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

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

July 2020

Abstract

COVID-19 has revealed several limitations of existing mechanisms for rationing scarce medical resources under emergency scenarios. Many argue that they abandon various ethical values such as equity by discriminating against disadvantaged communities. Illustrating that these limitations are aggravated by a restrictive choice of mechanism, we formulate pandemic rationing of medical resources as a new application of market design and propose a reserve system as a resolution. We develop a general theory of reserve design, introduce new concepts such as cutoff equilibria and smart reserves, extend previously-known ones such as sequential reserve matching, and relate these concepts to current debates.

JEL codes: D45, D47, I14
Keywords: ethical rationing, reserve system, COVID-19, vaccine, ventilator

*This version supercedes: "Leaving No Ethical Value Behind: Triage Protocol Design for Pandemic Rationing," NBER Working Paper 26951, April 2020. We are grateful for input from several experts in bioethics, emergency healthcare, and webinar participants at Johns Hopkins University, 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, and UK Virtual Seminars in Economic Theory. Nikhil Agarwal, Alex Rees-Jones, Robert Truog, Iván Werning, and Doug White provided helpful comments. Feedback from Eric Budish, David Delacrétaz, Fuhito Kojima, Govind Persad, and Alex Teytelboym was particularly valuable.

†Pathak: Department of Economics, Massachusetts Institute of Technology, and NBER, email: ppathak@mit.edu, Sönmez: Department of Economics, Boston College, email: sonmezt@bc.edu, Ünver: Department of Economics, Boston College, email: unver@bc.edu, Yenmez: Department of Economics, Boston College, email: bumin.yenmez@bc.edu.

"*If you fail to plan, you are planning to fail.*"

— Benjamin Franklin

1 Introduction

Agencies responsible for public health, medical and emergency preparedness at the international, national, regional, and hospital level all have proposed guidelines for allocating scarce medical resources in crisis situations. These situations range from wartime triage medicine to public health emergencies, such as influenza pandemics and COVID-19. In 2005, the World Health Organization warned of a severe influenza pandemic and advised countries to develop pandemic guidelines and triage protocols (WHO, 2005). While the nature of the resources in short supply depends on the crisis situation, for influenza and COVID-19, they may include ventilators, antiviral medications, and potential vaccines.

How to implement a rationing system during a crisis situation presents a complicated question rife with ethical concerns. Rationing guidelines typically start by articulating several different ethical principles. These principles include equity, which is fair distribution of benefits and burdens; utilitarianism, which involves 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; non-discrimination, which is requiring that certain individual characteristics such as gender, race, and age play no role allocation. Rationing guidelines also emphasize procedural values, such as accountability, reasonableness, and transparency (see, e.g., WHO (2007); Prehn and Vawter (2008); Emanuel et al. (2020); Truog, Mitchell and Daley (2020)).

After articulating these values, regional or national guidelines propose a way to operationalize them with an allocation mechanism. The most common system used in these guidelines is a *priority system* which places patients into a single priority order, and allocates all units based on this single priority. For example, 2018 US Centers for Disease Control Vaccine Allocation guideline place patients into one of four tiers based on their role in providing homeland and national security, providing health care and community support services, maintaining critical infrastructure, and being a member of the general population (CDC, 2018). In some cases priority orders are obtained through an objective scoring method, resulting in a *priority point system*. This is especially common for rationing of ICU beds and ventilators. For example, Piscitello et al. (2020) reports that 19 states have adopted priority point systems based on the Sequential Organ Failure Assessments (SOFA) score, a scoring method that measures severity of organ dysfunction in six organ systems where higher aggregate scores are associated with elevated mortality risk.

The COVID-19 pandemic has spurred renewed interest in these guidelines and has revealed several important limitations of the existing allocation mechanisms. Whether it is about rationing of ventilators, antiviral drugs or vaccines, a common theme in many of these debates is that existing guidelines and allocation mechanisms have given up on various important ethical values. For example, advocates for disadvantaged groups voice opposition to priority point systems which use survival probabilities based on various proxies such as the SOFA score. They

argue that these criteria are discriminatory for they fail to acknowledge pre-existing discrimination in access to health care (Schmidt, 2020). From March-June 13, 2020, the age-adjusted COVID-19 hospitalization rate of non-hispanic blacks is 178 per 100,000, which is more than four times that of non-hispanic whites (CDC, 2020). Moreover, disparate access to testing for disadvantaged groups have also increased COVID-19 prevalence in these communities, further elevating their disadvantage. Similarly, disabilities' advocates voice opposition to rationing plans based solely on survival probabilities. Some of them reject a detailed triage protocol in favor of random selection (Ne'eman, 2020). In their view, such criteria are inherently discriminatory for certain types of disabled patients.

In this paper, we argue that several shortcomings of the existing rationing systems are a direct consequence of restricting allocation mechanisms to priority systems. This restriction imposes a single priority order of individuals for allocation of all units. This limiting feature of a priority system compromises its ability to represent a variety of ethical considerations. A remedy must therefore break this characteristic. We propose an alternative way to accommodate multiple ethical principles through a *reserve system*. In a reserve system, units are divided into multiple categories each representing an ethical value or a balance of multiple ethical values. Rather than relying on a single priority order of individuals for allocation of all units, a distinct priority order is used to prioritize individuals for units of each category. This heterogeneity allows for accommodating the desired ethical values without the need to aggregate them into a single metric or into a strict lexicographic hierarchy.

The flexibility of a reserve system over a priority system can help reach compromises for several ongoing polarizing debates. For example, many argue that disadvantaged groups or under-represented minorities should be given priority access for a COVID-19 vaccine. This is a significant departure from the recommended priority tiers in the 2018 CDC Vaccine allocation guideline. Others question whether prioritizing these groups would erode public trust in vaccination. For more on this debate, see Ducharme (2020), McCaughey (2020), and Twohey (2020). While there is no middle ground for compromise under a priority system, a compromise can be reached through a reserve system by giving disadvantaged communities preferential treatment for a fraction of vaccines. In Section 5, we present other examples of how a reserve system can help reach compromises in other debates.

Reserve systems have an important precedent in medicine for deceased donor kidney allocation. Until 2014, the U.S. Organ Procurement and Transplantation Network (OPTN) used a priority system to allocate deceased donor kidneys (OPTN, 2014). After establishing medical compatibility, patients were ordered by priority type, and then priority points, with ties broken by waiting time. In 2014, the OPTN system changed to include a reserve (Israni et al., 2014). In the reserve, 20% of the highest quality kidneys are reserved for adults with the highest 20% expected post-transplant survival score (EPTS). A priority points system is used for the remaining kidneys. Parallel to our motive for introducing reserve systems (i.e. better integration of various ethical principles), OPTN added this reserve to better reflect the ethical principle of utilitarianism, which was not as prominently captured in the prior system.[1]

[1]The New Kidney Allocation System (KAS) Frequently Asked Questions, Organ Procurement and Transplantation Network. Available at `https://optn.transplant.hrsa.gov/media/1235/kas_faqs.pdf` (last accessed on

In addition to formulating medical rationing as a new application of market design, our paper also contributes to the theoretical matching literature by developing a general theory of reserve systems. Beyond our primary application, our results have direct implications for a wide range of applications such as affirmative action in allocation of public positions or public school seats and assignment of immigration visas. In the remaining part of this section, we summarize our contributions to the theory of reserve design.

In our formulation of medical rationing, there is a number of identical medical units to be allocated to a set of patients. These units can be ICU beds, ventilators, anti-virals, vaccines, or other scarce vital items. An allocation is a matching of patients either to a reserve category if the patient receives a unit or the empty set if she does not. Hence, by formulating the medical rationing problem as an application of reserve design, we do away with the restriction of mapping all ethical considerations on to a single dimension. Under our formulation, units are placed into one of a set of reserve categories, each with a distinct priority order of patients. A reserve system is a generalization of a priority system because units in different reserve categories can prioritize patients in different ways.

Our formal approach is axiomatic and lays the foundations of reserve systems by specifying three basic principles any system has to satisfy. First, patients should only receive units for which they are qualified. Second, no unit from any reserve category should stay idle as long as there is an eligible patient for that category. Finally, for each category, units should be allocated based on the priority order of individuals in the category. We see these principles as minimal requirements, which can be justified on both normative and positive grounds.

In many real-life applications of reserve systems it is customary to announce the outcome of a reserve system though a specification of the *cutoff individual at each category*, i.e., the lowest priority individual who has gained admission through each category.[2] The vector of cutoff individuals identify the *budget set* for each individual in the sense that they can see through which (if any) category they can receive a unit. Motivated by this observation, we formalize the notion of a *cutoff equilibrium*, a notion akin to a competitive equilibrium. A cutoff equilibrium is a vector of cutoff individuals together with a matching, where each patient is matched with a category in her budget set, and any category that has not filled its quota has a cutoff of $\emptyset$ (the counterpart of a price of zero in our model). Our first result not only rationalizes the prevalence of the use of cutoffs in real-life applications but also the plausibility of our three axioms: A matching satisfies these three properties if and only if it is supported at a cutoff equilibrium.

Although our first main result provides a full characterization of matchings that satisfy our three properties, it leaves open the question of how to find matchings supported at cutoff equilibria. Just like computing all competitive equilibria often presents challenges except in special cases, so does the computation of all cutoff equilibria. Our second main result is that matchings that are supported at cutoff equilibria can be computed by constructing a hypothetical two-sided matching market where each patient has strict preferences between categories. This construction is hypothetical because in the original problem the patient only cares about obtaining a resource and is therefore indifferent between all units. We show that a matching satisfies our

7/20/2020).

[2]See, for example, Appendix Figure A1.

three properties if and only if it can be computed as the outcome of the deferred acceptance algorithm (Gale and Shapley, 1962) for this hypothetical market.

Our link to a hypothetical two-sided matching market provides a great deal of flexibility to compute outcomes by changing the way patients rank categories under these artificial preferences. This flexibility has an important drawback: unless there is a systematic way to construct preferences, practitioners may not value the ability to construct artificial preferences at this level of generality. For instance, it may be difficult to explain 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 these categories in the reverse order. We therefore focus on an intuitive subclass of matchings, known as sequential reserve matchings, in which reserve categories are processed in sequence in a particular order. The set of sequential reserve matchings is a refinement of the set of cutoff matchings. Within this class, we show that the earlier a category is processed, the higher is the maximum cutoff. Intuitively this means that the earlier a category is processed the more competitive it becomes.

We then turn to a special environment where there is a baseline priority order. This special environment, which is the focus of much of the earlier literature on reserve systems, is widespread in real-life applications where the baseline order may depend on scores on standardized exams, results of random lotteries, or time of application. For medical rationing the baseline order may depend on an objective measure of expected health outcome such as the SOFA score.[3] There is an unreserved category in which all patients are beneficiaries and the priority order is the same as the baseline. Any other category is a preferential treatment category, with a set of beneficiaries, and all beneficiaries of a category are prioritized over patients who are not, but otherwise their relative priority order is the same as the baseline. If each patient is a beneficiary of at most one preferential treatment category, and there are not more than five categories, every beneficiary of the preferential treatment category who is matched when a category c is processed earlier is also matched when it is processed later. This result substantially generalizes earlier results on reserve processing to several categories, while also showing the limit of obtaining sharp formal results for more than five categories.

Finally, we turn to one possible shortcoming of certain sequential reserve matchings: some of the sequential reserve matchings can be Pareto dominated by others. This shortcoming is a direct consequence of the mechanical allocation of patients into reserve categories under sequential reserve matching when they clear the cutoffs for multiple categories. We therefore introduce an additional principle requiring a "smart" allocation of reserves in these situations, maximizing the accommodation of intended beneficiaries of reserves. This property together with non-wastefulness imply Pareto efficiency. We conclude our theoretical analysis by introducing an algorithm that produces a range of "smart" reserve matchings with the desired properties.

The rest of paper organized as follows. Section 2 presents some additional background on widespread use of priority systems for emergency rationing of medical resources and identifies

[3]The European Society of Intensive Care Medicine devised the SOFA score at a consensus meeting in October 1994 in Paris, France (Vincent et al., 1996). Each of six organ systems – lungs, liver, brain, kidneys, blood clotting, and blood pressure – is independently assigned a score of 1 to 4. The SOFA score sums these six scores, and sicker patients are assigned higher scores. While not initially designed as a prognostic score, subsequent research supports its use for that end (Jones, Trzeciak and Kline, 2009).

several shortcomings of these systems. Section 3 presents our formal model of reserve systems, states our main two characterization results, and introduces sequential reserve matchings. Section 4 specializes the model to one with a baseline priority order, develops comparative static results, and formulates and studies smart reserve matchings. It also relates our findings to earlier literature on reserve systems. Section 5 discusses how a reserve system can help with several medical rationing debates and some other practical aspects of medical rationing. The paper concludes in Section 6. All proofs are relegated to the Appendix.

2 Priority versus Reserve Systems for Pandemic Rationing

2.1 Background on Priority Systems

The most common allocation mechanism for medical rationing is a **priority system** where units are allocated to patients based on a single priority order. This priority order captures the ethical values guiding the allocation of the scarce medical resource. An example is the 2005 US National Vaccine Advisory Committee plan. This guideline places patients into the following tiers: Tier 1A: Health Care Workers, Tier 1B: Highest-Risk Groups, Tier 1C: Household contacts and pregnancy, Tier 1D: Pandemic responders, Tier 2A: Other high-risk groups, Tier 2B: Critical infrastructure groups, Tier 3: Other key government health care decision-makers, and Tier 4: Healthy patients between 2 to 64 years without any high-risk conditions. Within each group, all patients have equal priority. These guidelines were drafted following the 2004-05 influenza vaccine shortage in the United States (Temte, 2006).

In some of the applications, most notably for allocation of ventilators and ICU beds, the underlying priority order is obtained through a monotonic scoring function. Such a refinement of a priority system is called a **priority point system**. Under this system, each ethical value is 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 over medical resources are determined based on their point scores, with a lower score typically indicating a higher claim. Often a priority score is coarsened into tiers, and all patients in the same tier 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 is a prominent example. In the system, as a first step certain patients are deemed ineligible. The remaining patients are ordered based on estimated mortality risk, which is re-evaluated every 48 hours (Zucker et al., 2015). Mortality risk is measured by the Sequential Organ Failure Assessment (SOFA) score which places patients into priority tiers. 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 SOFA for ignoring multiple ethical values. These critics emphasize the need to integrate a variety of ethical values and advocate for a *multi-principle* approach, see, e.g., White et al. (2009) and Daugherty-Biddison et al. (2017). White et al. (2009) describe a multi-principle 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 have become the leading multi-principle priority point system for ventilators.[4]

Table 1. Illustration of Multi-Principle 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 of life	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 life through life's stages	Age 12-40	Age 41-60	Age 61-74	Age > 74

Notes: SOFA stands for sequential organ failure assessment. Multi-principle point system developed by White et al. (2009)

Table 1 shows how the ethical values of saving the most lives, saving the most years of life, and the life-cycle principle are integrated through an additive formula. As an example, consider a hypothetical patient with a SOFA score of seven. She obtains two points based on the ethical value of saving the most lives. If the patient has no chronic diseases or comorbidities and is between 61-74 years old, she obtains four more points based on the other two ethical values yielding a total of six. A patient with a lower total point score has a higher priority for the resource than a patient with a higher total point score.[5] More than half of US states use either a single- or multiple-principle priority point system (Whyte, 2020).

2.2 Challenges with Priority Systems

While practical, priority systems have a number of important limitations. A priority system may fail to integrate different ethical values because of incommensurability.[6] For example, the ethical values of saving the most lives and the life-cycle principle in Table 1 are incommensurable values, making it hard to interpret the role of these values under the priority point system in Table 1. In addition, various ethical values have implications on group composition, and a priority system lacks the flexibility to accommodate these considerations. In many cases, these challenges have led to the exclusion of some of ethical values all together. For these reasons, priority systems are

[4]As of April 24, 2020, California, Colorado, Massachusetts, New Jersey, Oklahoma, and Pennsylvania have adopted variants of this system. de Pu Kamp, Devine and Griffin (2020) reports that several hundred hospitals around the country have adopted this system.

[5]Most protocols specifies a tie-breaker between patients with identical total points, although the South Carolina protocol fails to provide one.

[6]In ethics, two values are incommensurable when they do not share a common standard of measurement.

often ill-defined for medical rationing and not usable. We elaborate on several of these points next, focusing on several debates on rationing of ventilators and ICU beds. Since virtually all states with guidelines recommend priority point systems, we present the shortcomings of these mechanisms.[7] Many of these shortcomings are shared by 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, there are some ethical principles where the claims of patients cannot be represented with a monotonic function. One example is group-based policies, such as those related to regions or gender. For example, the European Union has proposed balanced participation of women and men in political and public decision making by requiring that at least 40% of public offices are held by women and at least 40% are held by men (Dittmar, 2018; Rankin, 2020). For medical rationing, it is possible that a future pandemic is so devastating that it threatens a significant portion of the human race. In such a hypothetical crisis, a principle based on survival of the species may suggest a similar constraint. A guideline may recommend to allocate at least 40% of vital resources to female patients and at least 40% to male patients. Clearly, considerations based on group composition cannot be represented with a function that relies only on individual attributes. Similarly, a priority point system cannot accommodate a guideline that wishes to allocate resources to disabled citizens in proportion to their representation in society.

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 which involves "comparing apples to oranges." The debate on how rationing guidelines should compare claims of children versus adults illustrates this issue. Massachusetts guidelines state that indicators that feed into scores for adults are not reliable for children (Bateman et al., 2020). They explain that "scoring systems that are meaningful for adult critical care patients do not apply to pediatric patients or newborns." 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. An example of this phenomenon 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 New York State and Minnesota had to give up on this consideration, largely due to concerns about the extreme scenarios where no units may remain for the rest of the society. One of the reasons offered in Vawter et al. (2010) by the Minnesota Pandemic Working group is as follows:

[7] As of July 2020, the only exception is Michigan which recommends a priority system with a lexicographic hierarchy between several tiers of patients.

> ... 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.

The New York State Task Force also struggled with this topic, recognizing the need to provide "insurance" for frontline health workers (Zucker et al., 2015). However, they ultimately decided against it, stating:

> 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.

For both states, a limitation of priority systems resulted in completely giving up these ethical values. It is important to emphasize that these concerns are direct consequences of the mechanics of a priority system where providing preferential access to any group for any portion of the resources means giving preferential treatment for all of them.

Fourth, a single priority order struggles to integrate the principle of non-exclusion. 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."[8] 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 – there will be some patients in a lower-priority group who would never be treated during a shortage.

2.2.2 Implementation Issues in Existing Guidelines

These conceptual challenges with a priority point system are reflected in actual design challenges in several guidelines. We briefly describe three examples.

Massachusetts revised their critical care guidelines on April 7, 2020 for the COVID-19 pandemic. Like many guidelines, this document was the result of a committee consisting of medical experts and ethicists. The guidelines provided an adaption 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.

This clause provides no further description on 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.

[8] After widespread backlash, this plan was withdrawn on April 9, 2020. Carter (2020) warns that plans which discriminate against the disabled may violate the Americans with Disabilities Act.

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)

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 Arizona's June 6, 2020 update to their 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." Among the list of additional factors are whether patients are pediatric patients, first responders or health care 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) state that 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 are not followed. For example, Lee (2004) describes that the 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 Systems

Many of the challenges presented in Section 2.2 stem from one simple but limiting feature of a priority point system: it relies on a single priority ranking of patients that is identical for all units. A reserve system eliminates this feature of the mechanism because it allows for heterogeneity of patient claims over different units.

A **reserve system** has three main parameters. They are:

1. a division of all units into multiple segments referred to as *reserve categories*,[9]

2. number of units in each of these categories, and

3. specification of a *priority order* of the patients in each of these categories.

If reserve categories are processed sequentially, as in the case in several real-life applications of reserve systems, *processing order* of reserve categories can also be an additional parameter. For some (or all) of the reserve categories, there can also be *exclusion criteria*, based on the nature of the medical resource that is being rationed along with the patient's clinical assessment. The priority order of patients for each category also incorporates this information. Reserve categories can differ either based on the groups to receive higher priority or the combination of ethical principles to be invoked. The main idea is to use the associated priority order – which embeds ethical principles – when allocating units in each reserve category. Importantly, the priority order need not be the same between reserve categories.

A reserve system provides a natural resolution to the representation and implementation issues discussed in Sections 2.2. However, adoption of a reserve system without a clear understanding of one important aspect of its implementation can have unintended distributional consequences.[10] When a patient qualifies for a unit through multiple reserve categories, she does not care through which one she receives a unit. And yet this choice influences the outcome for other patients. We therefore next present a general theory of reserve systems.

3 A General Theory of Reserve Systems

While our primary application is rationing of scarce medical resources, in this section we provide a general model of reserve systems which has several other applications including affirmative action in school choice (Pathak and Sönmez, 2013; Dur et al., 2018; Correa et al., 2019), college admissions (Aygün and Bo, 2020; Baswana et al., 2019), assignment of government positions (Sönmez and Yenmez, 2019*a*,*b*), and skill diversity in immigration visa allocation (Pathak, Rees-Jones and Sönmez, 2020). Despite this generality, the terminology is tailored to our main application.

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 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

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

[10] See Dur et al. (2018) and Pathak, Rees-Jones and Sönmez (2020) for real-life examples of this phenomenon from school choice and H1-B visa allocation.

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 $x, y \in I \cup \{\emptyset\}$,

$$x \; \underline{\pi}_c \; y \qquad \Longleftrightarrow \qquad x = y \quad \text{or} \quad x \; \pi_c \; y.$$

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.

A **matching** $\mu : I \to \mathcal{C} \cup \{\emptyset\}$ is a function that maps each patient either to a category or to $\emptyset$ such that $\left|\mu^{-1}(c)\right| \leq r_c$ 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.

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,

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

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

In real-life applications of our model, it is important to allocate the 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 $c \in \mathcal{C}$,*

$$\mu(i) = c \qquad \Longrightarrow \qquad i \; \pi_c \; \emptyset.$$

Our first axiom formulates the idea that units should be awarded only to eligible individuals. For rationing of vital medical resources, any patient who is eligible for one category must also be eligible for any category. And if a patient is ineligible for all categories, then this patient can be dropped from the set of patients. Hence, compliance with eligibility requirements always holds for our main application.

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

$$i \; \pi_c \; \emptyset \quad \text{and} \quad \mu(i) = \emptyset \qquad \Longrightarrow \qquad \left|\mu^{-1}(c)\right| = r_c.$$

Our second axiom formulates the idea that no unit should be wasted. That is, if a unit remains idle, then there should not be any unmatched individual who is eligible for the unit. For rationing vital resources, each patient is eligible for all units, and therefore non-wastefulness in this context 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 $c \in \mathcal{C}$,*

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

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 are either explicitly or implicitly required. Hence, we see these three axioms as a minimal requirement for reserve systems. 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).

3.1 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. (See, for example, Appendix Figure A1.) 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 $f_c \ \underline{\pi}_c \ \emptyset$. Cutoffs in our model play the role of prices for exchange or production economies, where the last condition corresponds to prices being non-negative 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

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

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) $\mathcal{B}_i(f) \neq \emptyset \quad \Longrightarrow \quad \mu(i) \in \mathcal{B}_i(f)$.
2. For every category $c \in \mathcal{C}$,

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

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

1. each patient who has a non-empty 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 which 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 non-wasteful, 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 non-wasteful, and respects priorities.*

There can be multiple equilibrium cutoff vectors that supports 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

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

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

Here,

- $\overline{f}_c^{\mu}$ 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

- $\underline{f}_c^{\mu}$ 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,

$$\overline{f}_c^{\mu} \; \underline{\pi}_c \; \underline{f}_c^{\mu} \qquad \text{for any } c \in \mathcal{C}.$$

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 non-wasteful, and respects priorities. Then the pair (g, μ) is a cutoff equilibrium if, and only if,*

$$\overline{f}_c^{\mu} \; \underline{\pi}_c \; g_c \; \underline{\pi}_c \; \underline{f}_c^{\mu} \qquad \textit{for any } \; c \in \mathcal{C}.$$

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

3.2 Characterization via Deferred Acceptance Algorithm

Although Theorem 1 gives a full characterization of matchings that satisfy our three axioms, it leaves the question of how to find such a matching open. 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 follows 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 $i \in I$,

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

While 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^1_c 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 proposed 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:

Theorem 2. *A matching complies with eligibility requirements, is non-wasteful, and it 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, it also provides a concrete procedure to calculate all such matchings. Equivalently, Theorem 2 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.

3.3 Sequential Reserve Matching

An interpretation of the DA-induced matchings is helpful to motivate in 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 the DA algorithm 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, while there is a potentially large set of matchings that satisfy our three axioms, not all are necessarily obtained through an intuitive procedure. This may be a challenge especially in the context of medical rationing, since *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 Theorem 2.

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** $\varphi_{\triangleright}$, is a matching that is constructed as follows:

Suppose categories are ordered under $\triangleright$ as

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

Matching $\varphi_{\triangleright}$ 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 less than r_{c_1} eligible patients in I than 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 $I \setminus \cup_{k'=1}^{k-1} I^{k'}$ are matched with category c_k. If there are less than r_{c_k} eligible patients in $I \setminus \cup_{k'=1}^{k-1} I^{k'}$ 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 non-wasteful, and it respect priorities. Thus, it is DA-induced by Theorem 2. Indeed it corresponds to a very specific DA-induced matching.

Proposition 1. *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 $c, c' \in \mathcal{C}$,*

$$c \succ_i^{\triangleright} c' \quad \Longleftrightarrow \quad c \triangleright c'.$$

Then the sequential reserve matching $\varphi_{\triangleright}$ 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 2. *Fix a preferential treatment category $c \in \mathcal{C}$, 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* $c^* \in \mathcal{C} \setminus \{c, c'\}$

$$\hat{c} \rhd c^* \quad \Longleftrightarrow \quad \hat{c} \rhd' c^*.$$

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

$$\overline{f}_c^{\varphi_{\rhd'}} \underline{\pi}_c \overline{f}_c^{\varphi_{\rhd}}.$$

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 2. This result is intuitive because the earlier a category is processed, the larger is the set of patients who compete for these units in a setwise inclusion sense.

4 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, Yenmez and Yildirim (2013), the earlier market design literature on reserve systems exclusively considered this environment.[11] This priority order may depend on scores in a standardized exam, a random lottery, or arrival time of application. In our main application of pandemic resource allocation, it may depend on SOFA scores described in Section 2.1. 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.

To formulate this subclass, 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** 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 $I_c \subset I$ of beneficiaries and it is endowed with a priority order π_c with the following structure: for any pair of patients $i, i' \in I$,

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

[11] Including Hafalir, Yenmez and Yildirim (2013), several prior studies consider the allocation of heterogeneous objects and the baseline priority often depends on the object.

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:

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

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. $i \;\; \pi_c \;\; \emptyset$ for any of its beneficiaries $i \in I_c$, whereas
2. $\emptyset \;\; \pi_c \;\; i$ 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 H1-B visa allocation in the US.

4.1 Comparative Statics

Allocation rules based on sequential reserve matching are used in a wide range of practical applications. While an aspect that is 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 that is the relevant case for our main application of pandemic rationing.

We already know from Proposition 2 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 benefits from this comparative static exercise. The answer would be of course 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 2. 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

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

while 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:

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

and

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

In the following table, we demonstrate the construction of the two induced sequential reserve matchings by processing their mechanics in parallel:

	Order of Precedence $\triangleright$		*Order of Precedence* $\triangleright'$	
Step	*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

Thus the two sequential reserve matchings match the patients

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

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 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 $\triangleright$.

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 for our main application of soft reserves provided that there are at most five categories and each patient is a beneficiary of at most one preferential treatment category.

Proposition 3. *Assuming*

1. *there are at most five categories, and*
2. *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 $c^* \in \mathcal{C} \setminus \{c, c'\}$,*

$$\hat{c} \triangleright c^* \quad \Longleftrightarrow \quad \hat{c} \triangleright' c^*.$$

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

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

4.2 Smart Reserve Matching

Although virtually all practical applications of reserve systems are implemented through sequential reserve matching, this class of mechanisms may suffer from one 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 2. *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., $r_c = r_u = 1$). Since the reserves are hard, the resulting category-specific priority orders are given as follows:

$$i_1 \; \pi_u \; i_2 \; \pi_u \; \emptyset \qquad \textit{and} \qquad i_1 \; \pi_c \; \emptyset \; \pi_c \; i_2.$$

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

$$u \; \triangleright \; c.$$

Under matching $\varphi_{\triangleright}$, first patient i_1 is matched with the unreserved category u and subsequently the category-c unit is left idle since 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 $\varphi_{\triangleright'}(I) = \{i_1, i_2\}$. Since $\varphi_{\triangleright}(I) \subsetneq \varphi_{\triangleright'}(I)$, matching $\varphi_{\triangleright}$ is Pareto dominated by matching $\varphi_{\triangleright'}$.

Observe that Example 2 also illustrates that the cause of Pareto inefficiency is the myopic allocation of categories under sequential reserve matchings. Under matching $\varphi_{\triangleright}$, the more flexible unreserved unit is allocated to patient i_1 who is the only beneficiary of category-c. This results

in a suboptimal utilizations of reserves, which can be avoided with the concept of "smart reserve matching" we introduce below.

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

Definition 4. *A matching $\mu \in \mathcal{M}$ is* **maximal in beneficiary assignment** *if*

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

This axiom simply requires that the reserves should be maximally assigned to target beneficiaries to the extent it is feasible. It precludes the myopic assignment of categories to patients since 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 2 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 for soft reserves because sub-optimal utilization of reserves may receive heightened scrutiny. For example, consider a scenario with two preferential treatment categories, essential personnel and disadvantaged, each with one 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 only one of the reserves is assigned to members of the target beneficiary groups. This outcome could be seen problematic since 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 an algorithm that generates smart cutoff matchings:

Smart Reserve Matching Algorithm

Fix a parameter $n \in \{0, 1, \ldots, r_u\}$ that represents the number of unreserved units to be processed in the beginning of the algorithm.[12] 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

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

Proceed in two steps.

Step 1: Iteratively construct two sequences of patient sets $J_0^u \subseteq J_1^u \subseteq \ldots \subseteq J_{|I|}^u$, which determine patients to be matched with the unreserved category u in this step, and $J_0 \subseteq J_1 \subseteq \ldots \subseteq J_{|I|}$, which determine the patients to be matched with

[12] For $n = 0$, this algorithm is equivalent to the horizontal envelope algorithm in Sönmez and Yenmez (2020).

preferential treatment categories in $\mathcal{C} \setminus \{u\}$ that they are beneficiaries of, and a sequence of sets of matchings $\mathcal{M}_0 \supseteq \mathcal{M}_1 \supseteq \ldots \supseteq \mathcal{M}_{|I|}$ in $|I|$ substeps.

Define Step 1.(k) for any $k \in \{1, 2, \ldots, |I|\}$ 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

$$\mathcal{M}_0 = \arg\max_{\nu \in \mathcal{M}} \big| \cup_{c \in \mathcal{C} \setminus \{u\}} (\nu^{-1}(c) \cap I_c) \big|.$$

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

  $$J_k^u = J_{k-1}^u \cup \{i_k\}, \quad J_k = J_{k-1}, \quad \text{and} \quad \mathcal{M}_k = \big\{\mu \in \mathcal{M}_{k-1} : \mu(i_k) = u\big\}.$$

- 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

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

- 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.

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.

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

Lemma 2. *Consider either a soft reserve system or a hard reserve system induced by a baseline priority order π. For any $n \in \{0, 1, \ldots, r_u\}$ and any two smart reserve matchings $\sigma, \nu \in \mathcal{M}_S^n$,*

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

and moreover,

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

Lemma 2 states that 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.[13] By Lemma 2, for any $\sigma \in \mathcal{M}_S^n$,

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

Our second result on smart reserve matchings is as follows:

Proposition 4. *Consider either a soft reserve system or a hard reserve system induced by a baseline priority order π. For any $n \in \{0, 1, \ldots, r_u\}$, any smart reserve matching in $\mathcal{M}_S^n$ complies with eligibility requirements, is non-wasteful, 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.

Theorem 3. *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,*

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

Theorem 3 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.

[13] We describe a polynomial-time algorithm to construct I_S^n and a smart reserve matching in Appendix A.4.

Theorem 3 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.

4.3 Related Theoretical Literature

Our study of reserve systems contributes to literature in matching market design focused on distributional issues. Our main results of Theorems 1-3 as well as Propositions 1 and 2 have no antecedents in the literature, and they are novel to this paper. In contrast, Propositions 3 and 4 extend previously known ideas to our model.

While Theorem 1 is novel and our paper is the first one to formally introduce the notion of cutoff equilibrium for reserve systems, due to its intuitive appeal 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, Pathak and Sönmez, 2020), assignment of government positions in India (Sönmez and Yenmez, 2019*a*,*b*), college admissions in Brazil (Aygün and Bo, 2020), and H1-B visa allocation in the US for years 2006-2008 (Pathak, Rees-Jones and Sönmez, 2020). 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. While 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), Azevedo and Leshno (2016), and Leshno and Lo (2020)).

In addition to presenting a characterization of outcomes that satisfy three basic axioms, Theorem 2 also provides a procedure to calculate all cutoff matchings. While 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 algorithm for Engineering Colleges in India (Baswana et al., 2019).

Our characterizations in Theorems 1 and 2 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 2. This can be mitigated by filling reserves in a "smart" way. In Section 4.2 we introduce a class of smart reserve matching algorithms which 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 horizontal envelope algorithm introduced by Sönmez and Yenmez (2020). under exclusively "minimum guarantee" type reserves. Moreover, for $n = 0$ the smart reserve matching algorithm becomes equivalent to the horizontal envelope algorithm.

Of the two analytical results that extend previously known ideas in the literature, Proposition 3 extends the comparative static results of Dur et al. (2018); Dur, Pathak and Sönmez (2020); Pathak, Rees-Jones and Sönmez (2020). What is perhaps more novel is Example 1, which shows that the comparative static exercise in Proposition 3 fails to hold once there are more than five categories. Another implication of this situation is that the proof of Proposition 3 is considerably more involved than its predecessors. Proposition 4, on the other hand, extends the analysis on horizontal envelope algorithm in Sönmez and Yenmez (2020) to the more general smart reserve matching algorithm.

In addition to above described papers which directly relate to our analysis, there are also several others that have examined allocation in the presence of various distributional constraints such as minimum-guarantee reserves (or lower quotas), upper quotas, and regional quotas. Some of the most related ones include Abdulkadiroğlu (2005), Biro et al. (2010), Kojima (2012), Budish et al. (2013), Hafalir, Yenmez and Yildirim (2013), Westkamp (2013), Ehlers et al. (2014), Echenique and Yenmez (2015), Kamada and Kojima (2015), Kamada and Kojima (2017) Kamada and Kojima (2018), Aygün and Turhan (2020), Bo (2016), Dogan (2016), Kominers and Sönmez (2016), and Fragiadakis and Troyan (2017).

Our paper also introduces the medical rationing into the market design literature. By considering a real-world resource allocation problem, we contribute to the study of formal properties of specific allocation processes in the field and the study of alternative mechanisms. Studies in this vein include those on entry-level labor markets (Roth, 1984; Roth and Peranson, 1999), school choice (Balinski and Sönmez, 1999; Abdulkadiroğlu and Sönmez, 2003), spectrum auctions (Milgrom, 2000), kidney exchange (Roth, Sönmez and Ünver, 2004), internet auctions (Edelman, Ostrovsky and Schwarz, 2007; Varian, 2007), course allocation (Sönmez and Ünver, 2010; Budish, 2011), cadet-branch matching (Sönmez and Switzer, 2013; Sönmez, 2013), assignment of airport arrival slots (Schummer and Vohra, 2013; Schummer and Abizada, 2017), and refugee resettlement (Jones and Teytelboym, 2017; Delacrétaz, Kominers and Teytelboym, 2019; Andersson, 2019).

5 Additional Design Considerations for Pandemic Rationing

5.1 Dynamics of Allocation

In Sections 3 and 4, 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?*

Since most guidelines are on rationing of vaccines, ventilators, ICU beds, and anti-viral drugs or treatments, we focus our discussion on these four cases.

5.1.1 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-serve 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. While many have criticized first-come-first-serve allocation because of biases it induces based on access to health care (e.g., Kinlaw and Levine (2007)), reserve categories can be designed to mitigate these biases, even if priority is first-come-first-serve 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 2005-2008, H-1B immigration visas in the US 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, Rees-Jones and Sönmez, 2020).

5.1.2 Ventilator/ICU Bed Allocation

Since 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: since a ventilator is durable and assigned to a patient for a period, it can be interpreted as a good which 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) describes 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). While patient data needs 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.

5.1.3 Anti-viral Drugs or Treatments

For anti-virals 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-serve 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 5.1.1. Alternatively, drug assignment can be batched within pre-specified time-windows. Drugs can then be assigned based on expectations of the number of patients in each category over this time window. Since 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, some of the authors assisted with the design of the system used at the University of Pittsburgh Medical Center to allocate the anti-viral drug remdesivir in May 2020 with this implementation. The system had special provisions for hardest hit and essential personnel and used lotteries for prioritization (see White et al. (2020) for more details on this system.)

5.2 Procedural Dimensions of a Reserve System

A common theme of several medical rationing guidelines is the importance of the following procedural dimensions: transparency and community engagement; accountability and responsi-

bility; and adaptability and flexibility (WHO, 2007; Prehn and Vawter, 2008). We next briefly describe some virtues of a reserve system along these dimensions.

Task forces which may include medical personnel, ethicists, and lawyers often decide on prioritization. There is a widespread consensus about the need to accommodate these different perspectives. For example, a John Hopkins study on community perceptions of medical rationing states (Biddison et al., 2013):

> Both groups felt strongly that no single principle could adequately balance the competing aims and values triggered by allocation decisions. Some felt that a combination of principles should be used.

We think that a clear link between a reserve category and a particular ethical principle may help with community engagement and facilitate compromises between different stakeholders. For example, suppose that a task force values reciprocity and instrumental value, but several other goals as well. Then they could establish a reserve category for essential personnel for a fraction of resources and use other reserve categories to balance these other goals.

The transparency associated with reserve categories may be one reason systems are often used in settings that involve community engagement. For example, following debates between the pro-neighborhood and pro-choice factions, Boston's school assignment system established a reserve where half of each school's seats prioritize applicants from the walk zone in 1999. Likewise, reserves were developed as part of India's affirmative action system after more than a decade of community involvement summarized in the 1979 Mandal Commission Report and formulated in the landmark 1992 *Indra Sahwney* Supreme Court case. Similarly, after widespread community outreach, OPTN introduced a reserve category for kidney allocation in 2014. This category gives preferential treatment to patients with the highest expected benefit for 20% of the highest quality deceased-donor kidneys. These reserve-system precedents suggest that stakeholders find reserve categories easy to interpret.

The salience of reserve categories also helps with accountability and fostering public trust. In a priority point system, by contrast, it may be more difficult to explain exactly how an allocation rule reflects a particular value when points are aggregated across several dimensions. Moreover, our result relating reserve systems to cutoff equilibrium also simplify the process of explanation in situations where a patient is denied a resource. For example, if a patient is denied an anti-viral, a clinician can show that the patient's score was below the cutoffs for any category in which the patient is eligible.

Reserve systems also allow for adaptability and flexibility because of the way the allocation problem is divided into smaller pieces. Within each piece, a prioritization decision can be adjusted without the need to alter the categories. Emanuel et al. (2020) emphasizes that "prioritization guidelines should differ by intervention and should respond to changing scientific evidence." In particular, vaccines and anti-viral drugs may rely on different ethical principles than ventilators and ICU beds. We see this dimension of flexibility as particularly valuable given the need for guidelines to evolve during different stages of a public health emergency. Moreover, for certain resources, criteria may evolve as clinical information becomes available. For example, when there is no information on clinical effectiveness of an anti-viral, it will be difficult chose a

priority to maximize expected health outcomes because those are unknown; this may motivate a lottery within the category. As clinical data emerge, however, these priorities can change, without requiring a change in the category.[14]

5.3 Potential Reserve Categories in Current Debates

We next briefly describe some possible reserve categories and their relationship to some current debates.

5.3.1 Essential Personnel Category

As we discussed in Section 2.2.1, there is widespread agreement on the desirability of prioritizing essential personnel. Yet, Michigan is the only state to provide clear priority for essential personnel. Several states struggled to find a way to prioritize essential personnel, which led some to completely abandon it, others to be vague, and still others to use essential personnel status only as a tie-breaker. Adoption of a reserve system that includes a category giving preferential treatment to essential personnel provides a tool to accommodate the underlying ethical principles of reciprocity and instrumental value.

5.3.2 Disadvantaged Reserve Category

Public health emergencies can have differential impact across communities, and there are calls for rationing guidelines to respond to differential incidence. For example, growing evidence suggests that COVID-19 is hitting communities of color harder than other groups (see, e.g., Price-Haywood et al. (2020) and Sequist (2020)). These concerns have motivated criticisms of existing rationing guidelines. Indeed, shortly after Massachusetts released their revised crisis standards of care in April 2020, Manchanda, Couillard and Sivashanker (2020) criticize them for exacerbating biases in expected health outcomes driven by discrimination in access to health care or other social inequalities. For vaccines, Schmidt (2020) argues that rationing guidelines should give priority to groups that have been structurally and historically disadvantaged.[15] In fact, Melinda Gates, a major benefactor of the Global Alliance for Vaccines and Immunization, emphasized that after health care workers, there should be tiered vaccine allocation to "black people next, quite honestly, and many other people of color" (Ducharme, 2020). A reserve system offers flexibility to accommodate these concerns. A portion of scarce resources could be set aside in the form of a *disadvantaged category* based on legally-permissible measures of disadvantage.

[14] With either lottery or non-lottery based prioritization, it is also possible to use a reserve system to measure the effectiveness of the rationed medical resource (see Abdulkadiroglu et al. (2017) and Abdulkadiroglu et al. (2019)).

[15] In fact, following circulation of the first draft of this paper, Schmidt (2020) advocated for a reserve system for vaccines.

5.3.3 Disabled Reserve Categories

Disabilities rights advocates have opposed rationing plans based on expected health outcomes using survival probabilities because such criteria are inherently discriminatory. Ne'eman (2020) argues that provisions that exclude certain groups can undermine overall trust in the medical system "based on a well-founded fear of being sacrificed for the greater good."[16] Persad (2020) recounts that several prefer either random selection or minimal triage that completely ignores any differences in likelihood or magnitude of benefit, or the likely quantity of resources required for benefit. A reserve system allows for a resolution of this dispute. In particular, a *disabled category* can be established for disabled patients reserving some of the units for these groups. If the representatives of these groups reach a decision to implement random lottery within disabled patients for these units, this can be implemented under a reserve system without interfering with the priority order for units in other categories.

5.3.4 Good Samaritan Reciprocity Category

Another possible category is a *Good Samaritan reciprocity category*, which reserves a fraction of resources for those who have verified Good Samaritan acts. In such a category, a small fraction of resources are reserved for those who have saved lives through their past Good Samaritan acts. These could be participants for clinical trials on vaccine or treatment development (Emanuel et al., 2020), altruistic donors who have donated their kidneys to a stranger, or people who have donated large quantities of blood over the years. Good Samaritan status can also be provided for compatible patient-donor pairs who voluntarily participate in kidney exchange even though they do not have to, and save another patient's life who was incompatible with his/her donor. This type of incentive could save a large number of lives. Sönmez, Ünver and Yenmez (2020) estimate 180 additional kidney patients could receive living donor transplants for every 10 percent of compatible pairs who can be incentivized to participate in kidney exchange. Community engagement exercises can determine which acts "deserve" a Good Samaritan status. In addition to the ethical value of reciprocity, this category can also be motivated from incentives. The mere existence of a modest reserve of this nature may mitigate more persistent and ongoing crises in other healthcare domains through these incentives.

5.3.5 Expanding Supply through a Reserve System

While our analysis pertains to the rationing problem of a single entity, like a hospital system, it can be extended to multiple entities. This extension would allow for considerations that can also reduce waste in the system. For example, hospitals in the system can "loan" their unused units to the system, say to a *virtual hospital* that consists of excess units loaned to the system, and they can earn credit from the system for future use of the units at the virtual hospital when they have a shortage. Hospitals can be incentivized to loan their unused units to the virtual hospital,

[16]The Disability Rights Education and Defense Fund argues that the probability of survival should not be considered as long as it is positive (DREDF, 2020). That is, they claim that the "use of a disability to deny or limit an individual's access to health care or to provide them a lower priority in accessing scarce resources or supplies constitutes a clear violation of disability nondiscrimination law."

if their patients receive some priority for some of the units in the virtual hospital. There can be a specific reserve category where priorities may depend on credits earned by the hospitals, while for another category priorities may be determined by clinical criteria only.

6 Conclusion

Because of the anticipated and ongoing shortage of key medical resources during public health emergencies, several leaders in the medical ethics community have made important recommendations regarding medical rationing. These recommendations reflect compromises between several ethical principles – maximizing lives, maximizing life-years, life-cycle considerations, instrumental values, reciprocity, protecting to the sickest, and non-exclusion. We have argued that a reserve system offers additional flexibility to balance competing objectives.

The theory of reserve systems we develop is based on three simple criteria. Our formal results relate these abstract principles to a cutoff equilibrium concept, and we provide a complete characterization of all cutoff equilibria through a hypothetical two-sided matching market. We then focus on an intuitive class of reserve matchings and analyze comparative static properties within this class. Finally, we develop several formal results for scenarios where there is a baseline priority order.

Aside from our main motivation of pandemic rationing, our results can be directly applied to several other resource allocation problems with reserves. These include immigration visa allocation in the United States (Pathak, Rees-Jones and Sönmez, 2020), affirmative action in school choice systems in Boston, Chicago, New York City and Chile (Dur et al., 2018; Dur, Pathak and Sönmez, 2020; Correa et al., 2019), affirmative action for public school and government positions in India (Aygün and Turhan, 2017; Baswana et al., 2019; Sönmez and Yenmez, 2019*a*,*b*), and diversity plans for college admissions in Brazil (Aygün and Bo, 2020). We leave explorations of these connections to future research.

We hope that the reserve system we have analyzed will rarely be deployed. However, even if rationing guidelines are rarely applied, their mere existence reflects a statement of values. Several aspects of the design, including those related to essential personnel, disadvantaged and disabled communities, adults and children run the risk of upsetting social harmony. For example, Fink (2020) describes a risk with poorly designed systems excluding certain principles: "at the end you have got a society at war with itself. Some people are going to be told they don't matter enough." We have shown that a reserve system provides policy makers with an additional tool to navigate these complex challenges. Although confronting scarcity in life-and-death situations is dire, we also hope some of our ideas, such as creating a Good Samaritan category, hold the potential to make progress on longstanding medical problems beyond immediate public health emergencies.

Appendix A Proofs (Online Appendix)

A.1 Proofs of Results in Section 3

Proof of Theorem 1.

Part 1. Suppose matching $\mu \in \mathcal{M}$ complies with eligibility requirements, is non-wasteful, 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. Since μ complies with eligibility requirements, then $f_c \; \pi_c \; \emptyset$. On the other hand, if $|\mu^{-1}(c)| < r_c$, then $f_c = \emptyset$. Therefore, in either case $f_c \; \underline{\pi}_c \; \emptyset$. We showed that $f \in \mathcal{F}$, i.e., 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. Since μ complies with eligibility requirements, $i \; \pi_c \; f_c = \emptyset$. Thus, $\mu(i) \in \mathcal{B}_i(f)$.

(b) We show that $\mathcal{B}_i(f) \neq \emptyset \implies \mu(i) \in \mathcal{B}_i(f)$. 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 $c \in \mathcal{B}_i(f)$. Since 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, $f_c = \emptyset$. Since $c \in \mathcal{B}_i(f)$, $i \; \pi_c \; f_c = \emptyset$. These two statements together with $\mu(i) = \emptyset$ contradict non-wastefulness 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 non-wastefulness, and respects priorities.

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

Non-wastefulness: 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

$\mathcal{B}_i(f) = \emptyset$. 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 non-wasteful.

Respect of Priorities: Let patient $i \in I$ be such that for some category $c \in \mathcal{C}$, $\mu(i) = c$ while 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, $\mathcal{B}_j(f) = \emptyset$. In particular, $f_c \; \pi_c \; j$. Since $\mu(i) = c$, 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, $i \; \pi_c \; j$. ■

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

Claim 1. $\overline{f}^{\mu}$ is the maximum equilibrium cutoff vector supporting μ, i.e., $(\overline{f}^{\mu}, \mu)$ is a cutoff equilibrium and for every cutoff equilibrium (f, μ), $\overline{f}^{\mu}_c \; \underline{\pi}_c \; f_c$ for every $c \in \mathcal{C}$.

Proof. We prove the claim in two parts.
Part 1.We show that $(\overline{f}^{\mu}, \mu)$ is a cutoff equilibrium:

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

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

By this definition $\overline{f}^{\mu} \in \mathcal{F}$. 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, $i \; \underline{\pi}_c \; \overline{f}^{\mu}_c \in \{\emptyset, \min_{\pi_c} \mu^{-1}(c)\}$, and hence, $\mu(i) \in \mathcal{B}_i(\overline{f}^{\mu})$, 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 $\mu(i) \notin \mathcal{B}_i(\overline{f}^{\mu})$. Thus, $\mu(i) = \emptyset$ by Condition 1(a). Let $c \in \mathcal{C}$. If $|\mu^{-1}(c)| < r_c$, then $\overline{f}^{\mu}_c = \emptyset \; \pi_c \; i$ by non-wastefulness 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, $\overline{f}^{\mu}_c = \min_{\pi_c} \mu^{-1}(c) \; \pi_c \; i$. In either case, we have $c \notin \mathcal{B}_i(\overline{f}^{\mu})$. Thus, we get $\mathcal{B}_i(\overline{f}^{\mu}) = \emptyset$, showing that Condition 1(b) also holds for $(\overline{f}^{\mu}, \mu)$, and hence, completing the proof that $(\overline{f}^{\mu}, \mu)$ is a cutoff equilibrium.

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

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

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

Proof. We prove the claim in two parts.

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

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

$$\underline{f}^{\mu}_{c} = \begin{cases} \min_{\pi_c} \left\{ i \in \mu(I) : i \; \pi_c \; x_c \right\} & \text{if } x_c \neq \emptyset \\ \emptyset & \text{otherwise} \end{cases}$$

where x_c is defined as

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

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

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

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 non-wastefulness of μ. Thus, $x_c = \emptyset$. This implies $\underline{f}^{\mu}_{c} = \emptyset$ 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 $i \; \underline{\pi}_c \; \overline{f}^{\mu}_{c}$, since we showed in Claim 1 that $(\overline{f}^{\mu}, \mu)$ 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, $\overline{f}^{\mu}_{c} = \underline{f}^{\mu}_{c} = \emptyset$. Since $i \; \pi_c \; \emptyset$, $c = \mu(i) \in \mathcal{B}_i(\underline{f}^{\mu})$ showing that Condition 1(a) holds for i.
- If $|\mu^{-1}(c)| = r_c$: Then $\overline{f}^{\mu}_{c} = \min_{\pi_c} \mu^{-1}(c) \; \pi_c \; x_c$: as otherwise
  - if $x_c \in I$, then $\mu(x_c) = \emptyset$ (by definition of x_c) and yet $c \in \mathcal{B}_{x_c}(\overline{f}^{\mu})$, a contradiction to $(\overline{f}^{\mu}, \mu)$ being a cutoff equilibrium;
  - if $x_c = \emptyset$, then (i) $x_c \; \pi_c \; \overline{f}^{\mu}_{c} \in I$ contradicts $\overline{f}^{\mu}$ being a cutoff vector, and (ii) $x_c = \overline{f}^{\mu}_{c}$ contradicts $|\mu^{-1}(c)| = r_c$. Thus we cannot have $x_c \; \underline{\pi}_c \; \overline{f}^{\mu}_{c}$ in this case either.

  Thus,

  $$\overline{f}^{\mu}_{c} \; \underline{\pi}_c \; \min_{\pi_c} \{ i \in I : i \; \pi_c \; x_c \} = \underline{f}^{\mu}_{c}.$$

  Then

  $$i \; \underline{\pi}_c \; \overline{f}^{\mu}_{c} \; \underline{\pi}_c \; \underline{f}^{\mu}_{c},$$

  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 $i \; \pi_c \; \emptyset = \underline{f}^{\mu}_{c}$. By definition of x_c, $i \in \mu(I)$, i.e., $\mu(i) \neq \emptyset$.
- if $x_c \neq \emptyset$: Then $i \; \underline{\pi}_c \; \underline{f}^{\mu}_{c} \; \pi_c \; x_c$, which in turn implies that $\mu(i) \neq \emptyset$ by definition of $\underline{f}^{\mu}_{c}$ 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 $(\underline{f}^\mu, \mu)$ is a cutoff equilibrium.

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

Suppose to the contrary of the claim $\underline{f}^\mu_c \ \pi_c \ f_c$ for some $c \in \mathcal{C}$. Now, $\underline{f}^\mu_c$ is a patient, because $f_c \ \underline{\pi}_c \ \emptyset$ by the definition of a cutoff vector. By definition of $\underline{f}^\mu_c$, $x_c \neq \emptyset$ and $\underline{f}^\mu_c \ \pi_c \ x_c$. We cannot have $x_c \ \underline{\pi}_c \ f_c$, 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 $f_c \ \pi_c \ x_c$. Since x_c is eligible for c, $f_c \in I$. Furthermore, $f_c \in \mu(I)$, since $c \in \mathcal{B}_{f_c}(f)$ and (f, μ) is a cutoff equilibrium. Therefore, $f_c \in \{j \in \mu(I) : j \ \pi_c \ x_c\}$. Since $\underline{f}^\mu_c = \min_{\pi_c}\{j \in \mu(I) : j \ \pi_c \ x_c\}$, we have $f_c \ \underline{\pi}_c \ \underline{f}^\mu_c$, contradicting $\underline{f}^\mu_c \ \pi_c \ f_c$. Therefore, such a category c cannot exist, and hence, $\underline{f}^\mu$ is the minimum equilibrium cutoff vector supporting μ. $\diamond$

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

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. Since 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, $f_{c'} \ \underline{\pi}_{c'} \ h_{c'} \ \pi_{c'} \ j = g_{c'}$, 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 Condition 1(b) holds in the definition of cutoff equilibrium for (g', μ).

Since (g, μ) is a cutoff equilibrium (specifically Conditions 1(a) and 1(b) of the definition) imply that for every $i \in I$, $\mu(i) = \emptyset \iff \mathcal{B}_i(g) = \emptyset$. 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) that we showed above 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 Condition 2 is also satisfied proving that for every $c \in C$, $g'_c = \emptyset$ 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. $\diamond$ ■

Proof of Theorem 2.

***Sufficiency*:** We first prove that any DA-induced matching complies with eligibility requirements, is non-wasteful, 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.

Non-wastefulness: 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, 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: $\left|\mu^{-1}(c)\right| = r_c$. Hence, matching μ is non-wasteful.

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 $i \; \pi_c \; \emptyset$. Since $\mu(i') = \emptyset$, 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 $c \in \mathcal{C}$. Since 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. Since μ respects priorities, j has a lower priority than any patient in $\mu^{-1}(c)$, who also applied to c in Step 1. Furthermore, since μ is non-wasteful, $\left|\mu^{-1}(c)\right| = r_c$ (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. Since μ respects priorities and is non-wasteful, 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 non-wastefulness of μ, $\left|\mu^{-1}(c)\right| = r_c$, 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 $\succ$. ■

Proof of Proposition 1. Let $\triangleright \in \Delta$ be a precedence order and $\varphi_\triangleright$ be the associated sequential reserve matching. We show that $\varphi_\triangleright$ 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 $c, c' \in \mathcal{C}$,

$$c \succ_i' c' \iff c \triangleright c'.$$

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.

Since for any category $c \in \mathcal{C}$ and any patient $i \in I$ who is ineligible for category c, $\emptyset \; \pi_c \; c$, 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 $\varphi_\triangleright$. We show this by induction. Suppose $\triangleright$ orders categories as $c_1 \triangleright c_2 \triangleright \ldots \triangleright c_{|\mathcal{C}|}$. As the inductive assumption, for $k > 0$, suppose for categories $c_1, \ldots, c_{k-1}$, 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 $\varphi_\triangleright$.

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 since every patient who is not tentatively accepted by a category $c_1, \ldots, c_{k-1}$ 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 $c_1, \ldots, c_{k-1}$ 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 $\varphi_{\triangleright}$. This ends the induction.

Therefore, we conclude that $\varphi_{\triangleright}$ is DA-induced from patient preference profile $\succ^{\triangleright}$. ■

Proof of Proposition 2. Let $J^{\triangleright}, J^{\triangleright'} \subseteq I$ be the sets of patients remaining just before category c is processed under the sequential reserve matching procedure induced by $\triangleright$ and $\triangleright'$, respectively. Since 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), $\overline{f}_c^{\varphi_{\triangleright'}} = \emptyset = \overline{f}_c^{\varphi_{\triangleright}}$.

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

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

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

$$\overline{f}_c^{\varphi_{\triangleright'}} \; \underline{\pi}_c \; \overline{f}_c^{\varphi_{\triangleright}}.$$

■

A.2 Proofs of Lemma 2, Proposition 4, and Theorem 3 in Section 4

Proof of Lemma 2. By the definition of the smart reserve matching algorithm induced by assigning n unreserved units subsequently at the beginning, in Step 1.(k) for every $k \in \{0, 1, \ldots, |I|\}$, and for every 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 $\max\{0, q - |J^u_{|I|} \cup J_{|I|}|\}$ units. Since we assign the patients in $I \setminus (J^u_{|I|} \cup J_{|I|})$ based on priority according to π to the remaining units, we have $\sigma(I)$ is the same patient set regardless of the matching $\sigma \in \mathcal{M}^n_S$ we choose. Since the remaining $r_u - n$ unreserved units are assigned to the lowest priority patients that are matched in any $\sigma \in \mathcal{M}^n_S$, 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 $\sigma \in \mathcal{M}^n_S$ 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 $I \setminus (J^u_{|I|} \cup J_{|I|})$. This concludes proving that each of $\sigma(I)$ and $\sigma^{-1}(u)$ is the same patient set regardless of which $\sigma \in \mathcal{M}^n_S$ we choose.

■

Proof of Proposition 4. For any $n \in \{0, 1, \ldots, r_u\}$, we prove that every smart reserve matching in $\mathcal{M}^n_S$ complies with eligibility requirements, is non-wasteful, 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.

Non-wastefulness: Suppose to the contrary of the claim that there exists some $\sigma \in \mathcal{M}^n_S$ 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 $\sigma \in \mathcal{M}^n_S$ 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: $j \in J^u_{|I|} \cup J_{|I|}$ or not. We show that either case leads to a contradiction, showing that σ respects priorities.

- If $j \in J^u_{|I|} \cup J_{|I|}$: 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 $i' \in I \setminus \{i, j\}$. Since $i, j \in I_{\sigma(j)}$, and we match i instead of j with $\sigma(j)$, $\hat{\sigma}$ is a matching that is maximal in beneficiary assignment as well. Since $i \; \pi \; j$, i is processed before j in Step 1. Let i be processed in some Step 1.(k). Since $\sigma(i) = \emptyset$, $i \notin J_k \cup J^u_k$. 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. Since $\hat{\sigma} \in \mathcal{M}_{k-1}$ and $\hat{\sigma}(i) = u$ this implies $i \in J^u_k$. This contradicts $i \notin J^u_k \cup J_k$.
  - 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. Since $i \in I_{\hat{\sigma}(i)}$ 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. Since 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 $\sigma(j) = u$. Since 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. ■

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

Lemma 3 (Mendelsohn and Dulmage Theorem, 1958)**.** *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 $\nu \in \mathcal{M}^b$ then there exists a matching in $\mathcal{M}^b$ that matches all patients in J and at least $|\nu^{-1}(c)|$ units of each category $c \in \mathcal{C} \setminus \{u\}$.*

See for example page 266 of Schrijver (2003) 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 non-empty list $A = (i_1, \ldots, i_{\overline{m}})$ of patients such that

$$\begin{array}{rcll} \big[i_1 \notin I_{\mu(i_1)} \text{ or } \mu(i_1) = u\big] & \& & \big[i_1 \in I_{\nu(i_1)} \text{ and } \nu(i_1) \neq u\big], & \\ \big[i_m \in I_{\mu(i_m)} \text{ and } \mu(i_m) = \nu(i_{m-1})\big] & \& & \big[i_m \in I_{\nu(i_m)} \text{ and } \nu(i_m) \neq u\big] & \text{for every } m \in \{2, 3, \ldots, \overline{m} - 1\}, \\ \big[i_{\overline{m}} \in I_{\mu(i_{\overline{m}})} \text{ and } \mu(i_{\overline{m}}) = \nu(i_{\overline{m}-1})\big] & \& & \big[i_{\overline{m}} \notin I_{\nu(i_{\overline{m}})} \text{ or } \nu(i_{\overline{m}}) = u\big]. & \end{array}$$

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 $i_{\overline{m}}$ 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 1. *If μ and $\nu \in \mathcal{M}$ are two matchings such that for every $c \in \mathcal{C}\backslash\{u\}$, $|\mu^{-1}(c) \cap I_c| = |\nu^{-1}(c) \cap I_c|$ and there exists some $i \in \left(\cup_{c\in\mathcal{C}\backslash\{u\}} \nu^{-1}(c) \cap I_c\right) \setminus \left(\cup_{c\in\mathcal{C}\backslash\{u\}} \mu^{-1}(c) \cap I_c\right)$, 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 3. By Proposition 4 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 $\overline{f}_u^{\sigma_{r_u}} \underline{\pi} \overline{f}_u^{\mu}$:

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 $\sigma \in \mathcal{M}_S^{r_u^*}(I^*, r^*)$ 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 $k \in \left\{1, \ldots, \left|\nu^{-1}(u) \setminus \sigma^{-1}(u)\right|\right\}$, 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 $|\sigma^{-1}(u)| < |\nu^{-1}(u)|$. Since both σ and ν are maximal in beneficiary assignment for (I^*, r^*), then the 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 $\sigma \in \mathcal{M}^{r^*_u}_S(I^*, r^*)$ 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, since $\sigma(j) \neq u$, either

(i) all units of the unreserved category are assigned under σ to patients with higher π-priority than j, or

(ii) some unreserved category units are still available when j is processed.

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

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

$$\nu(i) = u \iff \sigma(i) = u. \tag{3}$$

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

$$i \,\pi\, j, \quad \sigma(i) \neq u, \quad \text{and} \quad i \notin I_{\nu(i)}, \tag{4}$$

(a) If such a patient i does not exist, then let $\hat{\nu} = \nu$ and $i = j$.

(b) If such a patient i exists, then let her be the highest π-priority patient with the property in Equation 4.

We show that $i \in I_{\sigma(i)}$. Consider the smart reserve matching algorithm. Since $\nu(j) = u$ and $i \,\pi\, j$, just before i is processed in Step 1, there is still at least one unreserved unit available by Equation 3. Since we are processing all unreserved units first and since $\sigma(i) \neq u$, 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. Since 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. $\sigma(i') = u \iff \nu(i') = u$ (by Equation 3 and $i \,\underline{\pi}\, j$),
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 4).

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, since $\sigma \in \mathcal{M}_{k-1}(I^*, r^*)$, 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^*), since σ is. By Lemma 3, there exists a matching $\tilde{\nu} \in \mathcal{M}^b(I^*, r^*)$[17] 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 $\nu(i^*) = u \implies \tilde{\nu}(i^*) = \emptyset$ 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^*), since 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 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) $\hat{\nu}(i') = u$ or $\hat{\nu}(i') = \emptyset$.
   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, i.e., 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 $\hat{\nu}$.
   If $\hat{\sigma}(i') \neq u$, then modify $\hat{\nu}$ by assigning an unreserved unit to i' instead of j: $\hat{\nu}(j) = \emptyset$ and $\hat{\nu}(i') = u$. Otherwise, do not modify $\hat{\nu}$ any further.

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

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 $\hat{\nu}$. Consider the beneficiary alternating path A we discovered above starting with patient i and ending with patient i' from $\hat{\sigma}$ to $\hat{\nu}$: Suppose $A = (i, i_2, \ldots, i_{\overline{m}-1}, i')$. Define

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

Observe that $\sigma' \in \mathcal{M}(I', r')$. The existence of σ' shows that it is feasible to match every patient in $J^u_{|I^*|}(I^*, r^*) \setminus \{i\}$ with u and it is feasible 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 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 $\hat{\nu}$ to (I', r') be ν'. Observe that ν' is a matching for (I', r'). Moreover, it is maximal in beneficiary assignment for (I', r'), since $\hat{\nu}$ is maximal in beneficiary assignment for (I^*, r^*).
Now one of the two following cases holds for ν:

(a) If $\nu(i') \neq u$: Recall that while $\nu(j) = u$, we updated $\hat{\nu}$ so that $\hat{\nu}(i') = u$ and $\hat{\nu}(j) = \emptyset$. Thus, $\nu'(j) = \emptyset$ as well. Since $i \; \pi \; j$, 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 $\nu(i') = u$: Since $i \; \pi \; i'$, 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 $\hat{\nu}(i) = \emptyset$: Recall that $\hat{\nu}(j) = u$. We modify $\hat{\nu}$ further that $\hat{\nu}(i) = u$ and $\hat{\nu}(j) = \emptyset$. 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 ν' be the restrictions of σ and $\hat{\nu}$ to (I', r'), respectively. Since, $\sigma(i) = \hat{\nu}(i) = u$ both σ' and ν' are matchings for (I', r'). Since 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 ν' 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 $\overline{f}_u^{\sigma_{r_u}} \; \underline{\pi} \; \overline{f}_u^{\mu} = \emptyset$. On the other hand, if $|\mu^{-1}(u)| = r_u$, Claim 1 implies that $\overline{f}_u^{\sigma_{r_u}} = \min_\pi \sigma_{r_u}^{-1}(u) \; \underline{\pi} \; \overline{f}_u^{\mu} = \min_\pi \mu^{-1}(u)$.

Proof of $\overline{f}_u^{\mu} \overline{\pi} \overline{f}_u^{\sigma_0}$:

We prove the following claim first.

Claim 2. For any set of patients $I^* \subseteq I$ and any capacity vector $r^* \leq r$, suppose $\nu \in \mathcal{M}(I^*, r^*)$ 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,

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

According to the baseline priority order π, for any $k \in \{1, \ldots, |J' \setminus J|\}$, 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.[18] 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 ν in beneficiary assignment for I^*. Thus, $|J \setminus J'| = |J' \setminus J|$, which implies $J' \setminus J \neq \emptyset$.

  By Lemma, 3 there exists a matching $\hat{\nu} \in \mathcal{M}^b(I^*, r^*)$ that matches patients only with preferential treatment categories that they are beneficiaries of such that $\hat{\nu}$ matches patients in J' and $|\sigma^{-1}(c) \cap I_c|$ units reserved for every preferential treatment category $c \in \mathcal{C} \setminus \{u\}$.[19] Since both ν and σ are maximal in beneficiary assignment for (I^*, r^*), then only patients in J' should be matched under $\hat{\nu}$ and no other patients (as otherwise ν would not be maximal in beneficiary assignment for (I^*, r^*)).

  Since $\hat{\nu}(i) = \emptyset$, $i \in I_{\sigma(i)}$, and $\sigma(i) \neq u$, by Observation 1, there exists a beneficiary alternating path A starting with i from $\hat{\nu}$ to σ and ending with a patient $i' \in I_{\hat{\nu}(i')}$ (and $\hat{\nu}(i') \neq u$ 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

[18] This claim's proof is similar to the proof of Proposition 1 Part 2 in Sönmez and Yenmez (2020).

[19] Although both σ and ν 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 ν and apply Lemma 3 to see such a matching $\hat{\nu}$ exists.

matching algorithm with all unreserved units processed last. Since 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 $\hat{c} = \hat{\nu}(i')$ 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 ν' be the restriction of $\hat{\nu}$ to (I', r'). As $\hat{\nu}(i) = \emptyset$, $\hat{\nu}(i') = \hat{c} \neq u$ such that $i \in I_{\hat{c}}$ and the capacity of $\hat{c}$ is reduced by 1 in the smaller economy, $\nu' \in \mathcal{M}(I', r')$, 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 $\hat{\nu}$ using the beneficiary alternating path A we found before. Recall that A is the beneficiary alternating path from $\hat{\nu}$ to σ beginning with i and ending with i'. Suppose $A = (i, i_2, \ldots, i_{\overline{m}-1}, i')$. Define

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

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, $\sigma' \in \mathcal{M}^0_S(I', r')$.

By the inductive assumption, the claim holds for σ' and ν' 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 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 non-wasteful). 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, i.e., the units assigned to the non-beneficiaries 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 $\overline{f}^{\mu}_u = \overline{f}^{\sigma_0}_u = \emptyset$. If $|\mu^{-1}(u)| = r_u$ then $|\sigma_0^{-1}(u)| \leq r_u$ and $\overline{f}^{\mu}_u = \min_\pi \mu^{-1}(u) \; \underline{\pi} \; \overline{f}^{\sigma_0}_u \in \{\emptyset, \min_\pi \sigma_0^{-1}(u)\}$.

- 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.

■

A.3 Proof of Proposition 3 in Section 4

In this subsection, we first show some lemmas that we will use in the proof of Proposition 3.

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 $\tau : I \to (\mathcal{C} \setminus \{u\}) \cup \{\emptyset\}$ 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 $i \in I_g$, i.e., 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 4. *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* $i \in I(2) \cap I_c$*, then* $i' \,\pi\, i$*.*

Proof of Lemma 4. The first statement in Lemma 4 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 $|I'(2) \setminus I(2)| = |I(2) \setminus I'(2)|$ since $\mu(\tilde{I}) \setminus \mu'(\tilde{I}) = I'(2) \setminus I(2)$ and $\mu'(\tilde{I}) \setminus \mu(\tilde{I}) = I(2) \setminus I'(2)$.

The second statement in Lemma 4 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 since 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 4, 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 4, $|I'(2) \setminus I(2)| = |I(2) \setminus I'(2)| \geq 1$ because $I(2) \setminus I'(2)$ has at least one patient. By the second statement in Lemma 4, $I'(2) \setminus I(2) \subseteq \tilde{I}_c$. Therefore, there exists $i' \in I'(2) \setminus I(2) = \mu(\tilde{I}) \setminus \mu'(\tilde{I})$ such that $i' \in \tilde{I}_c$. Since $i \in \mu'(\tilde{I})$, $i' \notin \mu'(\tilde{I})$, 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 4 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 4. Since $\mu(\tilde{I}_c) \subsetneq \tilde{I}_c$, there are at least r_c patients in $\tilde{I}_c$. Therefore, $\mu'^{-1}(c) \subseteq \tilde{I}_c$, which implies that $i \in \mu'^{-1}(u)$. Since $i' \in I'(2)$ is equivalent to $i' \notin \mu'(\tilde{I})$, we get $i \,\pi\, i'$.

The fifth statement in Lemma 4 follows from $i, i' \in \tilde{I}_c$, $i' \in \mu(\tilde{I})$, and $i \notin \mu(\tilde{I})$. ■

Lemma 5. *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* $\mu = \varphi^{\tilde{I}}_{\triangleright}$*,*
- $I'(2) = \tilde{I} \setminus \mu'(\tilde{I})$ *where* $\mu' = \varphi^{\tilde{I}}_{\triangleright'}$*, 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 5. 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 5 hold trivially. Suppose that $|\tilde{I}_{c'}| < r_{c'}$ for the rest of the proof.

The first statement in Lemma 5 follows as in the proof of the first statement in Lemma 4.

The second statement in Lemma 5 holds because if $i \in \mu^{-1}(c')$ and $i \in \tilde{I}_{c'}$, then $i \in \mu'(\tilde{I})$ since $|\tilde{I}_{c'}| < r_{c'}$. 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 since 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 5 is the same as the proof of the third statement in Lemma 4.

The fourth statement in Lemma 5 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 5. 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 5 is the same as the proof of the fifth statement in Lemma 4. ■

Lemma 6. *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* $\mu = \varphi^{\tilde{I}}_{\triangleright}$,
- $I'(3) = \tilde{I} \setminus \mu'(\tilde{I})$ *where* $\mu' = \varphi^{\tilde{I}}_{\triangleright'}$, *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 6. The first statement in Lemma 6 follows as in the proof of the first statement in Lemma 4. Likewise, the fourth statement in Lemma 6 follows as in the proof of the fifth statement in Lemma 4.

To show the other two statements, we use Lemma 4 and Lemma 5. Let $\hat{\triangleright}, \hat{\triangleright}' \in \Delta(\{c, c'\})$ be such that

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

Let $I(2) = \tilde{I} \setminus \varphi^{\tilde{I}}_{\hat{\triangleright}}(\tilde{I})$ and $I'(2) = \tilde{I} \setminus \varphi^{\tilde{I}}_{\hat{\triangleright}'}(\tilde{I})$. 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 $k = 1, \ldots, r_{c^*}$. 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 $k = 1, \ldots, r_{c^*}$. 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 $\ell = 1, \ldots, k$. 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

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

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 $\triangleright$ while the later 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. Since $i'_k \in J'_k \setminus J_k$, 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 since $i_k \in J_k \cap J'_k$. Therefore, $c^* \neq u$. Since 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.

Since $I(3) = J_{r_{c^*}}$ and $I'(3) = J'_{r_{c^*}}$, Statements 2 and 3 in Lemma 6 follow from Statements (a) and (b) above, respectively. ■

Lemma 7. *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* $i' \pi i$, *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* $i' \pi i$.

Proof of Lemma 7. The first statement in Lemma 7 follows as in the proof of the first statement in Lemma 4. Likewise, the fourth statement in Lemma 7 follows as in the proof of the fifth statement in Lemma 4.

To prove the other two statements, we use Lemma 6. Let $\hat{\triangleright}, \hat{\triangleright}' \in \Delta(\{c, c', c^*\})$ be such that

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

Let $I(3) = \tilde{I} \setminus \varphi_{\hat{\triangleright}}^{\tilde{I}}(\tilde{I})$ and $I'(3) = \tilde{I} \setminus \varphi_{\hat{\triangleright}'}^{\tilde{I}}(\tilde{I})$. 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 7. 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 $\pi_{\tilde{c}}$. Since $\mu(\tilde{I}_c) \subsetneq \tilde{I}_c$, $r_{\tilde{c}}$ 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 $k = 1, \ldots, r_{\tilde{c}}$. 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 $k = 1, \ldots, r_{\tilde{c}}$. 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 $i' \pi i$,

(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 $i' \pi i$

(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 6.

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 $\ell = 1, \ldots, k$. 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 $i'_k \in J'_k \cap J_k$. 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

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

Since $|J'_k| = |J_k|$ 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

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

Since $J_{k+1} \setminus J'_{k+1} \subseteq J_k \setminus J'_k$, 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 $i'_k \in J'_k \cap J_k$. 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), (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 $i_k \; \pi \; i$ 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. Since $i'_k \in J'_k \setminus J_k$, 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 $i' \in J'_k \setminus J_k$ since $J'_{k+1} \setminus J_{k+1} = ((J'_k \setminus J_k) \setminus \{i'_k\}) \cup \{i_k\}$ 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} since $J_{k+1} \subseteq J_k$. 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 since there are no category-$\tilde{c}$ patients in $J'_{k+1} \cup J_{k+1}$ we get $i' \; \pi \; i$.

Since $I(4) = J_{r_{\tilde{c}}}$ and $I'(4) = J'_{r_{\tilde{c}}}$, Statement 2 in Lemma 7 follows from Statement (a) and Statement 3 in Lemma 7 follows from Statements (b) and (c). ■

Lemma 8. *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 $\mu = \varphi^{\tilde{I}}_{\triangleright}$,*
- *$I'(5) = \tilde{I} \setminus \mu'(\tilde{I})$ where $\mu' = \varphi^{\tilde{I}}_{\triangleright'}$, 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 8 follows as in the proof of the first statement in Lemma 4.

To prove the second statement, we use Lemma 7. Let $\hat{\triangleright}, \hat{\triangleright}' \in \Delta(\{c, c', c^*, \tilde{c}\})$ be such that

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

Let $I(4) = \tilde{I} \setminus \varphi^{\tilde{I}}_{\hat{\triangleright}}(\tilde{I})$ and $I'(4) = \tilde{I} \setminus \varphi^{\tilde{I}}_{\hat{\triangleright}'}(\tilde{I})$. 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, $r_{\hat{c}}$ patients with the highest priority with respect to $\pi_{\hat{c}}$ 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 $\pi_{\hat{c}}$. Since $\mu(\tilde{I}_c) \subsetneq \tilde{I}_c$, $r_{\hat{c}}$ 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 $k = 1, \ldots, r_{\hat{c}}$. 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 $k = 1, \ldots, r_{\hat{c}}$. 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 $i' \; \pi_{\hat{c}} \; i$

by mathematical induction on k. These results trivially hold for $k = 1$ by Statements 2, 3, and 4 in Lemma 7.

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 $\ell = 1, \ldots, k$. 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

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

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, $i_k \; \pi_{\hat{c}} \; i$ for any $i \in J_{k+1}$, which together with Statement (b) for k imply Statement (b) for $k+1$.

Since $I(5) = J_{r_{\hat{c}}}$ and $I'(5) = J'_{r_{\hat{c}}}$, Statement 2 in Lemma 8 follows from Statement (a). ■

Proof of Proposition 3. 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 $\varphi^{\tilde{I}}_{\hat{\triangleright}}(\tilde{I}_c) = \tilde{I}_c$ then the result is proven. Therefore, assume that $\varphi^{\tilde{I}}_{\hat{\triangleright}}(\tilde{I}_c) \subsetneq \tilde{I}_c$ 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 4 and 5, we obtain $\varphi^{\tilde{I}}_{\hat{\triangleright}'}(\tilde{I}_c) \subseteq \varphi^{\tilde{I}}_{\hat{\triangleright}}(\tilde{I}_c)$.
- If $k = 3$, then by Lemma 6, we obtain $\varphi^{\tilde{I}}_{\hat{\triangleright}'}(\tilde{I}_c) \subseteq \varphi^{\tilde{I}}_{\hat{\triangleright}}(\tilde{I}_c)$.
- If $k = 4$, then by Lemma 7, we obtain $\varphi^{\tilde{I}}_{\hat{\triangleright}'}(\tilde{I}_c) \subseteq \varphi^{\tilde{I}}_{\hat{\triangleright}}(\tilde{I}_c)$.
- If $k = 5$, then by Lemma 8, we obtain $\varphi^{\tilde{I}}_{\hat{\triangleright}'}(\tilde{I}_c) \subseteq \varphi^{\tilde{I}}_{\hat{\triangleright}}(\tilde{I}_c)$.

These imply that

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

completing the proof. ■

A.4 A Polynomial Time Method for Smart Reserve Matching Procedure

Consider the following algorithm for any n:

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. 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. The solution finds the maximum number of patients who can be matched with a preferential treatment category that they are a 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 $\kappa \gg \epsilon > 0$ such that $\epsilon < 1$ and $\kappa > |I|$ would work. For $k = 1, \ldots, |I|$ we repeat the following substep given J_{k-1}^u, J_{k-1}:

Step 1.(k). Suppose i_k is the patient who is prioritized k^{th} in I according to π.

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 $W_{i,c} \in \mathbb{R}$ 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 $W_{i,x} := \epsilon$.
- If $x \in \mathcal{C}$ and i is not temporarily eligible for x as explained in (i.A), define $W_{i,x} := -\epsilon$.
- If $x = \emptyset$, define $W_{i,x} := 0$.

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

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

using a polynomial algorithm such as the *Hungarian algorithm* (Kuhn, 1955).

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

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

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

E. Otherwise,

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

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 $J^u_{|I|}$ 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) using a polynomial *augmenting alternating paths algorithm* (for example see Hopcroft and Karp, 1973; Karzanov, 1973).

(b) Modify σ as follows:

One at a time assign the remaining units unmatched in σ to the remaining highest priority patient in $I \setminus (J^u_{|I|} \cup J_{|I|})$ 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.

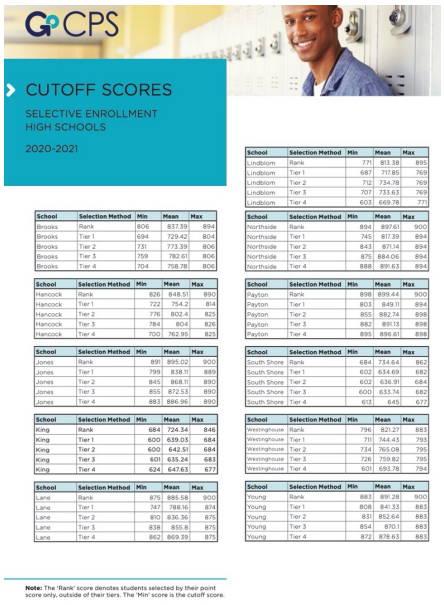
Step 3. Define

$$I^n_S := \sigma(I).$$

Matching σ is a smart reserve matching with n unreserved category units processed first.

The difference between this algorithm and the procedure we gave in the text is that we do not have to construct the matching sets $\mathcal{M}_k$ in every Step 1(k), as this is an NP-complete problem to solve. Instead, we solve appropriately constructed polynomial-time optimization problems in $|I|$ in each step to see whether desired matching exists in each step. As there are polynomial number of steps in $|I|$, the resulting algorithm becomes polynomial.

Appendix B Cutoffs in Real-life Applications



RAJASTHAN PUBLIC SERVICE COMMISSION, AJMER

DATE: 23-11-2012

THE CANDIDATES BEARING THE FOLLOWING ROLL NO. FOR THE RAJASTHAN STATE & SUBORDINATE SERVICES COMBINED COMPETITIVE (PRELIMINARY) EXAMINATION, 2012 HELD ON 14-06-2012 ARE DECLARED **PROVISIONALITY QUALIFIED FOR ADMISSION TO THE MAIN EXAMINATION** IF ANY CANDIDATE IS FOUND THAT HE/SHE DOES NOT FULFILL THE CONDITIONS OF ELIGIBILITY PRESCRIBED AS PER ADVERTISEMENT/RULES, **THE COMMISSION SHALL REJECT HIS/HER CANDIDATURE AT ANY STAGE.**

CUT OFF MARKS

CATEGORY		CUT OFF MARKS
GEN	GEN	218.89
	FEM	185.22
	WD	141.80
	DV	162.31
SC	GEN	210.79
	FEM	151.73
	WD	123.68
	DV	141.94
TSP SC	GEN	184.02
	FEM	
ST	GEN	218.91
	FEM	168.86
	WD	138.16
	DV	191.59
TSP ST	GEN	162.10
	FEM	140.59
OBC	GEN	218.91
	FEM	185.22
	WD	142.16
	DV	164.76
SBC	GEN	207.74
	FEM	171.36
BL/LV		140.68
LD		Already pass in respective category
HI		140.91
NG		Already pass in respective category
DC		140.27
EX		143.09

Quest Education

Figure A1. Examples of Cutoffs in Reserve Systems from Chicago's Affirmative Action System in 2020 and Indian Civil Service Assignment in 2012

References

Abdulkadiroğlu, Atila. 2005. "College Admissions with Affirmative Action." *International Journal of Game Theory*, 33(4): 535–549.

Abdulkadiroğlu, Atila, and Tayfun Sönmez. 2003. "School Choice: A Mechanism Design Approach." *American Economic Review*, 93: 729–747.

Abdulkadiroglu, Atila, Joshua D Angrist, Yusuke Narita, and Parag A Pathak. 2017. "Research Design Meets Market Design: Using Centralized Assignment for Impact Evaluation." *Econometrica*, 85(5): 1373–1432.

Abdulkadiroglu, Atila, Joshua D. Angrist, Yusuke Narita, and Parag A. Pathak. 2019. "Breaking Ties: Regression Discontinuity Design Meets Market Design." Cowles Foundation Discussion Paper 2170.

Andersson, Tommy. 2019. "Refugee Matching as a Market Design Application." *Future of Economic Design*, , ed. J.-F. Laslier, H. Moulin, M.R. Sanver and W.S. Zwicker, 445–450. Springer.

Aygün, Orhan, and Bertan Turhan. 2017. "Large-Scale Affirmative Action in School Choice: Admissions to IITs in India." *American Economic Review, Papers and Proceedings*, 107(5): 210–13.

Aygün, Orhan, and Bertan Turhan. 2020. "Dynamic Reserves in Matching Markets." *Journal of Economic Theory*, 188: 1050–1069.

Aygün, Orhan, and Inacio Bo. 2020. "College Admission with Multidimensional Reserves: The Brazilian Affirmative Action Case." *AEJ: Microeconomics*, forthcoming.

Azevedo, Eduardo, and Jacob Leshno. 2016. "A Supply and Demand Framework for Two-Sided Matching Markets." *Journal of Political Economy*, 124(5): 1235–1268.

Balinski, Michel, and Tayfun Sönmez. 1999. "A Tale of Two Mechanisms: Student Placement." *Journal of Economic Theory*, 84: 73–94.

Baswana, Surender, Partha Pratim Chakrabarti, Sharat Chandran, Yashodhan Kanoria, and Utkarsh Patange. 2019. "Centralized admissions for engineering colleges in India." 49: 307–396, C3.

Bateman, Scot, Monica Bharel, Paul Biddinger, Peter A. DePergola, Lachlan Forrow, Eric Goralnick, Joshua Kornbluth, Abraar Karan, Christine Mitchell, Emily Rubin, Dean Shapley, Sally Vitali, Michael Wagner, Matthias Walz, and Ellen Weinstein. 2020. "Crisis Standards of Care Planning Guidance for the COVID-19 Pandemic." April 7.

Biddison, Elizabeth, Howard Gwon, Monica Schoch-Spana, Robert Cavalier, Douglas B. White, Timothy Dawson, Peter B. Terry, Alex John London, Alan Regenberg, Ruth Faden, and Eric S. Toner. 2013. "The Community Speaks: Understanding Ethical Values in Allocation of Scarce Lifesaving Resources during Disasters." *Annals of the American Thoracic Society*, 11: 777–783.

Biro, P., T. Fleiner, R.W. Irving, and D.F. Manlove. 2010. "The College Admissions Problem with Lower and Common Quotas." *Theoretical Computer Science*, 411(34-26): 3136–3153.

Bo, Inacio. 2016. "Fair Implementation of Diversity in School Choice." *Games and Economic Behavior*, 97: 54–63.

Budish, Eric. 2011. "The Combinatorial Assignment Problem: Approximate Competitive Equilibrium from Equal Incomes." *Journal of Political Economy*, 119(6): 1061–1103.

Budish, Eric, Yeon-Koo Che, Fuhito Kojima, and Paul R. Milgrom. 2013. "Designing Random Allocation Mechanisms: Theory and Applications." *American Economic Review*, 103(2): 585–623.

Carter, Stephen. 2020. "Ventilator Rationing Guidelines Are Discriminatory." Bloomberg, April 10.

CDC. 2018. "Interim Planning Guidance on Allocating and Targeting Pandemic Influenza Vaccine During an Influenza Pandemic." Centers for Disease Control, Available at: `https://www.cdc.gov/flu/pandemic-resources/pdf/2018-Influenza-Guidance.pdf`, Last accessed: July 1, 2020.

CDC. 2020. "Covid-19 in Racial and Ethnic Minority Groups." Available at: `https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/racial-ethnic-minorities.html`, Last accessed: July 20.

Correa, Jose, Rafael Epstein, Juan Escobar, Ignacio Rios, Bastian Bahamondes, Carlos Bonet, Natalie Epstein, Nicolas Aramayo, Martin Castillo, Andres Cristi, and Boris Epstein. 2019. "School Choice in Chile." *EC '19*, 325–343. New York, NY, USA:ACM.

Daugherty-Biddison, Lee, Howard Gwon, Alan Regenberg, Monica Schoch-Spana, and Eric Toner. 2017. "Maryland Framework for the Allocation of Scarce Life-Sustaining Medical Resources in a Catastrophic Public Health Emergency." `https://www.law.umaryland.edu/media/SOL/pdfs/Programs/Health-Law/MHECN/ASR%20Framework_Final.pdf`.

Delacrétaz, David, Scott Duke Kominers, and Alexander Teytelboym. 2019. "Matching Mechanisms for Refugee Resettlement." Working paper.

de Pu Kamp, Majlie, Curt Devine, and Drew Griffin. 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.

DHS, Arizona. 2020. "Covid-19 Addendum: Allocation of Scarce Resources in Acute Care Facilities." Arizona Department of Health Services, June 6, Available: http://azpha.wildapricot.org/resources/covid-19-addendum.pdf. Last accessed: July 4, 2020.

Dittmar, Pia. 2018. "Gender Quotas - Balancing Act Between Efficient Policies and Gap Fillers." EU-Logos Athena.

Dogan, Battal. 2016. "Responsive Affirmative Action in School Choice." *Journal of Economic Theory*, 165: 69–105.

DREDF. 2020. "Preventing Discrimination in the Treatment of Covid-19 Patients: The Illegality of Medical Rationing on the Basis of Disability." March 25, Available at https://dredf.org/wp-content/uploads/2020/03/DREDF-Policy-Statement-on-COVID-19-and-Medical-Rationing-3-25-2020.pdf.

Ducharme, Jamie. 2020. "Melinda Gates Lays out her Biggest Concern for the Next Phase of the Covid-19 Pandemic." Time Magazine, June 4.

Dur, Umut, Parag Pathak, and Tayfun Sönmez. 2020. "Explicit vs. Statistical Preferential Treatment in Affirmative Action: Theory and Evidence from Chicago's Exam Schools." *Journal of Economic Theory*, 187: 104996.

Dur, Umut, Scott Kominers, Parag Pathak, and Tayfun Sönmez. 2018. "Reserve Design: Unintended Consequences and The Demise of Boston's Walk Zones." *Journal of Political Economy*, 126: 2457–2479.

Echenique, Federico, and M. Bumin Yenmez. 2015. "How to Control Controlled School Choice." *American Economic Review*, 105(8): 2679–2694.

Edelman, Benjamin, Michael Ostrovsky, and Michael Schwarz. 2007. "Internet Advertising and the Generalized Second-Price Auction: Selling Billions of Dollars Worth of Keywords." *American Economic Review*, 97(1): 242–259.

Ehlers, Lars, Isa E. Hafalir, M. Bumin Yenmez, and Muhammed A. Yildirim. 2014. "School Choice with Controlled Choice Constraints: Hard Bounds versus Soft Bounds." *Journal of Economic Theory*, 153: 648–683.

Emanuel, Ezekial J., Govind Persad, Ross Upshur, Beatriz Thome, Michael Parker, Aaron Glickman, Cathy Zhang, Conor Boyle, Maxwell Smith, and James Philips. 2020. "Fair Allocation of Scarce Medical Resources in the Time of Covid-19." *New England Journal of Medicine*, March 23.

Fink, Sheri. 2020. "U.S. Civil Rights Office Rejects Rationing Medical Care Based on Disability, Age." New York Times, March 29.

Fragiadakis, Daniel, and Peter Troyan. 2017. "Improving Matching under Hard Distributional Constraints." *Theoretical Economics*, 12(2): 863–908.

Gale, David, and Lloyd S. Shapley. 1962. "College Admissions and the Stability of Marriage." *American Mathematical Monthly*, 69: 9–15.

Hafalir, Isa E., M. Bumin Yenmez, and Muhammed A. Yildirim. 2013. "Effective Affirmative Action in School Choice." *Theoretical Economics*, 8(2): 325–363.

Hopcroft, John E., and Richard M. Karp. 1973. "An $n^{5/2}$ algorithm for maximum matchings in bipartite graphs." *SIAM Journal on Computing*, 2: 225–231.

Israni, A. K., N. Salkowski, S. Gustafson, J. J. Snyder, J. J. Friedewald, R. N. Formica, X. Wang, E. Shteyn, W. Cherikh, D. Stewart, C. J. Samana, A. chung, A. Hart, and B. L. Kasiske. 2014. "New National Allocation Policy for Deceased Donor Kidneys in the United States and Possible Effects on Patient Outcomes." *Journal of the American Society of Nephrology*, 25(8): 1842–1848.

Jones, A. E., S. Trzeciak, and J. A. Kline. 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." *Critical Care Medicine*, 37: 1649–1654.

Jones, Will, and Alexander Teytelboym. 2017. "The Local Refugee Match: Aligning Refugees' Preferences with the Capacities and Priorities of Localities." *Journal of Refugee Studies*, 31(2): 152–178.

Kamada, Yuichiro, and Fuhito Kojima. 2015. "Efficient Matching under Distributional Constraints: Theory and Applications." *American Economic Review*, 105(1): 67–99.

Kamada, Yuichiro, and Fuhito Kojima. 2017. "Stability Concepts in Matching with Distributional Constraints." *Journal of Economic Theory*, 168: 107–142.

Kamada, Yuichiro, and Fuhito Kojima. 2018. "Stability and Strategy-proofness for Matching with Constraints: A Necessary and Sufficient Condition." *Theoretical Economics*, 13(2): 761–794.

Karzanov, A. V. 1973. "An exact estimate of an algorithm for finding a maximum flow, applied to the problem on representatives." *Problems in Cybernetics*, 5: 66–70.

Kinlaw, K, and R Levine. 2007. "Ethical Guidelines in Pandemic Influenza: Recommendations of the Ethics Subcommittee of the Advisory Committee to the Director." Centers for Disease Control and Prevention.

Kojima, Fuhito. 2012. "School Choice: Impossibilities for Affirmative Action." *Games and Economic Behavior*, 75(2): 685–693.

Kominers, Scott Duke, and Tayfun Sönmez. 2016. "Matching with Slot-specific Priorities: Theory." *Theoretical Economics*, 11(2): 683–710.

Kuhn, Harold W. 1955. "The Hungarian Method for the assignment problem." *Naval Research Logistics Quarterly*, 2: 83–97.

Lee, Thomas. 2004. "Rationing Influenza Vaccine." *New England Journal of Medicine*, 351(23).

Leshno, Jacob, and Irene Lo. 2020. "The Cutoff Structure of Top Trading Cycles in School Choice." University of Chicago, Working Paper.

Manchanda, Emily Cleveland, Cheri Couillard, and Karthik Sivashanker. 2020. "Inequity in Crisis Standards of Care." New England Journal of Medicine, May 13.

McCaughey, Betsy. 2020. "The lunatic drive for racial quotas for COVID-19 vaccines." New York Post, July 16.

Mendelsohn, N. S., and A. L. Dulmage. 1958. "Some generalizations of the problem of distinct representatives." *Canadian Journal of Mathematics*, 10: 230–241.

Milgrom, Paul R. 2000. "Putting Auction Theory to Work: The Simultaneous Ascending Auction." *Journal of Political Economy*, 108: 245–272.

Ne'eman, Ari. 2020. "I Will Not Apologize for My Needs." *New York Times*, March 23.

OPTN. 2014. "Organ Procurement and Transplantation Network Policies." Technical Report.

Pathak, Parag, Alex Rees-Jones, and Tayfun Sönmez. 2020. "Immigration Lottery Design: Engineered and Coincidental Consequences of H-1B Reforms." NBER Working Paper 26767.

Pathak, Parag, and Tayfun Sönmez. 2013. "Recommendation on Algorithm Processing, Public Testimony to EAC." Available at: `https://economics.mit.edu/files/16966`.

Persad, Govind. 2020. "Why Disability Law Permits Evidence-Based Triage in a Pandemic." *The Yale Law Journal Forum*, 130: 26–50.

Piscitello, Gina M., William D. Miller, Esha M. Kapania, Juan C. Rojas, Mark Siegler, and William F. Parker. 2020. "Systematic Review of the Variation in U.S. State Ventilator Allocation Guidelines during a Public Health Crisis." JAMA-Network Open.

Prehn, Angela W., and Dorothy Vawter. 2008. "Ethical Guidance for Rationing Scarce Health-Related Resources in a Severe Influenza Pandemic: Literature and Plan Review." Minnesota Pandemic Ethics Project.

Price-Haywood, Eboni G., Jeffrey Burton, Daniel Fort, and Leonardo Seoane. 2020. "Hospitalization and Mortality among Black Patients and White Patients with Covid-19." New England Journal of Medicine, June 25.

Rankin, Jennifer. 2020. "EU revises plans for mandatory quotas of women on company boards." The Guardian, March 5.

Roth, Alvin E. 1984. "The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory." *Journal of Political Economy*, 92: 991–1016.

Roth, Alvin E., and Elliott Peranson. 1999. "The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design." *American Economic Review*, 89: 748–780.

Roth, Alvin E., Tayfun Sönmez, and M. Utku Ünver. 2004. "Kidney Exchange." *Quarterly Journal of Economics*, 119: 457–488.

Schmidt, Harald. 2020. "Vaccine Rationing and the Urgency of Social Justice in the Covid-19 Response." Hastings Center Report, 50: 1-4.

Schoch-Spana, Monica, Joseph Fitzgerald, Bradley R. Kramer, and UMPC Task Force. 2005. "Implications for Biosecurity and Public Health Preparedness." *Biosecurity and Bioterrorism*, 3(3).

Schrijver, Alexander. 2003. *Combinatorial Optimization: Polyhendra and Efficiency.* Vol. A, Springer.

Schummer, James, and Azar Abizada. 2017. "Incentives in Landing Slot Problems." *Journal of Economic Theory*, 170: 29–55.

Schummer, James, and Rakesh V. Vohra. 2013. "Assignment of Arrival Slots." *American Economic Journal: Microeconomics*, 5(2): 164–185.

Sequist, Thomas. 2020. "The Disproportionate Impact of Covid-19 on Communities of Color." New England Journal of Medicine, July 6.

Sönmez, Tayfun. 2013. "Bidding for Army Career Specialties: Improving the ROTC Branching Mechanism." *Journal of Political Economy*, 121(1): 186–219.

Sönmez, Tayfun, and M. Bumin Yenmez. 2019*a*. "Affirmative Action in India via Vertical and Horizontal Reservations." Boston College Working Paper 977.

Sönmez, Tayfun, and M. Bumin Yenmez. 2019*b*. "Constitutional Implementation of Vertical and Horizontal Reservations in India: A Unified Mechanism for Civil Service Allocation and College Admissions." Boston College Working Paper 978.

Sönmez, Tayfun, and M. Bumin Yenmez. 2020. "Affirmative Action with Overlapping Reserves." Boston College Working Paper 990.

Sönmez, Tayfun, and M. Utku Ünver. 2010. "Course Bidding at Business Schools." *International Economic Review*, 51(1): 99–123.

Sönmez, Tayfun, and Tobias Switzer. 2013. "Matching with (Branch-of-Choice) Contracts at the United States Military Academy." *Econometrica*, 81(2): 451–488.

Sönmez, Tayfun, M. Utku Ünver, and M. Bumin Yenmez. 2020. "Incentivized Kidney Exchange." *American Economic Review*, 110(7): 2198–2224.

Temte, Jonathan L. 2006. "Preparing for an Influenza Pandemic: Vaccine Prioritization." *Family Practice Management*, 32–34.

Truog, Robert D., Christine Mitchell, and George Q. Daley. 2020. "The Toughest Triage – Allocating Ventilators in a Pandemic." *New England Journal of Medicine*, March 23.

Twohey, Megan. 2020. "Who should get coronavirus vaccine first? U.S. considers race in a Coronavirus Plans." New York Times, July 9.

Varian, Hal R. 2007. "Position Auctions." *International Journal of Industrial Organization*, 25(6): 1163–1178.

Vawter, Dorothy E., J. Eline Garrett, Karen G. Gervais, Angela Witt Prehn, Debra A. DeBruin, Carol A. Tauer, Elizabeth Parilla, Joan Liaschenko, and Mary Faith Marshall. 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.

Vincent, J. L., R. Moreno, J. Takala, S. Willatts, A. De Mendonça, H. Bruining, C. K. Reinhart, P. M. Suter, and L. G. Thijs. 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 Medicine*, 22: 707–710.

Westkamp, Alexander. 2013. "An Analysis of the German University Admissions System." *Economic Theory*, 53(3): 561–589.

White, Douglas B., Mitchell H. Katz, John M. Luce, and Bernard Lo. 2009. "Who Should Receive Life Support During a Public Health Emergency? Using Ethical Principles to Improve Allocation Decisions." *Annals of Internal Medicine*, 150(2): 132–138.

White, Douglas, Mark Schmidhofer, Erin McCreary, Ryan Bariola, Graham Michael Synyder, Rachel Sackrowitz, Naudia N. Jonassaint, Jessica Daley, Donald M. Yealy, Tayfun Sönmez, M. Utku Ünver, Parag Pathak, Govind Persad, and Robert Truog. 2020. "Model Hospital Policy for Fair Allocation of Medications to Treat Covid-19." University of Pittsburgh, Department of Critical Care Medicine, Available at: `https://www.ccm.pitt.edu/?q=content/model-hospital-policy-fair-allocation-medications-treat-covid-19`. Last accessed: July 4, 2020.

WHO. 2005. "WHO Global Influenza Preparedness Plan." Available at https://www.who.int/csr/resources/publications/influenza/WHO_CDS_CSR_GIP_2005_5.pdf. Last accessed: July 29, 2020.

WHO. 2007. "Ethical considerations in developing a public health response to pandemic influenza." Available at: http://www.who.int/csr/resources/publications/WHO_CDS_EPR_GIP_2007_2/en/index.html.

Whyte, Liz Essley. 2020. "State Policies May Send People with Disabilities to the Back of the Line for Ventilators." April 8, https://publicintegrity.org/health/coronavirus-and-inequality/state-policies-may-send-people-with-disabilities-to-the-back-of-the-line-for-ventilators/.

Zucker, Howard A, Karl P. Adler, Donald P. Berens, Rock Brynner, Karen Butler, Yvette Calderon, Carolyn Corcoran, Nancy Neveloff Dubler, Paul J. Edelson, Joseph J. Fins, Rev. Francis H. Geer, Samuel Gorovitz, Cassandra Henderson, Hassan Khouli, Fr. Joseph W. Koterski, Hugh Maynard-Reid, John Murnane, Karen Porter, Robert Swidler, Sally True, Stuart C. Sherman, Susia A. Han, and Valier Gutmann Koch. 2015. "Ventilator Allocation Guidelines." New York State Task Force on Life and the Law, New York Department of Health.