assumption $f(0)<0$ restricts δ, the probability of needing a supplementary visit, from being too large. That aligns with practical return probabilities, which are estimated to be 6–20% (Yamamoto 2014, Uscher-Pines et al. 2016, Shi et al. 2018). Lastly, Proposition 7 derives the optimal subsidy subject to Scenario VF.

Proposition 7 **(Optimal b Subject to Scenario VF)**. *Define*

$$\tilde{b}^{VF}=\sqrt{\frac{\theta_f(r_v-r_f)-(1-\delta)\theta_f(R-T)}{-(1-\delta)\mu+\Lambda+\theta_v/R}}-R+T.$$

If $r_v-r_f-(1-\delta)(R-T)\leq 0$, $(1-\delta)\mu-\Lambda-\theta_v/R>0$, and $\tilde{b}^{VF}\in[\underline{b}_{VF},\bar{b}_{VF}]$, then $b^{VF*}=\tilde{b}^{VF}$. Otherwise, b^{VF*} takes the value of $\bar{b}_{VF}$ or $\underline{b}_{VF}$ whichever yields the higher net revenue.

Upon characterizing the optimal subsidy within each given scenario, we determine the one yielding the greatest net revenue as the globally optimal subsidy. To achieve this, we establish several thresholds for μ to streamline the selection of the optimal scenario. We denote $\underline{\mu}_X$ (where $X=NV,NF,VF$) as the minimum capacity necessary to ensure the occurrence of Scenario X. Additionally, we introduce $\bar{\mu}$ as the threshold, above which no subsidies are needed. Precisely, when $\mu\geq\bar{\mu}$, the provider can attain Regions F and V (see Corollary EC.1) without any subsidy. The detailed expressions for these capacity thresholds are available in E-Companion EC.1.3.3. By utilizing these thresholds, we can specify which scenario, as defined in Corollary 1, is optimal based on the size of the system, μ. Proposition 8 summarizes the optimal scenario for any μ.

Proposition 8. *The optimal scenario for a given service capacity μ is as follows.*

1. *When* $\mu \le \min\{\underline{\mu}_{NV}, \underline{\mu}_{NF}\}$*, Scenario B is optimal, and* $b^* = 0$.
2. *When* $\min\{\underline{\mu}_{NV}, \underline{\mu}_{NF}\} < \mu < \max\{\underline{\mu}_{NV}, \underline{\mu}_{NF}\}$*:*
   - *If* $\underline{\mu}_{NV} < \underline{\mu}_{NF}$*, Scenario NF is optimal, and* $b^* = b^*_{NF}$.
   - *If* $\underline{\mu}_{NV} > \underline{\mu}_{NF}$*, Scenario NV is optimal, and* $b^* = b^*_{NV}$.
3. *When* $\max\{\underline{\mu}_{NV}, \underline{\mu}_{NF}\} \le \mu < \underline{\mu}_{VF}$*, the optimal scenario is either NV or NF.*
4. *When* $\underline{\mu}_{VF} \le \mu < \bar{\mu}$*, the optimal scenario is either NV, NF, or VF.*
5. *When* $\mu \ge \bar{\mu}$*,* $b^* = 0$.

Proposition 8 shows that transportation subsidy only benefits the provider when the system size is neither too small nor too large, i.e., in cases 2 to 4 where $\min\{\underline{\mu}_{NV}, \underline{\mu}_{NF}\} < \mu < \bar{\mu}$.

5.2.2. Impact on Access to Care Offering transportation subsidies helps the provider increase her revenues and improves the accessibility of the in-person channel for patients. However, its impact on *overall* patient access to care is unclear. In our analysis, we operationalize the overall patient access to care by the total effective arrival rate of patients who access the healthcare service without balking. i.e, $\lambda_f^* + \lambda_v^*$, where λ_f^* and λ_v^* are the effective arrival rates in equilibrium under the provider's optimal joint decision b^* and (μ_f^*, μ_v^*). As discussed in Çakıcı and Mills (2024), the overall patient access to care can be viewed as a measure of social welfare. If such a subsidy, despite increasing the provider's revenue, reduces the overall patient access, its adoption deserves caution. Section 6 will investigate this issue using model parameters informed by real-world data. Here, we first derive sufficient conditions to ensure that the total access rate does *not* decline when utilizing a subsidy.

PROPOSITION 9. *Suppose the provider offers a transportation subsidy and jointly optimizes it with capacity allocation. With all model parameters fixed, the total access rate will not decrease under any of the following conditions.*

1. *Large system.*
2. *Medium-sized system or small system with* $\mu \ge \Lambda + \theta_f [R-T]^{-1}$*, if*

$$\frac{r_v}{r_f} \le 1 \quad or \quad \frac{r_v}{r_f} \ge (1+\delta) \max\left\{1, \frac{\Lambda}{\mu - \frac{\theta_f}{R-T} - \frac{\theta_v}{(1-\delta)R}}\right\}. \tag{13}$$

3. *Small system with* $\mu < \Lambda + \theta_f [R-T]^{-1}$*, if*

$$\frac{r_v}{r_f} \le \min\left\{1, \frac{(1+\delta)(\mu - \frac{\theta_f}{R-T})}{\mu - \frac{(\sqrt{\delta\theta_f} + \sqrt{\theta_v})^2}{R-\delta T}}\right\} \quad or \quad \frac{r_v}{r_f} \ge \frac{(1+\delta)(\mu - \frac{\theta_f}{R-T})}{\mu - \frac{\theta_f}{R-T} - \frac{\theta_v}{(1-\delta)R}}.$$

A large system has no incentive to provide transportation subsidy, and thus the access rate will not be influenced. As for medium-sized or small systems, when the ratio between r_v and r_f is sufficiently small or large, the total access rate will not decrease when adopting the optimal subsidy. To see this, let us consider the case where r_v/r_f is small enough. In this case, the in-person channel is preferable as it generates higher marginal revenue. Thus, the provider will tend not to utilize the virtual channel and we will have $\lambda_v^*(0) = 0$ (let $\lambda_f^*(b)$ and $\lambda_v^*(b)$ denote the effective arrival rates under optimal capacity allocation with b). After utilizing the optimal subsidy, the provider achieves higher revenue at a lower net payment per patient, which implies a larger access rate. To be more specific, we note that $r_f\lambda_f^*(b^*) + r_f\lambda_v^*(b^*) \geq (r_f - b^*)\lambda_f^*(b^*) + (r_v - \delta b^*)\lambda_v^*(b^*) \geq r_f\lambda_f^*(0)$, implying $\lambda_f^*(b^*) + \lambda_v^*(b^*) \geq \lambda_f^*(0)$ because $r_f \geq r_v$. A similar argument can explain why when r_v/r_f is large enough, the total access rate will not decrease when the optimal subsidy is utilized.

5.3. When the Government Provides a Transportation Subsidy

In this subsection, we study the scenario where the subsidy is paid by the government and, therefore, incurs no cost to the provider. The provider's objective is the same as in (P), while patient equilibrium behavior is governed by Corollary EC.1 under a given subsidy b.

$$\begin{aligned} \max_{\mu_f,\ \mu_v \geq 0} \quad & r_f\lambda_f + r_v\lambda_v \\ s.t. \quad & \mu_f + \mu_v = \mu, \\ & (\lambda_f, \lambda_v) \text{ is defined in Corollary EC.1.} \end{aligned} \tag{P2}$$

In (P2), the provider's optimal capacity allocation can be analyzed similarly to the base model (P), with T replaced by $T - b$. All results from Theorem 2 can be readily applied.

Transportation subsidies make it easier for patients to access the in-person channel and improve patient utilities. As the provider bears no cost, one may expect transportation subsidies paid by the government would always increase patient access to care. However, the provider can capitalize on the improved patient utility and adjust her capacity allocation decisions, which affect the patient mix she serves. It is uncertain whether social welfare, measured by the total access rate, necessarily improves under government interventions.

Similar to Proposition 9, we derive the conditions under which the total access rate will *not* decline when the government provides transportation subsidies.

Proposition 10. *Suppose that the government provides transportation subsides at no cost to the provider. With all model parameters fixed, the total access rate will not decrease under any of the following conditions.*

1. *Large system.*
2. *Medium-sized system or a small system with $\mu \geq \Lambda + \theta_f[R-T]^{-1}$, if*

$$\frac{r_v}{r_f} \leq \frac{(1+\delta)\Lambda}{\min\left\{(1+\delta)\Lambda, \mu - \frac{(\sqrt{\delta\theta_f}+\sqrt{\theta_v})^2}{R-\delta T+\delta b}\right\}} \quad \text{or} \quad \frac{r_v}{r_f} \geq (1+\delta)\max\left\{1, \frac{\Lambda}{\mu - \frac{\theta_f}{R-T} - \frac{\theta_v}{(1-\delta)R}}\right\}.$$

3. *Small system with* $\mu < \Lambda + \theta_f[R-T]^{-1}$, *if*

$$\frac{r_v}{r_f} \le \min\left\{1, \frac{(1+\delta)(\mu - \frac{\theta_f}{R-T})}{\mu - \frac{(\sqrt{\delta\theta_f}+\sqrt{\theta_v})^2}{R-\delta T}}\right\} \quad or \quad 1 < \frac{r_v}{r_f} \le \frac{(1+\delta)\{\Lambda, \mu - \frac{\theta_f}{R-T+b}\}}{\mu - \frac{(\sqrt{\delta\theta_f}+\sqrt{\theta_v})^2}{R-\delta T+\delta b}} \quad or \quad \frac{r_v}{r_f} \ge \frac{(1+\delta)(\mu - \frac{\theta_f}{R-T})}{\mu - \frac{\theta_f}{R-T} - \frac{\theta_v}{(1-\delta)R}}.$$

Intuitively, patient access to care is less likely to be hurt when the government funds transportation subsidies to support in-person visits at no cost to the provider. Indeed, the conditions presented in Proposition 10 are less stringent than those in Proposition 9. However, there is still no theoretical guarantee that patient access to care will always improve.

6. Numerical Analysis

In this section, we investigate the potential practical impact of transportation subsidies on provider operations and patient access to care, by calibrating our model using real-world data and previous empirical studies. Our purpose is not to provide decision support for any specific setting, but to project the potential practical implications of our analytical results. To populate our model, we obtain realistic values for the following parameters: transportation cost T, waiting cost rates θ_f and θ_v, probability of requiring a supplementary in-person visit after completing a virtual visit δ, and reimbursement rates r_f and r_v.

For transportation costs, we distinguish between rural and non-rural areas. A recent research report by the Rural Health Research Gateway (Ostmo and Rosencrans 2022) finds that rural residents travel, on average, 17.8 miles for care compared to 8.1 miles for non-rural residents. Rural residents spend an average of 34.2 minutes traveling for care, compared to 25.5 minutes for non-rural residents. To convert these transportation efforts into monetary values, we use the 2023 IRS standard mileage rates to calculate travel costs and use wage to measure the value of time, following the classic economics literature (Becker 1965). Specifically, we consider a millage rate of \$0.655 per mile and use the average hourly wage of \$35 reported by the U.S. Bureau of Labor Statistics (2025). We set the transportation costs, respectively, for rural and non-rural areas as:

$$T_{\text{rural}} = 17.8 \cdot 0.655 + \frac{34.2 \cdot 35}{60} \approx \$32 \quad \text{and} \quad T_{\text{non-rural}} = 8.1 \cdot 0.655 + \frac{25.5 \cdot 35}{60} \approx \$20.$$

We set the in-person waiting cost per hour to be $\theta_f = 35$, matching the hourly wage value. In a call center study, Hathaway et al. (2021) find that callers experience three to six times less discomfort per unit of time while waiting for callbacks than while waiting in a queue. This suggests that virtual waiting (akin to waiting for callbacks) is less costly than in-person waiting (similar to waiting in a queue). Following this rationale, we set $\theta_v = 12$, about one-third of θ_f.

The probability of requiring a supplementary in-person visit is set as $\delta = 5\%$, which is consistent, though conservative, with previous empirical findings (Gordon et al. 2017, Ashwood et al. 2017, Shi et al. 2018, Li et al. 2021). Furthermore, we use the out-of-pocket cost for an outpatient care visit as a proxy for the utility of seeing the provider. Such costs range from \$32 to \$175 per visit (Rohatsch 2025). In our numerical analysis, we use a moderate value of $R = \$60$.

To set the reimbursement rates, we consider the most-billed office visit CPT code 99214, reimbursed at about \$100 per visit (Kim 2023). Accordingly, we set $r_f = \$100$. Telehealth reimbursements are typically equal or lower than in-person rates. To be concrete, we consider that virtual visits are reimbursed at 80% of the in-person rate, following the reimbursement policy of Point32Health (Point32Health, Inc. 2024), one of the top twenty health plans in the U.S. Given $\delta = 5\%$, the expected reimbursement for a virtual visit is $r_v = 80 + 100 \times 5\% = \$85$. Lastly, we consider a practice with an average of 20 patient arrivals per 8-hour day, i.e., the hourly arrival rate is set as $\Lambda = 2.5$.

Based on these parameters, Figure 5 plots how the proportion of patients who receive services, i.e., the relative access rate $(\lambda_f^* + \lambda_v^*)/\Lambda$, changes with the total service capacity μ with and without adopting the transportation subsides. The higher the proportion, the better the overall patient access to care. We observe that in both the rural and non-rural cases, the dashed curve may drop below the solid one, suggesting that patient access to care may decrease when the provider offers transportation subsidies. This is in particular the case in rural areas (see the range where $\mu \approx 3.2$-3.3), where transportation subsidies make the in-person channel particularly attractive to patients because the transportation cost is high. By offering transportation subsidies, the provider can actually achieve a higher revenue with fewer total patient visits through the in-person channel––which, in turn, hurts overall patient access to care.

Figure 6 illustrates how the relative access rate changes at different levels of transportation subsidies offered by the government. The total service capacity is fixed at $\mu = 3.25$. One striking observation is on the left part of Figure 6: in rural areas, patient access to care may decline with government-funded transportation subsidy at no cost to the provider! This arises from the changes in the provider's optimal operational strategy. When travel costs are high and no subsidies are available, it is challenging to draw sufficient patients to the in-person channel with a higher reimbursement rate. As a result, the provider primarily operates the virtual channel with a lower reimbursement rate to serve more patients. Conversely, with subsidies, it becomes easier to attract patients to the more "profitable" channel. Even though fewer patients are served, the overall revenue could still increase. Hence, the provider adjusts her operational strategy, which may lead to a decrease in the overall access.

7. Conclusion

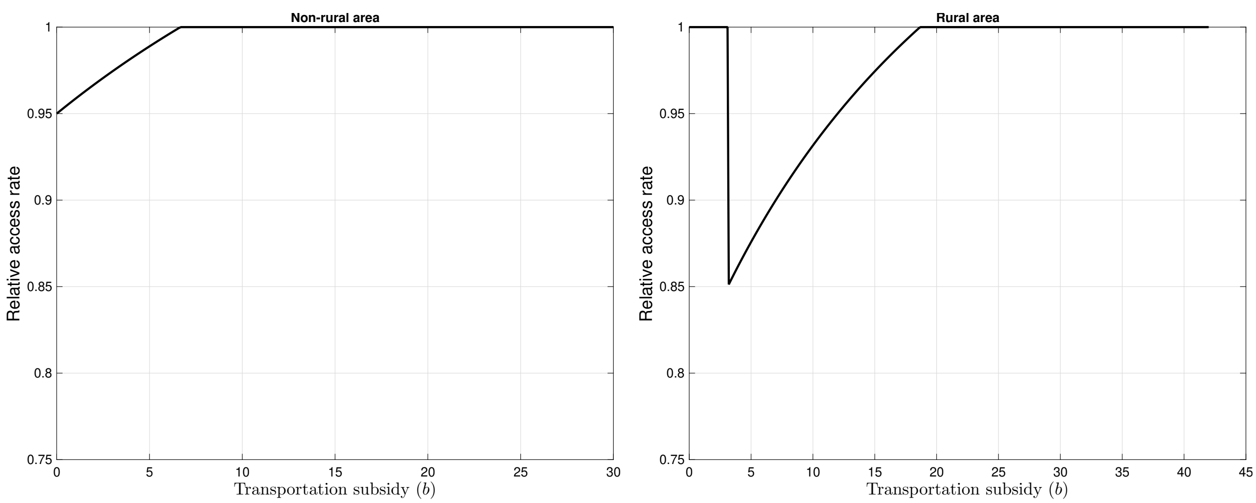
The use of telemedicine skyrocketed during the COVID-19 pandemic, offering convenience and reduced travel costs. However, it may not fully address patient needs, sometimes necessitating

Figure 5 **Comparison of access rates with and without provider-funded subsidies for rural and non-rural areas.**

Figures/CaseStudy_1-eps-converted-to.pdf

Figures/CaseStudy_2-eps-converted-to.pdf

Figure 6 **Comparison of access rates with government-funded subsidies for rural and non-rural areas.**



supplementary in-person visits. As the pandemic has subsided, in-person care regains favor. Transportation support—funded by providers or the government—emerges as a key enabler. Meanwhile, policy discussions are shifting from pay parity, originally designed to promote telemedicine adoption, to pay equity, which aims to align reimbursement with provider effort and patient outcomes. In this changing landscape for telemedicine practice, we study how an outpatient care provider can optimally balance virtual and in-person services and determine whether and how to engage with transportation subsidies.

Specifically, we develop a stylized queueing-game model to capture the operations of a single revenue-maximizing provider serving patients who make strategic decisions regarding service channels. We find that the optimal strategy depends critically on the size of the system. In small and large systems, focusing on a single channel proves more efficient, and offering transportation support may not improve revenues. However, medium-sized systems may benefit from offering both channels in combination with transportation support.

Our study highlights the impact of transportation support on patient access to care. Paradoxically, offering transportation support may reduce overall patient access to care, even when funded entirely by the government at no cost to the provider. This occurs because, by offering or with free transportation support, a provider can achieve higher revenue by shifting demand to the more "lucrative" service channel, serving fewer patients overall. However, when the payment gap between the two channels is large enough, the increased patient utilities due to transportation support will help the provider attract more visits, and thus both the provider's revenue and the total access rate will not decrease. To put this discussion in the context of a concrete reimbursement regime, consider a fee-for-service model. Recall that $r_f = r_f^{\text{FFS}}$ and $r_v = r_v^{\text{FFS}} + \delta r_f^{\text{FFS}}$ in Remark 2. The first condition in (13) is $r_v/r_f \leq 1$, which translates to $r_v^{\text{FFS}} \leq (1-\delta) r_f^{\text{FFS}}$. That is, if the virtual visit reimbursement is low enough relative to in-person care, the adoption of transportation support would not hurt overall patient access to care. This implication supports the ongoing policy discussion that advocates payment equity rather than parity for telemedicine (Shachar et al. 2020). In the post-pandemic era, proper financial incentives can help ensure that access to care is not negatively affected as providers reinvest in in-person care.

There are several interesting avenues for future research. First, a different operational lever—patient prioritization—may be considered. Another possible direction is to consider a setting where returning patients are served in a separate channel with dedicated capacity. Lastly, the design of reimbursement policies to align provider incentives with societal goals is not the focus of this work, and would be a fruitful direction to explore in the future.

References

Adida, E., F. Bravo. 2023. Primary care first initiative: Impact on care delivery and outcomes. *Manufacturing & Service Operations Management*, 25 (4), 1471-1488.

Ashwood, J.S., A. Mehrotra, D. Cowling, L. Uscher-Pines. 2017. Direct-to-consumer telehealth may increase access to care but does not decrease spending. *Health Affairs*, 36 (3), 485-491.

Baron, O., X. Chen, Y. Li. 2022. Omnichannel services: The false premise and operational remedies. *Management Science*, 69 (2), 865-884.

Bavafa, H., L.M. Hitt, C. Terwiesch. 2018. The impact of E-visits on visit frequencies and patient health: Evidence from primary care. *Management Science*, 64 (12), 5461-5480.

Bavafa, H., S. Savin, C. Terwiesch. 2021. Customizing primary care delivery using e-visits. *Production and Operations Management*, 30 (11), 4306-4327.

Becker, Gary S. 1965. A theory of the allocation of time. *The economic journal*, 75 (299), 493-517.

Bokolo, A.J. 2020. Use of telemedicine and virtual care for remote treatment in response to COVID-19 pandemic. *Journal of Medical Systems*, 44 (7), 1-9.

Çakıcı, Ö.E., A.F. Mills. 2021. On the role of teletriage in healthcare demand management. *Manufacturing & Service Operations Management*, 23 (6), 1483-1504.

Çakıcı, Ö.E., A.F. Mills. 2024. Telehealth in acute care: Pay parity and patient access. *Manufacturing & Service Operations Management*, .

Cordine, J., J. Fowkes, R. Malani, L. Medford-Davis. 2022. Patients love telehealth—physicians are not so sure. *McKinsey Insights*, .

Gordon, A.S., W.C. Adamson, A.R. DeVries. 2017. Virtual visits for acute, nonurgent care: a claims analysis of episode-level utilization. *Journal of medical Internet research*, 19 (2), e35.

Graham, E. 2022. Health care providers working to streamline transportation services for patients., National Center for Mobility Management, https://nationalcenterformobilitymanagement.org/blog/health-care-providers-transportation-services-for-patients/.

Green, L.V., S. Savin. 2008. Reducing delays for medical appointments: A queueing approach. *Operations Research*, 56 (6), 1526-1538.

Guan, X., K.J. Li, H. Wu. 2025. A blessing or a curse? teletriage service in healthcare. *Management Science*, (Forthcoming).

Hassin, R., M. Haviv. 2003. *To Queue or Not to Queue: Equilibrium Behavior in Queueing Systems*, vol. 59. Springer Science & Business Media.

Hassin, R., R. Roet-Green. 2020. On queue-length information when customers travel to a queue. *Manufacturing & Service Operations Management*, 23 (4), 989-1004.

Hathaway, B.A., S.M. Emadi, V. Deshpande. 2021. Don't call us, we'll call you: An empirical study of caller behavior under a callback option. *Management Science*, 67 (3), 1508-1526.

Huang, F., P. Guo, Y. Wang. 2022. Modeling patients' illness perception and equilibrium analysis of their doctor shopping behavior. *Production and Operations Management*, 31 (3), 1216-1234.

Khairat, S., X. Lin, S. Liu, Z. Man, T. Zaman, B. Edson, R. Gianforcaro. 2021. Evaluation of patient experience during virtual and in-person urgent care visits: Time and cost analysis. *Journal of Patient Experience*, 8 1-8.

Kim, T. 2023. The most-billed and most error-prone cpt codes of 2023., https://www.medcentral.com/coding-reimbursement/the-most-billed-and-most-error-prone-codes-of-2023.

Li, K.Y., Z. Zhu, S. Ng, C. Ellimoottil. 2021. Direct-to-consumer telemedicine visits for acute respiratory infections linked to more downstream visits: Study examines the association between telemedicine and downstream health care utilization. *Health Affairs*, 40 (4), 596-602.

Liu, N. 2016. Optimal choice for appointment scheduling window under patient no-show behavior. *Production and Operations Management*, 25 (1), 128-142.

Liu, N., W. van Jaarsveld, S. Wang, G. Xiao. 2023. Managing outpatient service with strategic walk-ins. *Management Science*, 69 (10), 5904-5922.

Liu, N., S. Ziya. 2014. Panel size and overbooking decisions for appointment-based services under patient no-shows. *Production and Operations Management*, 23 (12), 2209-2223.

Medicaid and CHIP Payment and Access Commission. 2019. Medicaid coverage of non-emergency medical transportation., https://www.macpac.gov/publication/medicaid-coverage-of-non-emergency-medical-transportation/.

Monaghesh, E., A. Hajizadeh. 2020. The role of telehealth during COVID-19 outbreak: A systematic review based on current evidence. *BMC Public Health*, 20 (1), 1-9.

Naor, P. 1969. The regulation of queue size by levying tolls. *Econometrica*, 15-24.

O'Reilly-Jacob, M., P. Mohr, M. Ellen, C. Petersen, C. Sarkisian, S. Attipoe, E. Rich. 2021. Digital health & low-value care. *Healthcare*, vol. 9. Elsevier, 100533.

Ostmo, P., J. Rosencrans. 2022. Travel burden to receive health care. Rural Health Research Gateway, https://www.ruralhealthresearch.org/assets/4993-22421/travel-burden-recap.pdf.

Point32Health, Inc. 2024. Telehealth reimbursement update for BH services., https://www.point32health.org/provider/news/telehealth-reimbursement-update-for-bh-services/.

Rajan, B., T. Tezcan, A. Seidmann. 2019. Service systems with heterogeneous customers: Investigating the effect of telemedicine on chronic care. *Management Science*, 65 (3), 1236-1267.

Reed, M., J. Huang, I. Graetz, E. Muelly, A. Millman, C. Lee. 2021. Treatment and follow-up care associated with patient-scheduled primary care telemedicine and in-person visits in a large integrated health system. *JAMA Network Open*, 4 (11), e2132793-e2132793.

Rohatsch, R. 2025. How much does urgent care cost?, https://www.solvhealth.com/health/how-much-does-urgent-care-cost.

Rosenthal, E. 2021. Telemedicine is a tool, not a replacement for your doctor's touch. New York Times, https://www.nytimes.com/2021/04/29/opinion/virtual-remote-medicine-covid.html.

Shachar, C., J. Engel, G. Elwyn. 2020. Implications for telehealth in a postpandemic future: regulatory and privacy issues. *Jama*, 323 (23), 2375-2376.

Sharma, S., Y. Xu, M.K. Gupta, C. Courcoubetis. 2020. Reducing non-urgent visits and emergency department congestion: Pricing or improving perceptions? *Available at SSRN 3729019*, .

Shi, Z., A. Mehrotra, C.A. Gidengil, S.J. Poon, L. Uscher-Pines, K.N. Ray. 2018. Quality of care for acute respiratory infections during direct-to-consumer telemedicine visits for adults. *Health Affairs*, 37 (12), 2014-2023.

Shumsky, R. A, E.J. Pinker. 2003. Gatekeepers and referrals in services. *Management Science*, 49 (7), 839-856.

U.S. Bureau of Labor Statistics. 2025. Economic news release., https://www.bls.gov/news.release/empsit.t19.htm.

Uscher-Pines, L., R. Malsberger, L. Burgette, A. Mulcahy, A. Mehrotra. 2016. Effect of teledermatology on access to dermatology care among medicaid enrollees. *JAMA dermatology*, 152 (8), 905-912.

Wang, X., Z. Zhang, J. Zhao, Y. Shi. 2019. Impact of telemedicine on healthcare service system considering patients' choice. *Discrete Dynamics in Nature and Society*, 2019 1-16.

Wong, M.Y.Z., D.V. Gunasekeran, S. Nusinovici, C. Sabanayagam, K. K. Yeo, C.-Y. Cheng, Y.-C. Tham. 2021. Telehealth demand trends during the COVID-19 pandemic in the top 50 most affected countries: Infodemiological evaluation. *JMIR Public Health and Surveillance*, 7 (2), e24445.

Yamamoto, D.H. 2014. Assessment of the feasibility and cost of replacing in-person care with acute care telehealth services, Alliance for Connected Care, December 2014.

Zacharias, C., M. Armony. 2016. Joint panel sizing and appointment scheduling in outpatient care. *Management Science*, 63 (11), 3978-3997.

Zang, L., Y. Hu, R. Roet-Green, S. Sun. 2024. The impact of information-granularity and prioritization on patients' care modality choice., Stanford University GSB Research Paper No. 4733392.

Zychlinski, N. 2024. Managing queues with reentrant customers in support of hybrid healthcare. *Stochastic Systems*, 14 (2), 167-190.

E-Companion

EC.1. Additional Technical Results

EC.1.1. Additional Technical Results for Section 3

Definition EC.1.

- Region B:

$$\frac{\theta_v}{\mu_v} + \frac{\delta\theta_f}{\mu_f} \geq R - \delta T, \tag{EC.1}$$

$$\mu_f \leq \frac{\theta_f}{R - T}. \tag{EC.2}$$

- Region V:

$$\frac{\theta_v}{\mu_v - \Lambda} - \frac{(1-\delta)\theta_f}{\mu_f - \delta\Lambda} \leq (1-\delta)T \quad \text{for } \mu_f > \delta\Lambda, \tag{EC.3}$$

$$\frac{\theta_v}{\mu_v - \Lambda} + \frac{\delta\theta_f}{\mu_f - \delta\Lambda} \leq R - \delta T, \tag{EC.4}$$

$$\mu_f > \delta\Lambda, \tag{EC.5}$$

$$\mu_v > \Lambda. \tag{EC.6}$$

- Region F:

$$\mu_f \geq \Lambda + \frac{\theta_f}{R - T}, \tag{EC.7}$$

$$\frac{\theta_v}{\mu_v} - \frac{(1-\delta)\theta_f}{\mu_f - \Lambda} \geq (1-\delta)T. \tag{EC.8}$$

- Region BVF:

$$\mu_v \geq \frac{\theta_v}{(1-\delta)R}, \tag{EC.9}$$

$$\mu_f - \delta\mu_v \geq \frac{\theta_f}{R - T} - \frac{\delta\theta_v}{(1-\delta)R}, \tag{EC.10}$$

$$\mu_f + (1-\delta)\mu_v \leq \Lambda + \frac{\theta_f}{R - T} + \frac{\theta_v}{R}. \tag{EC.11}$$

- Region BF:

$$\mu_f \geq \frac{\theta_f}{R - T}, \tag{EC.12}$$

$$\mu_f \leq \Lambda + \frac{\theta_f}{R - T}, \tag{EC.13}$$

$$\mu_v \leq \frac{\theta_v}{(1-\delta)R}. \tag{EC.14}$$

- Region BV:

$$\frac{\theta_v}{\mu_v} + \frac{\delta\theta_f}{\mu_f} \le R - \delta T, \tag{EC.15}$$

$$\frac{\theta_v}{\mu_v - \Lambda} + \frac{\delta\theta_f}{\mu_f - \delta\Lambda} \ge R - \delta T \text{ for } \mu_v > \Lambda \text{ and } \mu_f > \delta\Lambda, \tag{EC.16}$$

$$\mu_f - \delta\mu_v \le \frac{\theta_f}{R-T} - \frac{\delta\theta_v}{(1-\delta)R}. \tag{EC.17}$$

- Region VF:

$$\frac{\theta_v}{\mu_v} - \frac{(1-\delta)\theta_f}{\mu_f - \Lambda} \le (1-\delta)T \quad \text{for } \mu_f > \Lambda, \tag{EC.18}$$

$$\frac{\theta_v}{\mu_v - \Lambda} - \frac{(1-\delta)\theta_f}{\mu_f - \delta\Lambda} \ge (1-\delta)T \quad \text{for } \mu_v > \Lambda, \tag{EC.19}$$

$$\mu_f + (1-\delta)\mu_v \ge \Lambda + \frac{\theta_f}{R-T} + \frac{\theta_v}{R}, \tag{EC.20}$$

$$\mu_f > \delta\Lambda. \tag{EC.21}$$

EC.1.2. Additional Technical Results for Section 4

$\underline{\mu}^{BV}$ is the minimum μ required to attain Region BV. It is obtained by the solution of μ to

$$R - \delta T - \frac{\theta_v}{\Phi(\mu)} - \frac{\delta\theta_f}{\mu - \Phi(\mu)} = 0,$$

where

$$\Phi(\mu) = \max\left\{\frac{\theta_v}{(1-\delta)R}, \frac{\mu(\sqrt{\delta\theta_v\theta_f} - \theta_v)}{\delta\theta_f - \theta_v}\right\}. \tag{EC.22}$$

EC.1.3. Additional Technical Results for Section 5

EC.1.3.1. Patient equilibrium with transportation subsidy

Definition EC.2.

- Region B:

$$\frac{\theta_v}{\mu_v} + \frac{\delta\theta_f}{\mu_f} \ge R - \delta T + \delta b,$$
$$\mu_f \le \frac{\theta_f}{R - T + b}.$$

- Region V:

$$\frac{\theta_v}{\mu_v - \Lambda} - \frac{(1-\delta)\theta_f}{\mu_f - \delta\Lambda} \le (1-\delta)(T-b) \quad \text{for } \mu_f > \delta\Lambda,$$
$$\frac{\theta_v}{\mu_v - \Lambda} + \frac{\delta\theta_f}{\mu_f - \delta\Lambda} \le R - \delta T + \delta b,$$
$$\mu_f > \delta\Lambda,$$
$$\mu_v > \Lambda.$$

- Region F:

$$\mu_f \geq \Lambda + \frac{\theta_f}{R-T+b},$$
$$\frac{\theta_v}{\mu_v} - \frac{(1-\delta)\theta_f}{\mu_f - \Lambda} \geq (1-\delta)(T-b).$$

- Region BVF:

$$\mu_v \geq \frac{\theta_v}{(1-\delta)R},$$
$$\mu_f - \delta\mu_v \geq \frac{\theta_f}{R-T+b} - \frac{\delta\theta_v}{(1-\delta)R},$$
$$\mu_f + (1-\delta)\mu_v \leq \Lambda + \frac{\theta_f}{R-T+b} + \frac{\theta_v}{R}.$$

- Region BF:

$$\mu_f \geq \frac{\theta_f}{R-T+b},$$
$$\mu_f \leq \Lambda + \frac{\theta_f}{R-T+b},$$
$$\mu_v \leq \frac{\theta_v}{(1-\delta)R}.$$

- Region BV:

$$\frac{\theta_v}{\mu_v} + \frac{\delta\theta_f}{\mu_f} \leq R - \delta T + \delta b,$$
$$\frac{\theta_v}{\mu_v - \Lambda} + \frac{\delta\theta_f}{\mu_f - \delta\Lambda} \geq R - \delta T + \delta b \text{ for } \mu_v > \Lambda \text{ and } \mu_f > \delta\Lambda,$$
$$\mu_f - \delta\mu_v \leq \frac{\theta_f}{R-T+b} - \frac{\delta\theta_v}{(1-\delta)R}.$$

- Region VF:

$$\frac{\theta_v}{\mu_v} - \frac{(1-\delta)\theta_f}{\mu_f - \Lambda} \leq (1-\delta)(T-b) \quad \text{for } \mu_f > \Lambda,$$
$$\frac{\theta_v}{\mu_v - \Lambda} - \frac{(1-\delta)\theta_f}{\mu_f - \delta\Lambda} \geq (1-\delta)(T-b) \quad \text{for } \mu_v > \Lambda,$$
$$\mu_f + (1-\delta)\mu_v \geq \Lambda + \frac{\theta_f}{R-T+b} + \frac{\theta_v}{R},$$
$$\mu_f > \delta\Lambda.$$

Corollary EC.1. *The seven regions of (μ_f, μ_v) defined in Definition EC.2 are mutually exclusive and collectively exhaustive. And in each region, there exists a unique equilibrium strategy such that*

1. *When $(\mu_f, \mu_v) \in$ Region B, there exists a unique equilibrium strategy – strategy B – such that $\lambda_f = \lambda_v = 0$.*

2. *When* $(\mu_f, \mu_v) \in$ *Region* V*, there exists a unique equilibrium strategy – strategy* V *– such that* $\lambda_f = 0$ *and* $\lambda_v = \Lambda$*.*

3. *When* $(\mu_f, \mu_v) \in$ *Region* F*, there exists a unique equilibrium strategy – strategy* F *– such that* $\lambda_f = \Lambda$ *and* $\lambda_v = 0$*.*

4. *When* $(\mu_f, \mu_v) \in$ *Region* BVF*, there exists a unique equilibrium strategy – strategy* BVF *– such that*

$$\lambda_f = \mu_f - \delta\mu_v - \frac{\theta_f}{R-T+b} + \frac{\delta\theta_v}{(1-\delta)R}$$

and

$$\lambda_v = \mu_v - \frac{\theta_v}{(1-\delta)R}.$$

5. *When* $(\mu_f, \mu_v) \in$ *Region* BF*, there exists a unique equilibrium strategy – strategy BF – such that*

$$\lambda_f = \mu_f - \frac{\theta_f}{R-T+b}$$

and $\lambda_v = 0$*.*

6. *When* $(\mu_f, \mu_v) \in$ *Region* BV*, there exists a unique equilibrium strategy – strategy* BV *– such that* $\lambda_f = 0$ *and*

$$\lambda_v = \frac{\mathcal{B} - \sqrt{\mathcal{D}^2 + 4\delta^2\theta_f\theta_v}}{2\mathcal{A}},$$

where

$$\mathcal{A} = (R - \delta T + \delta b)\delta,$$

$$\mathcal{B} = (R - \delta T + \delta b)(\mu_f + \delta\mu_v) - \delta(\theta_f + \theta_v),$$

and

$$\mathcal{D} = (R - \delta T + \delta b)(\mu_f - \delta\mu_v) - \delta(\theta_f - \theta_v).$$

7. *When* $(\mu_f, \mu_v) \in$ *Region* VF*, there exists a unique equilibrium strategy – strategy VF – such that*

$$\lambda_f = \Lambda - \frac{-\tilde{\mathcal{B}} + \sqrt{\tilde{\mathcal{D}}^2 + 4\theta_f\theta_v}}{2\tilde{\mathcal{A}}}$$

and

$$\lambda_v = \frac{-\tilde{\mathcal{B}} + \sqrt{\tilde{\mathcal{D}}^2 + 4\theta_f\theta_v}}{2\tilde{\mathcal{A}}},$$

where

$$\tilde{\mathcal{A}} = (1-\delta)(T-b),$$

$$\tilde{\mathcal{B}} = (\theta_f + \theta_v) + (T-b)(\mu_f - \Lambda - (1-\delta)\mu_v),$$

and

$$\tilde{\mathcal{D}} = (\theta_f - \theta_v) + (T-b)(\mu_f - \Lambda + (1-\delta)\mu_v).$$

EC.1.3.2. The upper bounds and lower bounds for the optimal b in each scenario We first derive the upper bounds.

For Scenario NV, recall that in Corollary 1, we have

$$\lambda_f^{NV} = \min\{\Lambda, \mu - \frac{\theta_f}{R - T + b}\}.$$

To make sure λ_f^{NV} just to hit Λ, we have

$$b \leq \frac{\theta_f}{\mu - \Lambda} - R + T$$

when $\mu > \Lambda$. Since

$$\frac{\theta_f}{\mu - \Lambda} - R + T < \bar{b},$$

then we have

$$\bar{b}_{NV} = \begin{cases} r_f, & \text{if } \mu \leq \Lambda, \\ \min\{r_f, \frac{\theta_f}{\mu-\Lambda} - R + T\}, & \text{otherwise.} \end{cases} \tag{EC.23}$$

Then we derive the upper bound for Scenario NF. According to the formulation of λ_f^{NF} derived in Corollary 1, we have

$$\lambda_v^{NF} \leq \min\left\{\Lambda, \frac{\mu}{1+\delta} - \frac{(\sqrt{\delta\theta_f} + \sqrt{\theta_v})^2}{(1+\delta)(R - \delta T + \delta b)}\right\}. \tag{EC.24}$$

When

$$b = \frac{(\sqrt{\delta\theta_f} + \sqrt{\theta_v})^2}{\delta[\mu - (1+\delta)\Lambda]} - \frac{R}{\delta} + T,$$

the RHS of (EC.24) just hits Λ if $\mu > (1+\delta)\Lambda$. Then we finally have

$$\bar{b}_{NF} = \begin{cases} \frac{r_v}{\delta}, & \text{if } \mu \leq (1+\delta)\Lambda, \\ \min\{\frac{r_v}{\delta}, \frac{(\sqrt{\delta\theta_f} + \sqrt{\theta_v})^2}{\delta[\mu-(1+\delta)\Lambda]} - \frac{R}{\delta} + T, & \text{if } (1+\delta)\Lambda < \mu \leq (1+\delta)\Lambda + \frac{\theta_v}{(1-\delta)R}\}, \\ \min\{\frac{r_v}{\delta}, \bar{b}, \frac{(\sqrt{\delta\theta_f} + \sqrt{\theta_v})^2}{\delta[\mu-(1+\delta)\Lambda]} - \frac{R}{\delta} + T\}, & \text{otherwise.} \end{cases} \tag{EC.25}$$

As for the upper bound for Scenario VF, note that when $b = \bar{b}$, λ_v^{VF} just hits Λ. Thus we have

$$\bar{b}_{VF} = \begin{cases} \min\{r_f, \frac{r_v}{\delta}\}, & \text{if } \mu \leq (1+\delta)\Lambda + \frac{\theta_v}{(1-\delta)R}, \\ \min\{r_f, \frac{r_v}{\delta}, \frac{\theta_f}{\mu-(1+\delta)\Lambda - \frac{\theta_v}{(1-\delta)R}} - R + T\}, & \text{otherwise.} \end{cases} \tag{EC.26}$$

Now we derive the lower bounds, which serve as the smallest b that enables achieving each scenario.

The smallest b that enables achieving Scenario NV is to make sure $\lambda_f^{NV} \geq 0$. Thus we have

$$\underline{b}_{NV} = \frac{\theta_f}{\mu} - R + T. \tag{EC.27}$$

The smallest b that enables achieving Scenario NF is to make sure $\lambda_v^{NF} \geq 0$. Thus we have

$$\underline{b}_{NF} = \frac{\theta_v}{\delta\Phi(\mu)} + \frac{\theta_f}{\mu - \Phi(\mu)} - \frac{R}{\delta} + T, \tag{EC.28}$$

where $\Phi(\mu)$ is defined in (EC.22).

The smallest b that enables achieving Scenario VF is to make sure $\lambda_v^{VF} \geq 0$. Thus we have

$$\underline{b}_{VF} = \begin{cases} \infty, & \text{if } \mu \leq \Lambda + \frac{\theta_v}{(1-\delta)R}, \\ \frac{\theta_f}{\mu - \Lambda - \frac{\theta_v}{(1-\delta)R}} - R + T, & \text{otherwise.} \end{cases} \tag{EC.29}$$

EC.1.3.3. The thresholds for μ The following analysis narrows the options for the optimal scenario under a given total service capacity μ. Specifically, it is the capacity needed for a specific scenario, when b is set to be the profit margin.

The smallest capacity required to assure Scenario NV is defined by letting μ achieve Region BF when $b = r_f$, i.e.,

$$\underline{\mu}_{NV} = \frac{\theta_f}{R - T + r_f}.$$

The smallest capacity required to assure Scenario NF is defined by letting μ achieve Region BV when $b = r_v/\delta$, i.e., $\underline{\mu}_{NF} = \underline{\mu}^{BV}(r_v/\delta)$, where $\underline{\mu}^{BV}(r_v/\delta)$ is similar to $\underline{\mu}^{BV}$ which is defined in E-Companion EC.1.2. Specifically, $\underline{\mu}_{NF}$ is achieved by solving the following equation for μ:

$$R - \delta T + r_v - \frac{\theta_v}{\Phi(\mu)} - \frac{\delta\theta_f}{\mu - \Phi(\mu)} = 0,$$

where $\Phi(\mu)$ is defined in (EC.22).

The smallest capacity required to assure Scenario VF is defined by letting μ achieve Region VF when $b = \min\{r_f, r_v/\delta\}$, i.e.,

$$\underline{\mu}_{VF} = \Lambda + \frac{\theta_f}{R - T + \min\{r_f, \frac{r_v}{\delta}\}} + \frac{\theta_v}{(1-\delta)R}.$$

We also define

$$\bar{\mu} = (1+\delta)\Lambda + \frac{\theta_f}{R-T} + \frac{\theta_v}{(1-\delta)R}.$$

When $\mu \geq \bar{\mu}$, the provider is able to achieve Regions V and F without any subsidy. In other words, there is no need to consider the use of transportation subsidy if $\mu \geq \bar{\mu}$.

EC.2. Incorporating Heterogeneity

We consider that the travel cost t is a random variable which follows a uniform distribution. To simplify the notations, we normalize the scale of t such that $t \sim U[0,1]$. This normalization will not influence the results as we can also normalize R, θ_f and θ_v.

Suppose that the effective arrival rates are (λ_f, λ_v). For any $t \in [0,1]$, we have

$$U_f^t(\lambda_f, \lambda_v) = R - t - \frac{\theta_f}{\mu_f - \lambda_f - \delta\lambda_v},$$

$$U_v^t(\lambda_f, \lambda_v) = R - t - \frac{\delta\theta_f}{\mu_v - \lambda_v},$$

$$U_b^t(\lambda_f, \lambda_v) = 0.$$

It is straightforward to have the following Lemma.

Lemma EC.1.

- *If $U_f^t(\lambda_f, \lambda_v) \geq 0$ then $U_f^{t'}(\lambda_f, \lambda_v) > 0$ for any $t' < t$; and if $U_f^t(\lambda_f, \lambda_v) \leq 0$ then $U_f^{t'}(\lambda_f, \lambda_v) < 0$ for any $t' > t$.*
- *If $U_v^t(\lambda_f, \lambda_v) \geq 0$ then $U_v^{t'}(\lambda_f, \lambda_v) > 0$ for any $t' < t$; and if $U_v^t(\lambda_f, \lambda_v) \leq 0$ then $U_v^{t'}(\lambda_f, \lambda_v) < 0$ for any $t' > t$.*
- *If $U_f^t(\lambda_f, \lambda_v) \geq U_v^t(\lambda_f, \lambda_v)$, then $U_f^{t'}(\lambda_f, \lambda_v) > U_v^{t'}(\lambda_f, \lambda_v)$ for any $t' < t$; and if $U_f^t(\lambda_f, \lambda_v) \leq U_v^t(\lambda_f, \lambda_v)$, then $U_f^{t'}(\lambda_f, \lambda_v) < U_v^{t'}(\lambda_f, \lambda_v)$ for any $t' > t$.*

Following Lemma EC.1, Theorem EC.1 below establishes the existence and uniqueness of the equilibrium strategy for patients, given any service capacity allocation (μ_f, μ_v). We first define the following regions.

Definition EC.3.

- Region B:

$$\mu_f \leq \frac{\theta_f}{R}, \tag{EC.30}$$

$$\frac{\delta\theta_f}{\mu_f} + \frac{\theta_v}{\mu_v} \geq R. \tag{EC.31}$$

- Region V:

$$\theta_v\mu_f - (1-\delta)\theta_f\mu_v \leq [\delta\theta_v - (1-\delta)\theta_f]\Lambda, \tag{EC.32}$$

$$\frac{\delta\theta_f}{\mu_f - \delta\Lambda} + \frac{\theta_v}{\mu_v - \Lambda} \leq R - \delta, \tag{EC.33}$$

$$\mu_f > \delta\Lambda, \tag{EC.34}$$

$$\mu_v > \Lambda. \tag{EC.35}$$

- Region F:

$$\mu_f \geq \frac{\theta_f}{R-1} + \Lambda, \tag{EC.36}$$

$$\frac{\theta_v}{\mu_v} - \frac{(1-\delta)\theta_f}{\mu_f - \Lambda} \geq 1-\delta. \tag{EC.37}$$

- Region BVF:

$$\mu_v > \frac{\theta_v}{(1-\delta)R}, \tag{EC.38}$$

$$\Lambda + \frac{2\delta\theta_v\Lambda}{(1-\delta)\left[R\Lambda - \delta\mu_v - \sqrt{(R\Lambda-\delta\mu_v)^2 + \frac{4\delta\theta_v\Lambda}{(1-\delta)}}\right]} + \frac{2\delta\theta_f\Lambda}{2\delta(\mu_f-\Lambda) + (1-\delta)\left[R\Lambda + \delta\mu_v - \sqrt{(R\Lambda-\delta\mu_v)^2 + \frac{4\delta\theta_v\Lambda}{(1-\delta)}}\right]}$$
$$> \frac{R\Lambda + \delta\mu_v - \sqrt{(R\Lambda-\delta\mu_v)^2 + \frac{4\delta\theta_v\Lambda}{(1-\delta)}}}{2\delta} \quad \text{for } (1-\delta)\left[R\Lambda + \delta\mu_v - \sqrt{(R\Lambda-\delta\mu_v)^2 + \frac{4\delta\theta_v\Lambda}{(1-\delta)}}\right] > 2\delta(\Lambda - \mu_f), \tag{EC.39}$$

and, when $(1-\delta)\theta_f - \delta\theta_v > 0$:

$$[(1-\delta)\theta_f - \delta\theta_v]\, R\Lambda - \delta\,[(1-\delta)\theta_f\mu_v - \theta_v\mu_f] - \frac{[(1-\delta)\theta_f - \delta\theta_v]^2\,\Lambda}{(1-\delta)(\mu_f - \delta\mu_v)} > 0 \quad \text{for } (1-\delta)\theta_f\mu_v - \theta_v\mu_f \geq 0, \tag{EC.40}$$

$$\mu_f > \delta\mu_v \quad \text{for } (1-\delta)\theta_f\mu_v - \theta_v\mu_f \geq 0; \tag{EC.41}$$

when $(1-\delta)\theta_f - \delta\theta_v < 0$:

$$[(1-\delta)\theta_f - \delta\theta_v]\, R\Lambda - \delta\,[(1-\delta)\theta_f\mu_v - \theta_v\mu_f] - \frac{[(1-\delta)\theta_f - \delta\theta_v]^2\,\Lambda}{(1-\delta)(\mu_f - \delta\mu_v)} > 0 \quad \text{for } \mu_f < \delta\mu_v, \tag{EC.42}$$

$$(1-\delta)\theta_f\mu_v - \theta_v\mu_f < 0. \tag{EC.43}$$

- Region BF:

$$\mu_f > \frac{\theta_f}{R} \tag{EC.44}$$

$$\mu_f < \frac{\theta_f}{R-1} + \Lambda, \tag{EC.45}$$

$$\mu_v \leq \frac{\theta_v}{(1-\delta)R}. \tag{EC.46}$$

- Region BV:

$$\frac{\delta\theta_f}{\mu_f} + \frac{\theta_v}{\mu_v} < R, \tag{EC.47}$$

$$\frac{\delta\theta_f}{\mu_f - \delta\Lambda} + \frac{\theta_v}{\mu_v - \Lambda} > R - \delta \quad \text{for } \mu_f > \delta\Lambda \text{ and } \mu_v > \Lambda, \tag{EC.48}$$

and, when $(1-\delta)\theta_f - \delta\theta_v > 0$:

$$[(1-\delta)\theta_f - \delta\theta_v]\, R\Lambda - \delta\,[(1-\delta)\theta_f\mu_v - \theta_v\mu_f] - \frac{[(1-\delta)\theta_f - \delta\theta_v]^2\,\Lambda}{(1-\delta)(\mu_f - \delta\mu_v)} \leq 0 \quad \text{for } \mu_f > \delta\mu_v, \tag{EC.49}$$

$$(1-\delta)\theta_f\mu_v - \theta_v\mu_f \geq 0; \tag{EC.50}$$

when $(1-\delta)\theta_f - \delta\theta_v < 0$:

$$[(1-\delta)\theta_f - \delta\theta_v]\, R\Lambda - \delta\,[(1-\delta)\theta_f\mu_v - \theta_v\mu_f] - \frac{[(1-\delta)\theta_f - \delta\theta_v]^2\,\Lambda}{(1-\delta)(\mu_f - \delta\mu_v)} \le 0 \quad \text{for } (1-\delta)\theta_f\mu_v < \theta_v\mu_f, \tag{EC.51}$$

$$\mu_f < \delta\mu_v \quad \text{for } (1-\delta)\theta_f\mu_v < \theta_v\mu_f. \tag{EC.52}$$

- Region VF:

$$\theta_v\mu_f - (1-\delta)\theta_f\mu_v > [\delta\theta_v - (1-\delta)\theta_f]\,\Lambda, \tag{EC.53}$$

$$\frac{\theta_v}{\mu_v} - \frac{(1-\delta)\theta_f}{\mu_f - \Lambda} < 1-\delta \quad \text{for } \mu_f > \Lambda, \tag{EC.54}$$

$$\Lambda + \frac{2\delta\theta_v\Lambda}{(1-\delta)\left[R\Lambda - \delta\mu_v - \sqrt{(R\Lambda-\delta\mu_v)^2 + \frac{4\delta\theta_v\Lambda}{(1-\delta)}}\right]} + \frac{2\delta\theta_f\Lambda}{2\delta(\mu_f-\Lambda) + (1-\delta)\left[R\Lambda + \delta\mu_v - \sqrt{(R\Lambda-\delta\mu_v)^2 + \frac{4\delta\theta_v\Lambda}{(1-\delta)}}\right]}$$
$$\le \frac{R\Lambda + \delta\mu_v - \sqrt{(R\Lambda-\delta\mu_v)^2 + \frac{4\delta\theta_v\Lambda}{(1-\delta)}}}{2\delta}, \tag{EC.55}$$

$$(1-\delta)\left[R\Lambda + \delta\mu_v - \sqrt{(R\Lambda-\delta\mu_v)^2 + \frac{4\delta\theta_v\Lambda}{(1-\delta)}}\right] > 2\delta(\Lambda - \mu_f). \tag{EC.56}$$

THEOREM EC.1. *The seven regions of (μ_f, μ_v) defined above are mutually exclusive and collectively exhaustive. And in each region, there exists a unique equilibrium strategy such that*

1. *When $(\mu_f, \mu_v) \in$ Region B, there exists a unique equilibrium strategy – strategy B – such that $\lambda_f = \lambda_v = 0$.*

2. *When $(\mu_f, \mu_v) \in$ Region V, there exists a unique equilibrium strategy – strategy V – such that $\lambda_v = \Lambda$.*

3. *When $(\mu_f, \mu_v) \in$ Region F, there exists a unique equilibrium strategy – strategy F – such that $\lambda_f = \Lambda$.*

4. *When $(\mu_f, \mu_v) \in$ Region BVF, there exists a unique equilibrium strategy – strategy BVF – such that $\lambda_f + \lambda_v \le \Lambda$.*

   The effective arrival rates (λ_f, λ_v) are

$$\lambda_f = \tfrac{1}{2}\left[\mu_f - \delta\mu_v - \sqrt{(R\Lambda - \mu_f)^2 + 4\theta_f\Lambda} + \delta\sqrt{\left(\frac{R\Lambda}{\delta} - \mu_v\right)^2 + \frac{4\theta_v\Lambda}{\delta(1-\delta)}}\right]$$
$$\lambda_v = \tfrac{1}{2}\left[\frac{R\Lambda}{\delta} + \mu_v - \sqrt{\left(\frac{R\Lambda}{\delta} - \mu_v\right)^2 + \frac{4\theta_v\Lambda}{\delta(1-\delta)}}\right].$$

5. *When $(\mu_f, \mu_v) \in$ Region BF, there exists a unique equilibrium strategy – strategy BF – such that $\lambda_f \le \Lambda$, $\lambda_v = 0$. The effective arrival rate λ_f is*

$$\lambda_f = \tfrac{1}{2}\left[R\Lambda + \mu_f - \sqrt{(R\Lambda - \mu_f)^2 + 4\theta_f\Lambda}\right].$$

6. *When* $(\mu_f, \mu_v) \in$ *Region* BV*, there exists a unique equilibrium strategy – strategy BV – such that* $\lambda_f = 0,\ \lambda_v \le \Lambda$*. The effective arrival rate* λ_v *can be derived by solving*

$$\frac{R\Lambda}{\delta} - \frac{\theta_v \Lambda}{\delta(\mu_v - \lambda_v)} - \frac{\theta_f \Lambda}{\mu_f - \delta\lambda_v} = \lambda_v$$

7. *When* $(\mu_f, \mu_v) \in$ *Region* VF*, there exists a unique equilibrium strategy – strategy VF – such that* $\lambda_f + \lambda_v = \Lambda$.

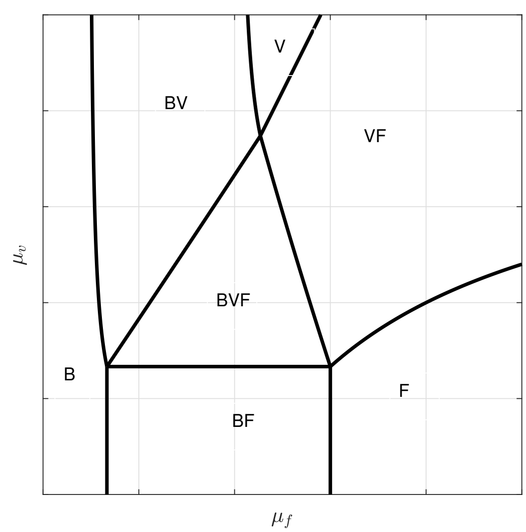
The effective arrival rates (λ_f, λ_v) *can be derived from solving the equations* $\lambda_f + \lambda_v = \Lambda$ *and*

$$\Lambda - \frac{\theta_v \Lambda}{(1-\delta)(\mu_v - \lambda_v)} + \frac{\theta_f \Lambda}{\mu_f - \Lambda + (1-\delta)\lambda_v} = \lambda_v.$$

Theorem EC.1 establishes that all possible capacity allocations fall into seven mutually exclusive and collectively exhaustive cases. Letting the x-axis represent μ_f and the y-axis represent μ_v, Figure EC.1 illustrates these seven regions.

Although the formulations are more complex in the case of heterogeneous travel costs, the structure of the seven regions remains very similar to that described in Theorem 1.

Figure EC.1 Illustration of the seven equilibrium regions in the case of uniformly distributed travel costs.



EC.3. Proofs of Analytical Results

Proof of Theorem 1: By Definition EC.1, it is easy to see (EC.1) v.s. (EC.15) divide the (μ_f, μ_f) space; (EC.2) v.s. (EC.12) divide the (μ_f, μ_f) space; (EC.3) with (EC.6) and v.s. (EC.19) with (EC.21) divide the (μ_f, μ_f) space; (EC.4) with (EC.5) and (EC.6) v.s. (EC.16) divide the (μ_f, μ_f) space; (EC.7) v.s. (EC.13) divide the (μ_f, μ_f) space; (EC.8) with $\mu_f \geq \Lambda$ v.s. (EC.18) divide the (μ_f, μ_f) space; (EC.9) v.s. (EC.14) divide the (μ_f, μ_f) space; (EC.10) v.s. (EC.17) divide the (μ_f, μ_f) space; (EC.11) v.s. (EC.20) divide the (μ_f, μ_f) space. Thus the seven regions of (μ_f, μ_v) defined in Definition EC.1 are mutually exclusive and collectively exhaustive.

We conduct the analysis for each region separately:

Region B: Since

$$\frac{\theta_v}{\mu_v} + \frac{\delta\theta_f}{\mu_f} \geq R - \delta T \quad \text{and} \quad \mu_f \leq \frac{\theta_f}{R-T},$$

we get that $U_f(0,0) \leq 0$ and $U_v(0,0) \leq 0$ and therefore, there exists a unique equilibrium such that $\lambda_f = \lambda_v = 0$.

Region V: Since

$$\frac{\theta_v}{\mu_v - \Lambda} - \frac{(1-\delta)\theta_f}{\mu_f - \delta\Lambda} \leq (1-\delta)T, \quad \frac{\theta_v}{\mu_v - \Lambda} + \frac{\delta\theta_f}{\mu_f - \delta\Lambda} \leq R - \delta T,$$
$$\mu_f \geq \delta\Lambda, \quad \text{and} \quad \mu_v \geq \Lambda,$$

we get that $U_f(0,\Lambda) \leq U_v(0,\Lambda)$ and $U_v(0,\Lambda) \geq 0$ and therefore, there exists a unique equilibrium such that $\lambda_f = 0$ and $\lambda_v = \Lambda$.

Region F: Since

$$\mu_f \geq \Lambda + \frac{\theta_f}{R-T} \quad \text{and} \quad \frac{\theta_v}{\mu_v} - \frac{(1-\delta)\theta_f}{\mu_f - \Lambda} \geq (1-\delta)T,$$

we get that $U_f(\Lambda,0) \geq U_v(\Lambda,0)$ and $U_f(\Lambda,0) \geq 0$ and therefore, there exists a unique equilibrium such that $\lambda_f = \Lambda$ and $\lambda_v = 0$.

Region BVF: The equality $U_v(\lambda_f,\lambda_v) = U_f(\lambda_f,\lambda_v) = 0$ yields the following conditions.

$$\lambda_v = \mu_v - \frac{\theta_v}{(1-\delta)R} \quad \text{and} \quad \lambda_f + \delta\lambda_v = \mu_f - \frac{\theta_f}{R-T}.$$

Then we have

$$\lambda_f = \mu_f - \frac{\theta_f}{R-T} - \delta\mu_v + \frac{\delta\theta_v}{(1-\delta)R}.$$

The constraints of Region BVF:

$$\mu_v \geq \frac{\theta_v}{(1-\delta)R}, \quad \mu_f - \delta\mu_v \geq \frac{\theta_f}{R-T} - \frac{\delta\theta_v}{(1-\delta)R}, \quad \mu_f + (1-\delta)\mu_v \leq \Lambda + \frac{\theta_f}{R-T} + \frac{\theta_v}{R},$$

guarantee $\lambda_v \geq 0$, $\lambda_f \geq 0$ and $\lambda_v + \lambda_f \leq \Lambda$. Thus, there exists a unique equilibrium such that $\lambda_v + \lambda_f \leq \Lambda$.

Region BF: The fact that $U_f(\lambda_f, 0) = 0$ yields the following result:

$$\lambda_f = \mu_f - \frac{\theta_f}{R-T}.$$

The constraints of Region BF:

$$\frac{\theta_f}{R-T} \leq \mu_f \leq \Lambda + \frac{\theta_f}{R-T}, \quad \text{and} \quad \mu_v \leq \frac{\theta_v}{(1-\delta)R}$$

guarantee $0 \leq \lambda_f \leq \Lambda$ and $U_v(\lambda_f, 0) \leq 0$. Thus, there exists a unique equilibrium such that $0 \leq \lambda_f \leq \Lambda$ and $\lambda_v = 0$.

Region BV: The fact that $U_v(0, \lambda_v) = 0$ yields the following conditions:

$$R - \delta T - \frac{\theta_v}{\mu_v - \lambda_v} - \frac{\delta\theta_f}{\mu_f - \delta\lambda_v} = 0, \tag{EC.57}$$

$$\mu_v - \lambda_v > 0 \quad \text{and} \quad \mu_f - \delta\lambda_v > 0.$$

Next, we show that

1) There is only one λ_v which solves (EC.57) and satisfies $\mu_f - \delta\lambda_v > 0$ and $\mu_v - \lambda_v > 0$.

2) With the constraints of Region BV, we must have $0 \leq \lambda_v \leq \Lambda$ and $U_f(0, \lambda_v) \leq 0$.

Thus, there exists a unique equilibrium such that $0 \leq \lambda_v \leq \Lambda$ and $\lambda_f = 0$.

We start with proving 1): There is only one λ_v which solves (EC.57) and satisfies $\mu_f - \delta\lambda_v > 0$ and $\mu_v - \lambda_v > 0$.

Specifically, by solving (EC.57), we have

$$\lambda_v = \frac{\mathcal{B} - \sqrt{\mathcal{D}^2 + 4\delta^2\theta_f\theta_v}}{2\mathcal{A}}, \tag{EC.58}$$

where

$$\mathcal{A} = (R - \delta T)\delta, \quad \mathcal{B} = (R - \delta T)(\mu_f + \delta\mu_v) - \delta(\theta_f + \theta_v),$$

$$\mathcal{C} = (R - \delta T)\mu_f\mu_v - \theta_v\mu_f - \delta\theta_f\mu_v, \quad \mathcal{D} = (R - \delta T)(\mu_f - \delta\mu_v) - \delta(\theta_f - \theta_v).$$

Let us rewrite (EC.57) as

$$(R - \delta T)\delta\lambda_v^2 - [(R - \delta T)(\mu_f + \delta\mu_v) - \delta(\theta_f + \theta_v)]\lambda_v + [(R - \delta T)\mu_f\mu_v - \theta_v\mu_f - \delta\theta_f\mu_v] = 0,$$

or alternatively, $\mathcal{A}\lambda_v^2 - \mathcal{B}\lambda_v + \mathcal{C} = 0$. Note that

$$\mathcal{B}^2 - 4\mathcal{A}\mathcal{C} = [(R - \delta T)(\mu_f - \delta\mu_v) - \delta(\theta_f - \theta_v)]^2 + 4\delta^2\theta_f\theta_v = \mathcal{D}^2 + 4\delta^2\theta_f\theta_v \geq 0.$$

Then, (EC.57) has two possible solutions:

$$\lambda_v^{BV.1} = \frac{\mathcal{B} + \sqrt{\mathcal{D}^2 + 4\delta^2\theta_f\theta_v}}{2\mathcal{A}} \quad \text{and} \quad \lambda_v^{BV.2} = \frac{\mathcal{B} - \sqrt{\mathcal{D}^2 + 4\delta^2\theta_f\theta_v}}{2\mathcal{A}}.$$

There are also two cases for $\mathcal{D}$:

- **Case 1:** $\mathcal{D} > 0$

—If $\lambda_v = \lambda_v^{BV.1}$,

$$\lambda_v^{BV.1} > \frac{\mathcal{B} + \mathcal{D}}{2\mathcal{A}} = \frac{(R-\delta T)\mu_f - \delta\theta_f}{(R-\delta T)\delta}.$$

Then, we have $\mu_f - \delta\lambda_v^{BV.1} < \frac{\delta\theta_f}{R-\delta T}$, which contradicts with either (EC.57) or $\mu_f - \delta\lambda_v > 0$.

—If $\lambda_v = \lambda_v^{BV.2}$,

$$\lambda_v^{BV.2} < \frac{\mathcal{B} - \mathcal{D}}{2\mathcal{A}} = \frac{(R-\delta T)\mu_v - \theta_v}{(R-\delta T)}.$$

Then, we have $\mu_v - \lambda_v^{BV.2} > \frac{\theta_v}{R-\delta T} > 0$, and $\lambda_v^{BV.2} < \frac{\mu_f}{\delta}$ (since $\mathcal{D} > 0$), which validates (EC.57), $\mu_v > \lambda_v$ and $\mu_f - \delta\lambda_v > 0$.

- **Case 2:** $\mathcal{D} \le 0$

—If $\lambda_v = \lambda_v^{BV.1}$,

$$\lambda_v^{BV.1} > \frac{\mathcal{B} - \mathcal{D}}{2\mathcal{A}} = \frac{(R-\delta T)\mu_v - \theta_v}{(R-\delta T)}.$$

Then, we have $\mu_v - \lambda_v^{BV.1} < \frac{\theta_v}{R-\delta T}$, which contradicts with (EC.57) or $\lambda_v < \mu_v$.

—If $\lambda_v = \lambda_v^{BV.2}$,

$$\lambda_v^{BV.2} < \frac{\mathcal{B} + \mathcal{D}}{2\mathcal{A}} = \frac{(R-\delta T)\mu_f - \delta\theta_f}{(R-\delta T)\delta}.$$

Then, we have $\mu_f - \delta\lambda_v^{BV.2} > \frac{\delta\theta_f}{R-\delta T} > 0$, and $\lambda_v^{BV.2} < \mu_v$ (by $\mathcal{D} \le 0$), which validates (EC.57), $\mu_v > \lambda_v$ and $\mu_f - \delta\lambda_v > 0$.

In sum, $\lambda_v = \lambda_v^{BV.2}$ is the only solution which solves (EC.57) and satisfies $\mu_f - \delta\lambda_v > 0$ and $\mu_v - \lambda_v > 0$.

Next, we move on to proving 2): With the constraints of Region BV, we must have $0 \le \lambda_v \le \Lambda$ and $U_f(0, \lambda_v) \le 0$; specifically, we prove the following:

2.1) $\lambda_v \ge 0$. Recall the constraints of Region BV:

$$\frac{\theta_v}{\mu_v} + \frac{\delta\theta_f}{\mu_f} \le R - \delta T, \quad \mu_f - \delta\mu_v \le \frac{\theta_f}{R-T} - \frac{\delta\theta_v}{(1-\delta)R},$$

$$\frac{\theta_v}{\mu_v - \Lambda} + \frac{\delta\theta_f}{\mu_f - \delta\Lambda} \ge R - \delta T \text{ for } \mu_v > \Lambda \text{ and } \mu_f > \delta\Lambda.$$

Then we have $\mathcal{A} > 0$, $\mathcal{B} > 0$ and $\mathcal{C} \ge 0$. Since $\mathcal{B}^2 - 4\mathcal{A}\mathcal{C} \ge 0$, then we have $\mathcal{B} - \sqrt{\mathcal{B}^2 - 4\mathcal{A}\mathcal{C}} \ge 0$. Thus $\lambda_v \ge 0$.

2.2) $\lambda_v \le \Lambda$. If $\mu_v \le \Lambda$ or $\mu_f \le \delta\Lambda$, then $\lambda_v < \Lambda$ since $\mu_v > \lambda_v$ and $\mu_f > \delta\lambda_v$.

If $\mu_v > \Lambda$ and $\mu_f > \delta\Lambda$, by (EC.57) and the constraint

$$\frac{\theta_v}{\mu_v - \Lambda} + \frac{\delta\theta_f}{\mu_f - \delta\Lambda} \geq R - \delta T \text{ for } \mu_v > \Lambda \text{ and } \mu_f > \delta\Lambda,$$

we must have $\lambda_v \leq \Lambda$.

2.3) $U_f(0, \lambda_v) \leq 0$. By (EC.57) we have

$$\delta(R - T - \frac{\theta_f}{\mu_f - \delta\lambda_v}) = \frac{\theta_v}{\mu_v - \lambda_v} - (1-\delta)R.$$

Suppose

$$U_f(0, \lambda_v) = R - T - \frac{\theta_f}{\mu_f - \delta\lambda_v} > 0,$$

we must have

$$\frac{\theta_v}{\mu_v - \lambda_v} - (1-\delta)R > 0.$$

Then we should have

$$\mu_f - \delta\lambda_v > \frac{\theta_f}{R-T}$$

and

$$\mu_v - \lambda_v < \frac{\theta_v}{(1-\delta)R},$$

which contradict with

$$\mu_f - \delta\mu_v \leq \frac{\theta_f}{R-T} - \frac{\delta\theta_v}{(1-\delta)R}.$$

Thus we must have $U_f(0, \lambda_v) \leq 0$.

Region VF: The fact that $\lambda_v + \lambda_f = \Lambda$ and $U_f(\lambda_f, \lambda_v) = U_v(\lambda_f, \lambda_v)$ yield the following conditions

$$\frac{\theta_v}{\mu_v - \lambda_v} - \frac{(1-\delta)\theta_f}{\mu_f - \Lambda + (1-\delta)\lambda_v} = (1-\delta)T, \tag{EC.59}$$

$$\frac{\Lambda - \mu_f}{1-\delta} < \lambda_v < \mu_v.$$

Recall the constraints of Region VF:

$$\begin{aligned}
&\frac{\theta_v}{\mu_v} - \frac{(1-\delta)\theta_f}{\mu_f - \Lambda} \leq (1-\delta)T \text{ for } \mu_f > \Lambda, \\
&\frac{\theta_v}{\mu_v - \Lambda} - \frac{(1-\delta)\theta_f}{\mu_f - \delta\Lambda} \geq (1-\delta)T \text{ for } \mu_v > \Lambda, \\
&\mu_f + (1-\delta)\mu_v \geq \Lambda + \frac{\theta_f}{R-T} + \frac{\theta_v}{R}, \text{ and} \\
&\mu_f \geq \delta\Lambda.
\end{aligned}$$

Since the left-hand side of (EC.59) is increasing in λ_v when $(\Lambda-\mu_f)/(1-\delta)<\lambda_v<\mu_v$. By the first two constraints of Region VF, there must be a unique λ_v such that $0\le\lambda_v\le\Lambda$ solving (EC.59). Suppose that $U_v(\Lambda-\lambda_v,\lambda_v)<0$, by (EC.59) we have

$$U_f(\Lambda-\lambda_v,\lambda_v)=R-T-\frac{\theta_f}{\mu_f-\Lambda+(1-\delta)\lambda_v}<0$$

and

$$(1-\delta)R-\frac{\theta_v}{\mu_v-\lambda_v}<0.$$

Then we should have

$$\mu_f+(1-\delta)\mu_v<\Lambda+\frac{\theta_f}{R-T}+\frac{\theta_v}{R}$$

which contradicts with the third constraint of Region VF. Thus we must have $U_f(\Lambda-\lambda_v,\lambda_v)=U_v(\Lambda-\lambda_v,\lambda_v)\ge 0$.

Thus, there exists a unique equilibrium such that $\lambda_f+\lambda_v=\Lambda$. Specifically, solving (EC.59) gives

$$\lambda_v=\frac{-\tilde{\mathcal{B}}+\sqrt{\tilde{\mathcal{D}}^2+4\theta_f\theta_v}}{2\tilde{\mathcal{A}}}, \tag{EC.60}$$

where

$$\tilde{\mathcal{A}}=(1-\delta)T,\quad \tilde{\mathcal{B}}=(\theta_f+\theta_v)+T(\mu_f-\Lambda-(1-\delta)\mu_v),$$

$$\tilde{\mathcal{C}}=\frac{\theta_v(\mu_f-\Lambda)}{(1-\delta)}-T\mu_v(\mu_f-\Lambda)-\theta_f\mu_v,\quad\text{and}\quad \tilde{\mathcal{D}}=(\theta_f-\theta_v)+T(\mu_f-\Lambda+(1-\delta)\mu_v).$$

Q.E.D.

Proof of Proposition 1: It directly follows Definition EC.1 and Theorem 1. Q.E.D.

Proof of Proposition 2: Recall that in Region BVF, $0\le\lambda_f+\lambda_v\le\Lambda$. The provider's objective is to maximize $r_f\lambda_f+r_v\lambda_v$. Since in this region we have

$$\lambda_v=\mu_v-\frac{\theta_v}{(1-\delta)R},\quad\text{and}\quad \lambda_f=\mu_f-\delta\mu_v-\frac{\theta_f}{R-T}+\frac{\theta_v}{(1-\delta)R},$$

the objective function can be rewritten as

$$r_f\left(\mu_f-\delta\mu_v-\frac{\theta_f}{R-T}+\frac{\theta_v}{(1-\delta)R}\right)+r_v\left(\mu_v-\frac{\theta_v}{(1-\delta)R}\right),$$

which by substituting $\mu_v=\mu-\mu_f$ and removing the constants is equivalent to

$$\left[(1+\delta)r_f-r_v\right]\mu_f.$$

Therefore, the provider's problem is

$$\max_{0\le\mu_f\le\mu} \quad [(1+\delta)r_f - r_v]\,\mu_f \tag{EC.61}$$
$$s.t. \quad \mu-\mu_f \ge \frac{\theta_v}{(1-\delta)R},$$
$$\mu_f - \delta(\mu-\mu_f) \ge \frac{\theta_f}{R-T} - \frac{\delta\theta_v}{(1-\delta)R},$$
$$\mu_f + (1-\delta)(\mu-\mu_f) \le \Lambda + \frac{\theta_f}{R-T} + \frac{\theta_v}{R}.$$

The optimal solution then depends on the sign of the objective function and the boundaries of μ_f in (EC.61). Q.E.D.

Proof of Proposition 3: Before characterizing the optimal capacity allocation in Region BV, we first provide an auxiliary result. Let $\lambda_v^{BV}(\mu-\mu_v,\mu_v)$ denote the unique effective arrival rate to the virtual channel (which is defined in (EC.58) and here we denote it as λ_v^{BV}) by substituting μ_f by $\mu-\mu_v$. Lemma EC.2 proves that $\lambda_v^{BV}(\mu-\mu_v,\mu_v)$ is a unimodal function in μ_v, which is maximized at the unique value $\mu_v=\tilde{\mu}_v^{BV}$.

Lemma EC.2. *Given (μ_f,μ_v) which satisfies (EC.57) and (EC.15)–(EC.17), there exists a unique $\tilde{\mu}_v^{BV}$ such that $\lambda_v^{BV}(\mu-\mu_v,\mu_v)$ is increasing in μ_v when $\mu_v\le\tilde{\mu}_v^{BV}$ and $\lambda_v^{BV}(\mu-\mu_v,\mu_v)$ is decreasing in μ_v when $\mu_v\ge\tilde{\mu}_v^{BV}$.*

In Proposition 3, when

$$\mu \ge \Lambda(1+\delta) + \min\left\{\frac{\theta_f}{R-T}+\frac{\theta_v}{(1-\delta)R}, \frac{(\sqrt{\delta\theta_f}+\sqrt{\theta_v})^2}{R-\delta T}\right\},$$

the results are straightforward as Λ can be achieved.

When

$$\mu < \Lambda(1+\delta) + \min\left\{\frac{\theta_f}{R-T}+\frac{\theta_v}{(1-\delta)R}, \frac{(\sqrt{\delta\theta_f}+\sqrt{\theta_v})^2}{R-\delta T}\right\}, \tag{EC.62}$$

we observe that the optimal solution is $\tilde{\mu}_v^{BV}$ without constraints (EC.15)–(EC.17) (it directly follows from Lemma EC.2). Next, we have to regularize μ_v^{BV*} to make (EC.15)–(EC.17) valid. Recall the constraints (EC.15)–(EC.17):

$$\frac{\theta_v}{\mu_v}+\frac{\delta\theta_f}{\mu_f} \le R-\delta T, \quad \mu_f - \delta\mu_v \le \frac{\theta_f}{R-T} - \frac{\delta\theta_v}{(1-\delta)R},$$

$$\frac{\theta_v}{\mu_v-\Lambda}+\frac{\delta\theta_f}{\mu_f-\delta\Lambda} \ge R-\delta T \text{ for } \mu_v>\Lambda \text{ and } \mu_f>\delta\Lambda.$$

Note that $\tilde{\mu}_v^{BV}$ must satisfy the first condition, otherwise no revenue is generated; $\tilde{\mu}_v^{BV}$ must satisfy the third condition, by (EC.62). When $(\mu-\tilde{\mu}_v^{BV}, \tilde{\mu}_v^{BV}) \in$ Region BV, the second constraint is also satisfied, then $(\mu-\tilde{\mu}_v^{BV}, \tilde{\mu}_v^{BV})$ is the optimal solution, yielding

$$\lambda_v^{BV*} = \frac{\mu}{1+\delta} - \frac{(\sqrt{\delta\theta_f}+\sqrt{\theta_v})^2}{(1+\delta)(R-\delta T)}.$$

When $(\mu-\tilde{\mu}_v^{BV}, \tilde{\mu}_v^{BV}) \notin$ Region BV, the second constraint is not satisfied, i.e, the boundary solution would be the optimal one:

$$\frac{\mu}{(1+\delta)} - \frac{\theta_f}{(1+\delta)(R-T)} + \frac{\delta\theta_v}{(1+\delta)(1-\delta)R},$$

yielding

$$\lambda_v^{BV*} = \frac{\mu}{1+\delta} - \frac{\theta_f}{(1+\delta)(R-T)} - \frac{\theta_v}{(1+\delta)(1-\delta)R}.$$

Q.E.D.

Proof of Lemma EC.2: Recall that given (μ_f, μ_v) which satisfies (EC.57) and (EC.15)–(EC.17), we have

$$\lambda_v^{BV} = \frac{\mathcal{B} - \sqrt{\mathcal{D}^2 + 4\delta^2\theta_f\theta_v}}{2\mathcal{A}}.$$

By replacing μ_f with $\mu-\mu_v$, we have

$$\frac{\partial \mathcal{B}}{\partial \mu_v} = -(R-\delta T)(1-\delta) < 0, \quad \text{and} \quad \frac{\partial \mathcal{D}}{\partial \mu_v} = -(R-\delta T)(1+\delta) < 0.$$

Therefore, when $\mathcal{D} < 0$, λ_v^{BV} is decreasing in μ_v. Note that $\mathcal{A}$ is independent of μ_v.

When $\mathcal{D} \geq 0$,

$$\frac{\partial \lambda_v^{BV}}{\partial \mu_v} = -\frac{-(1-\delta)}{2\delta} + \frac{(1+\delta)\mathcal{D}}{2\delta\sqrt{\mathcal{D}^2 + 4\delta^2\theta_f\theta_v}}.$$

Since $\mathcal{D} \geq 0$, then $\frac{\partial \lambda_v^{BV}}{\partial \mu_v}$ is increasing in $\mathcal{D}$. And when $\mathcal{D} = 0$, $\frac{\partial \lambda_v^{BV}}{\partial \mu_v} = -\frac{-(1-\delta)}{2\delta} < 0$. Let $\bar{D}$ solves $\frac{\partial \lambda_v^{BV}}{\partial \mu_v} = 0$, we must have $\bar{D} > 0$. So when $0 \leq \mathcal{D} < \bar{D}$, $\frac{\partial \lambda_v^{BV}}{\partial \mu_v} < 0$; and when $\mathcal{D} > \bar{D}$, $\frac{\partial \lambda_v^{BV}}{\partial \mu_v} > 0$.

When

$$\mu_v = \tilde{\mu}_v^{BV} = \frac{\mu}{1+\delta} - \frac{\delta(\theta_f-\theta_v) + (1-\delta)\sqrt{\delta\theta_f\theta_v}}{(1+\delta)(R-\delta T)},$$

$\mathcal{D} = \bar{D}$. Since $\mathcal{D}$ is decreasing in μ_v, we can conclude that when $\mu_v \leq \tilde{\mu}_v^{BV}$, λ_v^{BV} is increasing in μ_v; when $\mu_v \geq \tilde{\mu}_v^{BV}$, λ_v^{BV} is decreasing in μ_v. Q.E.D.

Proof of Proposition 4: Before characterizing the optimal capacity allocation in Region VF, we first provide an auxiliary result. Let $\lambda_v^{VF}(\mu-\mu_v, \mu_v)$ denote the unique effective arrival rate to the virtual channel (which is defined in (EC.60) and here we denote it as λ_v^{VF}) by substituting μ_f by $\mu-\mu_v$. Lemma EC.3 proves that $\lambda_v^{VF}(\mu-\mu_v, \mu_v)$ is increasing in μ_v.

Lemma EC.3. *Given (μ_f, μ_v) which satisfies (EC.59) and (EC.18)–(EC.21), λ_v^{VF} is increasing in μ_v.*

Since $\lambda_v^{VF}(\mu - \mu_v, \mu_v)$ is increasing in μ_v and the objective function is monotone in λ_v^{VF}, the optimal solution must be on the boundaries if $r_v \neq r_f$. Q.E.D.

Proof of Lemma EC.3: Recall that given (μ_f, μ_v) which satisfies (EC.59) and (EC.18)–(EC.21), we have

$$\lambda_v^{VF} = \frac{-\tilde{\mathcal{B}} + \sqrt{\tilde{\mathcal{D}}^2 + 4\theta_f\theta_v}}{2\tilde{\mathcal{A}}}.$$

By replacing μ_f with $\mu - \mu_v$, we have

$$\frac{\partial \tilde{\mathcal{B}}}{\partial \mu_v} = -(2-\delta)T < 0, \quad \text{and} \quad \frac{\partial \tilde{\mathcal{D}}}{\partial \mu_v} = -\delta T < 0.$$

Then,

$$\frac{\partial(-\tilde{\mathcal{B}} + \sqrt{\tilde{\mathcal{D}}^2 + 4\theta_f\theta_v})}{\partial \mu_v} = (2-\delta)T - \frac{\delta T \tilde{\mathcal{D}}}{\sqrt{\mathcal{D}^2 + 4\theta_f\theta_v}},$$

which is decreasing in $\tilde{\mathcal{D}}$. When $\tilde{\mathcal{D}} \to \infty$,

$$\frac{\partial(-\tilde{\mathcal{B}} + \sqrt{\tilde{\mathcal{D}}^2 + 4\theta_f\theta_v})}{\partial \mu_v} \to 2(1-\delta)T > 0.$$

So $\lambda_v^{VF}(\mu - \mu_v, \mu_v)$ is increasing in μ_v. Q.E.D.

Proof of Theorem 2: As illustrated in Figure 3, each system size includes several equilibrium regions. The optimal capacity allocation for each size is the one that yields the greatest profit among the regions it includes.

1. **Small System** includes Regions B, BV, BVF, BF and F. Region B is dominated by Regions BV and BF. From Proposition 2 we know that the optimal solution in Region BVF is on the boundaries (in this case, either on the boundary with BV or on the boundary with BF). When Region F is applicable ($\Lambda \leq \mu - \frac{\theta_f}{R-T}$), it dominates Region BF. The optimal value function is $r_f\Lambda$ in Region F, and $r_f\left(\mu - \frac{\theta_f}{R-T}\right)$ in Region BF. In both cases the optimal allocation is $\mu_f = \mu$ and $\mu_v = 0$. The optimal value function in Regions F and BF is, therefore, $r_f\left(\min\{\Lambda, \mu - \frac{\theta_f}{R-T}\}\right)$. Comparing the latter with the optimal value function in Region BV yields the condition as stated.

2. **Medium-sized System** includes Regions B, BV, BVF, VF and F. Region B is dominated by Region BV. From Proposition 2 we know that the optimal solution in Region BVF is dominated by either BV or VF. Moreover, Region F is dominated by VF (as an extreme case in VF). Therefore, it suffices to compare the optimal value function in Regions BV and VF, as stated.

3. **Large System** includes Regions B, BV, V, VF and F. Region B is dominated by Region BV and Regions BV and VF are dominated by Region V (as an extreme case in VF). Therefore, the optimal region is VF, as stated. Since the solution of Region VF is a boundary one, it will utilize either the in-person channel or the virtual channel while supporting returning patients.

Q.E.D.

Proof of Corollary 1: First let us replace T by $T-b$ in the results of Theorem 2. Then, the optimal capacity allocation must lead to the six scenarios presented in the Corollary statement. The results in each scenario are straightforward. It is worth noting that

$$\frac{\delta(\theta_f-\theta_v)+(1-\delta)\sqrt{\delta\theta_f\theta_v}}{R-\delta T+\delta b}\leq\frac{\theta_f}{R-T+b}-\frac{\delta\theta_v}{(1-\delta)R}$$

is equivalent to the second constraint for $(\mu-\mu_v,\mu_v)$ in Proposition 3, i.e., $(\mu-\tilde{\mu}_v^{BV},\tilde{\mu}_v^{BV})\in$ Region BV. Q.E.D.

Proof of Lemma 1: When

$$\mu\geq(1+\delta)\Lambda+\frac{\theta_v}{R(1-\delta)}+\frac{\theta_f}{R-T+b},$$

the system behaves as a large one, and $\bar{b}$ is the value of b to make this inequality hold as an equality. Q.E.D.

Proof of Proposition 5: It can be verified that

$$(r_f-b)\left(\mu-\frac{\theta_f}{R-T+b}\right)$$

is a concave function of b, and the first order condition leads to the solution

$$\tilde{b}^{NV}=\sqrt{\frac{\theta_f(R-T+r_f)}{\mu}}-R+T.$$

Then, it is straightforward to see that $\tilde{b}^{NV}$ is the optimal b if it falls in $[\underline{b}_{NV},\ \bar{b}_{NV}]$. Otherwise, the optimal b takes the boundary values of $[\underline{b}_{NF},\ \bar{b}_{NF}]$. Q.E.D.

Proof of Proposition 6: We first demonstrate the assumption made in the proposition, which is equivalent to $f(0)<0$.

Assumption EC.1. *The following inequality holds:*

$$\frac{\delta(\theta_f-\theta_v)+(1-\delta)\sqrt{\delta\theta_f\theta_v}}{R-\delta T}<\frac{\theta_f}{R-T}-\frac{\delta\theta_v}{(1-\delta)R}.$$

Note that the sign of

$$\frac{\delta(\theta_f-\theta_v)+(1-\delta)\sqrt{\delta\theta_f\theta_v}}{(R-\delta T+\delta b)}-\frac{\theta_f}{R-T+b}+\frac{\delta\theta_v}{(1-\delta)R}$$

is as same as the sign of

$$\left(\delta(\theta_f-\theta_v)+(1-\delta)\sqrt{\delta\theta_f\theta_v}\right)(R-T+b)-\theta_f(R-\delta T+\delta b)+\frac{\delta\theta_v(R-\delta T+\delta b)(R-T+b)}{(1-\delta)R}, \tag{EC.63}$$

as $R-\delta T>0$ and $R-T+b>0$. It can be verified that (EC.63) is a quadratic convex function of b. Therefore, if Assumption EC.1 holds, (EC.63) is negative when $b=0$. Thus there exists a unique $b^0>0$, such that (EC.63) equals to 0 when $b=b^0$, and when $b\le b^0$,

$$\frac{\delta(\theta_f-\theta_v)+(1-\delta)\sqrt{\delta\theta_f\theta_v}}{R-\delta T+\delta b}\le\frac{\theta_f}{R-T+b}-\frac{\delta\theta_v}{(1-\delta)R},$$

and when $b>b^0$,

$$\frac{\delta(\theta_f-\theta_v)+(1-\delta)\sqrt{\delta\theta_f\theta_v}}{R-\delta T+\delta b}>\frac{\theta_f}{R-T+b}-\frac{\delta\theta_v}{(1-\delta)R}.$$

Let $\lambda_v^{NF}(b)$ denote the effective λ_v^{NF} in equilibrium under optimal capacity allocation with given b. According to (9) in Corollary EC.1, when $b\le b^0$, $\lambda_v^{NF}(b)$ is the first case of (9); when $b>b^0$, $\lambda_v^{NF}(b)$ is the second case of (9). We therefore have

$$\lambda_v^{NF}(b)=\begin{cases}\min\left\{\Lambda,\frac{\mu}{1+\delta}-\frac{(\sqrt{\delta\theta_f}+\sqrt{\theta_v})^2}{(1+\delta)(R-\delta T+\delta b)}\right\}, & \text{if } b\le b^0,\\ \min\left\{\Lambda,\frac{\mu}{1+\delta}-\frac{\theta_f}{(1+\delta)(R-T+b)}-\frac{\theta_v}{(1+\delta)(1-\delta)R}\right\}, & \text{otherwise.}\end{cases} \tag{EC.64}$$

Note that when $b<\bar{b}_{NF}$, $\lambda_v^{NF}(b)$ cannot hit Λ by the definition of $\bar{b}_{NF}$. Thus

$$\lambda_v^{NF}(b)=\begin{cases}\frac{\mu}{1+\delta}-\frac{(\sqrt{\delta\theta_f}+\sqrt{\theta_v})^2}{(1+\delta)(R-\delta T+\delta b)}, & \text{if } b\le b^0,\\ \frac{\mu}{1+\delta}-\frac{\theta_f}{(1+\delta)(R-T+b)}-\frac{\theta_v}{(1+\delta)(1-\delta)R}, & \text{otherwise.}\end{cases}$$

One can verify that both

$$(r_v-\delta b)\left(\frac{\mu}{1+\delta}-\frac{(\sqrt{\delta\theta_f}+\sqrt{\theta_v})^2}{(1+\delta)(R-\delta T+\delta b)}\right) \tag{EC.65}$$

and

$$(r_v-\delta b)\left(\frac{\mu}{1+\delta}-\frac{\theta_f}{(1+\delta)(R-T+b)}-\frac{\theta_v}{(1+\delta)(1-\delta)R}\right) \tag{EC.66}$$

are concave in b.

Let b_{NF}^1 and b_{NF}^2 denote the maximizer of $(r_v-\delta b)\lambda_v^{NF}(b)$, which takes the forms given in (EC.65) and (EC.66), receptively. Specifically,

$$b_{NF}^1=\frac{(\sqrt{\delta\theta_f}+\sqrt{\theta_v})\sqrt{R-\delta T+r_v}}{\delta\sqrt{\mu}}-\frac{R-\delta T}{\delta},$$

which comes from taking the first order condition of (EC.65);

$$b_{NF}^{2} = \sqrt{\frac{\theta_f(\delta R - \delta T + r_v)}{\delta\mu - \frac{\delta\theta_v}{(1-\delta)R}}} - R + T,$$

which comes from taking the first order condition of (EC.66).

Thus, we have

$$b^{NF*} = \begin{cases} b_{NF}^{1}, & \text{if } b_{NF}^{1} \le b^{0} \\ b_{NF}^{2}, & \text{otherwise.} \end{cases}$$

And it is straightforward to see that b^{NF*} is the optimal b if it falls in $[\underline{b}_{NF},\ \bar{b}_{NF}]$. Otherwise, the optimal b should take the boundary values of $[\underline{b}_{NF},\ \bar{b}_{NF}]$.

Q.E.D.

Proof of Proposition 7: The objective function is

$$(r_f - b)(\Lambda - \frac{1}{\delta}(\mu - \Lambda - \frac{\theta_v}{(1-\delta)R} - \frac{\theta_f}{R-T+b})) + (r_v - \delta b)\frac{1}{\delta}(\mu - \Lambda - \frac{\theta_v}{(1-\delta)R} - \frac{\theta_f}{R-T+b}).$$

Its first order derivative is

$$\frac{(1-\delta)\mu}{\delta} - \frac{\Lambda}{\delta} - \frac{\theta_v}{\delta R} + \frac{\theta_f(r_v - r_f) - (1-\delta)\theta_f(R-T)}{\delta(R-T+b)^2}. \tag{EC.67}$$

Its second order derivative is

$$-\frac{2\theta_f(r_v - r_f) - (1-\delta)\theta_f(R-T)}{\delta(R-T+b)^3}. \tag{EC.68}$$

If $r_v - r_f - (1-\delta)(R-T) \le 0$, (EC.68) is non-negative. Thus, the objective function is convex. Then the optimal solution must be on the boundary. i.e., either $\underline{b}_{VF}$ or $\bar{b}_{VF}$.

If $r_v - r_f - (1-\delta)(R-T) > 0$, (EC.68) is negative, then the objective function is concave. The first order condition leads to

$$\sqrt{\frac{\theta_f(r_v - r_f) - (1-\delta)\theta_f(R-T)}{-(1-\delta)\mu + \Lambda + \theta_v/R}} - R + T. \tag{EC.69}$$

When $(1-\delta)\mu - \Lambda - \frac{\theta_v}{R} < 0$, then (EC.69) is well-defined. Thus, the optimal b is (EC.69) if it falls in $[\underline{b}_{VF}, \bar{b}_{VF}]$. Otherwise, the optimal b takes the boundary values of $[\underline{b}_{VF},\ \bar{b}_{VF}]$.

If $r_v - r_f - (1-\delta)(R-T) > 0$ and $(1-\delta)\mu - \Lambda - \frac{\theta_v}{R} \ge 0$, (EC.69) is not well-defined. However, (EC.67) is always positive, indicating that the objective function is increasing in b. Thus the optimal solution is to set b as large as $\bar{b}_{VF}$. Q.E.D.

Proof of Proposition 8: Follows directly from the definition of the thresholds for μ. Q.E.D.

Proof of Proposition 9: The first fact is that, if the original optimal solution without subsidy and the joint optimal solution with subsidy are in the same scenario (region), then the access rate will not decrease. The reason is that, with optimal subsidy, the marginal revenue is decreased but the total revenue is increased. Thus, the access rate will not decrease.

Next, we examine each case.

1. Large system: in this case,

$$\mu \geq (1+\delta)\Lambda + \frac{\theta_f}{R-T} + \frac{\theta_v}{(1-\delta)R}.$$

It is able to achieve Region F and V and obtain the largest revenue $\max\{r_f, r_v\}\Lambda$ without subsidy. Thus, the optimal subsidy is 0, and the total access rate is still Λ.

2. Medium-sized system or small system with $\mu \geq \Lambda + \theta_f[R-T]^{-1}$: in this case,

$$\Lambda + \frac{\theta_f}{R-T} \leq \mu < (1+\delta)\Lambda + \frac{\theta_f}{R-T} + \frac{\theta_v}{(1-\delta)R}.$$

• If $r_v \leq r_f$, it is able to achieve Region F which generates the largest revenue $r_f\Lambda$ without subsidy. Thus, $b^* = 0$, and the total access rate is still Λ.

• If

$$\frac{r_v}{r_f} \geq (1+\delta)\max\left\{1, \frac{\Lambda}{\mu - \frac{\theta_f}{R-T} - \frac{\theta_v}{(1-\delta)R}}\right\}, \tag{EC.70}$$

then we have $r_v \geq (1+\delta)r_f$ since $\mu - \frac{\theta_f}{R-T} > \mu - \frac{\theta_f}{R-T} - \frac{\theta_v}{(1-\delta)R}$. Per Propositions 2 and 4, Scenario VF is dominated by Scenario NF for any b. Since Region F, which leads to full access rate Λ, can be achieved without b, then the only one possible situation for decreased access rate is that the original optimal solution is in Region F but the joint optimal solution with subsidy is in Region BV. Note that $r_f\Lambda$ is the revenue achieved in Region F, and

$$r_v\left(\frac{\mu}{1+\delta} - \frac{\theta_f}{(1+\delta)(R-T)} - \frac{\theta_v}{(1+\delta)(1-\delta)R}\right)$$

is the lower bound for the revenue in Region BV. When condition (EC.70) holds, the original optimal without subsidy cannot be in Region F. Thus the access rate will not decrease with subsidy.

3. Small system with $\mu < \Lambda + \theta_f[R-T]^{-1}$: in this case,

$$\max\left\{\frac{\theta_f}{R-T}, \underline{\mu}^{BV}\right\} \leq \mu < \Lambda + \frac{\theta_f}{R-T}.$$

• If

$$\frac{r_v}{r_f} \leq \min\left\{1, \frac{(1+\delta)(\mu - \frac{\theta_f}{R-T})}{\mu - \frac{(\sqrt{\delta\theta_f} + \sqrt{\theta_v})^2}{R-\delta T}}\right\}, \tag{EC.71}$$

then $r_v \leq r_f$. Per Propositions 2, Region BVF is dominated by Region BF for any b. Without subsidy, the revenue in Region BF is $r_f(\mu - \frac{\theta_f}{R-T})$, and

$$r_v\left(\frac{\mu}{1+\delta} - \frac{\left(\sqrt{\delta\theta_f}+\sqrt{\theta_v}\right)^2}{(1+\delta)(R-\delta T)}\right)$$

is the upper bound for the revenue in Region BV. Since (EC.71) holds, the original optimal solution without subsidy is in Region BF. Thus, the only one possible situation for decreased access rate is that the joint optimal solution with subsidy is in Region BV. Suppose that the joint optimal solution with subsidy is in Region BV, we must have $(r_v - b^*)\lambda_v^{BV*}(b^*) \geq r_f \lambda_f^{BF*}(0)$. Since $r_v \leq r_f$, we must have $\lambda_v^{BV*}(b^*) \geq \lambda_f^{BF*}(0)$, indicating non-decreased access rate.

- If (EC.70) holds, i.e.,

$$\frac{r_v}{r_f} \geq \frac{(1+\delta)(\mu - \frac{\theta_f}{R-T})}{\mu - \frac{\theta_f}{R-T} - \frac{\theta_v}{(1-\delta)R}},$$

then we have $r_v \geq (1+\delta) r_f$ since $\mu - \frac{\theta_f}{R-T} > \mu - \frac{\theta_f}{R-T} - \frac{\theta_v}{(1-\delta)R}$. Per Propositions 2 and 4, Scenario VF is dominated by Scenario NF for any b. Without subsidy, the revenue in Region BF is $r_f(\mu - \frac{\theta_f}{R-T})$, and

$$r_v\left(\frac{\mu}{1+\delta} - \frac{\theta_f}{(1+\delta)(R-T)} - \frac{\theta_v}{(1+\delta)(1-\delta)R}\right)$$

is the lower bound for the revenue in Region BV. Since (EC.70) holds, the original optimal solution without subsidy is in Region BV. Thus, the only one possible situation for decreased access rate is that the joint optimal solution with subsidy is in Region BF. Suppose that the joint optimal solution with subsidy is in Region BF, we must have $(r_f - b^*)\lambda_f^{BF*}(b^*) \geq r_v \lambda_v^{BV*}(0)$. Since $r_v > r_f$, we must have $\lambda_f^{BF*}(b^*) > \lambda_v^{BV*}(0)$, indicating non-decreased access rate.

Q.E.D.

Proof of Proposition 10: Following the arguments in the proof of Proposition 9, we know the access rate will not decrease under any of the following conditions.

1. Large system.
2. Medium-sized system or small system with $\mu \geq \Lambda + \theta_f[R-T]^{-1}$, if

$$\frac{r_v}{r_f} \leq 1 \quad \text{or} \quad \frac{r_v}{r_f} \geq (1+\delta)\max\left\{1, \frac{\Lambda}{\mu - \frac{\theta_f}{R-T} - \frac{\theta_v}{(1-\delta)R}}\right\}.$$

3. Small system with $\mu < \Lambda + \theta_f[R-T]^{-1}$, if

$$\frac{r_v}{r_f} \leq \min\left\{1, \frac{(1+\delta)(\mu - \frac{\theta_f}{R-T})}{\mu - \frac{(\sqrt{\delta\theta_f}+\sqrt{\theta_v})^2}{R-\delta T}}\right\} \quad \text{or} \quad \frac{r_v}{r_f} \geq \frac{(1+\delta)(\mu - \frac{\theta_f}{R-T})}{\mu - \frac{\theta_f}{R-T} - \frac{\theta_v}{(1-\delta)R}}.$$

The remaining conditions we have to examine are

1. Medium-sized system or a small system with $\mu \geq \Lambda + \theta_f[R-T]^{-1}$, if

$$1 < \frac{r_v}{r_f} \leq \frac{(1+\delta)\Lambda}{\min\left\{(1+\delta)\Lambda, \mu - \frac{(\sqrt{\delta\theta_f}+\sqrt{\theta_v})^2}{R-\delta T+\delta b}\right\}}.$$

2. Small system with $\mu < \Lambda + \theta_f[R-T]^{-1}$, if

$$1 < \frac{r_v}{r_f} \leq \frac{(1+\delta)\{\Lambda, \mu - \frac{\theta_f}{R-T+b}\}}{\mu - \frac{(\sqrt{\delta\theta_f}+\sqrt{\theta_v})^2}{R-\delta T+\delta b}}.$$

Next, we examine each case.

1. Medium-sized system or small system with $\mu \geq \Lambda + \theta_f[R-T]^{-1}$: in this case,

$$\Lambda + \frac{\theta_f}{R-T} \leq \mu < (1+\delta)\Lambda + \frac{\theta_f}{R-T} + \frac{\theta_v}{(1-\delta)R}.$$

If

$$1 < \frac{r_v}{r_f} \leq \frac{(1+\delta)\Lambda}{\min\left\{(1+\delta)\Lambda, \mu - \frac{(\sqrt{\delta\theta_f}+\sqrt{\theta_v})^2}{R-\delta T+\delta b}\right\}}, \tag{EC.72}$$

then we have $r_v > r_f$. Since (EC.72) holds, then Scenario NF is dominated by Scenario NV after b is utilized. Thus the access rate cannot decrease as Scenario NV and Scenario VF both achieve full access rate.

2. Small system with $\mu < \Lambda + \theta_f[R-T]^{-1}$: in this case,

$$\max\left\{\frac{\theta_f}{R-T}, \underline{\mu}^{BV}\right\} \leq \mu < \Lambda + \frac{\theta_f}{R-T}.$$

If

$$1 < \frac{r_v}{r_f} \leq \frac{(1+\delta)\{\Lambda, \mu - \frac{\theta_f}{R-T+b}\}}{\mu - \frac{(\sqrt{\delta\theta_f}+\sqrt{\theta_v})^2}{R-\delta T+\delta b}}, \tag{EC.73}$$

then we have $r_v > r_f$. Since (EC.73) holds, then Scenario NF is dominated by Scenario NV after b is utilized. Thus the access rate cannot decrease as: 1) the original optimal capacity allocation without b must be in Scenario NF or NV; 2) after b is utilized, Scenario VF achieves full access rate; 3) if the original optimal capacity allocation after b is utilized is in Scenario NF, then the access rate must be non-decreasing since $r_v > r_f$.

Q.E.D.

Proof of Corollary EC.1: The results follow from Theorem 1 by replacing T with $T - b$. Q.E.D.

Proof of Lemma EC.1: It directly follows the definitions of $U_v^t(\lambda_f, \lambda_v)$, $U_f^t(\lambda_f, \lambda_v)$ and $U_b^t(\lambda_f, \lambda_v)$. Q.E.D.

Proof of Theorem EC.1: We first show that the seven regions of (μ_f, μ_v) defined in Definition EC.3 are mutually exclusive and collectively exhaustive. By Definition EC.3, it is easy to see (EC.30) v.s. (EC.44) divide the (μ_f, μ_v) space; (EC.31) v.s. (EC.47) divide the (μ_f, μ_v) space; (EC.32) v.s. (EC.53) divide the (μ_f, μ_v) space; (EC.33) with (EC.34) and (EC.35) v.s. (EC.48) divide the (μ_f, μ_v) space; (EC.36) v.s. (EC.45) divide the (μ_f, μ_v) space; (EC.37) with $\mu_f > \Lambda$ (implied by (EC.36)) v.s. (EC.54) divide the (μ_f, μ_v) space; (EC.38) v.s. (EC.46) divide the (μ_f, μ_v) space; (EC.39) v.s. (EC.55) with (EC.56) divide the (μ_f, μ_v) space; (EC.40) with (EC.41) v.s. (EC.49) with (EC.50) divide the (μ_f, μ_v) space; (EC.42) with (EC.43) v.s. (EC.51) with (EC.52) divide the (μ_f, μ_v) space. Thus the seven regions of (μ_f, μ_v) defined in Definition EC.3 are mutually exclusive and collectively exhaustive.

In Region B, everyone chooses balking, i.e., $\lambda_f = 0$, and $\lambda_v = 0$. We should have $U_f^t(0,0) \le 0$ and $U_v^t(0,0) \le 0$, for $\forall$ t. By Lemma EC.1, we arrive at

$$U_f^0(0,0) \le 0$$

and

$$U_v^0(0,0) \le 0,$$

which yield (EC.30) and (EC.31), indicating that if (EC.30) and (EC.31) hold then we must have $U_f^t(0,0) \le 0$ and $U_v^t(0,0) \le 0$, for $\forall$ t.

In Region V, everyone chooses virtual channel, i.e., $\lambda_f = 0$, and $\lambda_v = \Lambda$. Then we have $U_v^t(0,\Lambda) \ge U_f^t(0,\Lambda)$ and $U_v^t(0,\Lambda) \ge 0$, for $\forall$ t. By Lemma EC.1, we must have

$$U_v^0(0,\Lambda) \ge U_f^0(0,\Lambda)$$

and

$$U_v^1(0,\Lambda) \ge 0,$$

which yield (EC.32) and (EC.33) with (EC.34) and (EC.35), indicating that if (EC.32), (EC.33), (EC.34) and (EC.35) hold then we must have $U_v^t(0,\Lambda) \ge U_f^t(0,\Lambda)$ and $U_v^t(0,\Lambda) \ge 0$, for $\forall$ t.

In Region F, everyone chooses face-to-face channel, i.e., $\lambda_f = \Lambda$, and $\lambda_v = 0$. Then we have $U_f^t(\Lambda,0) \ge U_v^t(\Lambda,0)$ and $U_f^t(\Lambda,0) \ge 0$, for $\forall$ t. By Lemma EC.1, we must have

$$U_f^1(\Lambda,0) \ge U_f^1(\Lambda,0)$$

and

$$U_f^1(\Lambda, 0) \geq 0,$$

which yield (EC.36) and (EC.37) with $\mu_f > \Lambda$ (which is implied by (EC.36)), indicating that if (EC.36) and (EC.37) hold then we must have $U_f^t(\Lambda, 0) \geq U_v^t(\Lambda, 0)$ and $U_f^t(\Lambda, 0) \geq 0$, for $\forall\ t$.

Denote $1/(\mu_f - \lambda_f - \delta\lambda_v)$ as w_f and $1/(\mu_v - \lambda_v)$ as w_v. Define three thresholds:

$$t_{fb} = R - \theta_f w_f, \tag{EC.74}$$

$$t_{vb} = \frac{R - \theta_v w_v}{\delta} - \theta_f w_f, \tag{EC.75}$$

$$t_{fv} = \frac{\theta_v w_v}{1-\delta} - \theta_f w_f. \tag{EC.76}$$

By their definitions, we have if $t < t_{fb}$ then $U_f > U_b$, and if $t > t_{fb}$ then $U_f < U_b$; if $t < t_{vb}$ then $U_v > U_b$, and if $t > t_{vb}$ then $U_v < U_b$; if $t < t_{fv}$ then $U_f > U_v$, and if $t > t_{fv}$ then $U_f < U_v$.

In Region BVF, some patients choose face-to-face channel, some choose virtual channel and some choose balking, i.e., $\lambda_f > 0$, $\lambda_v > 0$ and $\lambda_f + \lambda_b < \Lambda$. Then we have $U_f^t(\lambda_f, \lambda_v) \geq \max\{U_v^t(\lambda_f, \lambda_v), 0\}$ for some t, $U_v^t(\lambda_f, \lambda_v) \geq \max\{U_f^t(\lambda_f, \lambda_v), 0\}$ for some t, and $0 \geq \max\{U_f^t(\lambda_f, \lambda_v), U_v^t(\lambda_f, \lambda_v)$ for some t. By Lemma EC.1, we must have $0 < t_{fv} < t_{vb} < 1$. Then

$$t_{fv} = \frac{\lambda_f}{\Lambda}$$

and

$$t_{vb} = \frac{\lambda_f + \lambda_v}{\Lambda}.$$

Then we have

$$t_{vb} - t_{fv} = \frac{R}{\delta} - \frac{\theta_v}{\delta(1-\delta)(\mu_v - \lambda_v)} = \frac{\lambda_v}{\Lambda}.$$

$\lambda_v > 0$ is equivalent to

$$\max_{\lambda_v > 0}\{\frac{R}{\delta} - \frac{\theta_v}{\delta(1-\delta)(\mu_v - \lambda_v)} - \frac{\lambda_v}{\Lambda}\} > 0,$$

thus we can have (EC.38).

We can also have

$$t_{vb} = \frac{R}{\delta} - \frac{\theta_v}{\delta(\mu_v - \lambda_v)} - \frac{\theta_f}{\mu_f - \lambda_f - \delta\lambda_v} = \frac{\lambda_f + \lambda_v}{\Lambda}.$$

$\lambda_f < \Lambda - \lambda_v$ is equivalent to

$$\min_{\lambda_f < \Lambda - \lambda_v}\{\frac{R}{\delta} - \frac{\theta_v}{\delta(\mu_v - \lambda_v)} - \frac{\theta_f}{\mu_f - \lambda_f - \delta\lambda_v} - \frac{\lambda_f + \lambda_v}{\Lambda}\} < 0,$$

for any $0 < \lambda_v < \Lambda$ and when $\mu_f - \lambda_f - \delta\lambda_v > 0$. It is equivalent to

$$\max_{0<\lambda_v<\Lambda}\{\frac{R}{\delta} - \frac{\theta_v}{\delta(\mu_v - \lambda_v)} - \frac{\theta_f}{\mu_f - \Lambda + (1-\delta)\lambda_v} - 1\} < 0, \tag{EC.77}$$

when $\mu_f - \Lambda + (1-\delta)\lambda_v > 0$. One can show (EC.39) is equivalent to (EC.77) by taking the maximization of the LHS of (EC.77).

We can also have

$$t_{fv} = \frac{\theta_v}{(1-\delta)(\mu_v - \lambda_v)} - \frac{\theta_f}{\mu_f - \lambda_f - \delta\lambda_v} = \frac{\lambda_f}{\Lambda}.$$

$\lambda_f > 0$ is equivalent to

$$\max_{\lambda_f>0}\{\frac{\theta_v}{(1-\delta)(\mu_v - \lambda_v)} - \frac{\theta_f}{\mu_f - \lambda_f - \delta\lambda_v} - \frac{\lambda_f}{\Lambda}\} > 0,$$

for any $0 < \lambda_v < \Lambda$. It is equivalent to

$$\frac{\theta_v}{(1-\delta)(\mu_v - \lambda_v)} - \frac{\theta_f}{\mu_f - \delta\lambda_v} < 0, \tag{EC.78}$$

for any $0 < \lambda_v < \Lambda$. Thus we have if $(1-\delta)\theta_f - \delta\theta_v > 0$ then

$$\lambda_v > \frac{(1-\delta)\theta_f\mu_v - \theta_v\mu_f}{(1-\delta)\theta_f - \delta\theta_v},$$

and if $(1-\delta)\theta_f - \delta\theta_v < 0$ then

$$\lambda_v < \frac{(1-\delta)\theta_f\mu_v - \theta_v\mu_f}{(1-\delta)\theta_f - \delta\theta_v}.$$

We first consider the case of $(1-\delta)\theta_f - \delta\theta_v > 0$. In this case, if $(1-\delta)\theta_f\mu_v - \theta_v\mu_f < 0$ then (EC.78) holds; if $(1-\delta)\theta_f\mu_v - \theta_v\mu_f \geq 0$ then we have

$$\max_{\lambda_v > \frac{(1-\delta)\theta_f\mu_v - \theta_v\mu_f}{(1-\delta)\theta_f - \delta\theta_v}}\{\frac{R}{\delta} - \frac{\theta_v}{\delta(1-\delta)(\mu_v - \lambda_v)} - \frac{\lambda_v}{\Lambda}\} > 0.$$

One can show (EC.40) with (EC.41) is equivalent to it.

We then consider the case of $(1-\delta)\theta_f - \delta\theta_v < 0$. In this case, we must have $(1-\delta)\theta_f\mu_v - \theta_v\mu_f < 0$ and

$$\min_{\lambda_v < \frac{(1-\delta)\theta_f\mu_v - \theta_v\mu_f}{(1-\delta)\theta_f - \delta\theta_v}}\{\frac{R}{\delta} - \frac{\theta_v}{\delta(1-\delta)(\mu_v - \lambda_v)} - \frac{\lambda_v}{\Lambda}\} < 0.$$

One can show (EC.42) with (EC.43) is equivalent to it with $(1-\delta)\theta_f\mu_v - \theta_v\mu_f < 0$.

Now we are ready to derive λ_f and λ_v. Combining

$$t_{fv} = \frac{\lambda_f}{\Lambda}$$

and

$$t_{vb} = \frac{\lambda_f + \lambda_v}{\Lambda}$$

together with (EC.75) and (EC.76), we can solve the equations and obtain

$$\lambda_f = \tfrac{1}{2}\left[\mu_f - \delta\mu_v - \sqrt{(R\Lambda-\mu_f)^2 + 4\theta_f\Lambda} + \delta\sqrt{\left(\frac{R\Lambda}{\delta}-\mu_v\right)^2 + \frac{4\theta_v\Lambda}{\delta(1-\delta)}}\right] \tag{EC.79}$$

$$\lambda_v = \tfrac{1}{2}\left[\frac{R\Lambda}{\delta} + \mu_v - \sqrt{\left(\frac{R\Lambda}{\delta}-\mu_v\right)^2 + \frac{4\theta_v\Lambda}{\delta(1-\delta)}}\right]. \tag{EC.80}$$

In Region BF, some patients choose face-to-face channel and some choose balking, i.e., $0 < \lambda_f < \Lambda$, and $\lambda_v = 0$. Then we have $U_f^t(\lambda_f, 0) \geq \max\{U_v^t(\lambda_f, 0), 0\}$ for some t, and $0 \geq \max\{U_f^t(\lambda_f, 0), U_v^t(\lambda_f, 0)\}$ for some t. By Lemma EC.1, we must have $0 < t_{fb} < 1$ and $t_{vb} \leq t_{fv}$. Then

$$t_{fb} = \frac{\lambda_f}{\Lambda}.$$

Then we have

$$t_{fb} = R - \frac{\theta_f}{\mu_f - \lambda_f} = \frac{\lambda_f}{\Lambda}.$$

Note that $\lambda_f > 0$ is equivalent to

$$\max_{\lambda_f > 0}\{R - \frac{\theta_f}{\mu_f - \lambda_f} - \frac{\lambda_f}{\Lambda}\} > 0.$$

thus we can have (EC.44).

Note that $\lambda_f < \Lambda$ is equivalent to

$$\min_{\lambda_f < \Lambda}\{R - \frac{\theta_f}{\mu_f - \lambda_f} - \frac{\lambda_f}{\Lambda}\} < 0.$$

Thus we can have (EC.45).

Note that $\lambda_v = 0$ is equivalent to $t_{vb} \leq t_{fv}$, i.e.,

$$\min_{\lambda_v \geq 0}\{R - \frac{\theta_v}{(1-\delta)(\mu_v - \lambda_v)}\} \leq 0.$$

Thus we can have (EC.46).

Now we are ready to derive λ_f. By

$$t_{fb} = R - \frac{\theta_f}{\mu_f - \lambda_f} = \frac{\lambda_f}{\Lambda}$$

and (EC.74), we can solve the equation and obtain

$$\lambda_f = \tfrac{1}{2}\left[R\Lambda + \mu_f - \sqrt{(R\Lambda - \mu_f)^2 + 4\theta_f\Lambda}\right]. \tag{EC.81}$$

In Region BV, some patients choose virtual channel and some choose balking, i.e., $0 < \lambda_v < \Lambda$, and $\lambda_f = 0$. Then we have $U_v^t(0,\lambda_v) \geq \max\{U_f^t(0,\lambda_v), 0\}$ for some t, and $0 \geq \max\{U_f^t(0,\lambda_v), U_v^t(0,\lambda_v)$ for some t. By Lemma EC.1, we must have $t_{fv} \leq 0 < t_{vb} < 1$. Then

$$t_{vb} = \frac{\lambda_v}{\Lambda}.$$

Note that $\lambda_v > 0$ is equivalent to

$$\max_{\lambda_v > 0}\{\frac{R}{\delta} - \frac{\theta_v}{\delta(\mu_v - \lambda_v)} - \frac{\theta_f}{\mu_f - \delta\lambda_v} - \frac{\lambda_v}{\Lambda}\} > 0.$$

thus we can have (EC.47).

Note that $\lambda_v < \Lambda$ is equivalent to

$$\min_{\lambda_v < \Lambda}\{\frac{R}{\delta} - \frac{\theta_v}{\delta(\mu_v - \lambda_v)} - \frac{\theta_f}{\mu_f - \delta\lambda_v} - \frac{\lambda_v}{\Lambda}\} < 0.$$

thus we can have (EC.48).

Note that $\lambda_f = 0$ is equivalent to $t_{fv} \leq 0$, i.e.,

$$\min_{\lambda_f \geq 0}\{\frac{\theta_v}{(1-\delta)(\mu_v - \lambda_v)} - \frac{\theta_f}{\mu_f - \lambda_f - \delta\lambda_v}\} \leq 0.$$

Thus we can have

$$\frac{\theta_v}{(1-\delta)(\mu_v - \lambda_v)} - \frac{\theta_f}{\mu_f - \delta\lambda_v} \leq 0, \tag{EC.82}$$

for any $\lambda_v \in [0, \Lambda]$. Then we have, if $(1-\delta)\theta_f - \delta\theta_v > 0$ then

$$\lambda_v \leq \frac{(1-\delta)\theta_f\mu_v - \theta_v\mu_f}{(1-\delta)\theta_f - \delta\theta_v},$$

and if $(1-\delta)\theta_f - \delta\theta_v < 0$ then

$$\lambda_v \geq \frac{(1-\delta)\theta_f\mu_v - \theta_v\mu_f}{(1-\delta)\theta_f - \delta\theta_v}.$$

We first consider the case of $(1-\delta)\theta_f - \delta\theta_v > 0$. In this case, we must have $(1-\delta)\theta_f\mu_v - \theta_v\mu_f \geq 0$ and

$$\min_{\lambda_v \leq \frac{(1-\delta)\theta_f\mu_v - \theta_v\mu_f}{(1-\delta)\theta_f - \delta\theta_v}}\{\frac{R}{\delta} - \frac{\theta_v}{\delta(\mu_v - \lambda_v)} - \frac{\theta_f}{\mu_f - \delta\lambda_v} - \frac{\lambda_v}{\Lambda}\} \leq 0.$$

One can show (EC.49) with (EC.50) is equivalent to it with $(1-\delta)\theta_f\mu_v - \theta_v\mu_f \geq 0$.

We then consider the case of $(1-\delta)\theta_f - \delta\theta_v < 0$. In this case, if $(1-\delta)\theta_f\mu_v - \theta_v\mu_f < 0$ then (EC.82) holds; if $(1-\delta)\theta_f\mu_v - \theta_v\mu_f \geq 0$ then we have

$$\max_{\lambda_v \geq \frac{(1-\delta)\theta_f\mu_v - \theta_v\mu_f}{(1-\delta)\theta_f - \delta\theta_v}}\{\frac{R}{\delta} - \frac{\theta_v}{\delta(\mu_v - \lambda_v)} - \frac{\theta_f}{\mu_f - \delta\lambda_v} - \frac{\lambda_v}{\Lambda}\} \geq 0.$$

One can show (EC.51) with (EC.52) is equivalent to it.

Now we are ready to derive λ_v. By

$$t_{vb} = \frac{\lambda_v}{\Lambda}$$

and (EC.75), we can solve the following equation and obtain λ_v:

$$\frac{R\Lambda}{\delta} - \frac{\theta_v \Lambda}{\delta(\mu_v - \lambda_v)} - \frac{\theta_f \Lambda}{\mu_f - \delta\lambda_v} = \lambda_v \tag{EC.83}$$

In Region VF, some patients choose face-to-face channel and some choose virtual channel, i.e., $\lambda_f > 0$, $\lambda_v > 0$ and $\lambda_f + \lambda_v = \Lambda$. Then we have $U_f^t(\Lambda - \lambda_v, \lambda_v) \geq \max\{U_v^t(\Lambda - \lambda_v, \lambda_v), 0\}$ for some t, and $U_v^t(\Lambda - \lambda_v, \lambda_v) \geq \max\{U_f^t(\Lambda - \lambda_v, \lambda_v), 0\}$ for some t. By Lemma EC.1, we must have $0 < t_{fv} < 1 \leq t_{vb}$. Then

$$t_{fv} = \frac{\Lambda - \lambda_v}{\Lambda}.$$

We first have $\lambda_f = \Lambda - \lambda_v$.

Note that $\lambda_v < \Lambda$ is equivalent to

$$\max_{\lambda_v < \Lambda}\{\frac{\theta_v}{(1-\delta)(\mu_v - \lambda_v)} - \frac{\theta_f}{\mu_f - \Lambda + (1-\delta)\lambda_v} - \frac{\Lambda - \lambda_v}{\Lambda}\} > 0.$$

Thus we have (EC.53).

Note that $\lambda_v > 0$ is equivalent to

$$\min_{\lambda_v > 0}\{\frac{\theta_v}{(1-\delta)(\mu_v - \lambda_v)} - \frac{\theta_f}{\mu_f - \Lambda + (1-\delta)\lambda_v} - \frac{\Lambda - \lambda_v}{\Lambda}\} < 0.$$

Thus we have (EC.54).

Note that $\lambda_b = 0$ is equivalent to $t_{vb} \geq 1$, i.e.,

$$\max_{0<\lambda_v<\Lambda}\{\frac{R}{\delta} - \frac{\theta_v}{\delta(\mu_v - \lambda_v)} - \frac{\theta_f}{\mu_f - \Lambda + (1-\delta)\lambda_v} - 1\} \geq 0, \tag{EC.84}$$

and $\mu_f - \Lambda + (1-\delta)\lambda_v > 0$. One can show (EC.55) with (EC.56) is equivalent to (EC.84) with $\mu_f - \Lambda + (1-\delta)\lambda_v > 0$ by taking the maximization of the LHS of (EC.84).

By

$$t_{fv} = \frac{\Lambda - \lambda_v}{\Lambda}$$

and (EC.76), we can solve the following equation and obtain λ_v:

$$\Lambda - \frac{\theta_v \Lambda}{(1-\delta)(\mu_v - \lambda_v)} + \frac{\theta_f \Lambda}{\mu_f - \Lambda + (1-\delta)\lambda_v} = \lambda_v. \tag{EC.85}$$

Q.E.D.