



Management Science

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

Fair Allocation of Vaccines, Ventilators and Antiviral Treatments: Leaving No Ethical Value Behind in Healthcare Rationing

Parag A. Pathak, Tayfun Sönmez, M. Utku Ünver, M. Bumin Yenmez

To cite this article:

Parag A. Pathak, Tayfun Sönmez, M. Utku Ünver, M. Bumin Yenmez (2024) Fair Allocation of Vaccines, Ventilators and Antiviral Treatments: Leaving No Ethical Value Behind in Healthcare Rationing. *Management Science* 70(6):3999-4036. <https://doi.org/10.1287/mnsc.2022.00930>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2023 The Author(s)

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

Fair Allocation of Vaccines, Ventilators and Antiviral Treatments: Leaving No Ethical Value Behind in Healthcare Rationing

Parag A. Pathak,^{a,b,*} Tayfun Sönmez,^c M. Utku Ünver,^{c,d} M. Bumin Yenmez^c

^aDepartment of Economics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139; ^bNational Bureau of Economic Research, Cambridge, Massachusetts 02138; ^cDepartment of Economics, Boston College, Chestnut Hill, Massachusetts 02467; ^dDeakin University, Burwood, 3125 Victoria, Australia

*Corresponding author

Contact: ppathak@mit.edu,  <https://orcid.org/0000-0001-8621-3864> (PAP); sonmezt@bc.edu,  <https://orcid.org/0000-0002-9153-0466> (TS); unver@bc.edu,  <https://orcid.org/0000-0001-7693-1635> (MUÜ); bumin.yenmez@bc.edu (MBY)

Received: March 25, 2022

Revised: June 1, 2023

Accepted: June 24, 2023

Published Online in Articles in Advance:
September 20, 2023

<https://doi.org/10.1287/mnsc.2022.00930>

Copyright: © 2023 The Author(s)

Abstract. A *priority system* has traditionally been the protocol of choice for the allocation of scarce life-saving resources during public health emergencies. COVID-19 revealed the limitations of this allocation rule. Many argue that priority systems abandon ethical values, such as equity, by discriminating against disadvantaged communities. We show that a restrictive feature of the traditional priority system largely drives these limitations. Following *minimalist market design*, an institution design paradigm that integrates research and policy efforts, we formulate pandemic allocation of scarce life-saving resources as a new application of market design. Interfering only with the restrictive feature of the priority system to address its shortcomings, we formulate a *reserve system* as an alternative allocation rule. Our theoretical analysis develops a general theory of reserve design. We relate our analysis to debates during COVID-19 and describe the impact of our paper on policy and practice.

History: Accepted by Axel Ockenfels, behavioral economics and decision analysis.

 **Open Access Statement:** This work is licensed under a Creative Commons Attribution 4.0 International License. You are free to copy, distribute, transmit and adapt this work, but you must attribute this work as "Management Science. Copyright © 2023 The Author(s). <https://doi.org/10.1287/mnsc.2022.00930>, used under a Creative Commons Attribution License: <https://creativecommons.org/licenses/by/4.0/>.

Keywords: crisis standards of care • triage • multiprinciple allocation framework • minimalist market design • reserve system • COVID-19

1. Introduction

Agencies responsible for public health and emergency preparedness regularly design guidelines to allocate scarce life-saving resources in crisis situations. These situations range from wartime triage medicine to public health emergencies, such as influenza pandemics and COVID-19. Items in short supply include vaccines, ventilators, and antiviral treatments.

How to implement a rationing system for life-saving resources during a crisis presents a complicated question rife with ethical concerns. Rationing guidelines typically start by articulating several ethical principles. These principles include equity, which is fair distribution of benefits and burdens; utilitarianism, which is maximizing welfare; reciprocity, which is respecting contributions others have made in the past; instrumental valuation, which is respecting contributions others could make in the future; solidarity, which is fellowship with other members of society; and nondiscrimination, which is requiring that certain individual characteristics, such

as gender, race, and age, play no role in allocation protocols. Guidelines also emphasize procedural values, such as accountability, reasonableness, and transparency.¹

After articulating these ethical principles, guidelines describe how to operationalize them with an allocation rule. Traditionally, the most common rule suggested in these guidelines is a *priority system*, in which patients are placed into a single priority order, and allocation is made in order of priority.² In some cases, the priority order is derived from an objective scoring method, resulting in a *priority point system*. This practice is especially prominent for rationing of intensive care unit (ICU) beds and ventilators, where 19 states in the United States base the priority score on the Sequential Organ Failure Assessment (SOFA) score, a measure of organ dysfunction (Piscitello et al. 2020).³

The COVID-19 pandemic has spurred renewed interest in medical rationing guidelines and has revealed several important limitations of the existing allocation mechanisms. Whether it is rationing of ventilators,

antiviral drugs, or vaccines, a common theme in many debates is that existing guidelines and allocation mechanisms have given up on certain values. For example, advocates for disadvantaged groups criticize priority point systems that use the SOFA score. They argue that these criteria are discriminatory, for they fail to acknowledge pre-existing discrimination in access to healthcare (Schmidt 2020). Similarly, disability advocates argue against rationing plans based solely on survival probabilities because they are inherently discriminatory for certain types of disabled patients. Some even reject a detailed triage protocol in favor of random selection (Ne'eman 2020).

In this paper, we report on our integrated research and policy efforts on design of pandemic medical resource allocation protocols during COVID-19. In these efforts, we followed an institution design paradigm called *minimalist market design* (Sönmez 2023) and succeeded in informing policy in several jurisdictions. Under this framework, the main tasks of a design economist are the following:

1. Identify the mission of the institution. What are the policy objectives of policymakers, system operators, and other stakeholders?
2. Assess whether the existing institution (in our application, the priority system) is a good choice for the design objectives.
3. If the existing institution is not a good choice, then find the root causes of the failures and design an alternative institution that corrects them by only interfering with the root causes of the failures.
4. In case there are multiple “minimalist” designs through tasks 1–3, present a comprehensive analysis of these competing institutions.

At the outset of the pandemic, we observed that many shortcomings of existing protocols are directly tied to the restriction of allocation rules to the priority system. A priority system’s single priority order for all units impedes its ability to represent a variety of ethical considerations. As an alternative, we formulated and proposed the *reserve system* by directly addressing the root cause of the shortcomings of the priority system.

In a reserve system, units are divided into multiple categories, with each representing an ethical value or a balance of multiple ethical values. Rather than relying on a single priority order to allocate all units, a category-specific priority order prioritizes individuals for units in each category. This heterogeneity allows a rationing system to accommodate the desired ethical values without needing to aggregate them into a single metric or into a strict lexicographic hierarchy.

In some applications, minimalist market design prescribes a unique institution. Some examples are the U.S. Army’s branching system for the cadets (Greenberg et al. 2023) and Supreme Court judgments in India for affirmative action policies for government jobs (Sönmez and

Yenmez 2022). In these applications, the fourth task above under minimalist market design is redundant. In our application, in contrast, a reserve system is a large class of allocation rules that have the flexibility to accommodate a rich set of normative principles. However, when individuals can receive units from multiple categories, some of these normative implications may be hard to observe, even for experts, and policymakers may risk undermining policy objectives by choosing the wrong reserve system. Consequently, we present a comprehensive analysis of the reserve system in the formal part of our paper.

The rest of our paper is organized as follows. In Section 2, we detail how minimalist market design shaped our integrated research and policy efforts for pandemic allocation of scarce life-saving resources during the COVID-19 pandemic. In Section 3, we present a general model and analysis of the reserve system. This section also includes a detailed discussion of the earlier literature on reserve systems and relates our analytical results to earlier ones in the literature. In Section 4, we discuss the policy impact of our paper during and after the COVID-19 pandemic. We conclude in Section 5 and relegate resource-dependent design considerations, a theory of smart reserve systems, and the proofs of our formal results to the appendix.

2. A Case Study in Minimalist Market Design: Pandemic Medical Resource Allocation

Minimalist market design (Sönmez 2023) is a framework that integrates research and policy efforts to influence the design of real-life allocation mechanisms. Our paper and policy work on the allocation of scarce life-saving resources during the COVID-19 pandemic is an exercise in this institution design paradigm.

Traditionally, economists have rarely, if ever, been part of triage committees or task forces that design pandemic medical resource allocation systems (Pathak et al. 2020a). One reason for this lack of representation is the following: Whereas the primary normative focus of welfare economics largely revolves around preference *utilitarianism* (Li 2017), bioethics and public health scholars and practitioners frown upon a narrow focus on this principle. Preference utilitarianism is also the dominant normative objective for mainstream approaches in design economics. In contrast, the objectives of policymakers, system operators, and other stakeholders—which may include nonutilitarian principles—take central stage in *minimalist market design*. This cautious and respectful approach to institution design is at the heart of the policy impact of our paper—reported in Section 4—during and after the COVID-19 pandemic.

Sönmez (2023) identifies three main steps and, in some applications, a fourth step to achieve policy goals

within this framework. We begin by briefly reviewing these steps and how they fit into our integrated research and policy efforts during COVID-19.

The first step under minimalist market design is identifying the main objectives of the policymakers, stakeholders, and authorities that operate the system. The historical evolution of pandemic resource allocation guidelines for influenza and other highly contagious respiratory diseases holds key clues regarding these objectives, as do bioethics and law literatures on the topic. We first identify and document these objectives as fulfilling multiple ethical criteria in a balanced manner.

The second step is determining whether the institution in place satisfies these objectives. In our application of pandemic medical resource allocation, various forms of the priority system turn out to be the protocol of choice. These systems fail to satisfy key design objectives, as we discuss and document below.

The third step is proposing an alternative to the current institution through a *minimally invasive* intervention. To do so, it is important to identify and address the *root causes* of the inconsistency between objectives and practice. Reliance on a single priority list to formulate the claims of individuals on the scarce resources is the root cause of the failures of existing pandemic rationing protocols. In this respect, we formulate and advocate for the *reserve system*, which differs from the priority system in this key aspect only, and fulfills the design objectives for the allocation of various scarce life-saving resources (e.g., ventilators, ICU beds, therapeutics, or vaccines) through multiple priority lists.

In some settings, a unique minimalist intervention may eliminate the inconsistency between the objectives of stakeholders and practice. In those cases, the three main steps of minimalist market design prescribe a unique intervention. In other applications, however, some of the primary objectives of stakeholders may be incompatible with others.⁴ In such applications, a design economist may need to formulate compromises between these objectives. Finally, in other cases, multiple minimalist interventions may eliminate the inconsistency between the primary objectives and practice. The fourth step of the minimalist market design may be potentially valuable in these applications, especially in applications “in which issues of social, racial and distributive justice are particularly salient” (Hitzig 2020, p. 408). Pandemic medical resource allocation is one such application.

The goal of the fourth step in minimalist market design is maintaining *informed neutrality* (Li 2017) between reasonable normative principles in design proposals. Our formal analysis in Section 3 pertains to this aspect of our application. Because our proposed reserve system has many flexible ways to fulfill multiple ethical principles, we develop new theory to interpret and characterize this

flexibility. This allows us to propose alternative solutions to policymakers with different tradeoffs and helps them to avoid unintended consequences.

In the next four subsections, we discuss how each of these steps was carried out in the context of pandemic medical resource allocation and how our minimalist approach to institution design helped us achieve policy goals and shape aspects of COVID-19 resource-allocation guidelines or procedures virtually in real-time with the progress of this paper.

2.1. Ethical Principles for Allocation of Scarce Life-Saving Resources

In the early days of the COVID-19 pandemic, it became clear that medical systems in many countries would be overwhelmed in providing care for patients. The most urgently needed life-saving resources, such as intensive care unit beds and ventilators, were insufficient to meet expected demand, even in countries with advanced health systems. New vaccines and therapeutics had to be developed to combat the SARS-CoV-2 virus. Once developed, the initial demand for these resources would likely surpass the supply, which was bounded by production capacity. As a result, rationing of such resources would be inevitable.

The rationing of ICU beds and ventilators was already underway in Northern Italy (Rosenbaum 2020) when leading bioethicists laid out principles of the design of triage protocols in an issue of *The New England Journal of Medicine* (available online in March 2020). In this issue, Emanuel et al. (2020) urge healthcare providers to use multiple ethical principles in the design of triage protocols, rather than a single criterion. Because subjective value judgments could lead to different weights for each principle, this paper argues for adopting transparent, fair, and consistent allocation methods determined with the consensus of affected parties, clinicians, patients, public officials, and others. Truog et al. (2020) also advocate using multiple principles in determining rationing protocols.

Different advocates put forward specific principles, including the *utilitarian principle* of saving the most lives or saving the most life-years; *instrumental value*—that is, healthcare professionals and frontline workers who will be important in helping society cope with the pandemic should get priority; the *life-cycle principle*, in which younger people should be prioritized over older patients due to a fairness notion, which states that each person should have a similar expected duration of life; *equal treatment of equals*, whereby similar patients should have an equal chance for the resources; and *reciprocity*, in which patients who have done good deeds for the society in the past should have priority.

Although these papers mention these principles in detail explicitly, they are largely silent about their operationalization with a specific allocation protocol. We

Table 1. Illustration of Multiprinciple Strategy for Ventilator Allocation

Principle	Specification	Point system			
		1	2	3	4
Save the most lives	Prognosis for short-term survival (SOFA score)	SOFA < 6	SOFA 6–9	SOFA 10–12	SOFA > 12
Save the most years	Prognosis for long-term survival (medical assessment of comorbidities)	No comorbidities	Minor comorbidities	Major comorbidities	Severe comorbidities; likely death within 1 year
Life-cycle principle	Prioritize those who have had the least chance to live through life's stages	Age 12–40	Age 41–60	Age 61–74	Age > 74

Notes. SOFA stands for sequential organ failure assessment. Multiprinciple point system developed by White et al. (2009). A lower score gives higher priority.

have to look into the history of pandemic triage guidelines in the United States and older bioethics literature to find what allocation procedures were actually proposed or adopted.

2.2. Priority System for Allocation of Scarce Life-Saving Resources

Prior to our work, the most common allocation mechanism for pandemic medical rationing was a version of a *priority system*. Under this system, individuals are priority-listed based on various criteria reflecting the ethical values guiding the allocation of the scarce medical resource, and all units are allocated based on this single priority order.

In some applications, most notably for the allocation of ventilators and ICU beds, the underlying priority order is obtained through a monotonic scoring function. This refinement of a priority system is called a *priority point system*. Under this system, each ethical value is also represented with a monotonic function. Values are then integrated with an additive formula, which produces an aggregate point score for each patient. The claims of patients on scarce life-saving resources are determined based on their point scores, where a lower score may be associated with a higher claim or a lower claim, depending on the application. Sometimes, priority scores may be coarsened into multiple tiers, and all patients in the same tier may have the same claim. Tie-breaking within a tier is typically based on clinical criteria or lotteries.

A single-principle point system is a priority point system based on only one ethical value. The 2015 New York State Ventilator Guideline (Zucker et al. 2015) is a prominent example. In the system, eligible patients are ordered based on a measure of estimated mortality risk called the *SOFA score*, which is re-evaluated every 48 hours. In cases of excess demand among members of a given priority tier, New York and other proposals recommend random allocation—a lottery—among equal-priority patients (Zucker et al. 2015, Emanuel et al. 2020).

Several bioethicists and clinicians criticize single-principle priority point systems solely based on the SOFA score for ignoring multiple ethical values. These critics emphasize the need to integrate a variety of ethical values, and they advocate for a *multiprinciple* approach (e.g., White et al. 2009 and Daugherty-Biddison et al. 2017). White et al. (2009) describe a multiprinciple priority point system where several ethical values are placed on a numerical scale and summed up across ethical values to arrive at a single number. Variants of the system shown in Table 1 are widespread as the leading multiprinciple priority point system for ventilators,⁵ and more than half of U.S. states adopted either a single- or multiple-principle priority point system during the COVID-19 pandemic (Whyte 2020).

Although practical, priority systems have a number of important limitations. A priority system struggles to integrate different ethical values because some values are incommensurable. Incommensurable values do not share a common standard of measurement and cannot be compared with one another. For example, the values of saving the most lives and the life-cycle principle in Table 1 are scaled to numeric values, even though there is no natural scaling. In addition, many ethical values have implications for group composition, and a priority point system (or a priority system in general) lacks the flexibility to accommodate these considerations. In many cases, these challenges have led to the exclusion of some ethical values all together. We elaborate on several of these points next, focusing on examples of the debates on rationing of ventilators and ICU beds. Because virtually all states with guidelines recommend priority point systems, we present their shortcomings, but many of our points directly apply to the broader class of priority systems in general.

2.2.1. Failure to Represent the Desired Ethical Values. A priority point system requires that ethical values be mapped to a single linear order. However, some ethical principles involve claims of patients that

cannot be represented with a monotonic function. One example is group-based policies, such as those regarding providing access to certain groups of individuals, such as historically disadvantaged groups, at proportions equal to or higher than their representation in the population. As we will discuss in Section 2.2.2, concerns for access for these groups have been an important desired feature of medical resource allocation mechanisms during the COVID-19 pandemic for therapeutics and vaccines. This shortcoming of priority systems is exacerbated due to the fact that considerations based on group composition cannot be represented with a single function that relies only on individual attributes.

When constructing priority points and incorporating multiple ethical values, a priority point system norms or scales different and potentially incommensurable ethical principles into one dimension. These challenges are like the usual ones associated with aggregating social alternatives into a single ordering based on multiple inputs—a situation that involves “comparing apples to oranges.” The debate on how rationing guidelines should compare claims of children versus adults illustrates this issue. Initial Massachusetts guidelines state that indicators that feed into scores for adults are not reliable for children. They explain that “scoring systems that are meaningful for adult critical care patients do not apply to pediatric patients or newborns” (DPH 2020a, p. 17). As a result, the Massachusetts guidelines use a different scoring system for children. However, their point system then uses a single priority point system to evaluate all patients together. This decision ends up comparing the point scores of children with those of adults.

Third, the fact that all resources are ordered using a single uniform priority order can result in the exclusion of certain ethical values. A prominent example appears in the debate about prioritizing essential personnel. Many groups argue that essential personnel, and especially frontline healthcare workers, should receive priority allocation of scarce resources under triage scenarios. This view is also strongly endorsed by medical ethicists based on the backward-looking principle of *reciprocity* and the forward-looking principle of *instrumental value* (Emanuel et al. 2020). Nevertheless, states such as Minnesota and New York had to give up on this consideration in the past, largely due to concerns about extreme scenarios where no units may remain for the rest of the society. The Minnesota Pandemic Working group reasons that:

... It is possible that they [key workers] would use most, if not all, of the short supply of ventilators; other groups systematically would be deprived access. (Vawter et al. 2010, p. 59)

The New York State Task Force recognized the need to provide “insurance” for frontline health workers, but ultimately decided against such a priority:

Expanding the category of privilege to include all the workers listed above may mean that only health care workers obtain access to ventilators in certain communities. This approach may leave no ventilators for community members, including children; this alternative was unacceptable to the Task Force. (Zucker et al. 2015, p. 45)

For both states, the committees abandoned the ethical values of reciprocity and instrumental value because of the limitation of priority systems. In a priority system, providing preferential access to any group for any portion of the resources means giving preferential treatment for all units.

Fourth, a single priority order may struggle to integrate the principle of nonexclusion. This principle is the idea that every patient, no matter his or her circumstances, should have some hope of obtaining a life-saving resource. In the March 2020 Alabama rationing plan, individuals with severe or profound mental disabilities were considered “unlikely candidates for ventilator support.”⁶ Washington state guidelines recommend that hospital patients with “loss of reserves in energy, physical ability, cognition and general health” be switched to outpatient or palliative care (Fink 2020). In a priority system coupled with excess demand for available resources by the higher-priority groups—even without any explicit exclusion of certain types of individuals—some patients in a lower-priority group would never be treated during a shortage.

2.2.2. Implementation Issues in Existing Guidelines. The conceptual challenges with a priority point system are reflected in actual design challenges in several guidelines. We describe three examples.

Massachusetts Crisis Standards of Care guidance for the COVID-19 pandemic was developed in April 2020 by a committee consisting of medical experts and ethicists (DPH 2020a). The guidelines provided an adoption of the system described in Table 1 without the life-cycle consideration. However, after precisely spelling out the priority order with a table of numbers for each dimension, the document states:

Individuals who perform tasks that are vital to the public health response, including all those whose work directly to support the provision of care to others, should be given heightened priority. (DPH 2020a, p. 16)

This clause provides no further description of how heightened priority is to be implemented. This lack of transparency contrasts with the level of precision regarding other ethical principles and may reflect their inability to arrive at consensus, given the underlying priority point system.⁷

Crisis standards of care guidelines initially developed in Pittsburgh use a similar adaptation of the system

described in Table 1. They offer a vague description of tie-breakers:

In the event that there are ties in priority scores between patients, life-cycle considerations will be used as a tiebreaker, with priority going to younger patients, who have had less opportunity to live through life's stages. In addition, individuals who perform tasks that are vital to the public health response specifically, those whose work directly supports the provision of acute care to others will also be given heightened priority (e.g., as a tiebreaker between identical priority scores) (University of Pittsburgh 2020, p. 1)

In their adaptation of Table 1, the designers saw saving the most lives as more justified than either the life-cycle principle or the instrumental value principle. However, the guidelines did not choose between these two latter ethical values in the event of tie-breaking.

The third example is from the June 6, 2020, update to the Arizona allocation framework (DHS 2020). This document also offers a table prioritizing patients based on SOFA scores and whether a patient is expected to live or die within one or five years, despite successful treatment of illness. It then warns that "a situation could arise where limited resources are needed by two or more patients with the same triage priority scores" in which case "additional factors *may* be considered as priorities" (DHS 2020, p. 7). Among the list of additional factors are whether patients are pediatric patients; first responders or healthcare workers; single caretakers for minors or dependent adults; pregnant; or have not had an opportunity to experience life stages. There is no further detail on how multiple tie-breakers would be implemented.

Beyond these specific updates to guidelines during the COVID-19 pandemic, there are also concerns that incomplete descriptions have rendered such guidelines ineffective in other settings. During the 2004 shortage of the influenza vaccine, Schoch-Spana et al. (2005, p. 225) state that the Centers for Disease Control and Prevention (CDC) guidelines were too general and broad. Specifically:

Local providers thus faced gaps in the local supply of inactivated vaccine as well as the absence of *a priori* prioritization standards relevant to initial and evolving local conditions. Practitioners and local and state health authorities throughout the U.S. faced a similar predicament.

Despite these vagaries, some state departments of health penalized clinicians if protocols were not followed. For example, Lee (2004) describes that Massachusetts threatened a fine or prison time for whoever violates the CDC order on distribution during the 2004 flu shot shortage. The requirement to follow an incompletely specified system places clinicians in a difficult position.

2.3. Reserve System as a Minimalist Alternative to Priority System

Challenges presented in Sections 2.2.1 and 2.2.2 mainly stem from one limiting feature of a priority system: It relies on a single priority ranking of patients that is identical for all units. Thus, the root cause of its failure is the use of a uniform priority criteria for allocation of each and every unit of the scarce medical resource. This limitation is directly addressed under a reserve system, which allows for heterogeneity of patient claims over different units.

A *reserve system* has the following three main parameters:

1. A division of all units into multiple segments referred to as *reserve categories*,⁸
2. The number of units in each of these categories, and
3. A *priority order* of individuals for each of reserve category.

Observe that a priority system is a single-category reserve system. This aspect of the reserve system makes it easier to communicate with policymakers and experts in bioethics and public health who are familiar with the priority system.

Reserve categories can differ, either based on the groups to receive higher priority or the combination of ethical principles to be invoked. Thus, categories can directly or indirectly incorporate different ethical principles advocated for the rationing of life-saving resources during the COVID-19 pandemic (e.g., see Emanuel et al. 2020 and Truog et al. 2020). As a result, the reserve system minimally differs from the more restrictive priority system by allowing category-specific criteria to prioritize individuals.

Various forms of the reserve system have been the procedure of choice in the allocation of various scarce resources, especially when compromises need to be reached among different segments of society. In many settings, reserve categories and the number of resources reserved for each category were determined after years of community engagement. For example, the most extensive affirmative action program in the world takes place in India in the allocation of state jobs and public school seats. After decade-long community discussions, a reserve system was adopted for affirmative action (see the 1979 Mandal Commission Report for the summary of the efforts) and was formulated in the 1992 *Indra Sawhney v. Union of India* Supreme Court case (Sönmez and Yenmez 2022). Chicago public schools also adopted a reserve system in admission to highly selective high schools after U.S. federal law prohibited the use of racial criteria. Five reserve categories were adopted based on socioeconomic group specifications (Dur et al. 2020).

The priority order of patients for each reserve category also incorporates information on *eligibility criteria*. If every individual is eligible only through one category, then these categories can be processed in parallel to

determine who will receive the resources, and the reserve system is uniquely identified. However, in most applications, many individuals will be eligible through multiple reserve categories. For example, essential workers, such as healthcare professionals or frontline workers, can be eligible for an ICU bed through the essential-worker category and also through an unreserved category that is open to all patients. In such cases, it is important to specify which category will be used to satisfy the demand. This brings us to the fourth step of the minimalistic market design.

2.4. Designing Reserve Systems for Rationing of Life-Saving Resources

When category membership overlaps (i.e., when at least one individual is eligible to receive units from two or more categories), the three main parameters of the reserve system allow for multiple ways to operationalize a reserve system. Handling this subtle aspect of the reserve system and communicating it to policymakers and other stakeholders requires extra care. On the one hand, the overlapping nature of category memberships brings an additional degree of freedom in the design of reserve categories, allowing for a wider range of policies. On the other hand, this aspect of a reserve system may also be a potential source of loopholes in rationing guidelines and, thus, may result in unintended consequences (cf. Dur et al. 2018, Sönmez and Yenmez 2022, and Pathak et al. 2023a, b). Therefore, before a reserve system can be recommended and utilized for the rationing of scarce pandemic resources, it is important to understand the implications of this extra degree of freedom to help policymakers make informed decisions on which of these systems better serves their policy objectives. That is, the role of the fourth step of the minimalistic market design is to illustrate the range of possibilities, while still maintaining *informed neutrality* (Li 2017) between reasonable normative principles.

Starting with its introduction to market-design literature with Hafalir et al. (2013), virtually all studies on reserve systems have focused on a special setting where category-specific priorities are uniquely determined by an underlying baseline priority order, together with category membership. Our application on pandemic medical resource allocation, in contrast, calls for a more flexible structure on category-specific priority orders. Therefore, to maintain informed neutrality between reasonable normative principles in recommending reserve systems for various settings in scarce medical resource allocation, in Section 3, we develop a general theory of reserve systems based on a richer system of priority orders.

3. A General Theory of Reserve Systems

Reserve systems are common in applications where competing interest groups fail to agree on allocation criteria.

If each individual is eligible to receive units only from one of these categories, then the theory of priority systems directly extends to this most basic form of reserve systems. However, in most applications, at least some of the individuals are eligible to receive units from multiple categories, thereby resulting in multiplicities in implementation of “categorized” priority systems. In this section, we develop a general theory of reserve systems, with particular emphasis on the analytical structure and the distributional implications of this multiplicity.

The earlier literature on reserve systems considers a special setting, where all categories share a common baseline priority order, although various protected groups receive preferential treatment or exclusive access in some of the categories. In contrast, we develop a theory of reserve systems where no cross-category structure is imposed on priority orders.⁹

Our model allows priority orders at various categories to be completely independent from each other, although they do not need to be. Because it may not be possible to capture priorities for all ethical considerations through adjustments of a single baseline priority order, this level of generality is necessary for our main application.¹⁰ To the best of our knowledge, reserve systems have not been studied in this generality before.

3.1. The Primitives

Despite the generality of our model, the terminology is tailored to the main application of pandemic medical resource rationing. There is a set of *patients* I and q identical *medical units* to allocate. There is a set of *reserve categories* \mathcal{C} . For every category $c \in \mathcal{C}$, r_c units are reserved so that $\sum_{c \in \mathcal{C}} r_c = q$. It is important to emphasize that individual units are not associated with the categories in our model. The phrase “ r_c units are reserved” does not mean that specific units are set aside for category c . Rather, it means that for the purposes of accounting, a total of unspecified r_c units are attached to category c .

For every category $c \in \mathcal{C}$, there is a linear *priority order* π_c over the set of patients I and \emptyset . This priority order represents the relative claims of the patients on units in category c , as well as their eligibility for those units. For every category $c \in \mathcal{C}$ and patient $i \in I$, we say that patient i is *eligible* for category c if

$$i \pi_c \emptyset.$$

Given priority order π_c , we represent its weak order by $\underline{\pi}_c$. That is, for any $x, y \in I \cup \{\emptyset\}$,

$$x \underline{\pi}_c y \Leftrightarrow x = y \text{ or } x \pi_c y.$$

For our main application of pandemic rationing, π_c orders patients based on the balance of ethical principles guiding the allocation of units in category c .

3.2. The Outcome: A Matching

A *matching* $\mu : I \rightarrow \mathcal{C} \cup \{\emptyset\}$ is a function that maps each patient either to a category or to \emptyset such that $|\mu^{-1}(c)| \leq r_c$ for every category $c \in \mathcal{C}$. For any patient $i \in I$, $\mu(i) = \emptyset$ means that the patient does not receive a unit and $\mu(i) = c \in \mathcal{C}$ means that the patient receives a unit reserved for category c . Let \mathcal{M} denote the set of matchings.

Our formulation of the outcome of a reserve system involves a subtle, but important, modeling choice. In models where there is a single type of good and each individual has unit demand, the traditional way to describe the outcome of a reserve system is through a *choice rule*, which indicates the set of individuals who are awarded a unit for any set of applicants. Because each individual is indifferent between all units of the good, this formulation is often seen as sufficient for analysis. However, for our purposes, it is not.

One of the primary appeals of the reserve system is its ability to facilitate compromises between various interest groups by allocating scarce resources through multiple categories. Although individuals are indifferent between all units of the scarce good, they have potentially different claims on units from different categories. And, consequently, merely specifying who receives a unit may not always be sufficient. For various analytical exercises, it is also important to specify through which category individuals receive their units.¹¹ Hence, we deviate from the prior formulation and let an outcome embed information about the category.

For any matching $\mu \in \mathcal{M}$ and any subset of patients $I' \subseteq I$, let $\mu(I')$ denote the set of patients in I' who are matched with a category under matching μ . More formally,

$$\mu(I') = \{i \in I' : \mu(i) \in \mathcal{C}\}.$$

These are the patients in I' who are matched (or, equivalently, who are assigned units) under matching μ .

3.3. Axioms

In real-life applications of our model, it is important to allocate units to qualified individuals without wasting any and abiding by the priorities governing the allocation of these units. We next formulate this idea through three axioms:

Definition 1. A matching $\mu \in \mathcal{M}$ complies with *eligibility requirements* if, for any $i \in I$ and $c \in \mathcal{C}$,

$$\mu(i) = c \Rightarrow i \pi_c \emptyset.$$

Our first axiom formulates the idea that units in any category should be awarded only to eligible individuals. For most applications of rationing of life-saving resources, any patient who is eligible for one category can also be eligible for any category. In those applications, if a patient is ineligible for all categories, then this patient can, as well, be dropped from the set of

individuals. Hence, compliance with eligibility requirement vacuously holds in applications when all individuals are eligible for all categories. This includes most applications in pandemic resource allocation, although there can be examples where certain individuals may be eligible for only some of the categories.

Definition 2. A matching $\mu \in \mathcal{M}$ is *nonwasteful* if, for any $i \in I$ and $c \in \mathcal{C}$,

$$i \pi_c \emptyset \quad \text{and} \quad \mu(i) = \emptyset \Rightarrow |\mu^{-1}(c)| = r_c.$$

Our second axiom formulates the idea that no unit should go idle, as long as there is an eligible individual to award it. That is, if a unit remains idle, then there should not be any unmatched individual who is eligible for the unit. In those applications of rationing vital resources where each patient is eligible for all units, nonwastefulness corresponds to either matching all the units or all the patients.

Definition 3. A matching $\mu \in \mathcal{M}$ respects *priorities* if, for any $i, i' \in I$ and $c \in \mathcal{C}$,

$$\mu(i) = c \text{ and } \mu(i') = \emptyset \Rightarrow i \pi_c i'.$$

Our last axiom formulates the idea that, for each category, the units should be allocated based on the priority order of individuals in this category.

As far as we know, in every real-life application of a reserve system, each of these three axioms is either explicitly or implicitly required. Hence, we see these three axioms as minimal requirements for a reserve system. In the next two sections, we present two characterizations of matchings that satisfy these axioms, one through a notion akin to competitive equilibria and the second based on the celebrated deferred acceptance algorithm of Gale and Shapley (1962).

3.4. Cutoff Equilibria

In many real-life applications of reserve systems, the outcome is often publicized through a system that identifies the lowest-priority individual that qualifies for admission for each category. This representation makes it straightforward to verify that an allocation was computed following the announced policy because an individual can compare her priority to the announced cutoffs. Often, a cardinal representation of the priority order, such as a merit score or a lottery number, is used to identify these “cutoff” individuals. This observation motivates the following equilibrium notion.

For any category $c \in \mathcal{C}$, a *cutoff* f_c is an element of $I \cup \{\emptyset\}$ such that $f_c \pi_c \emptyset$. Cutoffs in our model play the role of prices for exchange or production economies, where the last condition corresponds to prices being nonnegative in those economies. We refer to a list of cutoffs $f = (f_c)_{c \in \mathcal{C}}$ as a *cutoff vector*. Let \mathcal{F} be the set of cutoff vectors.

Given a cutoff vector $f \in \mathcal{F}$, for any patient $i \in I$, define the *budget set* of patient i at cutoff vector f as

$$\mathcal{B}_i(f) = \{c \in \mathcal{C} : i \leq_c f_c\}.$$

A *cutoff equilibrium* is a pair consisting of a cutoff vector and a matching $(f, \mu) \in \mathcal{F} \times \mathcal{M}$ such that

1. For every patient $i \in I$,
 - a. $\mu(i) \in \mathcal{B}_i(f) \cup \{\emptyset\}$, and
 - b. $\mathcal{B}_i(f) \neq \emptyset \Rightarrow \mu(i) \in \mathcal{B}_i(f)$.
2. For every category $c \in \mathcal{C}$,

$$|\mu^{-1}(c)| < r_c \Rightarrow f_c = \emptyset.$$

A cutoff equilibrium is an analog of a competitive equilibrium for a reserve system. A cutoff vector-matching pair is a cutoff equilibrium if

1. Each patient who has a nonempty budget set is matched with a category in her budget set, and each patient who has an empty budget set remains unmatched, and
2. Each category that has not filled its quota under this matching has cutoff \emptyset .

The first condition corresponds to “preference maximization within budget set,” and the second condition corresponds to the “market clearing condition.”

Our first characterization result gives an equivalence between cutoff equilibrium matchings and matchings that satisfy our basic axioms.

Theorem 1. *For any matching $\mu \in \mathcal{M}$ that complies with eligibility requirements, is nonwasteful, and respects priorities, there exists a cutoff vector $f \in \mathcal{F}$ that supports the pair (f, μ) as a cutoff equilibrium. Conversely, for any cutoff equilibrium $(f, \mu) \in \mathcal{F} \times \mathcal{M}$, matching μ complies with eligibility requirements, is nonwasteful, and respects priorities.*

There can be multiple equilibrium cutoff vectors that support a matching at cutoff equilibria. Next, we explore the structure of equilibrium cutoff vectors.

For any matching $\mu \in \mathcal{M}$ and category $c \in \mathcal{C}$, define

$$\bar{f}_c^\mu = \begin{cases} \min_{\pi_c} \mu^{-1}(c) & \text{if } |\mu^{-1}(c)| = r_c \\ \emptyset & \text{otherwise} \end{cases} \quad \text{and,} \quad (1)$$

$$\underline{f}_c^\mu = \begin{cases} \min_{\pi_c} \left\{ i \in \mu(I) : i \leq_c \max_{\pi_c} ((I \setminus \mu(I)) \cup \{\emptyset\}) \right\} & \text{if } \max_{\pi_c} ((I \setminus \mu(I)) \cup \{\emptyset\}) \neq \emptyset \\ \emptyset & \text{otherwise.} \end{cases} \quad (2)$$

Here,

- \bar{f}_c^μ identifies
 - the lowest π_c -priority patient who is matched with category c if all category- c units are exhausted under μ , and
 - \emptyset if there are some idle category- c units under μ , whereas

- \underline{f}_c^μ identifies

- the lowest π_c -priority patient with the property that every weakly higher π_c -priority patient than her is matched under μ if some category- c eligible patient is unmatched under μ , and
- \emptyset if all category- c eligible patients are matched under μ .

Let μ be any matching that respect priorities. By construction,

$$\bar{f}_c^\mu \leq_c \underline{f}_c^\mu \quad \text{for any } c \in \mathcal{C}.$$

Our next result characterizes the set of cutoff vectors.

Lemma 1. *Let $\mu \in \mathcal{M}$ be a matching that complies with eligibility requirements, is nonwasteful, and respects priorities. Then, the pair (g, μ) is a cutoff equilibrium if, and only if,*

$$\bar{f}_c^\mu \leq_c g_c \leq_c \underline{f}_c^\mu \quad \text{for any } c \in \mathcal{C}.$$

An immediate corollary to Lemma 1 is that for each cutoff equilibrium matching μ , $\bar{f}_c^\mu = (\bar{f}_c^\mu)_{c \in \mathcal{C}}$ is a *maximum equilibrium cutoff vector* and $\underline{f}_c^\mu = (\underline{f}_c^\mu)_{c \in \mathcal{C}}$ is a *minimum equilibrium cutoff vector*.

Of these equilibrium cutoff vectors, the first one has a clear economic interpretation. The maximum equilibrium cutoff of a category indicates the *selectivity* of this particular category. The higher the maximum cutoff is, the more competitive it becomes to receive a unit through this category. This is also the cutoff that is typically announced in real-life applications of reserve systems, due to its clear interpretation. The interpretation of the minimum equilibrium cutoff of a category is more about the entire matching than the category itself, and, in some sense, it is artificially lower than the maximum equilibrium cutoff, due to individuals who are matched with other categories. All other equilibrium cutoffs between the two are also artificial in a similar sense. Therefore, for much of our analysis, we focus on the maximum equilibrium cutoff vector.

3.5. Characterization via Deferred Acceptance Algorithm

Although Theorem 1 gives a full characterization of matchings that satisfy our three axioms, it leaves open the question of how to find such a matching. In this section, we present a procedure to construct all such matchings utilizing the celebrated deferred-acceptance algorithm by Gale and Shapley (1962).

Consider the following hypothetical many-to-one matching market. The two sides of the market are the set of patients I and the set of categories \mathcal{C} . Each patient $i \in I$ can be matched with at most one category, whereas each category $c \in \mathcal{C}$ can be matched with as many as r_c patients. Category c is endowed with the linear order

π_c that is specified in the primitives of the original rationing problem.

Observe that in our hypothetical market, all the primitives introduced so far naturally follow from the primitives of the original problem. The only primitive of the hypothetical market that is somewhat “artificial” is the next one:

Each patient $i \in I$ has a strict preference relation \succ_i over the set $\mathcal{C} \cup \{\emptyset\}$, such that, for each patient $i \in I$,

$$c \succ_i \emptyset \Leftrightarrow \text{patient } i \text{ is eligible for category } c.$$

Whereas in the original problem, a patient is indifferent between all units (and therefore all categories as well), in the hypothetical market, she has strict preferences between the categories. This “flexibility” in the construction of the hypothetical market is the basis of our main characterization.

For each patient $i \in I$, let \mathcal{P}_i be the set of all preferences constructed in this way, and let $\mathcal{P} = \times_{i \in I} \mathcal{P}_i$.

Given a preference profile $\succ = (\succ_i)_{i \in I}$, the **Individual Proposing Deferred Acceptance Algorithm (DA)** produces a matching as follows.

Individual Proposing Deferred Acceptance Algorithm (DA)

Step 1: Each patient in I applies to her most preferred category among categories for which she is eligible. Suppose that I_c^1 is the set of patients who apply to category c . Category c tentatively assigns applicants with the highest priority according to π_c until all patients in I_c^1 are chosen or all r_c units are allocated, whichever comes first, and permanently rejects the rest. If there are no rejections, then stop.

Step k : Each patient who was rejected in Step $k - 1$ applies to her next preferred category among categories for which she is eligible, if such a category exists. Suppose that I_c^k is the union of the set of patients who were tentatively assigned to category c in Step $k - 1$ and the set of patients who just applied to category c . Category c tentatively assigns patients in I_c^k with the highest priority according to π_c until all patients in I_c^k are chosen or all r_c units are allocated, whichever comes first, and permanently rejects the rest. If there are no rejections, then stop.

The algorithm terminates when there are no rejections, at which point all tentative assignments are finalized.

A matching $\mu \in \mathcal{M}$ is called *DA-induced* if it is the outcome of DA for some preference profile $\succ \in \mathcal{P}$.

We are ready to present our next result:

Proposition 1. *A matching complies with eligibility requirements, is nonwasteful, and respects priorities if, and only if, it is DA-induced.*

Not only is this result a second characterization of matchings that satisfy our three basic axioms, but it

also provides a concrete procedure to calculate all such matchings. Equivalently, Proposition 1 provides us with a procedure to derive all cutoff equilibria. This latter interpretation of our characterization leads us to a refinement of cutoff equilibrium matchings explored in our next section.

3.6. Sequential Reserve Matching

An interpretation of the DA-induced matchings is helpful to motivate focusing a subset of these matchings. Recall that the hypothetical two-sided matching market constructed above relies on an artificial preference profile $(\succ_i)_{i \in I}$ of patients over categories. What this corresponds to under DA is that any patient i is considered for categories that deem her eligible in sequence, following the ranking of these categories under her artificial preferences \succ_i . Unless there is a systematic way to construct these preferences, it may be difficult to motivate adopting this methodology for real-life applications. For example, if a patient i is considered first for an unreserved category and then for an essential personnel category, whereas another patient j with similar characteristics is considered for them in the reverse order, it may be difficult to justify this practice. That is, a potentially large set of matchings satisfy our three axioms, but not all are necessarily obtained through an intuitive procedure. This may be a challenge, especially in the context of medical rationing, because *procedural fairness* is also an important ethical consideration in this context. Procedural fairness is the main motivation for our focus in a subset of these matchings.

In many real-life applications of reserve systems, institutions process reserve categories sequentially and allocate units associated with each category one at a time using its category-specific priority order. We next formulate matchings obtained in this way and relate them to our characterization in Proposition 1.

An *order of precedence* \triangleright is a linear order over the set of categories \mathcal{C} . For any two categories $c, c' \in \mathcal{C}$,

$$c \triangleright c',$$

means that category- c units are to be allocated before category- c' units. In this case, we say category c has *higher precedence* than category c' . Let Δ be the set of all orders of precedence.

For a given order of precedence $\triangleright \in \Delta$, the induced *sequential reserve matching* φ_\triangleright is a matching that is constructed as follows:

Suppose categories are ordered under \triangleright as

$$c_1 \triangleright c_2 \triangleright \dots \triangleright c_{|\mathcal{C}|}.$$

Matching φ_\triangleright is found sequentially in $|\mathcal{C}|$ steps:

Step 1: Following their priority order under π_{c_1} , the highest-priority r_{c_1} category- c_1 -eligible patients in I are matched with category c_1 . If there are fewer

than r_{c_1} eligible patients in I , then all of these eligible patients are matched with category c_1 . Let I^1 be the set of patients matched in Step 1.

Step k: Following their priority order under π_{c_k} , the highest-priority r_{c_k} category- c_k -eligible patients in $I \setminus \bigcup_{k'=1}^{k-1} I^{k'}$ are matched with category c_k . If there are less than r_{c_k} eligible patients in $I \setminus \bigcup_{k'=1}^{k-1} I^{k'}$, then all of these eligible patients are matched with category c_k . Let I^k be the set of patients matched in Step k .

Given an order of precedence $\triangleright \in \Delta$, the induced sequential reserve matching complies with eligibility requirements, is nonwasteful, and respects priorities. Thus, it is DA-induced by Proposition 1. Indeed, it corresponds to a specific DA-induced matching.

Proposition 2. Fix an order of precedence $\triangleright \in \Delta$. Let the preference profile $\succ^\triangleright \in \mathcal{P}$ be such that, for each patient $i \in I$ and pair of categories $c, c' \in \mathcal{C}$,

$$c \succ_i^\triangleright c' \Leftrightarrow c \triangleright c'.$$

Then, the sequential reserve matching φ_\triangleright is DA-induced from the preference profile \succ^\triangleright .

We conclude this section with a comparative static result regarding the maximum equilibrium cutoff vectors supporting sequential reserve matchings:

Proposition 3. Fix two distinct categories $c, c' \in \mathcal{C}$ and a pair of orders of precedence $\triangleright, \triangleright' \in \Delta$ such that:

- $c' \triangleright c$,
- $c \triangleright' c'$, and
- For any $\hat{c} \in \mathcal{C}$ and $c^* \in \mathcal{C} \setminus \{c, c'\}$

$$\hat{c} \triangleright c^* \Leftrightarrow \hat{c} \triangleright' c^*.$$

That is, \triangleright' is obtained from \triangleright by only changing the order of c with its immediate predecessor c' . Then,

$$\bar{f}_c^{\varphi_{\triangleright'}} \pi_c \bar{f}_c^{\varphi_{\triangleright}}.$$

Recall that the maximum equilibrium cutoff for a category is indicative of how selective the category is. Therefore, the earlier a category is processed under a sequential reserve matching, the more selective it becomes by Proposition 3. This result is intuitive because the earlier a category is processed, the larger the set of patients who compete for these units in a set-wise inclusion sense.

3.7. Reserve Systems Under a Baseline Priority Order

In many real-life applications of reserve systems, there is a baseline priority order π of individuals. Starting with Hafalir et al. (2013), the earlier market-design literature on reserve systems exclusively considered this environment. This priority order may depend on scores in a standardized exam, a random lottery, or arrival time of application. For ventilator, ICU, antiviral, or monoclonal antibody allocation, it may depend on SOFA scores

described in Section 2.2. This baseline priority order is used to construct the priority order for each of the reserve categories, although each category except one gives preferential treatment to a specific subset of individuals. For example, in our main application, these could be essential personnel or persons from disadvantaged communities. In this section, we focus on this subclass of reserve systems and present an analysis of reserve matching on this class. In contrast to Section 3, where our analysis is more general than earlier literature and contributions are more conceptual than technical, analysis in this section represents our deeper technical contributions in relation to earlier literature.

To formulate this subclass of reserve systems, we designate a *beneficiary group* $I_c \subseteq I$ for each category $c \in \mathcal{C}$. It is assumed that all patients in its beneficiary group are eligible for a category. That is, for any $c \in \mathcal{C}$ and $i \in I_c$,

$$i \pi_c \emptyset.$$

There is an all-inclusive category $u \in \mathcal{C}$, called the *unreserved* (or *open*) category, which has all patients as its set of beneficiaries and endowed with the baseline priority order. That is,

$$I_u = I \quad \text{and} \quad \pi_u = \pi.$$

Any other category $c \in \mathcal{C} \setminus \{u\}$, referred to as a *preferential treatment category*, has a more exclusive set $I_c \subsetneq I$ of beneficiaries, and it is endowed with a priority order π_c with the following structure: For any pair of patients $i, i' \in I$,

$$\begin{aligned} i \in I_c \quad \text{and} \quad i' \in I \setminus I_c &\Rightarrow i \pi_c i', \\ i, i' \in I_c \quad \text{and} \quad i \pi i' &\Rightarrow i \pi_c i', \text{ and} \\ i, i' \in I \setminus I_c \quad \text{and} \quad i \pi i' &\Rightarrow i \pi_c i'. \end{aligned}$$

Under π_c , beneficiaries of category c are prioritized over patients who are not, but otherwise, their relative priority order is induced by the baseline priority order π .

Let I_g , referred to as the set of *general-community patients*, be the set of patients who are each a beneficiary of the unreserved category only:

$$I_g = I \setminus \bigcup_{c \in \mathcal{C} \setminus \{u\}} I_c.$$

In particular, two types of such problems have widespread applications.

We say that a priority profile $(\pi_c)_{c \in \mathcal{C}}$ has *soft reserves* if, for any category $c \in \mathcal{C}$ and any patient $i \in I$,

$$i \pi_c \emptyset.$$

Under a soft reserve system, all individuals are eligible for all categories. This is the case, for example, in our main application of pandemic resource allocation.

We say that a priority profile $(\pi_c)_{c \in \mathcal{C}}$ has *hard reserves* if, for any preferential treatment category $c \in \mathcal{C} \setminus \{u\}$,

1. $i \in I_c \setminus \emptyset$ for any of its beneficiaries $i \in I_c$, whereas
2. $\emptyset \subsetneq I_c \setminus i$ for any patient $i \in I \setminus I_c$ who is not a beneficiary.

Under a hard reserve system, only the beneficiaries of a preferred treatment category are eligible for units in this category. This is the case, for example, in H-1B visa allocation in the United States.

Allocation rules based on sequential reserve matching are used in many practical applications. Although an aspect often ignored in practical applications, it is important to pay attention to the choice of the order of precedence in these problems, for it has potentially significant distributional implications. In this subsection, we focus on sequential reserve matching under soft reserves, as this case likely is the relevant case for our main application of pandemic rationing.¹²

We already know from Proposition 3 that the later a category is processed, the less competitive it becomes. A natural follow-up question is whether this also means that the beneficiaries of this category necessarily benefit from this comparative static exercise. The answer would be straightforward if each patient was a beneficiary of a single category. But this is not the case in our model, because even if each patient is a beneficiary of at most one preferential treatment category, they are also each a beneficiary of the unreserved category. Indeed, even if that was not the case, unless the reserves are hard, non-beneficiaries may still be matched with units from preferential treatment categories. So the answer to this question is not an immediate implication of Proposition 3. Under some assumptions, such as when there is only one preferential treatment category (Dur et al. 2018), this question is already answered in the affirmative. However, as we present in the next example, this is not always the case.

Example 1. Suppose there are $q=6$ medical units to be allocated in total. There are six categories: the unreserved category u and five preferential treatment categories $c, c', c^*, \hat{c}, \tilde{c}$, and each category has a single unit capacity.

Suppose there are seven patients. All patients are beneficiaries of the unreserved category u :

$$I_u = \{i_1, i_2, i_3, i_4, i_5, i_6, i_7\}.$$

The beneficiaries of preferential treatment categories c, c^* , and \tilde{c} are given as

$$I_c = \{i_1, i_3, i_6\}, \quad I_{c^*} = \{i_2, i_5\}, \quad I_{\tilde{c}} = \{i_4, i_7\},$$

whereas there are no beneficiaries of preferential treatment categories c' and \hat{c} : $I_{c'} = \emptyset$ and $I_{\hat{c}} = \emptyset$. There are also no general-community patients: $I_g = \emptyset$. Suppose π , the baseline priority order of patients, is given as

$$i_1 \pi i_2 \pi i_3 \pi i_4 \pi i_5 \pi i_6 \pi i_7.$$

Also assume that all patients are eligible for all preferential treatment categories besides the unreserved category u .

We consider two sequential reserve matchings based on the following two orders of precedence:

$$\triangleright : \quad c' \triangleright c \triangleright c^* \triangleright \hat{c} \triangleright \tilde{c} \triangleright u,$$

and

$$\triangleright' : \quad c \triangleright' c' \triangleright' c^* \triangleright' \hat{c} \triangleright' \tilde{c} \triangleright' u.$$

In Table 2 below, we demonstrate the construction of the two induced sequential reserve matchings by processing their mechanics in parallel:

Thus, the two sequential reserve matchings match patients

$$\varphi_\triangleright = \{i_1, i_2, i_3, i_4, i_5, i_7\} \quad \text{and} \quad \varphi_{\triangleright'} = \{i_1, i_2, i_3, i_4, i_5, i_6\}.$$

In this problem category- c' and \hat{c} units are treated as if they are of unreserved category u , as these two categories do not have any beneficiaries in the problem. We use the baseline priority order π to match them.

Under the first order of precedence \triangleright , the highest π -priority patient i_1 , who is also a category- c beneficiary, receives the first unit, which is reserved for category c' . As a result, i_3 , who is the next category- c beneficiary, receives the only category- c unit. In the end, units associated with categories c^* and \tilde{c} are matched with their highest- and lowest-priority beneficiaries i_2 and i_7 , respectively. The highest-priority beneficiary of \tilde{c} , patient i_4 , receives the category- \hat{c} unit, which is processed like the unreserved category and before \tilde{c} . Hence, the lowest-priority beneficiary of category c , i_6 , remains unmatched, as the last unit, which is reserved for the unreserved category, goes to i_5 . Thus,

$$\varphi_\triangleright(I_c) = \{i_1, i_3\},$$

is the set of matched category- c beneficiaries.

Under the second order of precedence \triangleright' that switches the order of c and c' , the selectivity of category c increases, as it is processed earlier: The highest-priority category- c patient i_1 receives its unit instead of i_3 . This leads to the units associated with categories c^* and \tilde{c} being matched with their lowest- and highest-priority beneficiaries i_5 and i_4 , respectively—this is a switch of roles for these categories with respect to \triangleright . This is

Table 2. Construction of the Two Induced Sequential Reserve Matchings in Example 1

Step	Order of precedence \triangleright		Order of precedence \triangleright'	
	Category	Patient	Category	Patient
1	c'	i_1	c	i_1
2	c	i_3	c'	i_2
3	c^*	i_2	c^*	i_5
4	\hat{c}	i_4	\hat{c}	i_3
5	\tilde{c}	i_7	\tilde{c}	i_4
6	u	i_5	u	i_6

because the highest-priority beneficiary of c^* , patient i_2 , is now matched with category c' , which is processed like the unreserved category before c^* . This enables the lowest-priority beneficiary of category c , patient i_6 , to be matched with the unreserved category, as she is prioritized higher than i_7 under the baseline priority order. Hence,

$$\varphi_{\triangleright'}(I_c) = \{i_1, i_3, i_6\},$$

is the set of matched category- c beneficiaries.

Thus,

$$\varphi_{\triangleright}(I_c) \subsetneq \varphi_{\triangleright'}(I_c),$$

although category c is processed earlier under \triangleright' than under \triangleright .

Example 1 shows that earlier positive comparative static results in the literature fail to extend in our more general model. Also, observe that our negative example holds, even though each patient is a beneficiary of at most one preferential treatment category. Nevertheless, a positive result holds, provided that there are at most five categories and each patient is a beneficiary of no more than one preferential treatment category.

Proposition 4. *Assuming (i) there are at most five categories and (ii) each patient is a beneficiary of at most one preferential treatment category, consider a soft reserve system induced by a baseline priority order. Fix a preferential treatment category $c \in \mathcal{C} \setminus \{u\}$, another category $c' \in \mathcal{C} \setminus \{c\}$, and a pair of orders of precedence $\triangleright, \triangleright' \in \Delta$ such that:*

- $c' \triangleright c$,
- $c \triangleright' c'$, and
- For any $\hat{c} \in \mathcal{C}$ and $c^* \in \mathcal{C} \setminus \{c, c'\}$,

$$\hat{c} \triangleright c^* \Leftrightarrow \hat{c} \triangleright' c^*.$$

That is, \triangleright' is obtained from \triangleright by only changing the order of c with its immediate predecessor c' . Then,

$$\varphi_{\triangleright'}(I_c) \subseteq \varphi_{\triangleright}(I_c).$$

Sequential reserve systems play a central role in practical applications, including in applications for pandemic medical resource allocation reported in Section 4. Nevertheless, even in the basic version of the problem considered in this section, sequential reserve systems may have some limitations when beneficiaries of preferential categories overlap (cf. Sönmez and Yenmez 2022). For instance, they may be wasteful for the case of hard reserves, as Example B.1 in Appendix B shows. Although soft reserves are preferable for allocation of most life-saving resources, in Appendix B, we further introduce a theory of nonwasteful reserve systems for this version of the problem. Rather than processing the preferential treatment categories sequentially, processing them in parallel in a *smart* way turns out to be the key aspect of these systems.

3.8. Related Theoretical Literature

Our formal analysis of reserve systems contributes to literature in matching market design focused on distributional issues.

For the special setting with baseline priorities that we also consider in Section 3.7, Hafalir et al. (2013) introduce reserve systems to the matching market-design literature. In addition to an open category, where the priority order is the same as the baseline priority order, in their analysis, there is a single preferential treatment category. Hafalir et al. (2013) further assume that the preferential treatment category is processed prior to the open category. This version of a reserve system is called the *minimum guarantee* policy, whereas the version where the open category is processed prior to the preferential treatment category, later introduced in Dur et al. (2018), is called the *over-and-above* policy.

The concept of *order of precedence* used in sequential reserve systems is formally introduced by Kominers and Sönmez (2016), albeit in a more general form for individual units (or slots). For the basic version of the model considered in Hafalir et al. (2013), Dur et al. (2018) present the following comparative statics result: The later a preferential treatment unit is processed, the weakly better it is for its beneficiaries. A corollary of this result is that, the later the preferential category is processed under a sequential reserve system, the weakly better it is for its beneficiaries. Proposition 4 shows that this result extends for as long as there are at most four preferential treatment categories (and thus five categories in total). However, by Example 1, the result fails when there are five or more preferential treatment categories.

Unlike Proposition 4 and our results on smart reserve systems in Appendix B, Theorem 1 and Propositions 1–3 assume no cross-category structure on priority orders. As such, these results have no direct antecedents in the literature.

Although Theorem 1 is novel and our paper is the first one to formally introduce the notion of cutoff equilibrium for reserve systems, the use of this notion is widespread in real-life applications of reserve systems. In particular, the outcomes of reserve systems are often announced together with the cutoffs that support them. Examples include admission to exam high schools in Chicago (Dur et al. 2020), assignment of government positions in India (Sönmez and Yenmez 2022), college admissions in Brazil (Aygün and Bó 2021), and H-1B visa allocation in the United States for years 2006–2008 (Pathak et al. 2023a). For the first three of these applications, the cutoffs are given in terms of exam or merit scores, whereas for the last application, the cutoffs are given in terms of the date of visa-application receipt. Although the concept of cutoff equilibrium for reserve systems is novel to our paper, cutoffs are used in simpler matching environments in the absence of distributional

considerations (see, e.g., Balinski and Sönmez 1999 and Azevedo and Leshno 2016).

In addition to presenting a characterization of outcomes that satisfies three basic axioms, Proposition 1 also provides a procedure to calculate all cutoff matchings. Although our characterization itself is novel, the use of the deferred acceptance algorithm to derive a specific cutoff matching is not. This technique to obtain cutoff levels for reserve systems has been employed in various real-life applications, including in school choice algorithms of Boston Public Schools (Dur et al. 2018) and Chile (Correa et al. 2019) and college admissions algorithms for engineering colleges in India (Baswana et al. 2019).

Our characterizations in Theorem 1 and Proposition 1 use three simple axioms. As such, the resulting matchings can fail to be Pareto-efficient in some applications. This failure has to do with rather mechanical and inflexible assignment of agents to categories, as presented by Example B.1 in Appendix B. This can be mitigated by filling reserves in a “smart” way. The class of **Smart Reserve Matching Algorithms*** in Appendix B does precisely this and, through maximal utilization of reserves, always generates a Pareto-efficient matching. The **Smart Reserve Matching Algorithm*** is a generalization of the meritorious horizontal choice rule by Sönmez and Yenmez (2022), who introduced the idea under exclusively “minimum guarantee”-type reserves.

In addition to the above-discussed papers, which directly relate to our analysis, several others examine allocation under various forms of distributional constraints, such as upper quotas and regional quotas. Some notable papers along these lines include Abdulkadiroğlu (2005), Biro et al. (2010), Kojima (2012), Ehlers et al. (2014), Echenique and Yenmez (2015), Kamada and Kojima (2015, 2017, 2018), Bo (2016), Doğan (2016, 2017), Goto et al. (2017), Fragiadakis and Troyan (2017), Combe (2018), Kojima et al. (2018), Tomoeda (2018), Ehlers and Morrill (2019), and Imamura (2020).

Finally, our paper also introduces medical rationing into the market-design literature. Follow-up research on this application of market design include Aziz and Brandl (2021), Delacrétaz (2021), Dur et al. (2021), Grigoryan (2021), and Akbarpour et al. (2023).

4. Policy Impact

A few days after the initial draft of this paper was circulated, *The Boston Globe* reported that revised guidelines for ventilator rationing in Massachusetts give preferences to medical personnel (Kowalczyk 2020). Given the prior debates in Minnesota and New York, we were curious to see how this priority was implemented. The new April 2020 Massachusetts state guidelines made *no* preference for medical personnel and were a version of White’s priority point system (DPH 2020a). *The Boston*

Globe picked up on ambiguous language about a possible preference for medical personnel in this document, despite the fact that the final document did not include preference for medical personnel.

Upon seeing this article, our research team contacted Dr. Robert Truog, who was mentioned in *The Boston Globe* article as a member of the working group and was one of the scholars who had called for new rationing guidelines in Truog et al. (2020), which inspired our initial work on this topic. Truog, who is the Director of the Center for Bioethics at Harvard Medical School, convened a meeting and also invited legal scholar Govind Persad and Dr. Douglas White. Persad is a faculty member at the University of Denver Sturm College of Law, and he is the second author of Emanuel et al. (2020), the main paper that inspired our current work. White is a critical care professor and the director of the Program on Ethics and Decision Making in Critical Illness at the University of Pittsburgh Medical Center (UPMC), and he has codeveloped the multiprinciple priority point system (cf. White et al. 2009), which is the main allocation procedure proposed in the bioethics literature using the multiple ethical principle framework. After the meeting, we teamed up with these three experts to advocate for the reserve system to broader bioethics and medical communities (Sönmez et al. 2021). Our new collaborators encouraged us to use the alternative name of “categorized priority system” to emphasize the close connection between our proposed reserve system and the traditional priority system. To the best of our knowledge, a reserve system has neither been proposed nor implemented in the field for pandemic rationing of life-saving resources prior to the COVID-19 pandemic.

Around the same time, we also started collaborating with University of Pennsylvania bioethicist Harald Schmidt and Massachusetts General Hospital Clinical Care Physician Dr. Emily Rubin. Schmidt was interested in utilizing the reserve system to mitigate disparities in healthcare access (Schmidt et al. 2020). Schmidt became a major proponent of the reserve system, and, as we discuss in Section 4.2, he played a key role in the adoption of the reserve system in several states during the COVID-19 vaccine rollout, which started in December 2020. Rubin was a member of the working group for determining state rationing guidelines in Massachusetts. As we discuss in Section 4.5, we later worked with her in designing and deploying a reserve system for allocation of monoclonal antibody therapies in the state (Rubin et al. 2021).

Following our interaction with various interdisciplinary groups and public health officials during the COVID-19 pandemic, a reserve system was recommended or deployed in several jurisdictions. We outline various design considerations for different life-saving resources that we learnt and developed through these

interactions in Appendix A. In all these collaborations, the close connection between the reserve system and the traditional priority system made it possible for our team to communicate the strengths of our proposal to various stakeholders in a simple and transparent way. Many jurisdictions were caught unprepared for the challenges posed by the COVID-19 pandemic. The flexibility offered with the policy levers in reserve system—that is, the selection of the reserve categories and the allocation criteria along with the number of units in these categories—made it easier to reach compromises in several committees and task forces.

We next report various outreach activities that we conducted during COVID-19 and the policy impact of our efforts in several jurisdictions due to these activities.

4.1. Antiviral Therapy Allocation in Pennsylvania

Gilead's Remdesivir was one of the first antiviral therapies granted emergency-use authorization by the U.S. Food and Drug Administration (FDA). Following this approval in May 2020, UPMC devised a new procedure to allot its allotment of Remdesivir, obtained from the federal government. Douglas White led the effort to construct reserve categories through community engagement, and three reserve categories were established, based on whether a patient is from a hardest-hit region (based on the Area Deprivation Index), whether the patient is an essential worker (using the Pennsylvania state definition), and whether the patient is expected to die within one year (based on various clinical indicators). The system was implemented as a weighted lottery, given projections on total demand, and more than 300 patients went through the procedure. Our team provided software to implement the rationing system. White et al. (2022) provide additional details.

The UPMC system was held up as a model policy for the allocation of scarce antiviral medications. The Pennsylvania Commonwealth later recommended it statewide (Commonwealth of Pennsylvania 2020).

4.2. NASEM Framework for Equitable Vaccine Allocation

By late summer of 2020, the focus shifted from ventilator and antiviral therapy rationing to the upcoming vaccine rollout. Starting with the beginning of the pandemic, there was an intense debate about equitable vaccine allocation.

In July 2020, the U.S. CDC and National Institutes of Health commissioned the National Academies of Sciences, Engineering, and Medicine (NASEM) to formulate recommendations on the equitable allocation of COVID-19 vaccines. In September 2020, NASEM's committee of distinguished experts solicited public comments on a discussion draft of their Framework for Equitable Allocation of COVID-19 Vaccine (NASEM 2020a).

In written and oral comments, our collaborator Harald Schmidt inquired about the mechanism to prioritize members of hard-hit communities. In preparation for this contingency and in collaboration with Schmidt, weeks earlier, we circulated a white paper illustrating how easily a traditional tiered priority system can be “modified” as a reserve system to include equity as a goal though an index of social vulnerability (Pathak et al. 2020c). This precise formulation through minimalist market design was brought to the attention of the committee as a possible mechanism to embed equity in their framework. Around the same time, Persad et al. (2020) endorsed our proposed reserve system for vaccine rollout in a *JAMA* viewpoint. Following the terminology adopted in Sönmez et al. (2021), they described a reserve system as a “categorized priority system,” writing:

Dividing the initial vaccine allotment into priority access categories and using medical criteria to prioritize within each category is a promising approach. For instance, half of the initial allotment might be prioritized for frontline health workers, a quarter for people working or living in high-risk settings, and the remainder for others. Within each category, preference could be given to people with high-risk medical conditions. Such a categorized approach would be preferable to the tiered ordering previously used for influenza vaccines, because it ensures that multiple priority groups will have initial access to vaccines. (Persad et al. 2020, p. 1602)

In October 2020, NASEM published their final Framework (NASEM 2020b). Following the recommendation in Persad et al. (2020) and using the same formulation in Pathak et al. (2020c), the NASEM Framework recommended a 10% over-and-above reserve for hard-hit areas, where hard-hit is measured by the CDC's Social Vulnerability Index (SVI). Their stated justification emphasized the flexibility of a reserve system:

The committee does not propose an approach in which, within each phase, all vaccine is first given to people in high SVI areas. Rather the committee proposes that the SVI be used in two ways. First as previously noted, a reserved 10 percent portion of the total federal allocation of Covid-19 vaccine may be reserved to target areas with a high SVI (defined as the top 25 percent of the SVI distribution within the state). (NASEM 2020b, p. 3–36)

NASEM's recommendations, together with those from the CDC's Advisory Committee on Immunization Practices, were made available to states and other jurisdictions in charge of vaccine distribution.

4.3. Symposium on Vaccine Allocation and Social Justice

Despite the NASEM recommendation, by the beginning of December 2020, only Tennessee had adopted a formal reserve system for the upcoming vaccine rollout. At this

point, we co-organized a symposium directed at state policymakers and academics on ethical aspects of vaccine allocation (Ariadne Labs 2020). Our co-organizers were Ariadne Labs, a joint healthcare advocacy center of Brigham and Women's Hospital and Harvard T. H. Chan School of Public Health. We invited several state officials and also the top vaccine official of Tennessee, Dr. Michelle Fiscus, to explain the adoption process of their system.

During the symposium, we developed a better understanding of the needs of states and other jurisdictions in operationalizing a vaccine-allocation system and acquainted with different state officials, who made presentations on how their task forces were working to come up with a vaccine-allocation plan or participated as listeners.

The symposium led to two important direct policy developments. Our symposium partner, Ariadne Labs, worked with Massachusetts to develop a vaccine-allocation guideline featuring a reserve system. Our team had several meetings with California officials and introduced them to the idea of a reserve system and how to operationalize it. Subsequently, Massachusetts and California each adopted a reserve system for the vaccine rollout in December 2020 and March 2021, respectively. By May 2021, more than a dozen states and jurisdictions adopted the reserve system during various phases of the vaccine rollout (Schmidt et al. 2021). We next elaborate on the allocation frameworks in some of these jurisdictions.

4.4. Vaccine-Allocation Plans at Various Jurisdictions

In December 2020, the federal government began allocation of doses of Pfizer and Moderna vaccine to state, tribal, local, and territorial public health agencies in proportion to their population.

In the vaccine-allocation context, a reserve is often formulated as a community reserve, in which extra units are awarded to a particular location, and the authorities at the location are in charge of distribution to individuals. A popular reserve category is a hard-hit reserve defined by geography, where units are awarded above and beyond the population proportion of that geographic area. Table 3 contains examples from several states and jurisdictions.¹³

In October 2020, as mentioned earlier, Tennessee became the first state to adopt a reserve system for COVID-19 vaccine allocation (TDH 2020a). The initial plan has the following reserve categories: 5% of vaccines will be distributed equitably among all 95 counties, 10% of vaccines will be reserved by the State for use in targeted areas with high vulnerability to morbidity and mortality from the virus, and 85% will be distributed among all 95 counties based upon their populations. Reserve sizes were subsequently revised for specific

COVID-19 vaccines in response to the differences in vaccine-storage requirements (TDH 2020b).

In December 2020, Massachusetts released their vaccine-allocation plan. The plan describes that 20% of vaccines will be reserved for communities that have experienced a disproportionate COVID-19 burden and high social vulnerability. The state adopted an over-and-above implementation, allocating additional units to hard-hit areas (MA DPH 2020a, Biddinger 2020). Within a community, the allocation is based on phase prioritization.

Several other jurisdictions also embedded equity in their vaccine rollout through over-and-above reserves for their hard-hit communities, but other forms of reserve systems also found applications in the field. For example, in February 2021, Connecticut announced a 25% community-level reserve for communities with high social vulnerability, formulated as a minimum guarantee, and in January, Pennsylvania started to use a 10% minimum guarantee reserve for healthcare personnel who are not affiliated with any hospital system.

In March 2021, California announced a particularly elaborate reserve system, in which 40% of vaccines are reserved for communities in the first quartile of the Healthy Places Index, an index of socioeconomic disadvantage, and 20% are reserved for communities in each of the other quartiles. Within each quartile, 30% of the vaccines reserved for the quartile are allocated based on age eligibility, and 70% of the vaccines reserved for the quartile are allocated based on sector eligibility. It is noteworthy to emphasize that California's reserve system utilizes category-specific prioritization criteria. As shown in Table 3, other jurisdictions that adopted reserve systems utilizing category-specific prioritization criteria include Richmond, VA, and Washington, DC.

Three other plans in the table are noteworthy for showcasing the flexibility of a reserve system. First, whereas most reserve systems come in an over-and-above form, in February 2021, Connecticut announced a 25% community-level reserve for communities with high social vulnerability, formulated as a minimum guarantee. Second, Richmond City and Henrico County Health Districts in Virginia proposed several reserve categories (Richmond City 2021a, b), some with category-specific prioritization criteria. Third, Washington, DC's plan also uses the flexibility of a reserve with categories based on age, worker type, qualifying medical conditions, and certain prioritized zip codes, as shown in the table.

Because the Richmond City and Henrico County plan uses the full flexibility of our model with different category-specific priorities, we give it in detail here (although the plan partially specifies the category-specific priorities).

Table 3. Some Notable Reserve Systems Adopted During COVID-19 Pandemic

Jurisdiction	Date	Policy
Panel A. Vaccine rollout		
California	March 2021	<ul style="list-style-type: none"> • 40% of the vaccines are reserved for communities in the hardest-hit/most socially vulnerable quartile, and 20% are reserved for each remaining quartile. • Within each quartile, 70% of the vaccines are reserved based on age eligibility, and 30% are reserved for sector eligibility.
Connecticut	February 2021	<ul style="list-style-type: none"> • 10% of the vaccines are reserved in an over-and-above form for hard-hit/socially vulnerable communities. • 25% of the vaccines will be reserved in a minimum guarantee form for communities that have high social vulnerability.
Massachusetts	March 2021	<ul style="list-style-type: none"> • 20% of the vaccines are reserved in an over-and-above form for hard-hit/socially vulnerable communities.
Minnesota	December 2020	<ul style="list-style-type: none"> • 7,000 and 10,000 doses are reserved in an over-and-above form for older people and school/child-care workers, respectively. • 40% of the vaccines are reserved for communities in the hardest-hit/most socially vulnerable quartile, and 20% are reserved for each remaining quartile.
New York	February 2021	<ul style="list-style-type: none"> • 35,000 vaccines are reserved in an over-and-above form for college students. • 10% of the vaccines received by each healthcare provider will be reserved in a minimum guarantee form for vaccination of Phase 1A non-hospital-affiliated healthcare personnel.
Pennsylvania	May 2021	<ul style="list-style-type: none"> • 5% of the vaccines are distributed equitably among all counties, 10% are reserved in an over-and-above form for hard-hit/socially vulnerable counties, and the remaining 85% are distributed according to population among all counties. • The hard-hit/socially vulnerable reserve is decreased to 5%.
Tennessee	April 2021	<ul style="list-style-type: none"> • In Phase 1b allocation, 50% of the vaccines are reserved for people 65 and older, 23% are reserved for frontline essential personnel, 23% are reserved for people in age group 16–64 with comorbidities, and the remaining 4% of the vaccines are reserved for people living in congregate-care setting and Phase-1a-eligible but yet unvaccinated people. • Within each category, preregistered individuals are prioritized using category-specific criteria and are offered according to this priority order.
Richmond & Henrico, Virginia	January 2021	<ul style="list-style-type: none"> • 20% of the vaccines are reserved for people 65 and older who living in prioritized zip codes and 20% of the vaccines are reserved for people 65 and older in any zip code. • 20% of the vaccines are reserved for people in the age group 18–64 with a qualifying medical condition living in prioritized zip codes, and 20% of the vaccines are reserved for people in the age group 18–64 with a qualifying medical condition living in any zip code. • 10% of the vaccines are reserved for eligible workers in the age group 18–64 living in prioritized zip codes, and the remaining 10% of the vaccines are reserved for eligible workers who are in the age group 18–64 living in any zip code.
Washington, DC	December 2020	<ul style="list-style-type: none"> • 20% of the vaccines are reserved for people 65 and older who living in prioritized zip codes and 20% of the vaccines are reserved for people 65 and older in any zip code. • 20% of the vaccines are reserved for people in the age group 18–64 with a qualifying medical condition living in prioritized zip codes, and 20% of the vaccines are reserved for people in the age group 18–64 with a qualifying medical condition living in any zip code. • 10% of the vaccines are reserved for eligible workers in the age group 18–64 living in prioritized zip codes, and the remaining 10% of the vaccines are reserved for eligible workers who are in the age group 18–64 living in any zip code.
Panel B. Allocation of therapeutic agents		
Massachusetts	March 2021	<ul style="list-style-type: none"> • Massachusetts Department of Public Health announces a reserve system for monoclonal antibodies with an over-and-above reserve for patients from hard-hit/socially vulnerable communities. Within each category, beneficiary patients are prioritized using an even lottery.
UPMC Medical System, Pennsylvania	November 2020	<ul style="list-style-type: none"> • A weighted lottery is used to allocate antiviral drug Remdesivir in a reserve system with essential personnel, hard-hit area, and non-end-of-life patient categories, in which, within each category, beneficiary patients are prioritized using an even lottery based on approximate total demand projections.
UPMC Medical System, Pennsylvania	May 2020	<ul style="list-style-type: none"> • A weighted lottery is used to allocate antiviral drug Remdesivir in a reserve system with essential personnel, hard-hit area, and non-end-of-life patient categories, in which, within each category, beneficiary patients are prioritized using an even lottery based on approximate total demand projections.

Notes. Compiled by authors from public documents. Source information available at: <https://www.covid19reservesystem.org/>.

4.4.1. Richmond and Henrico Reserve System for Phase 1b COVID-19 Vaccine Rollout. Categories, their share of COVID-19 vaccines, and factors that affect category-specific priorities were specified by the health districts as follows:

1. Phase 1a and Congregate Care (4% of units)
2. Adults age 65+ (50% of units)
 - Age (older residents have higher priority)
 - Race and ethnicity (Black, Hispanic/Latinx, and American Indian or Alaska Native residents have higher priority)
 - Burden of disease in the area where a person lives
 - Social Vulnerability Index Score of the area where a person lives
3. Frontline Essential Workers (23% of units)
 - Age (older residents have higher priority)
 - Race and ethnicity (Black, Hispanic/Latinx, and American Indian or Alaska Native residents have higher priority)
 - Burden of disease in the area where a person lives
 - Social Vulnerability Index Score of the area where a person lives
4. People ages 16–64 with Comorbidities (23% of units)
 - Age (older residents have higher priority)
 - Race and ethnicity (Black, Hispanic/Latinx, and American Indian or Alaska Native residents have higher priority)
 - Socioeconomic status (residents who are under and uninsured have higher priority).

4.5. Monoclonal Antibody Allocation in Massachusetts

In November 2020, the FDA granted an emergency use authorization for monoclonal antibody treatments. The Massachusetts Department of Public Health (MA DPH) assembled a committee of doctors, health center representatives, and ethicists to devise a plan for equitable distribution of doses delivered to Massachusetts. Emily Rubin, a member of the committee who had been following our progress since May 2020, inquired about whether our proposed reserve system can be deployed and operationalized. Following our interaction, the December guideline recommended the use of a reserve system for within-hospital allocation (MA DPH 2020b). Patients are placed into two tiers: (1) patients who are age ≥ 65 and patients aged 18 and older with body mass index ≥ 35 , and (2) all other eligible patients.

At each time interval (defined either as once or twice a day), there are two serial allocations for treatment spots. The first allocation of 80% of available doses, known as the “open allocation,” will be available to all patients (MA DPH 2020b). The second allocation of 20% of available doses are available to patients who live in a census tract with SVI $> 50\%$ or in a city or town

with a seven-day average COVID-19 incidence rate in the top quartile, as reported in the most recent MA DPH weekly COVID report (“hardest-hit”). For the open 80%, allocation takes place in order of tier, with lottery tie-breaking within tier. For the 20% reserve, allocation is first for patients from hardest-hit areas. For these patients, allocation takes place in order of tier, with lottery tie-breaking within tier. If units remain in the 20% reserve, then they are allocated to patients who are eligible for open units in order of tier, with lottery tie-breaking. Rubin et al. (2021) report the experience of the Mass General Brigham health system with the implementation of the resulting reserve system for allocation of monoclonal antibody therapies.

4.6. Postpandemic Policy Impact: Crisis Care Guidance in Oregon

After the pandemic, some states formed advisory committees to evaluate and reform their crisis standards of care, which will be adopted in a public health emergency or an overwhelming disaster. Oregon is one of these states.

In May 2023, the Oregon Health Authority (OHA) requested public input for a preliminary framework prepared by the Oregon Resource Allocation Advisory Committee (ORAAC 2023). In this document, the committee presented six criteria under consideration for crisis care triage and discussed how they might be used in a stand-alone fashion or combined as part of a multicriteria approach.

As their recommended methodology to implement multiple criteria, the committee recommended the reserve system:

In a major health emergency situation, ease of implementation is a feature that needs to be taken seriously. Models such as those described below can all be implemented readily via a methodological approach known as Categorized Priority System (sometimes also called a Reserve System). During the Covid-19 pandemic, systems that combined factors such as survivability, level of disadvantage and essential worker status were successfully developed for purposes including allocating vaccines, tests and treatments. Custom-made, free-of-charge software has been developed to facilitate implementation. (ORAAC 2023, p. 20)

As can be seen from this recommendation, two of the factors that contributed to the committee’s endorsement of reserve system are its practicality and the free-of-charge software that we made publicly available for its implementation.¹⁴

4.7. Other Endorsements for the Reserve System in Bioethics, Critical Care Medicine, and Law Literatures

Following our efforts in recommending or designing reserve systems for different settings, many members of

the medical community and civil rights advocates have also been receptive to the adoption of this methodology.

Some retrospective studies assessed the effectiveness of field applications of certain allocation methods in overcoming pre-existing racial inequities in access to healthcare. Wu et al. (2022) show that during the COVID-19 pandemic, at a large healthcare system, Black and Hispanic patients were less likely to receive monoclonal antibody treatment. This was a discrepancy that could not be explained through the refusal of the patients to take the treatment. They recommend exploring the use of a reserve system in the future to correct such disparities. Using a large data set from 208 U.S. hospitals in 2014 and 2015, Wasserman et al. (2020) show that SOFA scores were associated with an overestimated mortality among Black patients compared with White patients, and this was associated with a structural disadvantage for Black patients in crisis standards-of-care guidelines. They also recommend the adoption of reserve systems to overcome the problems associated with biased SOFA-score-based prioritization using social disadvantage-based reserve categories.

The editorial by Makhoul and Drolet (2021), in response to Rubin et al. (2021), characterizes a reserve system as a method that is backed by growing field evidence and is a pragmatic framework for equitable medical resource allocation, citing the success of its use during the COVID-19 pandemic. They also note that

As reserve systems become more prevalent, it is important to acknowledge and understand the psychological effects on participants. Not only do reserve systems enable policymakers to allocate resources equitably, but they also signal to participants that expert judgment has been used to design a system for maximal societal benefit. Participants eligible for prioritized categories (eg, patients from high-SVI zip codes) may feel more adequately safeguarded. (Makhoul and Drolet 2021, p. 2005)

Finally, emphasizing the policy roles of its “minimum guarantee” and “over-and-above” variants, Hellman and Nicholson (2021) advocate for reserve system as a “middle road” between saving the most lives and equity for citizens with disability.

The reserve can be operationalized in either of two ways. It can provide a ‘boost’ for a group, giving it extra resources, as might be justified in the case for health care workers. Or it can function as a “protective measure” to ensure that members of a group are not left out altogether. The order in which the reserve category is processed determines whether it functions as a boost or a protective measure. If the reserve category is processed first, it functions as a protective measure; if it is process[ed] second, it functions as a boost. In our view and consistent with the ADA’s [The Americans with Disabilities Act’s] emphasis on equal rather than

superior opportunity, and in their view also, the reserve for disability should function not as a boost but as a protective measure.

While a discussion of mathematics of the reserve system is beyond the scope of this paper, what is important to understand is that this model provides an algorithm that enables policymakers to allocate resources according to multiple principles—one that can be fashioned in advance and applied much like the algorithms in SOFA-type scoring. For our purposes, this method provides a middle road between the goals of saving the most lives and ensuring that benefits and harms are distributed fairly. (Hellman and Nicholson 2021, pp. 1278–1279)

5. Conclusion

In two different ways, this paper is about compromise between considerations that pull in different directions.¹⁵ The first way is in the context of the specific market-design application formulated in this paper. How can various competing ethical values be balanced to allocate scarce and vital life-saving resources in a disaster setting? The second way is methodological. How can the role of a design economist in advancing scholarly research in a given application be balanced with the role of a design economist in developing practical and realistic tools that have a shot at influencing policy? We answer both questions in this paper together by proposing a reserve system as a flexible allocation rule to balance various ethical values and as a procedure that only minimally differs from the traditional priority system.

As a compromise between multiple ethical principles in a pandemic medical resource allocation setting, we formulate and advocate for a reserve system, where each ethical principle is represented with a reserve category. The underlying ethical principle at any category c gives rise to a category-specific priority order π_c of individuals for category c , which is then used to allocate the units reserved for category c . When there is only one ethical value, the reserve system reduces to the traditional priority system, which had been prevalent in the field. During the COVID-19 pandemic, this close connection gave our team a distinct advantage to communicate our ideas with experts in medical ethics and officials in healthcare. By correctly identifying some of the challenges they face due to limitations of the priority system and addressing those with minimal interference, we found that many in these communities are receptive to our ideas. This broader strategy of institution redesign, minimalist market design (Sönmez 2023), has recently been effective in other settings as well, such as the 2021 reform of the U.S. Army’s branching system (Greenberg et al. 2023). There is, however, one important aspect of the current application that differentiates

it from the Army application, and it is this difference that motivated our formal analysis in Section 3.

In the case of the 2021 reform of the U.S. Army's branching system, Greenberg et al. (2023) show that there is a unique mechanism that satisfies the Army's design objectives. Thus, the paradigm of minimalist market design has a unique prescription for the Army's branching application. In contrast, in our current application, our proposed reserve system refers to a broad class of allocation rules. Although many elements of this class can be used to overcome the failures of the traditional priority system, different elements of the class can have disparate distributional implications. This richness can be both a blessing, by giving policymakers additional flexibility in the design, and a curse, by causing unintended consequences due to the adoption of a wrong element of the broad class. Indeed, in settings where reserve categories correlate due to a baseline priority order, such unintended consequences and resulting policy challenges have been prevalent in the field (cf. Dur et al. 2018, Sönmez and Yenmez 2022, and Pathak et al. 2023a, b). Our formal analysis in Section 3 provides policymakers and other stakeholders with the analytical tools that enable them to choose a reserve system that truly reflects their policy objectives.

Our paper has shown that the field of market design has matured to a stage where it can provide real-time guidance for society during a crisis. It has also demonstrated that minimalist market design is an effective paradigm to communicate the ideas of design economists with policymakers and experts in other fields when they are most useful in a timely manner.

Acknowledgments

The authors are grateful for input from several experts in bioethics, emergency healthcare, and webinar participants at Johns Hopkins University, London Business School, Stony Brook International Conference on Game Theory, Australasian Microeconomic Theory Seminars, the HELP! (HEaLth and Pandemics) Econ Working Group Seminar Series, Penntheon Webinar Series, Stockholm School of Economics Corona Economic Research Network Webinar Series, UK Virtual Seminars in Economic Theory, NBER Workshop on Market Design, Harvard-MIT Theory Workshop, Virtual Market Design Seminar Series, University of Tokyo Theory Seminar, CUHK-HKU-HKUST Theory Seminar, North Carolina State University Theory Seminar, University of Birmingham Seminar, Brazilian School of Economics Seminar, University of Southern California Theory Seminar, 54th Memorial Gilbert Lecture at University of Rochester, and the University of Alabama Culverhouse Seminar. Nikhil Agarwal, Alex Rees-Jones, Robert Truog, Iván Werning, and Douglas White provided helpful comments. Feedback from an anonymous Associate Editor, three anonymous referees, David Delacrétaz, Fuhito Kojima, Govind Persad, and Alex Teytelboym was particularly valuable. This version supersedes: "Leaving No Ethical Value Behind: Triage Protocol Design for Pandemic Rationing," NBER Working Paper 26951, April 2020.

Appendix A. Resource-Dependent Design Considerations

In this appendix, we offer some thoughts about how to implement a reserve system might depend on the resource that is rationed. In Section 3, we model pandemic rationing as a one-shot static reserve system. Several vital resources, however, must be rationed during a pandemic as patients in need present. Hence, it is important to formulate how our static model can be operationalized in an application where patient arrival and allocation are both dynamic. The adequate formulation depends on the specific characteristics of the rationed resource. Most notably, answers to the following two questions factor in the implementation details:

1. Is the resource fully consumed upon allocation, or is it durable, utilized over a period, and can it be re-allocated?
2. Is there immediate urgency for allocation?

Because most guidelines are on rationing of vaccines, ventilators, ICU beds, and antiviral drugs or treatments, we focus our discussion on these four cases.

A.1. Vaccine Allocation

A unit of a vaccine is consumed upon allocation, and reallocation of the unit is not possible. Moreover, there is no immediate urgency to allocate a vaccine. Hence, a large number of units can be allocated simultaneously. Therefore, vaccine allocation is an application of our model where our proposed reserve system can be implemented on a static basis as vaccines become available.

This is, however, not the only reasonable way a reserve system can be operationalized for vaccine allocation. In the United States, there is a tradition of distributing influenza vaccines at local pharmacies or healthcare providers on a first-come-first-served basis. This practice can be interpreted as a single-category special case of a reserve system where the priorities are based on the time of arrival. This practice can easily be extended to any sequential reserve matching system with multiple categories where the baseline priorities are determined by the time of arrival. Under this dynamic implementation of a reserve system, as a patient arrives to a healthcare provider, she is allocated a vaccine as long as there is availability in a category for which she is a beneficiary. If there are multiple such categories, the patient is assigned a unit from the category that has the highest precedence under the sequential reserve matching. Although many have criticized first-come-first-served allocation because of biases it induces based on access to healthcare (e.g., Kinlaw and Levine 2007), reserve categories can be designed to mitigate these biases, even if priority is first-come-first-served within each category. For example, there can be a reserve category for patients from rural areas. There is an important precedent for using a reserve system in this dynamic form. Between 2005 and 2008, H-1B immigration visas in the United States were allocated through a reserve system with general and advanced-degree reserve categories, where priority for each category was based on the application arrival time (Pathak et al. 2023a).

A.2. Ventilator/ICU Bed Allocation

Because the relevant characteristics of ventilators and ICU beds are identical in relation to our model, the implementation of reserve systems for these resources will be similar.

Therefore, we present the details of their implementation together. For simplicity in this subsection, we refer the resource in short supply as a ventilator.

A ventilator is durable and can be reassigned once its use by its former occupant is completed. Moreover, there is always urgency in allocation of this vital resource. These two features make direct static implementation of a reserve system impractical; implementation always has to be dynamic. One important observation on ventilator allocation is key to formulate the implementation: Because a ventilator is durable and assigned to a patient for a period, it can be interpreted as a good that is allocated at each instant. During the course of using a ventilator, a patient's clinical situation and her priority for one or more categories may change. Therefore, with the arrival of each new patient, the allocation of all units has to be reevaluated. As such, the following additional ethical and legal consideration has an important bearing on the design of a reserve system:

3. Can a patient be removed from a ventilator once she is assigned?

There is widespread debate on this issue in the United States. Piscitello et al. (2020) describe 25 states with protocols that discuss the ethical basis of re-assigning ventilators. As of June 2020, the majority of guidelines support ventilator withdrawal. If a ventilator can be withdrawn, the design is simpler (and effectively identical to static implementation with each new arrival). Although patient data need to be updated through the duration of ventilator use, no fundamental adjustment is needed for the design of the main parameters of the reserve system. Of course, in this scenario, it is possible that the category of the unit occupied by the patient may change over time. For example, a patient may initially be assigned a unit from the general category, even though she has sufficiently high priority for multiple categories, such as the general category and essential personnel category. At a later time, she may only have high enough priority for the latter category. In this case, the patient will continue using the ventilator, although for accounting purposes, she will start consuming a unit from a different category.

If a ventilator cannot be withdrawn, a reserve system can still be applied with a grandfathering structure to reflect the property rights of patients who are already assigned. In this case, the priority system has to give highest priority to occupants of the units from any category for as long as they can hold these units, despite a change in their clinical situation or arrival of patients who otherwise would have higher priority for these units.

A.3 Antiviral Drugs or Treatments

For antiviral drugs and treatments, the vital resource is consumed upon allocation (as in vaccine allocation), but there is typically urgency, and allocation decisions will need to be made as patients arrive (as in ventilator allocation). One possible dynamic implementation is based on first-come-first-served arrival within reserve categories. This would be akin to the dynamic allocation scenario for ventilators with a baseline priority structure that depends on patient arrival time, as described in Section A.1. Alternatively, drug assignment can be batched within prespecified time windows. Drugs can then be assigned based on expectations of the number of patients in each category over this time window. Because

drugs would be administered by a clinician, the relationship between a reserve system and cutoffs can be particularly valuable. A clinician can simply assign the treatment to a patient if she clears the cutoff for any reserve for which she is eligible. In fact, after the first version of our paper was circulated, our team assisted with the design of the system used at the University of Pittsburgh Medical Center to allocate the antiviral drug Remdesivir in May 2020 with this implementation (White et al. 2022). The system had special provisions for hardest hit and essential personnel and used lotteries for prioritization. In November 2020, we also assisted the Massachusetts working group on the equitable allocation of COVID-19 therapies to help design a reserve system for allocation of monoclonal antibodies (Rubin et al. 2021).

Appendix B. Smart Reserve Matching

Although virtually all practical applications of reserve systems are implemented through sequential reserve matching, this class of mechanisms may suffer from an important shortcoming: They may lead to Pareto inefficient outcomes, due to myopic processing of reserves. The following example illustrates both how this may happen, and also motivates a possible refinement based on smart processing of reserves.

Example B.1. Consider a hard reserve system induced by baseline priority order π . There are two patients $I = \{i_1, i_2\}$, who are priority ordered as

$$i_1 \pi i_2$$

under the baseline priority order π . There are two categories: an unreserved category u with an all-inclusive beneficiary set of $I = \{i_1, i_2\}$ and a preferential treatment category c with a beneficiary set $I_c = \{i_1\}$ of a single preferential treatment patient. Both categories have a capacity of one unit each (i.e., $r_c = r_u = 1$). Because the reserves are hard, the resulting category-specific priority orders are given as follows:

$$i_1 \pi_u i_2 \pi_u \emptyset \quad \text{and} \quad i_1 \pi_c \emptyset \pi_c i_2.$$

Consider the sequential reserve matching φ_\triangleright induced by the order of precedence \triangleright , where

$$u \triangleright c.$$

Under matching φ_\triangleright , first, patient i_1 is matched with the unreserved category u , and, subsequently, the category- c unit is left idle because no remaining patient is eligible for this preferential treatment category. Therefore,

$$\varphi_\triangleright = \begin{pmatrix} i_1 & i_2 \\ u & \emptyset \end{pmatrix},$$

resulting in the set of matched patients $\varphi_\triangleright(I) = \{i_1\}$.

Next, consider the sequential reserve matching $\varphi_{\triangleright'}$ induced by the order of precedence \triangleright' , where

$$c \triangleright' u.$$

Under matching $\varphi_{\triangleright'}$, first, patient i_1 is matched with the preferential treatment category c , and, subsequently, patient i_2 is matched with the unreserved category u . Therefore,

$$\varphi_{\triangleright'} = \begin{pmatrix} i_1 & i_2 \\ c & u \end{pmatrix},$$

resulting in the set of matched patients $\varphi_{\triangleright}(I) = \{i_1, i_2\}$. Because $\varphi_{\triangleright}(I) \subsetneq \varphi_{\triangleright^*}(I)$, matching φ_{\triangleright} is Pareto-dominated by matching $\varphi_{\triangleright^*}$.

Observe that Example B.1 also illustrates that the cause of Pareto inefficiency is the myopic allocation of categories under sequential reserve matchings. Under matching φ_{\triangleright} , the more flexible unreserved unit is allocated to patient i_1 , who is the only beneficiary of category c . This results in a suboptimal utilization of reserves, which can be avoided with the concept of “smart reserve matching” that we introduce below.

To this end, we first introduce a new axiom, which, together with nonwastefulness, imply Pareto efficiency.

Definition B.1. A matching $\mu \in \mathcal{M}$ is *maximal in beneficiary assignment* if

$$\mu \in \arg \max_{\nu \in \mathcal{M}} \left| \bigcup_{c \in \mathcal{C} \setminus \{u\}} (\nu^{-1}(c) \cap I_c) \right|.$$

This axiom simply requires that the reserves should be maximally assigned to target beneficiaries to the extent that it is feasible. It precludes the myopic assignment of categories to patients because the desirability of a matching depends on the structure of the matching as a whole, rather than the individual assignments it prescribes for each category.

It is worth noting that the inefficiency observed in Example B.1 is specific to the case of hard reserves and cannot happen for soft reserves, as in our main application of pandemic rationing. Nonetheless, maximality in beneficiary assignment is a desirable axiom in general, including for soft reserves, because suboptimal utilization of reserves may receive heightened scrutiny. For example, consider a scenario with two preferential treatment categories, essential personnel and disadvantaged, each with one unit of reserve. Suppose patient A is both essential personnel and disadvantaged, patient B is disadvantaged, and there are several other patients who are neither. One possible way to use these reserves is to assign patient A to the disadvantaged reserve, leaving no other preferential treatment patients available for the essential personnel reserve. In this case, the essential personnel reserve would be opened up to patients who are neither essential personnel nor disadvantaged. This could, in turn, mean that only one of the reserves is assigned to members of the target beneficiary groups. This outcome could be seen as problematic because an alternative, which assigns patient A to the essential personnel reserve (instead of the disadvantaged reserve) and patient B to the disadvantaged reserve, accommodates both reserves. By imposing maximality in beneficiary assignment, we avoid this shortcoming through a “smart” utilization of reserves.

Building on this insight, we next present a polynomial-time algorithm that generates smart cutoff matchings:

Smart Reserve Matching Algorithm

Fix a parameter $n \in \{0, 1, \dots, r_u\}$ that represents the number of unreserved units to be processed in the beginning of the algorithm.¹⁶ The remaining unreserved units are to be processed at the end of the algorithm.

Fix a baseline priority order π , and for the ease of description, relabel patients so that

$$i_1 \pi i_2 \pi \dots \pi i_{|I|}.$$

Step 0. Find a matching that is maximal and complies with eligibility requirements by temporarily deeming that a patient $i \in I$ is eligible for a category $c \in \mathcal{C} \setminus \{u\}$ if and only if $i \in I_c$, and no patient is eligible for unreserved category u .¹⁷ The solution finds the maximum number of patients who can be matched with a preferential treatment category that they are beneficiaries of. Denote the number of patients matched by this matching as n_b .

Step 1. Let $J_0^u = \emptyset$, $J_0 = \emptyset$. Fix parameters $\kappa \gg \varepsilon > 0$ such that $\varepsilon < 1$ and $\kappa > |I|$.

For $k = 1, \dots, |I|$, we repeat the following substep given J_{k-1}^u, J_{k-1} :

Step 1.(k).

i. If $|J_{k-1}^u| < n$ continue with (i.A), and otherwise continue with (ii).

A. Temporarily deem all patients in $J_{k-1}^u \cup \{i_k\}$ eligible only for category u and all other patients eligible only for the categories in $\mathcal{C} \setminus \{u\}$ that they are beneficiaries of.

B. For every pair $(i, x) \in I \times \mathcal{C} \cup \{\emptyset\}$, define a weight $W_{i,x} \in \mathbb{R}$ as follows:

- If $x \in \mathcal{C}$ and i is temporarily eligible for x as explained in (i.A),
 - if $i \in J_{k-1}^u \cup J_{k-1}$, then define $W_{i,x} := \kappa$,
 - otherwise, define $W_{i,x} := \varepsilon$.
- If $x \in \mathcal{C}$ and i is not temporarily eligible for x as explained in (i.A), define $W_{i,x} := -\varepsilon$.
- If $x = \emptyset$, define $W_{i,x} := 0$.

C. Solve the following *assignment problem* to find a matching¹⁸

$$\sigma \in \arg \max_{\mu \in \mathcal{M}} \sum_{i \in I} W_{i,\mu(i)}.$$

D. If $|\sigma(I)| = n_b + |J_{k-1}^u| + 1$, then define

$$J_k^u := J_{k-1}^u \cup \{i_k\} \quad \text{and} \quad J_k := J_{k-1},$$

and go to Step 1.(k+1) if $k < |I|$ and Step 2 if $k = |I|$.

E. Otherwise, go to (ii).

ii. Repeat (i) with the exception that i_k is temporarily deemed eligible only for the categories in $\mathcal{C} \setminus \{u\}$ that she is a beneficiary of in Part (ii.A). Parts (ii.B) and (ii.C) are the same as Parts (i.B) and (i.C), respectively, with the exception that weights are constructed with respect to the eligibility construction in (ii.A). Parts (ii.D) and (ii.E) are as follows:

D. If $|\sigma(I)| = n_b + |J_{k-1}^u|$ and $\sigma(i) \neq \emptyset$ for all $i \in J_{k-1} \cup \{i_k\}$, then define

$$J_k^u := J_{k-1}^u \quad \text{and} \quad J_k := J_{k-1} \cup \{i_k\},$$

and go to Step 1.(k+1) if $k < |I|$ and Step 2 if $k = |I|$.

E. Otherwise,

$$J_k^u := J_{k-1}^u \quad \text{and} \quad J_k := J_{k-1},$$

and go to Step 1.(k+1) if $k < |I|$ and Step 2 if $k = |I|$.

Step 2.

- (a) Find a matching σ as follows:
 - i. Temporarily deem all patients in $J_{|I|}^u$ eligible only for category u , all patients in $J_{|I|}$ eligible only for the categories in $\mathcal{C} \setminus \{u\}$ that they are beneficiaries of, and all other patients ineligible for all categories.
 - ii. Find a maximal matching σ among all matchings that comply with the temporary eligibility requirements defined in (i).¹⁹
- (b) Modify σ as follows:
 - One at a time, assign the remaining units unmatched in σ to the remaining highest-priority patient in $I \setminus (J_{|I|}^u \cup J_{|I|})$ who is eligible for the category of the assigned unit in the real problem in the following order:
 - i. the remaining units of the preferential treatment categories in $\mathcal{C} \setminus \{u\}$ in an arbitrary order, and
 - ii. the remaining units of the unreserved category u .

Every matching σ constructed in this manner is referred to as a *smart reserve matching* induced by assigning n unreserved units subsequently at the beginning of the algorithm. Let \mathcal{M}_S^n be the set of all reserve matchings for a given n .

The idea behind the **Smart Reserve Matching Algorithm** can be summarized as follows:

We sequentially process patients with respect to their priority under the baseline priority order for each $k = 1, 2, \dots, |I|$. Step 1.(k).i is invoked if some of the initial n unreserved category units are still available and not all of them are already committed to a subset J_{k-1}^u of patients among i_1, \dots, i_{k-1} .

When Step 1.(k).i is invoked, we commit to giving an unreserved unit to i_k if this assignment will not decrease the maximum number of patients who can make use of the preferential treatment categories that they are beneficiaries of, denoted as n_b in the algorithm, in addition to the subset J_{k-1} of patients already committed to these units among i_1, \dots, i_{k-1} . Otherwise, i_k gets a unit from a preferential treatment category in Step 1.(k).ii.

In case Step 1.(k).i is skipped, and we directly continue with Step 1.(k).ii: In this case, we commit to giving a preferential treatment category unit to i_k from a category she is beneficiary of if this assignment, in addition to matching all patients in J_{k-1} , does not decrease the maximum number of patients who can make use of the preferential treatment category that they are beneficiaries of, n_b . Otherwise, we skip patient i_k for now.

We invoke Step 2 after we commit n unreserved units to patients in $J_{|I|}^u$ and the maximum number of preferential treatment category units, n_b , to patients in $J_{|I|}$. In this step, one at a time, we give one remaining unit from an eligible category to the remaining patients, according to the baseline priority order.

We have the following technical result about the sets of patients matched under smart reserve matchings:

Lemma B.1. Consider either a soft reserve system or a hard reserve system induced by a baseline priority order π . For any $n \in \{0, 1, \dots, r_u\}$ and any two smart reserve matchings $\sigma, \nu \in \mathcal{M}_S^n$,

$$\sigma^{-1}(u) = \nu^{-1}(u), \text{ and } \cup_{c \in \mathcal{C} \setminus \{u\}} (\sigma^{-1}(c) \cap I_c) = \cup_{c \in \mathcal{C} \setminus \{u\}} (\nu^{-1}(c) \cap I_c),$$

and moreover,

$$\sigma(I) = \nu(I).$$

Lemma B.1 states that for a given baseline priority order π , every smart reserve matching for a given n matches the same set of patients with the unreserved category u , the same set of patients with preferential treatment categories in $\mathcal{C} \setminus \{u\}$ that they are beneficiaries of, and the same set of patients overall.

In a soft reserve system or a hard reserve system, for a given n , we denote the set patients matched in every smart reserve matching with n unreserved units processed first as I_S^n . By Lemma B.1, for any $\sigma \in \mathcal{M}_S^n$,

$$I_S^n = \sigma(I).$$

Our next result on smart reserve matchings is as follows:

Proposition B.1. Consider either a soft reserve system or a hard reserve system induced by a baseline priority order π . For any $n \in \{0, 1, \dots, r_u\}$, any smart reserve matching in \mathcal{M}_S^n complies with eligibility requirements, is nonwasteful, respects priorities, and is maximal in beneficiary assignment.

The choice of parameter n is not without a consequence. In particular, matchings produced by the algorithm with the lowest parameter, $n=0$, and the highest parameter, $n=r_u$, both have distinctive distributional consequences with potentially important policy implications.

Theorem B.1. Consider either a soft reserve system or a hard reserve system induced by a baseline priority order π . Let $\sigma_{r_u} \in \mathcal{M}_S^{r_u}$ be a smart reserve matching when all unreserved units are assigned first (i.e., $n=r_u$) and $\sigma_0 \in \mathcal{M}_S^0$ be a smart reserve matching when all unreserved units are assigned last (i.e., $n=0$). Then, for any cutoff equilibrium matching $\mu \in \mathcal{M}$ that is maximal in beneficiary assignment,

$$\bar{f}_u^{\sigma_{r_u}} \underline{\pi} \bar{f}_u^{\mu} \underline{\pi} \bar{f}_u^{\sigma_0}.$$

Theorem B.1 states that among all maximum equilibrium cutoff vectors that support maximal matchings in beneficiary assignment, the selectivity of the unreserved category is

- The most competitive for smart reserve matchings with $n=r_u$ —that is, when all unreserved category units are assigned in the beginning of the algorithm, and
- The least competitive for smart reserve matchings with $n=0$ —that is, when all unreserved category units are assigned at the end of the algorithm.

We believe Theorem B.1 can have important role in policy. The most widespread real-life applications of sequential reserve systems are in one of the two forms “over-and-above” or “minimum guarantee.” Theorem B.1 shows that both forms have their counterparts in smart reserve matching systems. Adopting a smart reserve matching, instead of a sequential reserve matching, therefore need not interfere with existing policies.

On a technical note, Theorem B.1 is silent about the qualifications of maximum equilibrium cutoffs of preferential treatment categories supporting smart reserve matchings. While keeping the maximality in beneficiary assignment, the patients who are matched with preferential treatment categories are uniquely

determined under the two extreme smart reserve matchings (and for any given n), respectively. However, there can be multiple ways of assigning these patients to different preferential treatment categories. Thus, the maximum equilibrium cutoffs of preferential treatment categories are not uniquely determined for these matchings.

Appendix C. Proofs

C.1. Proofs of the Results in Section 3.4

Proof of Theorem 1

Part 1. Suppose matching $\mu \in \mathcal{M}$ complies with eligibility requirements, is nonwasteful, and respects priorities. We construct a cutoff vector $f \in \mathcal{F}$ as follows: For each category $c \in \mathcal{C}$, define

$$f_c = \begin{cases} \min_{\pi_c} \mu^{-1}(c) & \text{if } |\mu^{-1}(c)| = r_c, \\ \emptyset & \text{otherwise.} \end{cases}$$

Fix a category $c \in \mathcal{C}$. If $|\mu^{-1}(c)| = r_c$, then $f_c \in \mu^{-1}(c)$ by construction. Because μ complies with eligibility requirements, then $f_c \neq \emptyset$. On the other hand, if $|\mu^{-1}(c)| < r_c$, then $f_c = \emptyset$. Therefore, in either case $f_c \neq \emptyset$. We showed that $f \in \mathcal{F}$ —that is, it is a well-defined cutoff vector.

Next, we show that (f, μ) is a cutoff equilibrium. Condition (2) in cutoff equilibrium definition is immediately satisfied because if, for any $c \in \mathcal{C}$, $|\mu^{-1}(c)| < r_c$, then $f_c = \emptyset$ by construction.

We next show that Condition (1) in cutoff equilibrium definition is also satisfied in two parts. Let $i \in I$.

a. We show that $\mu(i) \in \mathcal{B}_i(f) \cup \{\emptyset\}$. If $\mu(i) = \emptyset$, then we are done. Therefore, suppose $\mu(i) = c$ for some $c \in \mathcal{C}$. Two cases are possible:

- If $|\mu^{-1}(c)| = r_c$, then $f_c = \min_{\pi_c} \mu^{-1}(c)$, and, hence, $i \underline{\pi}_c f_c$. Thus, $c \in \mathcal{B}_i(f)$.

- If $|\mu^{-1}(c)| < r_c$, then $f_c = \emptyset$ by construction. Because μ complies with eligibility requirements, $i \pi_c f_c = \emptyset$. Thus, $\mu(i) \in \mathcal{B}_i(f)$.

b. We show that $\mathcal{B}_i(f) \neq \emptyset \Rightarrow \mu(i) \in \mathcal{B}_i(f)$. Suppose $\mathcal{B}_i(f) \neq \emptyset$; but to the contrary of the claim, suppose that $\mu(i) \notin \mathcal{B}_i(f)$. By Condition (1(a)) in the definition of a cutoff equilibrium, $\mu(i) = \emptyset$. Let $c \in \mathcal{B}_i(f)$. Because μ respects priorities, then for every $j \in \mu^{-1}(c)$, we have $j \pi_c i$. If $|\mu^{-1}(c)| = r_c$, then by construction, $f_c \in \mu^{-1}(c)$, and, hence, $f_c \pi_c i$, contradicting $c \in \mathcal{B}_i(f)$. We conclude that $|\mu^{-1}(c)| < r_c$. Then, by construction, $f_c = \emptyset$. Because $c \in \mathcal{B}_i(f)$, $i \pi_c f_c = \emptyset$. These two statements together with $\mu(i) = \emptyset$ contradict nonwastefulness of μ . Thus, $\mu(i) \in \mathcal{B}_i(f)$.

Hence, we showed that (f, μ) is a cutoff equilibrium.

Part 2. Conversely, suppose pair $(f, \mu) \in \mathcal{F} \times \mathcal{M}$ is a cutoff equilibrium. We will show that matching μ complies with eligibility requirements, is nonwastefulness, and respects priorities.

Compliance with eligibility requirements: Consider a patient $i \in I$. Because by Condition (1(a)) of cutoff equilibrium definition $\mu(i) \neq \emptyset$ implies $\mu(i) \in \mathcal{B}_i(f)$, we have $i \underline{\pi}_c f_c$. Because the cutoff satisfies $f_c \underline{\pi}_c \emptyset$, by transitivity of π_c , $i \pi_c \emptyset$. Therefore, μ complies with eligibility requirements.

Nonwastefulness: Let $i \in I$ be such that $\mu(i) = \emptyset$ and $i \pi_c \emptyset$ for some $c \in \mathcal{C}$. We show that $|\mu^{-1}(c)| = r_c$. Then, by Condition (1(a)) of the definition of a cutoff equilibrium for (f, μ) , we have

$B_i(f) = \emptyset$. In particular, $c \notin \mathcal{B}_i(f)$. Then, $f_c \pi_c i$, implying that $f_c \pi_c \emptyset$ and, hence, $|\mu^{-1}(c)| = r_c$. Thus, μ is nonwasteful.

Respect of Priorities: Let patient $i \in I$ be such that for some category $c \in \mathcal{C}$, $\mu(i) = c$, whereas for some patient $j \in I$, $\mu(j) = \emptyset$. We show that $i \pi_c j$, which will conclude that matching μ respects priorities. By Condition (1(b)) of cutoff equilibrium definition, $B_j(f) = \emptyset$. In particular, $f_c \pi_c j$. Because $\mu(i) = c$, by Condition (1(a)) of cutoff equilibrium definition, $c \in \mathcal{B}_i(f)$, implying that $i \underline{\pi}_c f_c$. By transitivity of π_c , $i \pi_c j$. \square

Proof of Lemma 1. We prove the lemma in three claims. Let matching $\mu \in \mathcal{M}$ comply with eligibility requirement, be nonwasteful, and respect priorities.

Claim 1. \bar{f}^μ is the maximum equilibrium cutoff vector supporting μ —that is, (\bar{f}^μ, μ) is a cutoff equilibrium and for every cutoff equilibrium (f, μ) , $\bar{f}_c^\mu \leq f_c$ for every $c \in \mathcal{C}$.

Proof. We prove the claim in two parts.

Part 1. We show that (\bar{f}^μ, μ) is a cutoff equilibrium:

We restate the definition of \bar{f}^μ given in Equation (1) in the main text: For every $c \in \mathcal{C}$,

$$\bar{f}_c^\mu = \begin{cases} \min_{\pi_c} \mu^{-1}(c) & \text{if } |\mu^{-1}(c)| = r_c \\ \emptyset & \text{otherwise.} \end{cases}$$

By this definition $\bar{f}^\mu \in \mathcal{F}$. Moreover, Condition (2) in the definition of a cutoff equilibrium is trivially satisfied.

We show that Condition (1(a)) holds next. Let $i \in I$. If $\mu(i) = \emptyset$, then Condition (1(a)) is satisfied for i . Suppose $\mu(i) = c$ for some $c \in \mathcal{C}$. We have $i \underline{\pi}_c \min_{\pi_c} \mu^{-1}(c)$. Moreover, $i \pi_c \emptyset$, as μ complies with eligibility requirements. Thus, $i \underline{\pi}_c \bar{f}_c^\mu \in \{\emptyset, \min_{\pi_c} \mu^{-1}(c)\}$, and, hence, $\mu(i) \in \mathcal{B}_i(\bar{f}^\mu)$, showing Condition (1(a)) is satisfied.

Finally, we show that Condition (1(b)) is satisfied. We prove its contrapositive. Let $i \in I$ be such that $\mu(i) \notin \mathcal{B}_i(\bar{f}^\mu)$. Thus, $\mu(i) = \emptyset$ by Condition (1(a)). Let $c \in \mathcal{C}$. If $|\mu^{-1}(c)| < r_c$, then $\bar{f}_c^\mu = \emptyset \pi_c i$ by nonwastefulness of μ . If $|\mu^{-1}(c)| = r_c$, then $j \pi_c i$ for every $j \in \mu^{-1}(c)$ as μ respects priorities; thus, $\bar{f}_c^\mu = \min_{\pi_c} \mu^{-1}(c) \pi_c i$. In either case, we have $c \notin \mathcal{B}_i(\bar{f}^\mu)$. Thus, we get $\mathcal{B}_i(\bar{f}^\mu) = \emptyset$, showing that Condition (1(b)) also holds for (\bar{f}^μ, μ) , and, hence, completing the proof that (\bar{f}^μ, μ) is a cutoff equilibrium.

Part 2. Let (f, μ) be a cutoff equilibrium. We prove that $\bar{f}_c^\mu \leq f_c$ for every $c \in \mathcal{C}$:

Suppose, for contradiction, that there exists some category $c \in \mathcal{C}$ such that $f_c \pi_c \bar{f}_c^\mu$. Then, $|\mu^{-1}(c)| = r_c$ as (f, μ) is a cutoff equilibrium and $f_c \pi_c \emptyset$, which follows from the fact that $\bar{f}_c^\mu \underline{\pi}_c \emptyset$. Thus, $\bar{f}_c^\mu = \min_{\pi_c} \mu^{-1}(c) \pi_c \emptyset$ by definition and μ complying with eligibility requirements. Then, for the patient $i = \bar{f}_c^\mu$, $\mu(i) = c \notin \mathcal{B}_i(f)$, contradicting that (f, μ) is a cutoff equilibrium. Thus, such a category c does not exist, and, hence, \bar{f}^μ is the maximum cutoff equilibrium vector supporting matching μ . \diamond

Claim 2. \underline{f}^μ is the minimum equilibrium cutoff vector supporting μ —that is, (\underline{f}^μ, μ) is a cutoff equilibrium, and for every cutoff equilibrium (f, μ) , $\underline{f}_c^\mu \leq f_c$ for every $c \in \mathcal{C}$.

Proof. We prove the claim in two parts.

Part 1. We show that (\underline{f}^μ, μ) is a cutoff equilibrium:

We restate \underline{f}^μ using its definition in Equation (2): For every $c \in \mathcal{C}$,

$$\underline{f}_c^\mu = \begin{cases} \min_{\pi_c} \{i \in \mu(I) : i \pi_c x_c\} & \text{if } x_c \neq \emptyset \\ \emptyset & \text{otherwise,} \end{cases}$$

where x_c is defined as

$$x_c = \max_{\pi_c} (I \setminus \mu(I)) \cup \{\emptyset\}.$$

For every $c \in \mathcal{C}$, because $x_c \pi_c \emptyset$, we have $\min_{\pi_c} \{i \in \mu(I) : i \pi_c x_c\} \pi_c \emptyset$, if $x_c \neq \emptyset$. Hence, $\underline{f}_c^\mu \pi_c \emptyset$, showing that $\underline{f}^\mu \in \mathcal{F}$.

We show that the conditions in the definition of a cutoff equilibrium are satisfied by (\underline{f}^μ, μ) .

Condition 2. Suppose $|\mu^{-1}(c)| < r_c$ for some $c \in \mathcal{C}$. For any $i \in I \setminus \mu(I)$, we have $\emptyset \pi_c i$ by nonwastefulness of μ . Thus, $x_c = \emptyset$. This implies $\underline{f}_c^\mu = \emptyset$ by its definition. Hence, Condition (2) is satisfied.

Condition 1(a). Let $i \in I$. If $\mu(i) = \emptyset$, then Condition (1(a)) is satisfied for i . Suppose $\mu(i) = c$ for some $c \in \mathcal{C}$. We have $i \pi_c \bar{f}_c^\mu$, because we showed in Claim 1 that (\bar{f}^μ, μ) is a cutoff equilibrium. Two cases are possible about $\mu^{-1}(c)$:

• If $|\mu^{-1}(c)| < r_c$: we showed in proving Condition (2) that $x_c = \emptyset$, thus, $\bar{f}_c^\mu = \underline{f}_c^\mu = \emptyset$. Because $i \pi_c \emptyset$, $c = \mu(i) \in \mathcal{B}_i(\underline{f}^\mu)$ showing that Condition (1(a)) holds for i .

- If $|\mu^{-1}(c)| = r_c$, then $\bar{f}_c^\mu = \min_{\pi_c} \mu^{-1}(c) \pi_c x_c$: as otherwise
 - If $x_c \in I$, then $\mu(x_c) = \emptyset$ (by definition of x_c), and yet $c \in \mathcal{B}_{x_c}(\bar{f}^\mu)$, a contradiction to (\bar{f}^μ, μ) being a cutoff equilibrium;
 - If $x_c = \emptyset$, then (i) $x_c \pi_c \bar{f}_c^\mu \in I$ contradicts \bar{f}^μ being a cutoff vector, and (ii) $x_c = \bar{f}_c^\mu$ contradicts $|\mu^{-1}(c)| = r_c$. Thus, we cannot have $x_c \pi_c \bar{f}_c^\mu$ in this case either.

Thus,

$$\bar{f}_c^\mu \pi_c \min_{\pi_c} \{i \in I : i \pi_c x_c\} = \underline{f}_c^\mu.$$

Then,

$$i \pi_c \bar{f}_c^\mu \pi_c \underline{f}_c^\mu,$$

implying $c = \mu(i) \in \mathcal{B}_i(\underline{f}^\mu)$ and showing that Condition (1(a)) holds for i .

Condition 1(b). Let $i \in I$ be such that $\mathcal{B}_i(\underline{f}^\mu) \neq \emptyset$. Let $c \in \mathcal{B}_i(\underline{f}^\mu)$.

- If $x_c = \emptyset$: Then, $i \pi_c \emptyset = \underline{f}_c^\mu$. By definition of x_c , $i \in \mu(I)$ —that is, $\mu(i) \neq \emptyset$.
- If $x_c \neq \emptyset$: Then, $i \pi_c \underline{f}_c^\mu \pi_c x_c$, which, in turn, implies that $\mu(i) \neq \emptyset$ by definition of \underline{f}_c^μ and x_c .

In either case, by Condition (1(a)), $\mu(i) \in \mathcal{B}_i(\underline{f}^\mu)$. Thus, Condition (1(b)) is satisfied for i .

These conclude, proving that (\underline{f}^μ, μ) is a cutoff equilibrium.

Part 2. Let (f, μ) be a cutoff equilibrium. We prove that $f_c \pi_c f_c^\mu$ for every $c \in \mathcal{C}$:

Suppose, to the contrary of the claim $f_c^\mu \pi_c f_c$, for some $c \in \mathcal{C}$. Now, f_c^μ is a patient because $f_c \pi_c \emptyset$ by the definition of a cutoff

vector. By definition of \underline{f}_c^μ , $x_c \neq \emptyset$ and $f_c^\mu \pi_c x_c$. We cannot have $x_c \pi_c f_c$, as otherwise, we have $c \in \mathcal{B}_{x_c}(f)$ however, by definition of x_c , $\mu(x_c) = \emptyset$, contradicting (f, μ) is a cutoff equilibrium. Thus, $f_c \pi_c x_c$. Because x_c is eligible for c , $f_c \in I$. Furthermore, $f_c \in \mu(I)$, because $c \in \mathcal{B}_f(f)$ and (f, μ) is a cutoff equilibrium. Therefore, $f_c \in \{j \in \mu(I) : j \pi_c x_c\}$. Because $\underline{f}_c^\mu = \min_{\pi_c} \{j \in \mu(I) : j \pi_c x_c\}$, we have $f_c \pi_c \underline{f}_c^\mu$, contradicting $\underline{f}_c^\mu \pi_c f_c$. Therefore, such a category c cannot exist, and, hence, \underline{f}^μ is the minimum equilibrium cutoff vector supporting μ . \diamond

Claim 3. For any two given cutoff equilibria (f, μ) and (g, μ) such that $f_c \pi_c g_c$ for every $c \in \mathcal{C}$, the pair (h, μ) is also a cutoff equilibrium, where $h \in \mathcal{F}$ satisfies for every $c \in \mathcal{C}$, $f_c \pi_c h_c \pi_c g_c$.

Proof. We can obtain cutoff vector h from g after a sequence of repeated applications of the following operation: Change the cutoffs of one of the categories $c \in \mathcal{C}$ of an input vector $f' \in \mathcal{F}$ so that its cutoff increases by one patient and gets closer to h_c than f'_c . We start with $f' = g$ to the sequence. We show that each iteration of this operation results with a new equilibrium cutoff vector g' supporting μ , and we use this g' as the input of the next iteration of the operation. Because the outcome vector gets closer to h at each step, the last cutoff vector of the sequence is h by finiteness of the patient set, and, thus, (h, μ) is a cutoff equilibrium:

Suppose $c' \in \mathcal{C}$ is such that $h_{c'} \pi_{c'} g_{c'}$. We prove that for cutoff vector $g' \in \mathcal{F}$ such that $g'_c = g_c$ for every $c \in \mathcal{C} \setminus \{c'\}$ and $g'_{c'} = \min_{\pi_{c'}} \{i \in I : i \pi_{c'} g_{c'}\}$, (g', μ) is a cutoff equilibrium. It is straightforward to show that $g' \in \mathcal{F}$. Observe also that $\mathcal{B}_i(g') = \mathcal{B}_i(g)$ for every $i \in I \setminus \{j\}$, where $j = g_{c'}$. Three cases are possible regarding j :

- If $j \notin I$: $j = \emptyset$.
- If $j \in I$ and $\mu(j) = c'$: Category $c' \in \mathcal{B}_j(f)$ as (f, μ) is a cutoff equilibrium. However, $f_{c'} \pi_{c'} h_{c'} \pi_{c'} j = g_{c'}$, contradicting that $c' \in \mathcal{B}_j(f)$. Therefore, this case cannot happen.
- If $j \in I$ and $\mu(j) \neq c'$: Observe that $\mu(j) \neq \emptyset$, as $c' \in \mathcal{B}_j(g)$ and (g, μ) is a cutoff equilibrium. Moreover, we have $\mu(j) \in \mathcal{B}_j(g)$, in turn, together with $\mu(j) \neq c'$ implying that $\mu(j) \in \mathcal{B}_j(g')$ as $g'_{\mu(j)} = g_{\mu(j)}$.

These and the fact that (g, μ) is a cutoff equilibrium (specifically its Condition (1(b))) show that $\mu(i) \in \mathcal{B}_i(g')$ for every $i \in I$ such that $\mathcal{B}_i(g') \neq \emptyset$, proving that Condition (1(b)) holds in the definition of cutoff equilibrium for (g', μ) .

Because (g, μ) is a cutoff equilibrium (specifically Conditions (1(a)) and (1(b)) of the definition) imply that for every $i \in I$, $\mu(i) = \emptyset \Leftrightarrow \mathcal{B}_i(g) = \emptyset$. Therefore, we have $\mu(i) = \emptyset$ for every $i \in I$ such that $\mathcal{B}_i(g') = \emptyset$, because $\mathcal{B}_i(g') \subseteq \mathcal{B}_i(g) = \emptyset$, where the set inclusion follows from the fact that the cutoffs have weakly increased for each category under g' . This and Condition (1(b)) imply that for all $i \in I$, $\mu(i) \in \mathcal{B}_i(g') \cup \{\emptyset\}$. Thus, Condition (1(a)) in the definition of a cutoff equilibrium is also satisfied by (g', μ) .

We show that Condition (2) is also satisfied, proving that for every $c \in \mathcal{C}$, $g'_c = \emptyset$ if $|\mu^{-1}(c)| < r_c$ to conclude that (g', μ) is a cutoff equilibrium. Suppose $|\mu^{-1}(c)| < r_c$ for some $c \in \mathcal{C}$. If $c \neq c'$, then $g'_c = g_c = \emptyset$, where the latter equality follows from (g, μ) being a cutoff equilibrium (specifically its Condition (2)). If $c = c'$, $f_{c'} = \emptyset \pi_{c'} g_{c'}$, where the first equality follows from (f, μ) being a cutoff equilibrium (specifically its Condition (2)). This contradicts $g \in \mathcal{F}$. So $c \neq c'$, completing the proof. \diamond

C.2. Proof of the Result in Section 3.5

Proof of Proposition 1

Sufficiency: We first prove that any DA-induced matching complies with eligibility requirements, is nonwasteful, and respects priorities. Let $\succ \in \mathcal{P}$ be a preference profile of patients over categories and \emptyset . Suppose $\mu \in \mathcal{M}$ is DA-induced from this preference profile.

Compliance with eligibility requirements: Suppose that $\mu(i) = c$ for some $c \in \mathcal{C}$. Then, i must apply to c in a step of the DA algorithm, and, hence, $c \succ_i \emptyset$. By construction of \succ_i , this means $i \pi_c \emptyset$. Therefore, matching μ complies with eligibility requirements.

Nonwastefulness: Suppose that $i \pi_c \emptyset$ and $\mu(i) = \emptyset$ for some category $c \in \mathcal{C}$ and patient $i \in I$. By construction of \succ_i , $c \succ_i \emptyset$ because i is eligible for c . As patient i is unmatched in μ , she applies to c in some step of the DA algorithm. However, c rejects i at this or a later step. This means that c should have been holding at least r_c offers from eligible patients at this step. From this step on, c always holds r_c offers, and eventually all of its units are assigned: $|\mu^{-1}(c)| = r_c$. Hence, matching μ is nonwasteful.

Respecting priorities: Suppose that $\mu(i) = c$ and $\mu(i') = \emptyset$ for two patients $i, i' \in I$ and a category $c \in \mathcal{C}$. For every category $c' \in \mathcal{C}$, $\pi_{c'}$ is used to choose eligible patients at every step of the DA algorithm. Therefore, $\mu(i) = c$ implies $i \pi_c \emptyset$. Because $\mu(i') = \emptyset$, then it must be either because $\emptyset \pi_c i'$ or because $i \pi_c i'$. In the first case, we get $i \pi_c i'$ as well because π_c is transitive. Therefore, matching μ respects priorities.

Necessity: We now prove that any matching $\mu \in \mathcal{M}$ with the three stated properties is DA-induced from some preference profile. We construct a candidate preference profile $\succ \in \mathcal{P}$ as follows:

- Consider a patient $i \in \mu^{-1}(c)$ where $c \in \mathcal{C}$. Because μ complies with eligibility requirements, i must be eligible for category c . Let i rank category c first in \succ_i . The rest of the ranking in \succ_i is arbitrary, as long as all eligible categories are ranked above the empty set.
- Consider an unmatched patient $i \in \mu^{-1}(\emptyset)$. Let i rank categories in any order in \succ_i such that only eligible categories are ranked above the empty set.

We now show that μ is DA-induced from preference profile \succ . In the induced DA algorithm under \succ , for every category $c' \in \mathcal{C}$, patients in $\mu^{-1}(c')$ apply to category c' first. Every unmatched patient $j \in \mu^{-1}(\emptyset)$ applies to her first-ranked eligible category according to \succ_j , if there is any. Suppose $c \in \mathcal{C}$ is this category. Because μ respects priorities, j has a lower priority than any patient in $\mu^{-1}(c)$, who also applied to c in Step 1. Furthermore, because μ is nonwasteful, $|\mu^{-1}(c)| = r_c$ (as there are unmatched eligible patients for this category, for example, j). Therefore, all unmatched patients in μ are rejected at the first step of the DA algorithm. Moreover, for every category $c' \in \mathcal{C}$, all patients in $\mu^{-1}(c')$ are tentatively accepted by category c' at the end of Step 1.

Each unmatched patient in $j \in \mu^{-1}(\emptyset)$ continues to apply, according to \succ_j to the other categories at which she is eligible. Because μ respects priorities and is nonwasteful, she is rejected from all categories for which she is eligible one at a time, because each of these categories $c \in \mathcal{C}$ continues to tentatively hold patients $\mu^{-1}(c)$ from Step 1 who all have higher priority than j , according to π_c , as μ respects priorities. Moreover, by nonwastefulness of μ , $|\mu^{-1}(c)| = r_c$, as there are unmatched eligible patients (for example, j) under μ .

As a result, when the algorithm stops, the outcome is such that, for every category $c' \in \mathcal{C}$, all patients in $\mu^{-1}(c')$ are matched with c' . Moreover, every patient in $\mu^{-1}(\emptyset)$ remains unmatched at the end. Therefore, μ is DA-induced from the constructed patient preferences \succ . \square

C.3. Proofs of the Results in Section 3.6

Proof of Proposition 2. Let $\triangleright \in \Delta$ be a precedence order and φ_\triangleright be the associated sequential reserve matching. We show that φ_\triangleright is DA-induced from preference profile $\succ^\triangleright = (\succ_i^\triangleright)_{i \in I}$.

For every patient $i \in I$, consider another strict preference relation \succ'_i such that all categories are ranked above the empty set, and, furthermore, for any $c, c' \in \mathcal{C}$,

$$c \succ'_i c' \Leftrightarrow c \triangleright c'.$$

Note that the relative ranking of two categories for which i is eligible is the same between \succ_i^\triangleright and \succ'_i .

We use an equivalent version of the DA algorithm as the one given in the text. Consider a Step k : Each patient i who is not tentatively accepted currently by a category applies to the best category that has not rejected her yet, according to \succ'_i . Suppose that I_c^k is the union of the set of patients who were tentatively assigned to category c in Step $k-1$ and the set of patients who just proposed to category c . Category c tentatively assigns eligible patients in I_c^k with the highest priority according to π_c until all patients in I_c^k are chosen or all r_c units are allocated, whichever comes first, and permanently rejects the rest.

Because for any category $c \in \mathcal{C}$ and any patient $i \in I$ who is ineligible for category c , $\emptyset \pi_c i$, the outcome of the DA algorithm when the preference profile is \succ^\triangleright and $\succ' = (\succ'_i)_{i \in I}$ are the same.

Furthermore, when the preference profile is \succ' , the DA algorithm works exactly like the sequential reserve procedure that is used to construct φ_\triangleright . We show this by induction. Suppose \triangleright orders categories as $c_1 \triangleright c_2 \triangleright \dots \triangleright c_{|\mathcal{C}|}$. As the inductive assumption, for $k > 0$, suppose for categories c_1, \dots, c_{k-1} , the tentative matches at the end of Step $k-1$ and final matches at the end under the DA algorithm from \succ' are identical to their matches in sequential reserve matching φ_\triangleright .

We next consider Step k of the DA algorithm from \succ' . Only patients who are rejected from category c_{k-1} apply in Step k of the DA algorithm, and they all apply to category c_k . Then, c_k uses its priority order π_{c_k} to tentatively accept the r_{c_k} highest-priority eligible applicants (and if there are less than r_{c_k} eligible applicants, all eligible applicants) and rejects the rest. Observe that because every patient who is not tentatively accepted by a category c_1, \dots, c_{k-1} applied to this category in Step k , none of these patients will ever apply to it again; and by the inductive assumption, no patient who is tentatively accepted in categories c_1, \dots, c_{k-1} will ever be rejected, and, thus, they will never apply to c_k either. Thus, the tentative acceptances by c_k will become permanent at the end of the DA algorithm. Moreover, this step is identical to Step k of the sequential reserve procedure under precedence order \triangleright and the same patients are matched with category c_k in φ_\triangleright . This ends the induction.

Therefore, we conclude that φ_\triangleright is DA-induced from patient preference profile \succ^\triangleright . \square

Proof of Proposition 3. Let $J^p, J^{p'} \subseteq I$ be the sets of patients remaining just before category c is processed under the

sequential reserve matching procedure induced by \triangleright and \triangleright' , respectively. Because c is processed earlier under \triangleright' and every other category preceding c and c' under \triangleright and \triangleright' are ordered in the same manner order, $J^{\triangleright} \subseteq J^{\triangleright'}$. Two cases are possible:

1. If $|\varphi_{\triangleright'}^{-1}(c)| < r_c$: Then, $J^{\triangleright} \subseteq J^{\triangleright'}$ implies $|\varphi_{\triangleright}^{-1}(c)| < r_c$. Therefore, by Equation (1), $\bar{f}_c^{\varphi_{\triangleright'}} = \emptyset = \bar{f}_c^{\varphi_{\triangleright}}$.

2. If $|\varphi_{\triangleright'}^{-1}(c)| = r_c$: Then, $J^{\triangleright} \subseteq J^{\triangleright'}$ implies,

$$\bar{f}_c^{\varphi_{\triangleright'}} = \min_{\pi_c} \varphi_{\triangleright'}^{-1}(c) \underline{\pi}_c \min_{\pi_c} \varphi_{\triangleright}^{-1}(c),$$

where the first equality follows by Equation (1). By the same equation, $\bar{f}_c^{\varphi_{\triangleright}} \in \{\emptyset, \min_{\pi_c} \varphi_{\triangleright}^{-1}(c)\}$, and by the definition of a cutoff vector, $\bar{f}_c^{\varphi_{\triangleright'}} \neq \emptyset$. Thus,

$$\bar{f}_c^{\varphi_{\triangleright'}} \underline{\pi}_c \bar{f}_c^{\varphi_{\triangleright}}. \quad \square$$

C.4. Proof of Result in Section 3.7

In this subsection, we first show some lemmas that we will use in the proof of Proposition 4, the only result in Section 3.7.

Fix a soft reserve system induced by the baseline priority order π . Suppose each patient is a beneficiary of at most one preferential treatment category.

First, we introduce some concepts.

We introduce function $\tau : I \rightarrow (\mathcal{C} \setminus \{u\}) \cup \{\emptyset\}$ to denote the preferential treatment category that a patient is beneficiary of, if there is such a category. That is, for any patient $i \in I$, if $i \in I_c$ for some $c \in \mathcal{C} \setminus \{u\}$, then $\tau(i) = c$, and if $i \in I_g$ —that is, i is a general-community patient—then $\tau(i) = \emptyset$.

For a category $c^* \in \mathcal{C}$, a set of patients $\tilde{I} \subseteq I$, and a patient $i \in \tilde{I}$, let $\text{rank}(i; \tilde{I}, \pi_{c^*})$ denote the rank of i among patients in \tilde{I} , according to π_{c^*} .

We consider *incomplete* orders of precedence. For a given subset of categories $\mathcal{C}^* \subseteq \mathcal{C}$, we define an *order of precedence on \mathcal{C}^** as a linear order on \mathcal{C}^* . Let $\Delta(\mathcal{C}^*)$ be the set of orders of precedence on \mathcal{C}^* .

We extend the definition of sequential reserve matchings to cover incomplete precedence orders and match a subset of patients $\tilde{I} \subseteq I$ as follows: A *sequential reserve matching induced by $\triangleright \in \Delta(\mathcal{C}^*)$ over \tilde{I}* is the outcome of the sequential reserve procedure, which processes only the categories in \mathcal{C}^* in the order of \triangleright to match only the patients in \tilde{I} and leaves all categories in $\mathcal{C} \setminus \mathcal{C}^*$ unmatched and patients in $I \setminus \tilde{I}$ unmatched. Let $\varphi_{\triangleright}^{\tilde{I}}$ denote this matching.

Lemma C.1. Suppose that $\tilde{I} \subseteq I$, $c \in \mathcal{C} \setminus \{u\}$, and $\triangleright, \triangleright' \in \Delta(\{u, c\})$ are such that

- \triangleright is given as $u \triangleright c$,
- \triangleright' is given as $c \triangleright' u$,
- $I(2) = \tilde{I} \setminus \mu(\tilde{I})$ where $\mu = \varphi_{\triangleright}^{\tilde{I}}$,
- $I'(2) = \tilde{I} \setminus \mu'(\tilde{I})$, where $\mu' = \varphi_{\triangleright'}^{\tilde{I}}$, and
- $\mu(\tilde{I}_c) \subsetneq \tilde{I}_c$.

Then, the following results hold:

1. $|I(2) \setminus I'(2)| = |I'(2) \setminus I(2)|$,
2. $I'(2) \setminus I(2) \subseteq \tilde{I}_c$,
3. $I(2) \setminus I'(2) \subseteq \tilde{I} \setminus \tilde{I}_c$,
4. If $i \in I(2) \setminus I'(2)$ and $i' \in I'(2)$, then $i \pi i'$, and
5. If $i' \in I'(2) \setminus I(2)$ and $i \in I(2) \cap \tilde{I}_c$, then $i' \pi i$.

Proof of Lemma C.1. The first statement in Lemma C.1 holds because under soft reserves, every patient is eligible for every category, which implies that $|\mu(\tilde{I})| = |\mu'(\tilde{I})|$. As a

result, $|\mu(\tilde{I}) \setminus \mu'(\tilde{I})| = |\mu'(\tilde{I}) \setminus \mu(\tilde{I})|$, which is equivalent to $|I'(2) \setminus I(2)| = |I(2) \setminus I'(2)|$ because $\mu(\tilde{I}) \setminus \mu'(\tilde{I}) = I'(2) \setminus I(2)$ and $\mu'(\tilde{I}) \setminus \mu(\tilde{I}) = I(2) \setminus I'(2)$.

The second statement in Lemma C.1 holds because if $i \in \mu^{-1}(u)$, then $\text{rank}(i; \tilde{I}, \pi) \leq r_u$. Therefore, $i \in \mu'(\tilde{I})$. Furthermore, every $i \in \mu^{-1}(c)$ is a category- c patient because there exists $j \in \tilde{I}_c$ such that $j \notin \mu(\tilde{I})$. As a result, we get

$$\tilde{I}_c \supseteq \mu^{-1}(c) \supseteq \mu(\tilde{I}) \setminus \mu'(\tilde{I}) = I'(2) \setminus I(2).$$

To prove the third statement in Lemma C.1, suppose for contradiction that there exists $i \in I(2) \setminus I'(2)$ such that $i \in \tilde{I}_c$. Therefore, $i \in \mu'(\tilde{I}) \setminus \mu(\tilde{I}) = I(2) \setminus I'(2)$. By the first statement in Lemma C.1, $|I'(2) \setminus I(2)| = |I(2) \setminus I'(2)| \geq 1$ because $I(2) \setminus I'(2)$ has at least one patient. By the second statement in Lemma C.1, $I'(2) \setminus I(2) \subseteq \tilde{I}_c$. Therefore, there exists $i' \in I'(2) \setminus I(2) = \mu(\tilde{I}) \setminus \mu'(\tilde{I})$ such that $i' \in \tilde{I}_c$. Because $i \in \mu'(\tilde{I})$, $i' \notin \mu'(\tilde{I})$, and $\tau(i) = \tau(i')$, we get

$$i \pi i'.$$

Likewise, $i' \in \mu(\tilde{I})$, $i \notin \mu(\tilde{I})$, and $\tau(i) = \tau(i')$ imply

$$i' \pi i.$$

The two displayed relations above contradict each other.

The fourth statement in Lemma C.1 is true because for every $i \in I(2) \setminus I'(2) = \mu'(\tilde{I}) \setminus \mu(\tilde{I})$, we know that $i \notin \tilde{I}_c$ by the third statement in Lemma C.1. Because $\mu(\tilde{I}_c) \subsetneq \tilde{I}_c$, there are at least r_c patients in \tilde{I}_c . Therefore, $\mu'^{-1}(c) \subseteq \tilde{I}_c$, which implies that $i \in \mu'^{-1}(u)$. Because $i' \in I'(2)$ is equivalent to $i' \notin \mu'(\tilde{I})$, we get $i \pi i'$.

The fifth statement in Lemma C.1 follows from $i, i' \in \tilde{I}_c$, $i' \in \mu(\tilde{I})$, and $i \notin \mu(\tilde{I})$. \square

Lemma C.2. Suppose that $c, c' \in \mathcal{C} \setminus \{u\}$ are different categories. Let $\tilde{I} \subseteq I$ and $\triangleright, \triangleright' \in \Delta(\{c, c'\})$ be such that

- \triangleright is given as $c' \triangleright c$,
- \triangleright' is given as $c \triangleright' c'$,
- $I(2) = \tilde{I} \setminus \mu(\tilde{I})$ where $\mu = \varphi_{\triangleright}^{\tilde{I}}$,
- $I'(2) = \tilde{I} \setminus \mu'(\tilde{I})$ where $\mu' = \varphi_{\triangleright'}^{\tilde{I}}$, and
- $\mu(\tilde{I}_c) \subsetneq \tilde{I}_c$.

Then, the following results hold:

1. $|I(2) \setminus I'(2)| = |I'(2) \setminus I(2)|$,
2. $I'(2) \setminus I(2) \subseteq \tilde{I}_c$,
3. $I(2) \setminus I'(2) \subseteq \tilde{I} \setminus \tilde{I}_c$,
4. If $i \in I(2) \setminus I'(2)$ and $i' \in I'(2)$, then $i \pi i'$, and
5. If $i' \in I'(2) \setminus I(2)$ and $i \in I(2) \cap \tilde{I}_c$, then $i' \pi i$.

Proof of Lemma C.2. If $|\tilde{I}_c| \geq r_{c'}$, then $\mu(\tilde{I}) = \mu'(\tilde{I})$ and, therefore, $I(2) = I'(2)$. Then, all the statements in Lemma C.2 hold trivially. Suppose that $|\tilde{I}_c| < r_{c'}$ for the rest of the proof.

The first statement in Lemma C.2 follows as in the proof of the first statement in Lemma C.1.

The second statement in Lemma C.2 holds because if $i \in \mu^{-1}(c')$ and $i \in \tilde{I}_{c'}$, then $i \in \mu'(\tilde{I})$ because $|\tilde{I}_{c'}| < r_{c'}$. If $i \in \mu^{-1}(c')$ and $i \notin \tilde{I}_{c'}$, then $\text{rank}(i; \tilde{I} \setminus \tilde{I}_{c'}, \pi) \leq r_{c'} - |\tilde{I}_{c'}|$. As a result $i \in \mu'(\tilde{I})$. These two statements imply that $\mu^{-1}(c') \subseteq \mu'(\tilde{I})$. Furthermore, every $i \in \mu^{-1}(c)$ is a category- c patient because there exists $i \in \tilde{I}_c$ such that $i \notin \mu(\tilde{I})$. As a result, we get that

$$I'(2) \setminus I(2) = \mu(\tilde{I}) \setminus \mu'(\tilde{I}) = \mu^{-1}(c) \setminus \mu'(\tilde{I}) \subseteq \mu^{-1}(c) \subseteq \tilde{I}_c.$$

The proof of the third statement in Lemma C.2 is the same as the proof of the third statement in Lemma C.1.

The fourth statement in Lemma C.2 is true because for every $i \in I(2) \setminus I'(2) = \mu'(\tilde{I}) \setminus \mu(\tilde{I})$, we know that $i \notin \tilde{I}_c$ by the third statement in Lemma C.2. Moreover, $\mu'^{-1}(c) \subseteq \tilde{I}_c$, as there exists $j \in \tilde{I}_c$ such that $j \notin \mu(\tilde{I})$, which implies that there are at least r_c category- c patients. This implies $i \in \mu'^{-1}(c')$. Furthermore, $i \notin \tilde{I}_{c'}$ because $\tilde{I}_{c'} \subseteq \mu(\tilde{I})$, which follows from $|\tilde{I}_{c'}| < r_{c'}$. Consider $i' \in I'(2)$. Then $i' \notin \mu'(\tilde{I})$, which implies that $i \pi i'$ because $\mu'(i) = c'$, $\tau(i) \neq c'$, and $\mu'(i') = \emptyset$.

The proof of the fifth statement in Lemma C.2 is the same as the proof of the fifth statement in Lemma C.1. \square

Lemma C.3. Suppose that $c \in \mathcal{C} \setminus \{u\}$ and $c', c^* \in \mathcal{C} \setminus \{c\}$ are different categories. Let $\tilde{I} \subseteq I$ and $\triangleright, \triangleright' \in \Delta(\{c, c', c^*\})$ be such that

- \triangleright is given as $c' \triangleright c \triangleright c^*$,
- \triangleright' is given as $c \triangleright' c' \triangleright' c^*$,
- $I(3) = \tilde{I} \setminus \mu(\tilde{I})$, where $\mu = \varphi_{\triangleright}^{\tilde{I}}$,
- $I'(3) = \tilde{I} \setminus \mu'(\tilde{I})$, where $\mu' = \varphi_{\triangleright'}^{\tilde{I}}$, and
- $\mu(\tilde{I}_c) \subsetneq \tilde{I}_c$.

Then, the following results hold:

1. $|I(3) \setminus I'(3)| = |I'(3) \setminus I(3)|$,
2. $I'(3) \setminus I(3) \subseteq \tilde{I}_c$,
3. $I(3) \setminus I'(3) \subseteq \tilde{I} \setminus \tilde{I}_c$, and
4. If $i' \in I'(3) \setminus I(3)$ and $i \in I(3) \cap \tilde{I}_c$, then $i' \pi i$.

Proof of Lemma C.3. The first statement in Lemma C.3 follows as in the proof of the first statement in Lemma C.1. Likewise, the fourth statement in Lemma C.3 follows as in the proof of the fifth statement in Lemma C.1.

To show the other two statements, we use Lemma C.1 and Lemma C.2. Let $\hat{\triangleright}, \hat{\triangleright}' \in \Delta(\{c, c'\})$ be such that

$$\hat{\triangleright} : c' \hat{\triangleright} c \quad \text{and} \quad \hat{\triangleright}' : c \hat{\triangleright}' c'.$$

Let $I(2) = \tilde{I} \setminus \varphi_{\hat{\triangleright}}^{\tilde{I}}(\tilde{I})$ and $I'(2) = \tilde{I} \setminus \varphi_{\hat{\triangleright}'}^{\tilde{I}}(\tilde{I})$. Then, $I(3) = I(2) \setminus \mu^{-1}(c^*)$ and $I'(3) = I'(2) \setminus \mu'^{-1}(c^*)$.

For both precedence orders \triangleright and \triangleright' under the sequential reserve matching procedure, consider the beginning of the third step, at which category c^* is processed. For \triangleright , the set of available patients is $I(2)$. For \triangleright' , the set of available patients is $I'(2)$. If $I(2) = I'(2)$, then all the statements hold trivially because in this case, we get $I(3) = I'(3)$. Therefore, assume that $I(2) \neq I'(2)$. For every precedence order, r_{c^*} patients with the highest priority with respect to π_{c^*} are chosen.

We consider each patient chosen under \triangleright and \triangleright' for category c^* one at a time in sequence with respect to the priority order π_{c^*} . For both precedence orders, there are r_{c^*} patients matched with c^* because $\mu(\tilde{I}_c) \subsetneq \tilde{I}_c$. Let i_k be the k^{th} patient chosen under \triangleright for c^* and i'_k be the k^{th} patient chosen under \triangleright' for c^* where $k = 1, \dots, r_{c^*}$. Let J_k be the set of patients available when we process \triangleright for the k^{th} patient and J'_k be the set of patients available when we process \triangleright' for the k^{th} patient where $k = 1, \dots, r_{c^*}$. For $k = 1$, $J_k = I(2)$ and $J'_k = I'(2)$. By definition, $J_{k+1} = J_k \setminus \{i_k\}$ and $J'_{k+1} = J'_k \setminus \{i'_k\}$. We show that

- a. $J'_k \setminus J_k \subseteq \tilde{I}_c$,
- b. $J_k \setminus J'_k \subseteq \tilde{I} \setminus \tilde{I}_c$, and
- c. If $i \in J_k \setminus J'_k$ and $i' \in (J_k \cap J'_k) \setminus \tilde{I}_c$, then $i \pi_{c^*} i'$.

by mathematical induction on k . These three claims trivially hold for $k = 1$ by statements (2), (3), and (4) in Claims 1 and 2.

Fix k . In the inductive step, assume that Statements (a), (b), and (c) hold for k . Consider $k + 1$. If $J_{k+1} = J'_{k+1}$, then the statements trivially hold. Assume that $J_{k+1} \neq J'_{k+1}$, which implies that $J_\ell \neq J'_\ell$ for $\ell = 1, \dots, k$. There are four cases depending on which sets i_k and i'_k belong to. We consider each case separately.

Case 1. $i_k \in J_k \setminus J'_k$ and $i'_k \in J'_k \cap J_k$. Then,

$$J'_{k+1} \setminus J_{k+1} = (J'_k \setminus \{i'_k\}) \setminus (J_k \setminus \{i_k\}) = J'_k \setminus J_k,$$

and

$$J_{k+1} \setminus J'_{k+1} = (J_k \setminus \{i_k\}) \setminus (J'_k \setminus \{i'_k\}) = ((J_k \setminus J'_k) \setminus \{i_k\}) \cup \{i'_k\}.$$

By Statement (a) of the inductive assumption for k , $J'_{k+1} \setminus J_{k+1} = J'_k \setminus J_k \subseteq \tilde{I}_c$. Therefore, Statement (a) holds for $k + 1$.

As $J'_k \neq J_k$ and $|J_k| = |J'_k|$ because of the soft-reserves condition, $J'_k \setminus J_k$ has at least one category- c patient. Moreover, this patient has higher priority with respect to π_{c^*} than any other category- c patient in $J'_k \cap J_k$ because the former is chosen under \triangleright , whereas the latter is not chosen under \triangleright . Therefore, i'_k cannot be a category- c patient. As a result, $J_{k+1} \setminus J'_{k+1} \subseteq \tilde{I} \setminus \tilde{I}_c$, so Statement (b) holds for $k + 1$.

To show Statement (c) for $k + 1$, observe that $J_{k+1} \setminus J'_{k+1} = ((J_k \setminus J'_k) \setminus \{i_k\}) \cup \{i'_k\}$ and $J_{k+1} \cap J'_{k+1} = (J_k \cap J'_k) \setminus \{i_k\}$. Therefore, Statement (c) for $k + 1$ follows from Statement (c) for k and the fact that $i'_k \pi_{c^*} i$ for any $i \in J_{k+1} \cap J'_{k+1}$.

Case 2. $i_k \in J_k \setminus J'_k$ and $i'_k \in J'_k \setminus J_k$. Then,

$$J'_{k+1} \setminus J_{k+1} = (J'_k \setminus \{i'_k\}) \setminus (J_k \setminus \{i_k\}) = (J'_k \setminus J_k) \setminus \{i'_k\},$$

and

$$J_{k+1} \setminus J'_{k+1} = (J_k \setminus \{i_k\}) \setminus (J'_k \setminus \{i'_k\}) = (J_k \setminus J'_k) \setminus \{i_k\}.$$

Therefore, $J'_{k+1} \setminus J_{k+1} \subseteq \tilde{I}_c$ and $J_{k+1} \setminus J'_{k+1} \subseteq \tilde{I} \setminus \tilde{I}_c$ by Statements (a) and (b) for k , respectively, implying Statements (a) and (b) for $k + 1$.

To show Statement (c) for $k + 1$, observe that $J_{k+1} \setminus J'_{k+1} = (J_k \setminus J'_k) \setminus \{i_k\}$ and $J_{k+1} \cap J'_{k+1} = J_k \cap J'_k$. Therefore, Statement (c) for $k + 1$ follows from Statement (c) for k trivially.

Case 3. $i_k \in J_k \cap J'_k$ and $i'_k \in J'_k \cap J_k$. In this case, $i_k = i'_k$, then

$$J'_{k+1} \setminus J_{k+1} = (J'_k \setminus \{i'_k\}) \setminus (J_k \setminus \{i_k\}) = J'_k \setminus J_k,$$

and

$$J_{k+1} \setminus J'_{k+1} = (J_k \setminus \{i_k\}) \setminus (J'_k \setminus \{i'_k\}) = J_k \setminus J'_k.$$

Therefore, Statements (a) and (b) for $k + 1$ follows from the respective statements for k .

To show Statement (c) for $k + 1$, observe that $J_{k+1} \setminus J'_{k+1} = J_k \setminus J'_k$ and $J_{k+1} \cap J'_{k+1} = (J_k \cap J'_k) \setminus \{i_k\}$. Therefore, Statement (c) for $k + 1$ follows from Statement (c) for k trivially.

Case 4. $i_k \in J_k \cap J'_k$ and $i'_k \in J'_k \setminus J_k$. We argue that this case is not possible. Because $i'_k \in J'_k \setminus J_k$, i'_k must be a category- c patient by Statement (a) for k . If $c^* = u$, then every patient in $J_k \setminus J'_k$ has a higher priority with respect to π than every

patient in $J_k \cap J'_k$, which cannot happen because $i_k \in J_k \cap J'_k$. Therefore, $c^* \neq u$. Because i'_k is a category- c patient, there must not be a category- c^* patient in J'_k . By Statement (c) for k , we know that every patient in $J_k \setminus J'_k$ has a higher priority with respect to π_{c^*} than every patient in $(J_k \cap J'_k) \setminus I_{c^*} = J_k \cap J'_k$. This is a contradiction to $i_k \in J_k \cap J'_k$. Therefore, Case 4 is not possible.

Because $I(3) = J_{r_c}$ and $I'(3) = J'_{r_c}$, Statements (2) and (3) in Lemma C.3 follow from Statements (a) and (b) above, respectively. \square

Lemma C.4. Suppose that $c \in \mathcal{C} \setminus \{u\}$ and $c', c^*, \tilde{c} \in \mathcal{C} \setminus \{c\}$ are different categories. Let $\tilde{I} \subseteq I$ and $\triangleright, \triangleright' \in \Delta(\{c, c', c^*, \tilde{c}\})$ be such that

- \triangleright is given as $c' \triangleright c \triangleright c^* \triangleright \tilde{c}$,
- \triangleright' is given as $c \triangleright' c' \triangleright' c^* \triangleright' \tilde{c}$,
- $I(4) = \tilde{I} \setminus \mu(\tilde{I})$, where $\mu = \varphi_{\triangleright}^{\tilde{I}}$,
- $I'(4) = \tilde{I} \setminus \mu'(\tilde{I})$, where $\mu' = \varphi_{\triangleright'}^{\tilde{I}}$, and
- $\mu(\tilde{I}_c) \subsetneq \tilde{I}_c$.

Then, the following results hold:

1. $|I(4) \setminus I'(4)| = |I'(4) \setminus I(4)|$,
2. $I(4) \setminus I'(4) \subseteq \tilde{I} \setminus \tilde{I}_c$,
3. If $i' \in I'(4) \setminus I(4)$, $i' \notin \tilde{I}_c$, and $i \in I(4)$, then $i' \succ i$, and
4. If $i' \in I'(4) \setminus I(4)$, $i' \in \tilde{I}_c$, and $i \in I(4) \cap \tilde{I}_c$, then $i' \succ i$.

Proof of Lemma C.4. The first statement in Lemma C.4 follows as in the proof of the first statement in Lemma C.3. Likewise, the fourth statement in Lemma C.4 follows as in the proof of the fifth statement in Lemma C.3.

To prove the other two statements, we use Lemma C.3. Let $\triangleright, \triangleright' \in \Delta(\{c, c', c^*\})$ be such that

$$\triangleright : c' \triangleright c \triangleright c^* \quad \text{and} \quad \triangleright' : c \triangleright' c' \triangleright' c^*.$$

Let $I(3) = \tilde{I} \setminus \varphi_{\triangleright}^{\tilde{I}}(\tilde{I})$ and $I'(3) = \tilde{I} \setminus \varphi_{\triangleright'}^{\tilde{I}}(\tilde{I})$. Then, $I(4) = I(3) \setminus \mu^{-1}(\tilde{c})$ and $I'(4) = I'(3) \setminus \mu'^{-1}(\tilde{c})$.

For both precedence orders \triangleright and \triangleright' under the sequential reserve matching procedure, consider the beginning of the fourth step, at which category \tilde{c} is processed. For \triangleright , the set of available patients is $I(3)$. For \triangleright' , the set of available patients is $I'(3)$. If $I(3) = I'(3)$, then $I(4) = I'(4)$, which implies all the statements in Lemma C.4. Therefore, assume that $I(3) \neq I'(3)$. For every precedence order, $r_{\tilde{c}}$ patients with the highest priority with respect to $\pi_{\tilde{c}}$ are chosen.

We consider each patient chosen under \triangleright and \triangleright' for category \tilde{c} one at a time in sequence with respect to the priority order $\pi_{\tilde{c}}$. Because $\mu(\tilde{I}_c) \subsetneq \tilde{I}_c$, $r_{\tilde{c}}$ patients are matched with \tilde{c} under both precedence orders. Let i_k be the k^{th} patient chosen under \triangleright for \tilde{c} and i'_k be the k^{th} patient chosen under \triangleright' for \tilde{c} , where $k = 1, \dots, r_{\tilde{c}}$. Let J_k be the set of patients available when we process \triangleright for the k^{th} patient and J'_k be the set of patients available when we process \triangleright' for the k^{th} patient, where $k = 1, \dots, r_{\tilde{c}}$. For $k = 1$, $J_k = I(3)$ and $J'_k = I'(3)$. By definition, $J_{k+1} = J_k \setminus \{i_k\}$ and $J'_{k+1} = J'_k \setminus \{i'_k\}$.

We show that

- a. $J_k \setminus J'_k \subseteq \tilde{I} \setminus \tilde{I}_c$,
- b. If $\tilde{c} = u$, $i' \in J'_k \setminus J_k$, $\tau(i') \neq c$, and $i \in J_k$, then $i' \succ i$,
- c. If $\tilde{c} \neq u$, $i' \in J'_k \setminus J_k$, $\tau(i') \neq c$, and $i \in J_k$, then $(J_k \cup J'_k) \cap I_{\tilde{c}} = \emptyset$ and $i' \succ i$

d. If $\tilde{c} \neq u$, then $(J'_k \setminus J_k) \cap I_{\tilde{c}} = \emptyset$.

by mathematical induction on k . These three claims trivially hold for $k = 1$ by Statements (2) and (3) in Lemma C.3.

Fix k . In the inductive step, assume that Statements (a), (b), (c), and (d) hold for k . Consider $k + 1$. If $J_{k+1} = J'_{k+1}$, then the statements trivially hold. Assume that $J_{k+1} \neq J'_{k+1}$, which implies that $J_{\ell} \neq J'_{\ell}$ for $\ell = 1, \dots, k$. There are four cases, depending on which sets i_k and i'_k belong to. We consider each case separately.

Case 1. $i_k \in J_k \setminus J'_k$ and $i'_k \in J'_k \setminus J_k$. When $|\{i' \in J'_k \setminus J_k : \tau(i') \neq c\}| \geq 1$, $i'_k \in J'_k \setminus J_k$ by Statements (b) and (c) for k . Therefore, $|\{i' \in J'_k \setminus J_k : \tau(i') \neq c\}| = 0$. Furthermore,

$$J'_{k+1} \setminus J_{k+1} = (J'_k \setminus \{i'_k\}) \setminus (J_k \setminus \{i_k\}) = J'_k \setminus J_k,$$

and

$$J_{k+1} \setminus J'_{k+1} = (J_k \setminus \{i_k\}) \setminus (J'_k \setminus \{i'_k\}) = ((J_k \setminus J'_k) \setminus \{i_k\}) \cup \{i'_k\}.$$

Because $|J'_k| = |J_k|$ and $J'_k \neq J_k$, we get $|J'_k \setminus J_k| \geq 1$. Therefore, $J'_k \setminus J_k$ has at least one category- c patient because $|\{i' \in J'_k \setminus J_k : \tau(i') \neq c\}| = 0$. Moreover, this patient has higher priority with respect to $\pi_{\tilde{c}}$ than any other category- c patient in $J'_k \cap J_k$ because the former patient is chosen under \triangleright and the latter is not, so i'_k cannot be category c . Therefore, Statement (a) for $k + 1$ follows from Statement (a) for k . Statements (b) and (c) trivially hold for $k + 1$ as well because

$$\{i' \in J'_{k+1} \setminus J_{k+1} : \tau(i') \neq c\} = \{i' \in J'_k \setminus J_k : \tau(i') \neq c\} = \emptyset.$$

Statement (d) for $k + 1$ follows from Statement (d) for k .

Case 2. $i_k \in J_k \setminus J'_k$ and $i'_k \in J'_k \setminus J_k$. Then,

$$J'_{k+1} \setminus J_{k+1} = (J'_k \setminus \{i'_k\}) \setminus (J_k \setminus \{i_k\}) = (J'_k \setminus J_k) \setminus \{i'_k\},$$

and

$$J_{k+1} \setminus J'_{k+1} = (J_k \setminus \{i_k\}) \setminus (J'_k \setminus \{i'_k\}) = (J_k \setminus J'_k) \setminus \{i_k\}.$$

Because $J_{k+1} \setminus J'_{k+1} \subseteq J_k \setminus J'_k$, Statement (a) for $k + 1$ follows from Statement (a) for k . Likewise, Statements (b), (c), and (d) for $k + 1$ follow from the corresponding statements for k because $J'_{k+1} \setminus J_{k+1} \subseteq J'_k \setminus J_k$, $J'_{k+1} \cup J_{k+1} \subseteq J'_k \cup J_k$, and $J_{k+1} \subseteq J_k$.

Case 3. $i_k \in J_k \cap J'_k$ and $i'_k \in J'_k \setminus J_k$. In this case, $i_k = i'_k$, then

$$J'_{k+1} \setminus J_{k+1} = (J'_k \setminus \{i'_k\}) \setminus (J_k \setminus \{i_k\}) = J'_k \setminus J_k,$$

and

$$J_{k+1} \setminus J'_{k+1} = (J_k \setminus \{i_k\}) \setminus (J'_k \setminus \{i'_k\}) = J_k \setminus J'_k.$$

In this case, Statements (a), (b), (c), and (d) for $k + 1$ follow from their respective statements for k .

Case 4. $i_k \in J_k \cap J'_k$ and $i'_k \in J'_k \setminus J_k$. In this case,

$$J'_{k+1} \setminus J_{k+1} = (J'_k \setminus \{i'_k\}) \setminus (J_k \setminus \{i_k\}) = ((J'_k \setminus J_k) \setminus \{i'_k\}) \cup \{i_k\},$$

and

$$J_{k+1} \setminus J'_{k+1} = (J_k \setminus \{i_k\}) \setminus (J'_k \setminus \{i'_k\}) = J_k \setminus J'_k.$$

Statement (a) for $k + 1$ follows from Statement (a) for k trivially.

Statement (b) for $k + 1$ follows from $i_k \succ i$ for any $i \in J_{k+1}$ and also from Statement (b) for k .

To show Statement (d) for $k + 1$, suppose that $\tilde{c} \neq u$. Then by Statement (d) for k , $J'_k \setminus J_k$ does not have a category- \tilde{c} patient.

Because $i'_k \in J'_k \setminus J_k$, this implies that there are no category- \tilde{c} patients in J'_k . Therefore, i_k does not have category \tilde{c} . We conclude that $J'_{k+1} \setminus J_{k+1} = ((J'_k \setminus J_k) \setminus \{i'_k\}) \cup \{i_k\}$ does not have a category \tilde{c} patient, which is the Statement (d) for $k+1$.

To show Statement (c) for $k+1$, suppose that $\tilde{c} \neq u$, $i' \in J'_{k+1} \setminus J_{k+1}$, $\tau(i') \neq c$, and $i \in J_{k+1}$. If $i' \neq i_k$, then $i' \in J'_k \setminus J_k$ because $J'_{k+1} \setminus J_{k+1} = ((J'_k \setminus J_k) \setminus \{i'_k\}) \cup \{i_k\}$ and Statement (c) for $k+1$ follows from Statement (c) for k because $i \in J_{k+1} \subseteq J_k$. Otherwise, suppose that $i' = i_k$. By Statement (d) for $k+1$, i' does not have category \tilde{c} , which implies that there are no category- \tilde{c} patients in J_k ; this in turn implies there are no category- \tilde{c} patients in J_{k+1} because $J_{k+1} \subseteq J_k$. Furthermore, by Statement (d) for $k+1$, there are no category- \tilde{c} patients in $J'_{k+1} \setminus J_{k+1}$. We conclude that there are no category- \tilde{c} patients in $J'_{k+1} \cup J_{k+1}$. Finally, $i' \pi_{\tilde{c}} i$ for any $i \in J_{k+1} = J_k \setminus \{i'\}$ and because there are no category- \tilde{c} patients in $J'_{k+1} \cup J_{k+1}$, we get $i' \pi_{\tilde{c}} i$.

Because $I(4) = J_{r_{\tilde{c}}}$ and $I'(4) = J'_{r_{\tilde{c}}}$, Statement (2) in Lemma C.4 follows from Statement (a) and Statement 3 in Lemma C.4 follows from Statements (b) and (c). \square

Lemma C.5. Suppose that $c \in \mathcal{C} \setminus \{u\}$ and $c', c^*, \tilde{c}, \hat{c} \in \mathcal{C} \setminus \{c\}$ are different categories. Let $\tilde{I} \subseteq I$ and $\triangleright, \triangleright' \in \Delta(\{c, c', c^*, \tilde{c}, \hat{c}\})$ be such that

- \triangleright be such that $c' \triangleright c \triangleright c^* \triangleright \tilde{c} \triangleright \hat{c}$,
- \triangleright' be such that $c \triangleright' c' \triangleright' c^* \triangleright' \tilde{c} \triangleright' \hat{c}$,
- $I(5) = \tilde{I} \setminus \mu(\tilde{I})$, where $\mu = \varphi_{\triangleright}^{\tilde{I}}$,
- $I'(5) = \tilde{I} \setminus \mu'(\tilde{I})$, where $\mu' = \varphi_{\triangleright'}^{\tilde{I}}$, and
- $\mu(\tilde{I}_c) \subsetneq \tilde{I}_c$.

Then, the following results hold:

1. $|I(5) \setminus I'(5)| = |I'(5) \setminus I(5)|$ and
2. $I(5) \setminus I'(5) \subseteq \tilde{I} \setminus \tilde{I}_c$.

Proof. The first statement in Lemma C.5 follows as in the proof of the first statement in Lemma C.1.

To prove the second statement, we use Lemma C.4. Let $\hat{\triangleright}, \hat{\triangleright}' \in \Delta(\{c, c', c^*, \tilde{c}\})$ be such that

$$\hat{\triangleright} : c' \hat{\triangleright} c \hat{\triangleright} c^* \hat{\triangleright} \tilde{c} \quad \text{and} \quad \hat{\triangleright}' : c \hat{\triangleright}' c' \hat{\triangleright}' c^* \hat{\triangleright}' \tilde{c}.$$

Let $I(4) = \tilde{I} \setminus \varphi_{\hat{\triangleright}}^{\tilde{I}}(\tilde{I})$ and $I'(4) = \tilde{I} \setminus \varphi_{\hat{\triangleright}'}^{\tilde{I}}(\tilde{I})$. Then, $I(5) = I(4) \setminus \mu^{-1}(\tilde{c})$ and $I'(5) = I'(4) \setminus \mu'^{-1}(\tilde{c})$.

For both precedence orders \triangleright and \triangleright' under the sequential reserve matching procedure, consider the beginning of the fifth step, at which category \hat{c} is processed. For \triangleright , the set of available patients is $I(4)$. For \triangleright' , the set of available patients is $I'(4)$. If $I(4) = I'(4)$, then we get $I(5) = I'(5)$, which implies all the statements. Therefore, assume that $I(4) \neq I'(4)$. For every precedence order, $r_{\tilde{c}}$ patients with the highest priority with respect to $\pi_{\tilde{c}}$ are chosen.

We consider each patient chosen under \triangleright and \triangleright' for category \hat{c} one at a time in sequence with respect to the priority order $\pi_{\tilde{c}}$. Because $\mu(\tilde{I}_c) \subsetneq \tilde{I}_c$, $r_{\tilde{c}}$ patients are matched with \hat{c} under both precedence orders. Let i_k be the k^{th} patient chosen under \triangleright for \hat{c} and i'_k be the k^{th} patient chosen under \triangleright' for \hat{c} , where $k = 1, \dots, r_{\tilde{c}}$. Let J_k be the set of patients available when we process \triangleright for the k^{th} patient and J'_k be the set of patients available when we process \triangleright' for the k^{th} patient, where $k = 1, \dots, r_{\tilde{c}}$. For $k=1$, $J_k = I(4)$ and $J'_k = I'(4)$. By definition, $J_{k+1} = J_k \setminus \{i_k\}$ and $J'_{k+1} = J'_k \setminus \{i'_k\}$.

We show that

- a. $J_k \setminus J'_k \subseteq \tilde{I} \setminus \tilde{I}_c$ and
- b. if $i' \in J'_k \setminus J_k$ and $i \in J_k \cap \tilde{I}_c$, then $i' \pi_{\tilde{c}} i$,

by mathematical induction on k . These results trivially hold for $k=1$ by Statements (2), (3), and (4) in Lemma C.4.

Fix k . In the inductive step, assume that Statements (a) and (b) hold for k . Consider $k+1$. If $J_{k+1} = J'_{k+1}$, then the statements trivially hold. Assume that $J_{k+1} \neq J'_{k+1}$, which implies that $J_{\ell} \neq J'_{\ell}$ for $\ell = 1, \dots, k$. There are four cases depending on which sets i_k and i'_k belong to. We consider each case separately.

Case 1. $i_k \in J_k \setminus J'_k$ and $i'_k \in J'_k \cap J_k$. Then,

$$J'_{k+1} \setminus J_{k+1} = (J'_k \setminus \{i'_k\}) \setminus (J_k \setminus \{i_k\}) = J'_k \setminus J_k,$$

and

$$J_{k+1} \setminus J'_{k+1} = (J_k \setminus \{i_k\}) \setminus (J'_k \setminus \{i'_k\}) = (J_k \setminus J'_k) \setminus \{i_k\} \cup \{i'_k\}.$$

If $i'_k \in \tilde{I}_c$, then we get a contradiction to Statement (b) for k . Therefore, $i'_k \notin \tilde{I}_c$, which implies that Statement (a) holds for $k+1$ by Statement (a) for k and the second displayed equation. Statement (b) for $k+1$ follows trivially from Statement (b) for k , the first displayed equation, and $J_k \supseteq J_{k+1}$.

Case 2. $i_k \in J_k \setminus J'_k$ and $i'_k \in J'_k \setminus J_k$. Then,

$$J'_{k+1} \setminus J_{k+1} = (J'_k \setminus \{i'_k\}) \setminus (J_k \setminus \{i_k\}) = (J'_k \setminus J_k) \setminus \{i'_k\}$$

And,

$$J_{k+1} \setminus J'_{k+1} = (J_k \setminus \{i_k\}) \setminus (J'_k \setminus \{i'_k\}) = (J_k \setminus J'_k) \setminus \{i_k\}.$$

In this case, Statements (a) and (b) for $k+1$ follow trivially from the corresponding statements for k .

Case 3. $i_k \in J_k \cap J'_k$ and $i'_k \in J'_k \cap J_k$. In this case, $i_k = i'_k$, then

$$J'_{k+1} \setminus J_{k+1} = (J'_k \setminus \{i'_k\}) \setminus (J_k \setminus \{i_k\}) = J'_k \setminus J_k,$$

and

$$J_{k+1} \setminus J'_{k+1} = (J_k \setminus \{i_k\}) \setminus (J'_k \setminus \{i'_k\}) = J_k \setminus J'_k.$$

In this case, Statements (a) and (b) for $k+1$ follow trivially from the corresponding statement for k .

Case 4. $i_k \in J_k \cap J'_k$ and $i'_k \in J'_k \setminus J_k$. In this case,

$$J'_{k+1} \setminus J_{k+1} = (J'_k \setminus \{i'_k\}) \setminus (J_k \setminus \{i_k\}) = (J'_k \setminus J_k) \setminus \{i'_k\} \cup \{i_k\},$$

and

$$J_{k+1} \setminus J'_{k+1} = (J_k \setminus \{i_k\}) \setminus (J'_k \setminus \{i'_k\}) = J_k \setminus J'_k.$$

Then, Statement (a) for $k+1$ follows from Statement (a) for k . Furthermore, $i_k \pi_{\tilde{c}} i$ for any $i \in J_{k+1}$, which together with Statement (b) for k imply Statement (b) for $k+1$.

Because $I(5) = J_{r_{\tilde{c}}}$ and $I'(5) = J'_{r_{\tilde{c}}}$, Statement 2 in Lemma C.5 follows from Statement (a). \square

Proof of Proposition 4. Let $|\mathcal{C}| \leq 5$. Let $\mathcal{C}^* = \{c^* \in \mathcal{C} : c^* \triangleright c'\}$ be the set of categories processed before c' under \triangleright and before c under \triangleright' . The orders of categories in \mathcal{C}^* are the same with respect to \triangleright and \triangleright' . Thus, just before category c' is processed under \triangleright and c is processed under \triangleright' , the same patients are matched in both sequential reserve matching procedures. Let \tilde{I} be the set of patients that are available at this point in either procedure.

Let $\hat{\triangleright}$ be the incomplete precedence order on $\mathcal{C} \setminus \mathcal{C}^*$ that processes categories in the same order as in \triangleright . Likewise, let $\hat{\triangleright}'$ be

the incomplete precedence order on $\mathcal{C} \setminus \mathcal{C}^*$ that processes categories in the same order as in \triangleright' .

If $\varphi_{\triangleright}'(\tilde{I}_c) = \tilde{I}_c$, then the result is proven. Therefore, assume that $\varphi_{\triangleright}'(\tilde{I}_c) \subsetneq \tilde{I}_c$ in the rest of the proof. Let $k = |\mathcal{C} \setminus \mathcal{C}^*|$ be the number of remaining categories.

• If $k=2$, then by Lemmas C.1 and C.2, we obtain $\varphi_{\triangleright}'(\tilde{I}_c) \subseteq \varphi_{\triangleright}'(\tilde{I}_c)$.

• If $k=3$, then by Lemma C.3, we obtain $\varphi_{\triangleright}'(\tilde{I}_c) \subseteq \varphi_{\triangleright}'(\tilde{I}_c)$.

• If $k=4$, then by Lemma C.4, we obtain $\varphi_{\triangleright}'(\tilde{I}_c) \subseteq \varphi_{\triangleright}'(\tilde{I}_c)$.

• If $k=5$, then by Lemma C.5, we obtain $\varphi_{\triangleright}'(\tilde{I}_c) \subseteq \varphi_{\triangleright}'(\tilde{I}_c)$.

These imply that

$$\varphi_{\triangleright'}(\tilde{I}_c) \subseteq \varphi_{\triangleright}(\tilde{I}_c),$$

completing the proof. \square

C.5. Proofs of Results in Appendix B

Before providing proofs for the results regarding the **Smart Reserve Matching Algorithm**, we state an equivalent, descriptive (but exponential-in-time) **Smart Reserve Matching Algorithm***, which uses an iterative sequence of matching sets $\mathcal{M}_0, \mathcal{M}_1, \dots, \mathcal{M}_{|I|}$ and enables us to prove our results more efficiently. Our algorithm in Appendix B does not require constructing these sets, and hence, it is polynomial in time.

Smart Reserve Matching Algorithm*

Fix a parameter $n \in \{0, 1, \dots, r_u\}$ that represents the number of unreserved units to be processed in the beginning of the algorithm.²⁰ The remaining unreserved units are to be processed at the end of the algorithm.

Fix a baseline priority order π , and for the ease of description, relabel patients so that

$$i_1 \pi i_2 \pi \dots \pi i_{|I|}.$$

Proceed in two steps.

Step 1. Iteratively construct two sequences of patient sets $J_0^u \subseteq J_1^u \subseteq \dots \subseteq J_{|I|}^u$, which determine patients to be matched with the unreserved category u in this step, and $J_0 \subseteq J_1 \subseteq \dots \subseteq J_{|I|}$, which determine the patients to be matched with preferential treatment categories in $\mathcal{C} \setminus \{u\}$ that they are beneficiaries of, and a sequence of sets of matchings $\mathcal{M}_0 \supseteq \mathcal{M}_1 \supseteq \dots \supseteq \mathcal{M}_{|I|}$ in $|I|$ substeps.

Define Step 1.(k) for any $k \in \{1, 2, \dots, |I|\}$ as follows:

If $k=1$, let

$$J_0^u = \emptyset, \quad J_0 = \emptyset,$$

and \mathcal{M}_0 be the set of all matchings that are maximal in beneficiary assignment; that is,

$$\mathcal{M}_0 = \arg \max_{v \in \mathcal{M}} \left| \bigcup_{c \in \mathcal{C} \setminus \{u\}} (v^{-1}(c) \cap I_c) \right|.$$

If $k > 1$, then sets of patients J_{k-1}^u and J_{k-1} and set of matchings \mathcal{M}_{k-1} are constructed in the previous substep, Step 1. ($k-1$).

Step 1.(k). Process patient i_k . Three cases are possible:

- If $|J_{k-1}^u| < n$ and there exists a matching in \mathcal{M}_{k-1} that matches patient i_k with the unreserved category

u , then define

$$J_k^u = J_{k-1}^u \cup \{i_k\}, \quad J_k = J_{k-1}, \quad \text{and}$$

$$\mathcal{M}_k = \{\mu \in \mathcal{M}_{k-1} : \mu(i_k) = u\}.$$

- Otherwise:

- If there exists a matching in \mathcal{M}_{k-1} that matches patient i_k with a preferential treatment category $c \in \mathcal{C} \setminus \{u\}$ that she is a beneficiary of—that is, $i_k \in I_c$ —then define

$$J_k^u = J_{k-1}^u, \quad J_k = J_{k-1} \cup \{i_k\}, \quad \text{and}$$

$$\mathcal{M}_k = \{\mu \in \mathcal{M}_{k-1} : \mu(i_k) \notin \{\emptyset, u\} \text{ and } i_k \in I_{\mu(i_k)}\}.$$

- Otherwise, define

$$J_k^u = J_{k-1}^u, \quad J_k = J_{k-1}, \quad \text{and} \quad \mathcal{M}_k = \mathcal{M}_{k-1}.$$

Step 2. For any matching $\mu \in \mathcal{M}_{|I|}$, construct a matching $\sigma \in \mathcal{M}$ as follows:

- Assign $\mu(i)$ to every patient $i \in J_{|I|} \cup J_{|I|}^u$.
- One at a time, iteratively assign the remaining units to the remaining highest-priority patient in $I \setminus (J_{|I|}^u \cup J_{|I|})$ who is eligible for the category of the assigned unit in the following order:

1. The remaining units of the preferential treatment categories in an arbitrary order, and
2. The remaining units of the unreserved category u .

Proof of Lemma B.1. By the **Smart Reserve Matching Algorithm*** induced by assigning n unreserved units subsequently at the beginning, in Step 1.(k) for every $k \in \{0, 1, \dots, |I|\}$, and for every matching $\mu \in \mathcal{M}_k$,

- $\mu(i) = u$ for every $i \in J_k^u$, and
- $\mu(i) \notin \{u, \emptyset\}$ and $i \in I_{\mu(i)}$ for every $i \in J_k$.

We show that for any $i \in I \setminus (J_{|I|}^u \cup J_{|I|})$, there is no matching $\mu \in \mathcal{M}_{|I|}$ such that $\mu(i) \notin \{u, \emptyset\}$ and $i \in I_{\mu(i)}$. Suppose contrary to the claim that such a patient i and matching μ exist. Patient i is processed in some Step 1.(k). We have $i \notin J_k^u \cup J_k \subseteq J_{|I|}^u \cup J_{|I|}$. We have $\mu \in \mathcal{M}_{k-1}$ as $\mathcal{M}_{k-1} \supseteq \mathcal{M}_{|I|}$. Then, $|J_{k-1}^u| = n$, as otherwise we can always match i with u , even if we cannot match her with a preferential treatment category that she is a beneficiary of when she is processed under a matching that is maximal in beneficiary assignment, contradicting $i \notin J_{|I|}^u \cup J_{|I|}$. But then, as $\mu(i) \notin \{u, \emptyset\}$ and $i \in I_{\mu(i)}$, we have $\mu \in \mathcal{M}_k$ and $i \in J_k$, contradicting again $i \notin J_{|I|}^u \cup J_{|I|}$.

Thus, in Step 2, no patient is matched with a preferential treatment categories that she is a beneficiary of. These prove $\cup_{c \in \mathcal{C} \setminus \{u\}} (\sigma^{-1}(c) \cap I_c)$ is the same set, regardless of the matching $\sigma \in \mathcal{M}_S^n$ we choose.

To prove that $\sigma^{-1}(u)$ is the same for every $\sigma \in \mathcal{M}_S^n$, we consider two cases (for Step 2):

- If we have a soft reserve system: Then, all patients are eligible for the remaining $\max\{0, q - |J_{|I|}^u \cup J_{|I|}| \}$ units. Because we assign the patients in $I \setminus (J_{|I|}^u \cup J_{|I|})$ based on priority according to π to the remaining units, we have $\sigma(I)$ is the same patient set regardless of the matching $\sigma \in \mathcal{M}_S^n$ we choose. Because the remaining $r_u - n$ unreserved units are assigned to the lowest-priority patients that are matched in any $\sigma \in \mathcal{M}_S^n$, if any, and $\mu^{-1}(u)$ is a fixed set, regardless of which matching

$\mu \in \mathcal{M}_{|I|}$ we choose (as we proved above), then $\sigma^{-1}(u)$ is the same set regardless of which $\sigma \in \mathcal{M}_S^n$ we choose.

- If we have a hard reserve system: Then, in Step 2, no patient is matched with a unit of a category that she is not a beneficiary of. Thus, the $r_u - n$ remaining unmatched units are assigned to the highest π -priority patients in $I \setminus (J_{|I|}^u \cup J_{|I|})$. This concludes proving that each of $\sigma(I)$ and $\sigma^{-1}(u)$ is the same patient set regardless of which $\sigma \in \mathcal{M}_S^n$ we choose. \square

Proof of Proposition B.1. For any $n \in \{0, 1, \dots, r_u\}$, we prove that every smart reserve matching in \mathcal{M}_S^n complies with eligibility requirements, is nonwasteful, respects priorities, and is maximal in beneficiary assignment.

Compliance with eligibility requirements: By construction, no patient is ever matched with a category for which she is not eligible during the procedure.

Nonwastefulness: Suppose to the contrary of the claim that there exists some $\sigma \in \mathcal{M}_S^n$ that is wasteful. Thus, there exists some category $c \in \mathcal{C}$ and a patient $i \in I$ such that $\sigma(i) = \emptyset$, $i \pi_c \emptyset$, and $|\mu^{-1}(c)| < r_c$. Then, in Step 2, patient i or another patient should have been matched with c , as we assign all remaining units to eligible patients, which is a contradiction.

Respect for priorities: Let $\sigma \in \mathcal{M}_S^n$ be a smart reserve matching. Suppose patients $i, j \in I$ are such that $i \pi j$ and $\sigma(i) = \emptyset$. We need to show either (i) $\sigma(j) = \emptyset$ or (ii) $i \notin I_{\sigma(j)}$ and $j \in I_{\sigma(j)}$, which is equivalent to $j \pi_{\sigma(j)} i$. Suppose $\sigma(j) \neq \emptyset$. Suppose to the contrary that $i \in I_{\sigma(j)}$ and $j \in I_{\sigma(j)}$. Consider the smart reserve matching procedure with n . Two cases are possible: $j \in J_{|I|}^u \cup J_{|I|}$ or not. We show that either case leads to a contradiction, showing that σ respects priorities.

• If $j \in J_{|I|}^u \cup J_{|I|}$: Consider the matching $\hat{\sigma}$ obtained from σ as follows: $\hat{\sigma}(i) = \sigma(j)$, $\hat{\sigma}(j) = \emptyset$, and $\hat{\sigma}(i') = \sigma(i')$ for every $i' \in I \setminus \{i, j\}$. Because $i, j \in I_{\sigma(j)}$, and we match i instead of j with $\sigma(j)$, $\hat{\sigma}$ is a matching that is maximal in beneficiary assignment as well. Because $i \pi j$, i is processed before j in Step 1. Let i be processed in some Step 1.(k). Because $\sigma(i) = \emptyset$, $i \notin J_k \cup J_k^u$. Then, $\hat{\sigma} \in \mathcal{M}_{k-1}$ as $\sigma \in \mathcal{M}_{k-1}$. Two cases are possible:

- If $\sigma(j) = u$: As j is matched with an unreserved unit in Step 1, then an unreserved unit is still to be allocated in the procedure when i is to be processed in Step 1.(k) before j . We try to match i with an unreserved unit first. Because $\hat{\sigma} \in \mathcal{M}_{k-1}$ and $\hat{\sigma}(i) = u$, this implies $i \in J_k^u$. This contradicts $i \notin J_k^u \cup J_k$.

- If $\sigma(j) \neq u$: Then, when i is to be processed in Step 1.(k), we are trying to match her (i) if it is possible, with unreserved category u first and if not, with a preferential treatment category that she is a beneficiary of; or (ii) directly with a preferential treatment category that she is a beneficiary of without sacrificing the maximality in beneficiary assignment. However, as $\sigma(i) = \emptyset$, we failed in doing either. Because $i \in I_{\hat{\sigma}}(i)$ and $\hat{\sigma}(i) \neq u$, at least there exists a matching in \mathcal{M}_{k-1} that would match i with a preferential treatment category that she is a beneficiary of. Hence, this contradicts $i \notin J_k^u \cup J_k$.

• If $j \notin J_{|I|}^u \cup J_{|I|}$: Therefore, j is matched in Step 2 of the **Smart Reserve Matching Algorithm*** with n unreserved units processed first. Because we match patients in Step 2 either with the preferential treatment categories that they are eligible but not beneficiary of or with the unreserved category u , and we assumed $j \in I_{\sigma(j)}$, then $\sigma(j) = u$. Because i is also available when j is matched, and $i \pi j$, patient i or another patient who

has higher π -priority than j should have been matched instead of j , which is a contradiction.

Maximality in beneficiary assignment: By construction $\mathcal{M}_S^n \subseteq \mathcal{M}_0$, which is the set of matchings that are maximal in beneficiary assignment in the smart reserve matching procedure with n . \square

The following lemma and concepts from graph theory will be useful in our next proof. We state the lemma as follows:

Lemma C.6 (Mendelsohn and Dulmage (1958) Theorem). *Let \mathcal{M}^b be the set of matchings that match patients with only preferential treatment categories that they are beneficiaries of and otherwise leave them unmatched. If there is a matching in \mathcal{M}^b that matches patients in some $J \subseteq I$ and there is another matching $v \in \mathcal{M}^b$, then there exists a matching in \mathcal{M}^b that matches all patients in J and at least $|v^{-1}(c)|$ units of each category $c \in \mathcal{C} \setminus \{u\}$.*

See, for example, Schrijver (2003, p. 266) for a proof of this result.

Let us define $I_\emptyset = \emptyset$ for notational convenience.

We define a *beneficiary alternating path* from μ to ν for two matchings in $\mu, \nu \in \mathcal{M}$ as a nonempty list $A = (i_1, \dots, i_m)$ of patients such that

$$\begin{aligned} & [i_1 \notin I_{\mu(i_1)} \text{ or } \mu(i_1) = u] \quad \& \quad [i_1 \in I_{\nu(i_1)} \text{ and } \nu(i_1) \neq u], \\ & [i_m \in I_{\mu(i_m)} \text{ and } \mu(i_m) = \nu(i_{m-1})] \quad \& \quad [i_m \in I_{\nu(i_m)} \text{ and } \nu(i_m) \neq u] \\ & \quad \quad \quad \text{for every } m \in \{2, 3, \dots, \bar{m} - 1\}, \\ & [i_{\bar{m}} \in I_{\mu(i_{\bar{m}})} \text{ and } \mu(i_{\bar{m}}) = \nu(i_{\bar{m}-1})] \quad \& \quad [i_{\bar{m}} \notin I_{\nu(i_{\bar{m}})} \text{ or } \nu(i_{\bar{m}}) = u]. \end{aligned}$$

A beneficiary alternating path begins with a patient i_1 , who is not matched with a preferential treatment category that she is a beneficiary of under μ , and ends with a patient $i_{\bar{m}}$, who is not matched with a preferential treatment category that she is a beneficiary of under ν . Everybody else in the path is matched under both matchings with a preferential treatment category that she is a beneficiary of. We state the following observation, which directly follows from the finiteness of categories and patients.

Observation C.1. *If μ and $\nu \in \mathcal{M}$ are two matchings such that for every $c \in \mathcal{C} \setminus \{u\}$, $|\mu^{-1}(c) \cap I_c| = |\nu^{-1}(c) \cap I_c|$ and there exists some $i \in (\cup_{c \in \mathcal{C} \setminus \{u\}} \nu^{-1}(c) \cap I_c) \setminus (\cup_{c \in \mathcal{C} \setminus \{u\}} \mu^{-1}(c) \cap I_c)$, then there exists a beneficiary alternating path from μ to ν beginning with patient i .*

We are ready to prove our last theorem:

Proof of Theorem B.1. By Proposition B.1 and Theorem 1, any $\sigma_0 \in \mathcal{M}_S^0$ and $\sigma_{r_u} \in \mathcal{M}_S^{r_u}$ are cutoff equilibrium matchings. Let $\mu \in \mathcal{M}$ be any other cutoff equilibrium matching that is maximal in beneficiary assignment.

We extend the definitions of our concepts to smaller economies: Given any $I^* \subseteq I$ and $r^* = (r_c^*)_{c \in \mathcal{C}}$ such that $r_c^* \leq r_c$ for every $c \in \mathcal{C}$, all properties and algorithms are redefined for this smaller economy (I^*, r^*) by taking the restriction of the baseline priority order π on I^* , and denoted using the argument (I^*, r^*) at the end of the notation. For example, $\mathcal{M}(I^*, r^*)$ denotes the set of matchings for (I^*, r^*) .

Proof of $\bar{f}_u^{\sigma_{r_u}} \pi \bar{f}_u^\mu$: We prove the following claim first.

Claim 1. For any set of patients $I^* \subseteq I$ and any capacity vector $r^* \leq r$, suppose matching $\nu \in \mathcal{M}(I^*, r^*)$ is maximal in beneficiary assignment for (I^*, r^*) . Let $\sigma \in \mathcal{M}_S^u(I^*, r^*)$ be a smart reserve matching with all unreserved units processed first. Then,

$$|\sigma^{-1}(u)| \geq |\nu^{-1}(u)|.$$

Moreover, according to the baseline priority order π , for any $k \in \{1, \dots, |\nu^{-1}(u) \setminus \sigma^{-1}(u)|\}$, let j_k be the k^{th} highest-priority patient in $\sigma^{-1}(u) \setminus \nu^{-1}(u)$ and j'_k be the k^{th} highest-priority patient in $\nu^{-1}(u) \setminus \sigma^{-1}(u)$, then,

$$j_k \pi j'_k.$$

Proof. Suppose to the contrary of the first statement $|\sigma^{-1}(u)| < |\nu^{-1}(u)|$. Because both σ and ν are maximal in beneficiary assignment for (I^*, r^*) , then there exists some patient $i \in (\cup_{c \in \mathcal{C}} \nu^{-1}(c) \cap I_c) \setminus (\cup_{c \in \mathcal{C}} \sigma^{-1}(c) \cap I_c)$. This patient is not committed to be matched in Step 1 of the **Smart Reserve Matching Algorithm*** with all unreserved units first, despite the fact that there exists at least one available unreserved unit when she was processed, which is a contradiction. Thus, $|\sigma^{-1}(u)| \geq |\nu^{-1}(u)|$.

For the rest of the proof, we use induction on the cardinality of I^* and on the magnitude of vector of category capacity vector r^* :

- For the base case when $I^* = \emptyset$ and $r_c^* = 0$ for every $c \in \mathcal{C}$, the claim holds trivially.
- As the inductive assumption, suppose that for all capacity vectors of categories bounded above by vector r^* and all subsets of I bounded above by cardinality k^* , the claim holds.
- Consider a set of patients $I^* \subseteq I$ such that $|I^*| = k^*$ and capacity vector of categories $r^* = (r_c^*)_{c \in \mathcal{C}}$. Let $\sigma \in \mathcal{M}_S^u(I^*, r^*)$ be a smart reserve matching for (I^*, r^*) with all unreserved units processed first and $\nu \in \mathcal{M}(I^*, r^*)$ be maximal in beneficiary assignment for (I^*, r^*) . If $\sigma^{-1}(u) \supseteq \nu^{-1}(u)$, then the claim for (I^*, r^*) is trivially true. Thus, suppose not. Then, there exists $j \in \nu^{-1}(u) \setminus \sigma^{-1}(u)$. Moreover, let j be the highest π -priority patient in $\nu^{-1}(u) \setminus \sigma^{-1}(u)$. We have two cases that we consider separately:

Case 1. There is no patient $i \in I^*$ such that $i \pi j$ and $i \in \sigma^{-1}(u) \setminus \nu^{-1}(u)$:

We show that this case leads to a contradiction, and, hence, it cannot hold.

When j is processed in Step 1 of the **Smart Reserve Matching Algorithm*** with all unreserved units processed first, because $\sigma(j) \neq u$, either

- All units of the unreserved category are assigned under σ to patients with higher π -priority than j , or
- Some unreserved category units are still available when j is processed.

Observe that (i) cannot hold, because it contradicts Case 1. Thus, (ii) holds.

Because j is the highest π -priority patient in $\nu^{-1}(u) \setminus \sigma^{-1}(u)$ and because we are in Case 1, for every $i \in I^*$ such that $i \pi j$, we have

$$\nu(i) = u \Leftrightarrow \sigma(i) = u. \quad (C.1)$$

We construct a new matching $\hat{\nu} \in \mathcal{M}(I^*, r^*)$ from ν and fix a patient $i \in I^*$ as follows. We check whether there exists a

patient $i \in I^*$ such that

$$i \pi j, \quad \sigma(i) \neq u, \quad \text{and} \quad i \notin I_{\nu(i)}, \quad (C.2)$$

- If such a patient i does not exist, then let $\hat{\nu} = \nu$ and $i = j$.
- If such a patient i exists, then let her be the highest π -priority patient with the property in Equation (C.2).

We show that $i \in I_{\sigma(i)}$. Consider the **Smart Reserve Matching Algorithm***. Because $\nu(j) = u$ and $i \pi j$, just before i is processed in Step 1, there is still at least one unreserved unit available by Equation (C.1). Because we are processing all unreserved units first and because $\sigma(i) \neq u$, it should be the case that we had to match i with a preferential treatment category that she is a beneficiary of. Thus, $i \in I_{\sigma(i)}$.

We create a new matching for (I^*, r^*) from ν , which we refer to as $\hat{\nu}$, such that $\hat{\nu}$ matches every patient in I^* exactly as under ν , except that $\hat{\nu}$ leaves patient j unmatched and matches i with category u instead. Because ν is maximal in beneficiary assignment for (I^*, r^*) , so is $\hat{\nu}$.

So far, we have for every $i' \in I^*$ such that $i' \pi i$,

1. $\sigma(i') = u \Leftrightarrow \nu(i') = u$ (by Equation (C.1) and $i \pi j$),
2. $\sigma(i') \in I_{\sigma(i')}$ (an unreserved unit is available before i is processed in Step 1 of the **Smart Reserve Matching Algorithm*** with all unreserved units processed first; thus, every patient processed before i is matched, if possible, with u , and, if not possible, with a preferential treatment category that she is a beneficiary of under the restriction of maximality in beneficiary assignment), and
3. $\hat{\nu}(i') \in I_{\hat{\nu}}(i')$ (by definition of i as the highest π -priority patient satisfying Equation (C.2)).

We also have $\hat{\nu}(i) = u$ and $\sigma(i) \neq u$.

Patient i is processed in some Step 1.(k) in the **Smart Reserve Matching Algorithm*** with all unreserved units processed first. As $\sigma(i) \neq u$, we have $i \notin J_k^u(I^*, r^*)$. On the other hand, because $\sigma \in \mathcal{M}_{k-1}(I^*, r^*)$, by Statements 1, 2, and 3 above, we have $\hat{\nu} \in \mathcal{M}_{k-1}(I^*, r^*)$ as well, and it matches i with u , contradicting $i \notin J_k^u(I^*, r^*)$. Therefore, Case 1(ii) cannot hold either.

Case 2. There is some $i \in \sigma^{-1}(u) \setminus \nu^{-1}(u)$ such that $i \pi j$:

Construct a matching $\hat{\sigma}$ from σ that it leaves every patient who is matched in Step 2 of the **Smart Reserve Matching Algorithm*** with all unreserved units processed first: For every $i^* \in I^*$, $\hat{\sigma}(i^*) = \sigma(i^*)$ if $i^* \in I_{\sigma(i)}$ and $\hat{\sigma}(i^*) = \emptyset$ otherwise. Clearly, $\hat{\sigma} \in \mathcal{M}(I^*, r^*)$ and is maximal in beneficiary assignment for (I^*, r^*) , because σ is. By Lemma C.6, there exists a matching $\tilde{\nu} \in \mathcal{M}^b(I^*, r^*)^{21}$ such that under $\tilde{\nu}$, all patients in $\cup_{c \in \mathcal{C} \setminus \{u\}} \tilde{\nu}^{-1}(c) \cap I_c$ are matched with the preferential treatment categories in $\mathcal{C} \setminus \{u\}$ that they are beneficiaries of, and, for every $c \in \mathcal{C} \setminus \{u\}$, $|\tilde{\nu}^{-1}(c) \cap I_c| = |\hat{\sigma}^{-1}(c) \cap I_c|$ (equality follows rather than \geq as dictated by the lemma, because $\hat{\sigma}$ is maximal in beneficiary assignment for (I^*, r^*)). For every $i^* \in I^*$, we have $\nu(i^*) = u \Rightarrow \tilde{\nu}(i^*) = \emptyset$ as $\hat{\nu} \in \mathcal{M}^b(I^*, r^*)$. We modify $\tilde{\nu}$ to obtain $\hat{\nu}$: For every $i^* \in \nu^{-1}(u)$, we set $\hat{\nu}(i^*) = u$, and for every $i^* \in I^* \setminus \nu^{-1}(u)$, we set $\hat{\nu}(i^*) = \tilde{\nu}(i^*)$. Clearly, $\hat{\nu} \in \mathcal{M}(I^*, r^*)$ and is maximal in beneficiary assignment for (I^*, r^*) , because ν is. We will work with $\hat{\sigma}$ and $\hat{\nu}$ instead of σ and ν from now on.

Recall that $\hat{\sigma}(i) = u$ and $\hat{\nu}(i) \neq u$. Two cases are possible: $i \in I_{\hat{\nu}}(i)$ or $\hat{\nu}(i) = \emptyset$.

1. If $i \in I_{\hat{\nu}}(i)$: Then, by Observation C.1, there exists a beneficiary alternating path A from $\hat{\sigma}$ to $\hat{\nu}$ beginning with i and

ending with some $i' \in I^*$ such that (i) $\hat{\sigma}(i') \in I_{\hat{\sigma}}(i')$ and $\hat{\sigma}(i') \neq u$, and (ii) $\hat{v}(i') = u$ or $\hat{v}(i') = \emptyset$.

By the existence of the beneficiary alternating path, it is possible to match either i or i' with a preferential treatment category that she is a beneficiary of and match the other one with u without changing the type of match of any other patient $i^* \in I^* \setminus \{i, i'\}$ has—that is, either i^* is matched with a preferential treatment category under both matchings or not. Yet, when i is processed in Step 1 of the **Smart Reserve Matching Algorithm*** with all unreserved units processed first, we chose i to be matched with u and i' with a preferential treatment category. This means

$$i \pi i'.$$

Let $\hat{c} \neq u$ be the category that i is matched with under \hat{v} .

If $\hat{\sigma}(i') \neq u$, then modify \hat{v} by assigning an unreserved unit to i' instead of j : $\hat{v}(j) = \emptyset$ and $\hat{v}(i') = u$. Otherwise, do not modify \hat{v} any further.

Consider the smaller economy (I', r') such that $I' = I^* \setminus \{i, i'\}$ and for every $c \in \mathcal{C}$, $r'_c = r_c^* - 1$ if $c \in \{\hat{c}, u\}$ and $r'_c = r_c^*$, otherwise.

We show that a smart reserve matching $\sigma' \in \mathcal{M}_S^{r_u'}(I', r')$ can be obtained from the original smart reserve matching $\sigma \in \mathcal{M}_S^{r_u}(I^*, r^*)$ and \hat{v} . Consider the beneficiary alternating path A we discovered above, starting with patient i and ending with patient i' from $\hat{\sigma}$ to \hat{v} : Suppose $A = (i, i_2, \dots, i_{m-1}, i')$. Define

$$\sigma'(i^*) = \sigma(i^*) \text{ for every } i^* \in I' \setminus \{i_2, \dots, i_{m-1}\} \text{ and}$$

$$\sigma'(i^*) = \hat{v}(i^*) \text{ for every } i^* \in \{i_2, \dots, i_{m-1}\}.$$

Observe that $\sigma' \in \mathcal{M}(I', r')$. The existence of σ' shows that it is feasible to match every patient in $J_{|I'|}^u(I^*, r^*) \setminus \{i\}$ with u , and it is feasible to match every patient in $J_{|I'|}(\hat{v}) \setminus \{i'\}$ with a preferential treatment category that she is a beneficiary of in (I', r') . Thus, the **Smart Reserve Matching Algorithm*** with all unreserved units processed first proceeds exactly in the same manner as it does for (I^*, r^*) , with the exception that it skips patients i and i' in the smaller economy (I', r') . Hence, $\sigma' \in \mathcal{M}_S^{r_u'}(I', r')$.

Let the restriction of matching \hat{v} to (I', r') be v' . Observe that v' is a matching for (I', r') . Moreover, it is maximal in beneficiary assignment for (I', r') , because \hat{v} is maximal in beneficiary assignment for (I^*, r^*) .

Now, one of the two following cases holds for v' :

- a. If $v'(i') \neq u$: Recall that while $v(j) = u$, we updated \hat{v} so that $\hat{v}(i') = u$ and $\hat{v}(j) = \emptyset$. Thus, $v'(j) = \emptyset$ as well. Because $i \pi j$, this together with the inductive assumption that the claim holds for (I', r') imply that the claim also holds for (I^*, r^*) , completing the induction.
- b. If $v'(i') = u$: Because $i \pi i'$, this together with the inductive assumption that the claim holds for (I', r') imply that the claim also holds for (I^*, r^*) , completing the induction.
- 2. If $\hat{v}(i) = \emptyset$: Recall that $\hat{v}(j) = u$. We modify \hat{v} further that $\hat{v}(i) = u$ and $\hat{v}(j) = \emptyset$. Consider the smaller economy (I', r') , where $I' = I^* \setminus \{i\}$, $r'_u = r_u^* - 1$, and $r'_c = r_c^*$ for every $c \in \mathcal{C} \setminus \{u\}$.

Let σ' and v' be the restrictions of σ and \hat{v} to (I', r') , respectively. Because $\sigma(i) = \hat{v}(i) = u$ both σ' and v' are matchings for (I', r') . Because the capacity of category u is decreased by one, σ' is a smart reserve matching with all unreserved units processed first for (I', r') . To see this, observe that the algorithm

proceeds as it does for (I^*, r^*) , with the exception that it skips i . Matching v' is maximal in beneficiary assignment for (I', r') . Therefore, by the inductive assumption, the claim holds for (I', r') . This together with the fact that $i \pi j$ imply the claim holds for (I^*, r^*) , completing the induction. \diamond

If $|\mu^{-1}(u)| < r_u$, then $\bar{f}_u^{\sigma_{r_u}} \underline{\pi} \bar{f}_u^{\mu} = \emptyset$. On the other hand, if $|\mu^{-1}(u)| = r_u$, Claim 1 implies that $\bar{f}_u^{\sigma_{r_u}} = \min_{\pi} \sigma_{r_u}^{-1}(u) \underline{\pi} \bar{f}_u^{\mu} = \min_{\pi} \mu^{-1}(u)$.

Proof of $\bar{f}_u^{\mu} \underline{\pi} \bar{f}_u^{\sigma_0}$. We prove the following claim first.

Claim 2. For any set of patients $I^* \subseteq I$ and any capacity vector $r^* \leq r$, suppose $v \in \mathcal{M}(I^*, r^*)$ is a matching that is maximal in beneficiary assignment for (I^*, r^*) . Let $\sigma \in \mathcal{M}_S^0(I^*, r^*)$ be a smart reserve matching with all unreserved units processed last,

$$J = \cup_{c \in \mathcal{C} \setminus \{u\}} (\sigma^{-1}(c) \cap I_c), \quad \text{and} \quad J' = \cup_{c \in \mathcal{C} \setminus \{u\}} (v^{-1}(c) \cap I_c).$$

According to the baseline priority order π , for any $k \in \{1, \dots, |J'| \setminus |J|\}$, let j_k be the k^{th} highest-priority patient in $J \setminus J'$ and j'_k be the k^{th} highest-priority patient in $J' \setminus J$, then

$$j_k \pi j'_k.$$

Proof. We use induction on the cardinality of I^* and on the magnitude of vector of capacities of categories r^* :

- For the base case when $I^* = \emptyset$ and $r_c^* = 0$ for every $c \in \mathcal{C}$, the claim holds trivially.
- In the inductive step, suppose for every capacity of categories bounded above by vector r^* and subsets of patients in I bounded above by cardinality k^* , the claim holds.
- Consider set of patients $I^* \subseteq I$ such that $|I^*| = k^*$ and capacity vector for categories $r^* = (r_c^*)_{c \in \mathcal{C}}$. If $J \setminus J' = \emptyset$, then the claim holds trivially. Suppose $J \setminus J' \neq \emptyset$. Let $i \in J \setminus J'$ be the highest priority patient in $J \setminus J'$ according to π .

We have $|J| = |J'|$ by maximality of σ and v in beneficiary assignment for I^* . Thus, $|J \setminus J'| = |J' \setminus J|$, which implies $J' \setminus J \neq \emptyset$.

By Lemma C.6, there exists a matching $\hat{v} \in \mathcal{M}^b(I^*, r^*)$ that matches patients only with preferential treatment categories that they are beneficiaries of, such that \hat{v} matches patients in J' and $|\sigma^{-1}(c) \cap I_c|$ units reserved for every preferential treatment category $c \in \mathcal{C} \setminus \{u\}$.²² Because both v and σ are maximal in beneficiary assignment for (I^*, r^*) , then only patients in J' should be matched under \hat{v} , and no other patients (as otherwise, v would not be maximal in beneficiary assignment for (I^*, r^*)).

Because $\hat{v}(i) = \emptyset$, $i \in I_{\hat{v}(i)}$, and $\sigma(i) \neq u$, by Observation C.1, there exists a beneficiary alternating path A starting with i from \hat{v} to σ and ending with a patient $i' \in I_{\hat{v}}(i')$ (and $\hat{v}(i') \neq u$ by its construction), and, yet, $i' \notin I_{\sigma(i)}$ or $\sigma(i') = u$.

Existence of the beneficiary alternating path shows that it is possible to match i or i' (but not both) with preferential treatment categories that they are beneficiaries of without affecting anybody else's status as committed or uncommitted in Step 1 of the **Smart Reserve Matching Algorithm*** with all unreserved units processed last. Because σ matches i with a preferential treatment category that she is a beneficiary of at the cost of patient i' , we have

$$i \pi i'.$$

Next, consider the smaller economy (I', r') , in which we remove (i) i and i' and set $I' = I^* \setminus \{i, i'\}$, (ii) one of the units associated with preferential treatment category $\hat{c} = \hat{v}(i')$ and set $r'_{\hat{c}} = r^*_{\hat{c}} - 1$, and (iii) keep the capacity of every other category $c \in \mathcal{C} \setminus \{\hat{c}\}$ the same and set $r'_c = r^*_c$.

Let v' be the restriction of \hat{v} to (I', r') . As $\hat{v}(i) = \emptyset$, $\hat{v}(i') = \hat{c} \neq u$ such that $i \in I_{\hat{c}}$ and the capacity of \hat{c} is reduced by one in the smaller economy, $v' \in \mathcal{M}(I', r')$, and, furthermore, it is maximal in beneficiary assignment for (I', r') .

We form a matching $\sigma' \in \mathcal{M}(I', r')$ by modifying σ and \hat{v} using the beneficiary alternating path A we found before. Recall that A is the beneficiary alternating path from \hat{v} to σ beginning with i and ending with i' . Suppose $A = (i, i_2, \dots, i_{\bar{m}-1}, i')$. Define

$$\sigma'(i^*) = \sigma(i^*) \text{ for every } i^* \in I' \setminus \{i_2, \dots, i_{\bar{m}-1}\} \text{ and}$$

$$\sigma'(i^*) = \hat{v}(i^*) \text{ for every } i^* \in \{i_2, \dots, i_{\bar{m}-1}\}.$$

Observe that $\sigma' \in \mathcal{M}(I', r')$. The existence of σ' shows that it is possible to match every patient in $J_{|I'|}(I^*, r^*) \setminus \{i\}$ with a preferential treatment category that she is a beneficiary of in (I', r') . Thus, the **Smart Reserve Matching Algorithm*** with all unreserved units processed last proceeds as it does for (I^*, r^*) , with the exception that it skips patients i and i' . Therefore, $\sigma' \in \mathcal{M}_S^0(I', r')$.

By the inductive assumption, the claim holds for σ' and v' for (I', r') . This completes the induction, as we already showed $i \pi i'$. \diamond

Thus, we showed that at the end of Step 1 of the **Smart Reserve Matching Algorithm*** with all unreserved units processed last, weakly lower-priority patients have remained uncommitted in $J' = I \setminus J_{|I|} = I \setminus \cup_{c \in \mathcal{C} \setminus \{u\}} (\sigma_0^{-1}(c) \cap I_c)$ than in $\hat{J} = I \setminus \cup_{c \in \mathcal{C} \setminus \{u\}} (\mu^{-1}(c) \cap I_c)$.

Two cases are possible:

- If we have a soft reserves system: As both σ_0 and μ are maximal in beneficiary assignment, an equal number of units are assigned to the highest π -priority patients in \hat{J} (by Step 2 of the **Smart Reserve Matching Algorithm***) and J^* (as by Theorem 1, μ respects priorities and is nonwasteful). Under σ_0 , the unreserved units are assigned last in order in Step 2 of the algorithm. On the other hand, the remainder of μ —that is, the units assigned to the nonbeneficiaries of preferential treatment categories and beneficiaries of u —can be constructed by assigning the rest of the units sequentially to the highest-priority patients in J^* one by one when unreserved units are not necessarily processed last.

Therefore, if $|\mu^{-1}(u)| < r_u$, then $|\sigma_0^{-1}(u)| < r_u$, in turn, implying $\bar{f}_u^\mu = \bar{f}_u^{\sigma_0} = \emptyset$. If $|\mu^{-1}(u)| = r_u$, then $|\sigma_0^{-1}(u)| \leq r_u$ and $\bar{f}_u^\mu = \min_{\pi} \mu^{-1}(u) \sqsubseteq \bar{f}_u^{\sigma_0} \in \{\emptyset, \min_{\pi} \sigma_0^{-1}(u)\}$.

- If we have a hard reserves system: The proof is identical as the above case with the exception that now only unreserved units are assigned, as both σ_0 and μ comply with eligibility requirements. \square

Endnotes

¹ Examples of this framework are in WHO (2007), Prehn and Vawter (2008), Emanuel et al. (2020), and Truog et al. (2020), among others.

² For example, 2018 U.S. Centers for Disease Control Influenza Vaccine Allocation guidelines place patients into one of four tiers based

on (1) provision of homeland and national security, (2) provision of healthcare and community support services, (3) maintenance of critical infrastructure, and (4) membership of the general population (CDC 2018).

³ The European Society of Intensive Care Medicine devised the SOFA score at a consensus meeting in October 1994 in Paris (Vincent et al. 1996). Each of six organ systems—lungs, liver, brain, kidneys, blood clotting, and blood pressure—is independently assigned a score of one to four. The SOFA score sums these six scores, and sicker patients are assigned higher scores. Although not initially designed as a prognostic score, some subsequent research supports its use for that end (Jones et al. 2009).

⁴ A well-known example in market-design literature is the incompatibility between the normative principles of *Pareto efficiency* and *no justified envy* in the context of school choice (Balinski and Sönmez 1999, Abdulkadiroğlu and Sönmez 2003).

⁵ de Puy Kamp et al. (2020) report that several hundred U.S. hospitals had adopted this system at the outset of the COVID-19 pandemic.

⁶ After widespread backlash, this plan was withdrawn on April 9, 2020.

⁷ Illustrating this tension, revised MA guidelines issued on October 20, 2020, no longer include any reference for heightened priority for these essential personnel groups (DPH 2020b).

⁸ This division is for accounting purposes only, and it does not attach a specific unit to a category.

⁹ Our analyses in Section 3.7 and Appendix B are two exceptions to this statement.

¹⁰ This more general structure is not only needed for our main application of medical resource allocation, but also in other applications. For example, in Germany, priority for college seats depends on applicant waiting time for (up to) 20% of the seats, on their performance in high school leaving exams for (up to) 20% of the seats, and on college-specific criteria for the rest of the seats (Westkamp 2013).

¹¹ As an illustration, consider affirmative action in India, where a percentage of government jobs are reserved for historically discriminated-against groups, such as Scheduled Castes (SCs). By the Supreme Court judgment *Indra Sawhney* (1992), units reserved for SCs cannot be awarded to candidates from other groups, even if there are not sufficiently many applicants from SCs. Moreover, by the same judgment, any member of SCs who is awarded one of these SC-reserve units needs to have lower merit score than anyone who receives any unreserved unit (cf. Sönmez and Yenmez 2022). These legal requirements cannot be verified by simply specifying who receives a unit. It is also necessary to specify through which category the unit is received.

¹² The case of hard reserves is much simpler to analyze and offers more general results. See theorem 2 in Pathak et al. (2020b).

¹³ Additional details on these and other plans can be found at <https://www.covid19reservesystem.org/>.

¹⁴ The free software is available from the first three co-authors upon request; its demo video can be seen at <https://www.covid19reserve-system.org/software>.

¹⁵ We are grateful to an anonymous referee for this insightful observation.

¹⁶ For $n=0$, this algorithm is equivalent to the meritorious horizontal algorithm in Sönmez and Yenmez (2022).

¹⁷ This is known as a *bipartite maximum cardinality matching problem* in graph theory, and many *augmenting alternating path algorithms* (such as those by Hopcroft and Karp 1973, Karzanov 1973) can solve it in polynomial time.

¹⁸ A polynomial-time algorithm such as the *Hungarian algorithm* (Kuhn 1955) solves this problem.

¹⁹ This problem can be solved by using a polynomial *augmenting alternating paths algorithm* (for example, see Hopcroft and Karp 1973 and Karzanov 1973).

²⁰ For $n=0$, this algorithm gives the meritorious horizontal choice rule in Sönmez and Yenmez (2022).

²¹ As defined in the hypothesis of the lemma, $\tilde{v} \in \mathcal{M}^b(I^*, r^*)$ means that for every $i^* \in I^*$ and $c \in C$, $i^* \in \tilde{v}^{-1}(c)$ implies $c \neq u$ and $i^* \in I_c$.

²² Although both σ and v may be matching some patients with categories that they are not beneficiaries of or with the unreserved category, we can simply leave those patients unmatched in σ and v and apply Lemma C.6 to see such a matching \hat{v} exists.

References

- Abdulkadiroğlu A (2005) College admissions with affirmative action. *Internat. J. Game Theory* 33(4):535–549.
- Abdulkadiroğlu A, Sönmez T (2003) School choice: A mechanism design approach. *Amer. Econom. Rev.* 93:729–747.
- Akbarpour M, Budish E, Dworczak P, Kominers SD (2023) An economic framework for vaccine prioritization. *Quart. J. Econom.*, ePub ahead of print June 12, <https://doi.org/10.1093/qje/qjad022>.
- Ariadne Labs (2020) Symposium: Vaccine allocation and social justice. Accessed January 2023, <https://www.ariadnelabs.org/2020/12/08/symposium-vaccine-allocation-and-social-justice/>.
- Aygün O, Bó I (2021) College admission with multidimensional privileges: The Brazilian affirmative action case. *Amer. Econom. J. Microeconom.* 13(3):1–28.
- Azevedo E, Leshno J (2016) A supply and demand framework for two-sided matching markets. *J. Polit. Econom.* 124(5):1235–1268.
- Aziz H, Brandl F (2021) Efficient, fair, and incentive-compatible healthcare rationing. Preprint, submitted February 8, <https://arxiv.org/abs/2102.04384>.
- Balinski M, Sönmez T (1999) A tale of two mechanisms: Student placement. *J. Econom. Theory* 84:73–94.
- Baswana S, Chakrabarti, PP, Chandran S, Kanoria Y, Patange U (2019) Centralized admissions for engineering Colleges in India. *INFORMS J. Appl. Anal.* 49(5):307–396.
- Biddinger P (2020) Interview. Conducted on December 18.
- Biro P, Fleiner T, Irving RW, Manlove DF (2010) The college admissions problem with lower and common quotas. *Theor. Comput. Sci.* 411(34–36):3136–3153.
- Bo I (2016) Fair implementation of diversity in school choice. *Games Econom. Behav.* 97:54–63.
- CDC (2018) Interim planning guidance on allocating and targeting pandemic influenza vaccine during an influenza pandemic. Accessed July 1, 2020, <https://www.cdc.gov/flu/pandemic-resources/pdf/2018-Influenza-Guidance.pdf>.
- Combe J (2018) Are affirmative action policies so impossible? *Rev. Econom. Polit.* 128(6):1123–1173.
- Commonwealth of Pennsylvania (2020) Ethical allocation framework for emerging treatments of COVID-19. Accessed January 2023, <https://www.health.pa.gov/topics/disease/coronavirus/Pages/Guidance/Ethical-Allocation-Framework.aspx>.
- Correa J, Epstein R, Escobar J, Rios I, Bahamondes B, Bonet C, Epstein N, et al. (2019) School choice in Chile. *EC'19 Proc. 2019 ACM Conf. Econom. Comput.* (Association for Computing Machinery, New York), 325–343.
- Daugherty-Biddison L, Gwon H, Regenberg A, Schoch-Spana M, Toner E (2017) Maryland framework for the allocation of scarce life-sustaining medical resources in a catastrophic public health emergency. Accessed January 2023, https://www.law.umaryland.edu/media/SOL/pdfs/Programs/Health-Law/MHECN/ASR\%20Framework_Final.pdf.
- de Puy Kamp M, Devine C, Griffin D (2020) As coronavirus cases grow, hospitals adopt a system to rank patients for treatment. CNN (April 3), <https://www.cnn.com/2020/04/03/health/coronavirus-hospital-ethics-ventilators-invs/index.html>.
- Delacrétaz D (2021) Processing reserves simultaneously. Working paper, University of Manchester, Manchester, UK.
- DHS Arizona (2020) COVID-19 addendum: Allocation of scarce resources in acute care facilities. Accessed July 4, 2020, <http://azpha.wildapricot.org/resources/covid-19-addendum.pdf>.
- Doğan B (2016) Responsive affirmative action in school choice. *J. Econom. Theory* 165:69–105.
- Doğan B (2017) How to control controlled school choice: Comment. *Amer. Econom. Rev.* 107(4):1362–1364.
- DPH (2020a) Crisis standards of care planning guidance for the COVID-19 pandemic. Accessed January 2023, https://d279m997dpfwgl.cloudfront.net/wp/2020/04/CSC_April-7_2020.pdf.
- DPH (2020b) Crisis standards of care planning guidance for the COVID-19 pandemic (revised). Accessed January 2023, <https://www.mass.gov/doc/crisis-standards-of-care-planning-guidance-for-the-covid-19-pandemic-0/download>.
- Dur U, Morrill T, Phan W (2021) Allocating medical resources during a pandemic. Working paper, North Carolina State University, Raleigh, NC.
- Dur U, Pathak PA, Sönmez T (2020) Explicit vs. statistical preferential treatment in affirmative action: Theory and evidence from Chicago's exam schools. *J. Econom. Theory* 187:104996.
- Dur U, Kominers SD, Pathak PA, Sönmez T (2018) Reserve design: Unintended consequences and the demise of Boston's walk zones. *J. Polit. Econom.* 126:2457–2479.
- Echenique F, Yenmez MB (2015) How to control controlled school choice. *Amer. Econom. Rev.* 105(8):2679–2694.
- Ehlers L, Morrill T (2019) (II)legal assignments in school choice. *Rev. Econom. Stud.* 87(4):1837–1875.
- Ehlers L, Hafalir IE, Yenmez MB, Yildirim MA (2014) School choice with controlled choice constraints: Hard bounds versus soft bounds. *J. Econom. Theory* 153:648–683.
- Emanuel EJ, Persad G, Upshur R, Thome B, Parker M, Glickman A, Zhang C, et al. (2020) Fair allocation of scarce medical resources in the time of COVID-19. *N. Engl. J. Med.* 382(21):2049–2055.
- Fink S (2020) U.S. Civil Rights Office rejects rationing medical care based on disability, age. *New York Times* (March 28), <https://www.nytimes.com/2020/03/28/us/coronavirus-disabilities-rationing-ventilators-triage.html>.
- Fragiadakis D, Troyan P (2017) Improving matching under hard distributional constraints. *Theor. Econom.* 12(2):863–908.
- Gale D, Shapley LS (1962) College admissions and the stability of marriage. *Amer. Math. Monthly* 69:9–15.
- Goto M, Kojima F, Kurata R, Tamura A, Yokoo M (2017) Designing matching mechanisms under general distributional constraints. *Amer. Econom. J. Microeconom.* 9(2):226–262.
- Greenberg K, Pathak PA, Sönmez T (2023) Redesigning the US Army's branching process: A case study in minimalist market design. Preprint, submitted March 15, <https://arxiv.org/abs/2303.06564>.
- Grigoryan A (2021) Effective, fair and equitable pandemic rationing. Preprint, submitted October 6, <http://dx.doi.org/10.2139/ssrn.3646539>.
- Hafalir IE, Yenmez MB, Yildirim MA (2013) Effective affirmative action in school choice. *Theor. Econom.* 8(2):325–363.
- Hellman D, Nicholson KM (2021) Rationing and disability: The civil rights and wrongs of state triage protocols. *Washington Lee Law Rev.* 78(4).
- Hitzig Z (2020) The normative gap: Mechanism design and ideal theories of justice. *Econom. Philos.* 36(3):407–434.

- Hopcroft JE, Karp RM (1973) An $n^{5/2}$ algorithm for maximum matchings in bipartite graphs. *SIAM J. Comput.* 2:225–231.
- Imamura K (2020) Meritocracy versus diversity. Working paper, Boston College, Chestnut Hill, MA.
- Jones AE, Trzeciak S, Kline JA (2009) The Sequential Organ Failure Assessment score for predicting outcome in patients with severe sepsis and evidence of hypoperfusion at the time of emergency department presentation. *Crit. Care Med.* 37: 1649–1654.
- Kamada Y, Kojima F (2015) Efficient matching under distributional constraints: Theory and applications. *Amer. Econom. Rev.* 105(1): 67–99.
- Kamada Y, Kojima F (2017) Stability concepts in matching with distributional constraints. *J. Econom. Theory* 168:107–142.
- Kamada Y, Kojima F (2018) Stability and strategy-proofness for matching with constraints: A necessary and sufficient condition. *Theor. Econom.* 13(2):761–794.
- Karzanov AV (1973) An exact estimate of an algorithm for finding a maximum flow, applied to the problem on representatives. *Problems Cybernetics* 5:66–70.
- Kinlaw K, Levine R (2007) Ethical guidelines in pandemic influenza: Recommendations of the Ethics Subcommittee of the Advisory Committee to the Director. Report, Centers for Disease Control and Prevention, Atlanta.
- Kojima F (2012) School choice: Impossibilities for affirmative action. *Games Econom. Behav.* 75(2):685–693.
- Kojima F, Tamura A, Yokoo M (2018) Designing matching mechanisms under constraints: An approach from discrete convex analysis. *J. Econom. Theory* 176:803–833.
- Kominers SD, Sönmez T (2016) Matching with slot-specific priorities: Theory. *Theor. Econom.* 11(2):683–710.
- Kowalczyk L (2020) Who gets a ventilator? New gut-wrenching state guidelines issued on rationing equipment. *Boston Globe* (April 7), <https://www.bostonglobe.com/2020/04/07/metro/massachusetts-officials-release-plan-ration-ventilators-icu-beds-if-need-arises/>.
- Kuhn HW (1955) The Hungarian Method for the assignment problem. *Naval Res. Logist. Quart.* 2:83–97.
- Lee T (2004) Rationing influenza vaccine. *N. Engl. J. Med.* 351: 2365–2366.
- Li S (2017) Ethics and market design. *Oxford Rev. Econom. Policy* 33(4):705–720.
- MA DPH (2020a) MA COVID-19 vaccine presentation. Accessed January 2023, <https://www.mass.gov/doc/ma-covid-19-vaccine-presentation-1292020/download>.
- MA DPH (2020b) Massachusetts Department of Public Health guidance for allocation of COVID-19 monoclonal antibody therapeutics. Accessed January 2023, <https://www.mass.gov/doc/guidance-for-allocation-of-covid-19-monoclonal-antibody-therapeutics/download>.
- Makhoul AT, Drolet BC (2021) Reserve system implementation during the COVID-19 pandemic. *Chest* 160(6):2005–2006.
- Mendelsohn NS, Dulmage AL (1958) Some generalizations of the problem of distinct representatives. *Canad. J. Math.* 10:230–241.
- NASEM (2020a) A framework for equitable allocation of vaccine for the novel coronavirus. Accessed January 2023, <https://www.nap.edu/resource/25917/25914.pdf>.
- NASEM (2020b) A framework for equitable allocation of vaccine for the novel coronavirus. Accessed January 2023, <https://www.nationalacademies.org/our-work/a-framework-for-equitable-allocation-of-vaccine-for-the-novel-coronavirus>.
- Ne'eman A (2020) I will not apologize for my needs. *New York Times* (March 23), <https://www.nytimes.com/2020/03/23/opinion/coronavirus-ventilators-triage-disability.html>.
- ORAAC (2023) Request for public input. Accessed January 2023, https://www.oregon.gov/oha/ORAACDocuments/ORAAC%20Request%20for%20Public%20Comment_EN.pdf.
- Pathak PA, Rees-Jones A, Sönmez T (2023a) Immigration lottery design: Engineered and coincidental consequences of H-1B reforms. *Rev. Econom. Statist.* Forthcoming.
- Pathak PA, Rees-Jones A, Sönmez T (2023b) Reversing reserves. *Management Sci.* Forthcoming.
- Pathak PA, Sönmez T, Ünver MU (2020a) Improving ventilator rationing through collaboration with experts on resource allocation. *JAMA Netw. Open* 3(6):e2012838.
- Pathak PA, Sönmez T, Ünver MU, Yenmez MB (2020b) Leaving no ethical value behind: Triage protocol design for pandemic rationing. NBER Working Paper 26951, National Bureau of Economic Research, Cambridge, MA.
- Pathak PA, Schmidt H, Solomon A, Song E, Sönmez T, Ünver MU (2020c) Do Black and Indigenous communities receive their fair share of vaccines under the 2018 CDC guidelines? NBER Working Paper 27817, National Bureau of Economic Research, Cambridge, MA.
- Persad G, Peek ME, Emanuel E (2020) Fairly prioritizing groups for access to COVID-19 vaccines. *JAMA* 324(16):1601–1602.
- Piscitello GM, Miller WD, Kapania EM, Rojas JC, Siegler M, Parker WF (2020) Systematic review of the variation in U.S. state ventilator allocation guidelines during a public health crisis. *JAMA Netw. Open* 3:e2012606.
- Prehn AW, Vawter D (2008) Ethical guidance for rationing scarce health-related resources in a severe influenza pandemic: Literature and plan review. Report, Minnesota Pandemic Ethics Project, St. Paul, MN.
- Richmond City (2021a) Phase 2 is now open in Richmond and Henrico! Accessed January 2023, <https://www.vdh.virginia.gov/richmond-city/2021/04/08/phase-2-is-now-open-in-richmond-and-henrico/>.
- Richmond City (2021b) Richmond and Henrico Health Districts expand vaccine eligibility to all Phase 1b groups. Accessed January 2023, <https://www.vdh.virginia.gov/richmond-city/phase-1b-expansion/>.
- Rosenbaum L (2020) Facing COVID-19 in Italy—Ethics, logistics, and therapeutics on the epidemic's front line. *N. Engl. J. Med.* 382(20): 1873–1875.
- Rubin E, Dryden-Peterson SL, Hammond SP, Lennes I, Letourneau AR, Pathak PA, Sönmez T, Ünver MU (2021) A novel approach to equitable distribution of scarce therapeutics: Institutional experience implementing a reserve system for allocation of COVID-19 monoclonal antibodies. *Chest* 160(6):2324–2331.
- Schmidt H (2020) Vaccine rationing and the urgency of social justice in the COVID-19 response. *Hastings Center Rep.* 50:1–4.
- Schmidt H, Pathak P, Sönmez T, Ünver MU (2020) COVID-19: How to prioritize worse-off populations in allocating safe and effective vaccines. *BMJ* 371:m3795.
- Schmidt H, Weintraub R, Williams MA, Miller K, Buttenheim A, Sadecki E, Wu H, et al. (2021) Equitable allocation of COVID-19 vaccines in the United States. *Nat. Med.* 27:1298–1307.
- Schoch-Spana M, Fitzgerald J, Kramer BR, UPMC Influenza Task Force (2005) Influenza vaccine scarcity 2004–05: Implications for biosecurity and public health preparedness. *Biosecurity Bioterrorism* 3(3):224–234.
- Schrijver A (2003) *Combinatorial Optimization: Polyhedral and Efficiency*, vol. A (Springer, Berlin, Heidelberg).
- Sönmez T (2023) Minimalist market design: A framework for economists with policy aspirations. Working paper, Boston College, Chestnut Hill, MA; <https://www.tayfunsonmez.net/wp-content/uploads/2023/03/Minimalist-Market-Design-February2023.pdf>.
- Sönmez T, Yenmez MB (2022) Affirmative action in India via vertical, horizontal, and overlapping reservations. *Econometrica* 90(3): 1143–1176.
- Sönmez T, Pathak PA, Ünver MU, Persad G, Truog RD, White DB (2021) Categorized priority systems: A new tool for fairly allocating scarce medical resources in the face of profound social inequities. *Chest* 159:1294–1299.

- TDH (2020a) COVID-19 vaccination plan, version 1.0, October 16. Accessed January 2023, <https://www.cdc.gov/vaccines/covid-19/downloads/tennessee-jurisdiction-executive-summary.pdf>.
- TDH (2020b) COVID-19 vaccination plan, version 3.0, December 30. Accessed January 2023, https://www.tn.gov/content/dam/tn/health/documents/cedep/novel-coronavirus/COVID-19_Vaccination_Plan.pdf.
- Tomoeda K (2018) Finding a stable matching under type-specific minimum quotas. *J. Econom. Theory* 176:81–117.
- Truog RD, Mitchell C, Daley GQ (2020) The toughest triage—Allocating ventilators in a pandemic. *N. Engl. J. Med.* 382:1973–1975.
- University of Pittsburgh (2020) Allocation of scarce critical care resources during a public health emergency. School of Medicine, Department of Critical Care Medicine.
- Vawter DE, Garrett JE, Gervais KG, Prehn AW, DeBruin DA, Tauer CA, Parilla E, et al. (2010) For the good of us all: Ethically rationing health resources in Minnesota in a severe influenza pandemic. Minnesota Pandemic Ethics Project, Minnesota Department of Health, St. Paul, MN.
- Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, Reinhart CK, et al. (1996) The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med.* 22:707–710.
- Wasserman D, Persad G, Millum J (2020) Setting priorities fairly in response to COVID-19: Identifying overlapping consensus and reasonable disagreement. *J. Law Biosci.* 7(1):lsaa044.
- Westkamp A (2013) An analysis of the German university admissions system. *Econom. Theory*, 53(3):561–589.
- White DB, Katz MH, Luce JM, Lo B (2009) Who should receive life support during a public health emergency? Using ethical principles to improve allocation decisions. *Ann. Intern. Med.* 150(2): 132–138.
- White DB, McCreary EK, Chang C-CH, Schmidhofer M, Bariola JR, Jonassaint NN, Persad G, et al. (2022) A multicenter weighted lottery to equitably allocate scarce COVID-19 therapeutics. *Amer. J. Respiratory Crit. Care Med.* 206(4):503–506.
- WHO (2007) Ethical considerations in developing a public health response to pandemic influenza. Accessed January 2023, http://www.who.int/csr/resources/publications/WHO_CDS_EPR_GIP_2007_2/en/index.html.
- Whyte LE (2020) State policies may send people with disabilities to the back of the line for ventilators. Accessed January 2023, <https://publicintegrity.org/health/coronavirus-and-inequality/state-policies-may-send-people-with-disabilities-to-the-back-of-the-line-for-ventilators/>.
- Wu E-L, Kumar RN, Moore WJ, Hall GT, Vysniauskaitė I, Kim K-YA, Angarone MP, et al. (2022) Disparities in COVID-19 monoclonal antibody delivery: A retrospective cohort study. *J. Gen. Intern. Med.* 37(10):2505–2513.
- Zucker HA, Adler KP, Berens DP, Brynner R, Butler K, Calderon Y, Corcoran C, et al. (2015) Ventilator allocation guidelines. New York State Task Force on Life and the Law, New York Department of Health, New York.