Patient Loyalty in Hospital Choice: Evidence from New York

Mariano Irace*

January 1, 2018

JOB MARKET PAPER - PLEASE SEE LATEST VERSION AT:
`http://sites.northwestern.edu/mih372`

Abstract

When choosing a hospital, patients favor facilities they have used in the past. Using data from New York, I investigate the sources of patient loyalty to hospitals. To distinguish persistent unobserved heterogeneity and state dependence, I exploit shocks that induce patients to try a new hospital: emergency hospitalizations and temporary hospital closures due to Hurricane Sandy. I find evidence of state dependence under minimal assumptions about the data generating process. State dependence has an impact on health outcomes by preventing the reallocation of patients to high quality hospitals. In the context of hospital choice for heart surgery, patients would switch to hospitals with lower risk-adjusted mortality absent state dependence, leading to a 3% reduction in expected mortality relative to the actual state of the world.

*I would like to express my sincere gratitude to Igal Hendel, Robert Porter, Leemore Dafny, and Gaston Illanes for their patience, motivation, and invaluable advice. This paper has benefited from conversations with Alex Torgovistky, Martin Gaynor, Amanda Starc, Christopher Ody, Frank Limbrock, Matthew Notowidigdo, Elena Prager, and Matias Escudero. I thank participants in the Industrial Organization Student Seminar Group and the Seminar in Industrial Organization at Northwestern University for helpful comments and suggestions. I gratefully acknowledge support from the Becker Friedman Institute's Health Economics Initiative. All errors are mine.

1. Introduction

This paper studies persistence in hospital choices of patients. The conventional wisdom is that patients patronize one hospital and use it for all of their medical care needs. This idea of patient loyalty is consistent with views of industry analysts about the business practices of hospitals. For example, many view maternity services as loss leaders: hospitals offer these services not because maternity patients are profitable *per se*, but because they expect mothers and their families to go back to the hospital for more profitable services in the future. Another example concerns "data blocking" activities. It has been argued (Miller and Tucker, 2009; Miller and Tucker, 2014; HITPC, 2015; Desai, 2016) that hospitals hinder the sharing of patient data with other providers for competitive reasons: patients might find it easier to leave a hospital and seek treatment elsewhere once their clinical data can follow them across providers.

As pointed out by Heckman (1981), the empirical observation that a patient repeatedly uses the same hospital can be explained by either unobserved heterogeneity or state dependence. In the first case, the patient has strong and persistent latent preferences for the hospital. In the second case, previous choices have a causal impact on the current decision. The stories in the previous paragraph rest implicitly on the idea that hospital choices of patients are "sticky", but their implications depend on whether preference heterogeneity or state dependence drives the stickiness in patients' behavior. If the observed choice persistence is due to unobserved heterogeneity, hospitals cannot control the evolution of patients' preferences. In this case, there are no dynamic incentives to invest in unprofitable service lines (e.g. maternity services) in the hope of developing long-standing relationships with patients. Under state dependence, on the other hand, hospitals' investments in loss leader services influence future demand. Similarly, if the continued use of a hospital reflects patients' latent preferences for that provider, then data blocking activities do not have any impact on future demand.

Although the idea of patient loyalty towards hospitals is consistent with the views of people in the industry, there is little supporting evidence. Few studies document persistence in the hospital choices of patients or analyze the determinants of persistence. In particular, whether persistence results from stable preferences or state dependence remains an open question. In this paper, I fill this gap by investigating empirically the determinants and implications of persistence in hospital choices of patients in the state of New York.

The study of persistence in hospital choices of patients is interesting for several reasons. Previous studies have documented the existence of state dependence in consumers' choices in a variety of industries. I provide additional insights about the sources of patients' preferences by analyzing whether patients display similar purchase patterns. Characterizing patient behavior is important to improve our understanding of the forces that drive the allocation of patients to hospitals and to inform the design of policies that influence patient demand. Although there is some market discipline in the hospital industry (in the sense that better performing hospitals have higher and increasing market shares), many patients choose hospitals that are far from the quality frontier (Chandra et al., 2016). I investigate whether state dependence plays a role in this process: does it stand in the way of the reallocation of patients to high quality providers? If so, then, for example, the benefits from better sorting of patients across facilities should be considered when evaluating policies aimed at achieving interoperability of hospitals' electronic health record (EHR) systems. Second, previous use of a hospital is a strong predictor of the current hospital choice of a patient. However, whether this preference is due to state dependence or unobserved heterogeneity matters for welfare analysis and policy evaluation. State dependence implies a less durable preference for a facility that has been used in the past: if the patient is forced to switch hospitals, then she has to pay a one time cost, while unobserved heterogeneity implies that the utility loss from going to a less preferred alternative is permanent. The distinction can affect policy conclusions. For example, if state dependence drives the persistence in patient behavior, the long-run welfare effect of excluding a hospital from an insurer's network will be lower than if persistence reflects unobserved heterogeneity.

In the first part of the paper, the primary concern is to distinguish between persistence in choice due to state dependence and persistence in choice due to unobserved preference heterogeneity. It is difficult for researchers to separate the sources of persistence in a credible way. Most previous studies have relied, at least in part, on functional form assumptions about the nature of preference heterogeneity for identification. The concern is that parametric assumptions might lead to overestimates of the magnitude of state dependence. This is a natural concern in my context, where I expect unobserved heterogeneity to be empirically relevant, given that there are multiple attributes of patients and hospitals that I cannot control for (such as religious affiliation and location of the workplace of the patient, and hospital amenities). Given this, the burden of proof is to show that there is state dependence. I use an event-study approach based on intuitive and easily interpretable assumptions about the nature of the data generating process (dgp). In particular, I exploit quasi-exogenous shocks that force patients to visit a hospital that they would otherwise not have chosen: emergency hospitalizations and temporary hospital closures due to a natural disaster.

In the first case, I find that patients who visit a new hospital during an emergency hospitalization are more likely to continue using that same facility in subsequent episodes than observationally similar patients. Patients who visit the same hospital they had been using before for the emergency hospitalization exhibit a higher repurchase rate, which indicates that unobserved heterogeneity is also empirically relevant. The observed patterns are similar across different types of emergencies, hospitals, and patients.

In the second case, I exploit the unexpected closures of three hospitals in New York City following Hurricane Sandy. I show that patients who needed hospital care during the time an affected hospital was closed for repairs were less likely to use the facility after its reopening than similar patients who did not have hospital visits during the unavailability window. Moreover, patients continue using the hospital they visited during the time their usual hospital was unavailable. I present this evidence in a transparent and simple way, without relying on complex estimators. Next, I cast this setting into the nonparametric framework of Torgovitsky (2016). In this model, each consumer has a type given by a vector of dynamic potential choices. The potential choices in a given period indicate the alternatives the consumer would choose under exogenous manipulations of the previous period choice. If I knew the distribution of types in the population, I could determine the proportion of consumers who exhibit state dependence. However, I do not know this distribution. Therefore, I consider all the distributions that could have generated the data. I restrict the set of admissible distributions to those that are consistent with the distribution of observables and satisfy an exclusion restriction (which takes the form of an independence restriction between preferences and the timing of hospital visits). Then, I search for the admissible distributions for which state dependence is lowest and highest. I am able to bound the proportion of patients who exhibit state dependence away from zero without imposing parametric assumptions about the nature of preference heterogeneity.

Given that there is state dependence in my setting, I next analyze its implications. The focus of most studies of switching costs has been to determine their impact on pricing, by analyzing whether the investment motive (reduce prices to expand the customer base) or the harvesting motive (raise prices to exploit locked-in consumers) dominates (Farrell and Klemperer, 2007). Given the institutional features of the US hospital industry, I expect dynamic pricing to be a less relevant issue in my context[1]. Instead, I focus on the allocative role of state dependence. Lock-in hinders the ability of patients to react to changes in the environment. In particular, lock-in might discourage a patient from seeking treatment in hospitals that are more suitable than her previous choice for treating her current medical condition: any quality differential must compensate for switching costs. Absent state dependence, more patients may obtain treatment at high-quality hospitals.

Hospital quality may be disease-specific. Moreover, quality measures are difficult to compare across different medical conditions. I therefore study the allocative role of state dependence in the context of hospital choice for a specific procedure: heart surgery (Coronary artery bypass grafting, CABG). Ideally, I would exploit the same quasi-exogenous shocks

[1]Dynamic considerations could affect quality choices of hospitals. For example, hospitals could invest in the quality of certain services to attract patients, and then exploit locked-in patients with low quality for other services.

used in the first part of the paper to conduct this analysis. However, small sample sizes prevent me from doing so. I therefore rely on observational data to recover the parameters of interest, following a more traditional but less transparent identification strategy.

I compare the allocation of patients to hospitals in the actual state of the world and in a counterfactual scenario where there is no state dependence. This counterfactual scenario can be thought of as resulting from a combination of policies that target the different sources of state dependence. The allocative role of state dependence is large: previous use of a hospital increases the probability that the patient chooses that facility for CABG more than three times above the baseline. Absent state dependence, patients would switch to higher quality hospitals: in this scenario, ex-ante expected mortality would be 3% lower than the observed mortality rate.

State dependence might arise from a variety of sources. First, patients face monetary and/or time costs of transferring medical records between providers. These switching costs result in part from the data blocking activities mentioned above. Improvements in the interoperability of EHRs and the diffusion of health information exchanges (HIE) have the potential to reduce these costs. More generally, the existence of switching costs is associated with relationship-specific investments that cannot be transferred seamlessly across providers.

Second, state dependence might originate from search and evaluation costs. Choosing a hospital is a complex activity. Patients need to collect information about different alternatives. Hospitals are complicated objects to evaluate, so cognitive limitations might be substantial. Search costs may be such that inertia is the efficient way to deal with moderate or temporary changes in the environment (Stigler and Becker, 1977). Then, state dependence might capture the use of heuristics by patients for choosing a hospital. The presence of search costs also suggests that a patient might not consider all possible alternatives in a choice occasion. Hospitals used in the recent past are more likely to be included in the patient's consideration set (Samuelson and Zeckhauser, 1988; Andrews and Srinivasan, 1995). The asymmetric position in the decision-making process of recently used hospitals means that they are more likely to be chosen.

Third, the presence of state dependence can be explained by learning costs. Uncertainty about the quality of hospitals leads risk averse patients to remain with familiar hospitals. Then, state dependence arises from the premium that patients are willing to pay for greater familiarity with a facility.

It is likely that state dependence arises from a combination of the factors mentioned above. For example, in a two-stage decision process, previous use of a hospital might have an impact on both the consideration and the evaluation stages: hospitals used in the recent past are more likely to enter the consideration set, and patients pay a cost conditional on switching hospitals due to the transfer of medical records between providers.

My analysis focuses on the overall impact of state dependence on patients' hospital choices, without distinguishing between the potential underlying mechanisms. However, disentangling the various sources of state dependence is important in order to craft policies to overcome inertia. Moreover, the welfare implications of eliminating state dependence will be different if it results from a tangible cost, as opposed to something that only affects choices. Data limitations prevent me from decomposing the sources of state dependence in a credible way in the current setting, but this is an interesting avenue for future work.

Previous studies have documented the presence of state dependence in consumers' choices of a variety of products (orange juice and margarine (Dubé et al., 2010), internet portals (Goldfarb, 2006a), pension funds (Luco, 2016; Illanes, 2016), health insurance plans (Nosal, 2012; Handel, 2013; Ericson, 2014; Polyakova, 2015; Ho et al., 2017), among others). There is a more limited number of studies of persistence in patients' choices of medical providers. The papers most closely related to my study are Jung et al. (2011), Shepard (2016), and Raval and Rosenbaum (2017). Jung et al. (2011) study the factors that affect hospital choices of employees at a large self-insured company. They use stated preference data from a survey: employees were asked to indicate the hospitals they would be most likely to consider if they needed to be hospitalized for a surgical procedure. The authors find that prior use of a hospital and patient satisfaction with a facility from prior experiences have a large effect on future hospital choices; the effect of prior use is smaller in cases where the previous admission occurred through the emergency department. While their analysis is based on hypothetical future choices, I study actual sequences of choices; moreover, I study what drives the observed persistence in choices. Shepard (2016) provides evidence of adverse selection against health insurance plans covering prestigious and costly hospitals. These plans attract consumers with strong preferences for these type of providers, particularly consumers who have used them in the past. These consumers are likely to choose these hospitals for all of their medical care needs, driving up costs for the insurer, which leads to exclusion of the facilities from the network. In this setting, previous use of a provider is useful for identifying patients with strong preferences for the hospital: whether patient loyalty is due to state dependence or to durable preference heterogeneity is irrelevant, and the empirical analysis does not attempt to separate them. Shepard (2016) emphasizes the effect of choice persistence on medical costs, while I consider how persistence might prevent a patient from switching to a hospital that is better at treating her current medical condition. Raval and Rosenbaum (2017) analyze patients' hospital choices for childbirth in Florida. They use a panel data fixed effects estimator to separate persistence in choice due to switching costs and persistence in choice due to unobserved preference heterogeneity. They consider women who have three children and switch hospitals between their first and second births. For identification, they compare the hospital choices (for the third birth) of women who attended the same two hospitals for the first two births but in different order. They find that approximately 40% of choice persistence reflects switching costs. My work differs from their study in two dimensions. First, I do not restrict attention to a particular medical condition, but study persistence in hospital choices of patients more generally (in particular, patients might seek treatment for different medical conditions over time). Second, I use a different identification strategy to separate the sources of choice persistence.

The remainder of the paper proceeds as follows. Section 2 describes the data used in the empirical analysis. Section 3 discusses the empirical challenges present in my setting. Sections 4 and 5 provide evidence of state dependence in hospital choices of patients by exploiting shocks that shift the loyalty state of patients: emergencies and temporary hospital closures, respectively. Section 6 quantifies the impact of state dependence on health outcomes in the context of hospital choice for cardiac surgery. Section 7 provides concluding remarks.

2. Data

For the empirical analysis, I use detailed patient level data on the universe of visits to hospitals in the state of New York for inpatient and outpatient care. The dataset was obtained from the Statewide Planning and Research Cooperative System (SPARCS). It includes data on inpatient discharges (IP) (1995-2015), ambulatory surgery visits (AS) (1995-2015), and emergency department visits (ED) (2003-2015). In this section, I provide a brief overview of the data. Details about the specific samples used for the different applications are discussed in the corresponding sections.

Each record in the dataset is a hospital visit. The data includes an encrypted patient identifier that allows me to track patients' visits over time, hospital and physician identifiers, patient demographics (age, gender, race, ethnicity, zip code of residence), admission and discharge dates, type and source of admission, discharge status, diagnosis and treatment information, primary payer, charges, length of stay, and indicators for mortality within 7, 15, 30, 180, and 365 days of the discharge date[2]. The patient identifier is missing in 1.7% of records; I exclude these observations in the analysis.

The dataset is therefore a panel that follows the hospital choices of patients in New York. Given that data on hospital visits is more complete for the period 2005-2015 and that information about hospitals characteristics before 2005 is scarcer, I analyze hospital choices of patients in the period 2005-2015. However, I use all the available data to create individual histories of hospital visits.

For the empirical analysis, I want different hospital visits of a patient to correspond to

[2]A complete list of variables can be found here: `https://www.health.ny.gov/statistics/sparcs/datadic.htm`

different episodes of care. I refer to the initial hospital visit of a patient to treat a certain condition as the index event, and I treat readmissions or visits for follow-up care as part of the same episode rather than as different episodes. In the latter case, persistence would be inflated by counting a visit for follow-up care to the same hospital where the patient originally received treatment as a repurchase. The general criterion I follow is to aggregate visits by a patient to the same hospital within a short period of time for related medical conditions into a single episode of care. If a visit is erroneously categorized as a readmission using this criterion, I expect the choice situation to be similar to the (erroneously) associated index visit and I do not want persistence to be driven by these cases. If anything, I prefer to err on the side of understating, rather than overstating, the extent of persistence. I tried different time windows (30, 60, 90, and 120 days) to identify readmissions; the qualitative results are robust to the use of different specifications. To identify hospital visits for related medical conditions, I use the Multi-level Clinical Classifications Software (CCS) developed by the Agency for Healthcare Research and Quality[3]. This classification system groups diagnoses (ICD-9-CM) into 18 Level 1 CCS categories (these are broad condition categories such as "Diseases of the circulatory system", "Diseases of the digestive system", and "Diseases of the respiratory system"). Visits assigned to the same Level 1 CCS category based on principal diagnosis are considered to be related to the same medical condition for aggregation purposes.

I combine the patient data with data on hospital characteristics from SPARCS, the New York State Department of Health, Institutional Cost Reports, Hospital Compare, and the American Hospital Association (AHA). I used Google Maps to calculate the driving distance from the geographic centroid of a patient's zip code of residence to the street address of each hospital; as some zip codes could not be located, information on travel distance is not available for a small fraction of records.

3. Conceptual framework

As mentioned in the introduction, several studies have documented, in a variety of contexts, that consumers who have purchased a product in the past are more likely to choose that same product in the current choice occasion than consumers who have not bought the product before. As Heckman (1981) points out, there are two explanations for this empirical regularity. First, preferences, prices, or constraints relevant to future choices are altered as a consequence of current purchase decisions; structural relationships of this sort are referred to as state dependence. Second, consumers differ according to some serially correlated unobserved propensity to make purchase decisions; the relationship between past purchases and current choice probabilities only arises if unobserved consumer differences are not properly

[3]https://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccsfactsheet.jsp

accounted for, in which case it is termed spurious state dependence (as it originates from the misspecification of the distribution of consumer preference heterogeneity). Therefore, we cannot take observed persistence in choices as conclusive evidence of state dependence. In the context of hospital choice, if I observe a patient who chooses the same hospital each time she needs medical care, it might be the case that she evaluates the characteristics (quality, convenience, etc) of different alternatives in each occasion and then decides to visit the same hospital.

In order to provide a reference point for the empirical analysis, I consider the identification problem in the context of a model of hospital choice. The discussion only intends to illustrate the empirical challenges that I face and the possible strategies to deal with them; each section of the empirical analysis will use a particular framework to address the questions of interest. Patients experience health shocks over time. These shocks determine the medical conditions for which a patient needs medical care. In each episode, a patient visits a hospital. I treat the incidence and timing of hospital visits as exogenous. In each episode, the patient chooses the hospital with the highest utility[4]. Let $h_{it} = j$ denote patient i's admission to hospital j in episode t. The utility that patient i obtains from alternative j in episode t is given by:

$$u_{ijt} = \alpha_{it} D_{ijt} + \beta_{it} Z_{jt} + \gamma_{it} \mathbb{I}(s_{it} = j) + \epsilon_{ijt} \tag{1}$$

where D_{ijt} is the travel distance from the patient's home to the hospital in episode t, Z_{jt} is a vector of hospital attributes in episode t, $\mathbb{I}(x)$ takes the value one if x is true and zero otherwise, s_{it} is the loyalty state of the patient in episode t (which summarizes the history of her past hospital visits), and ϵ_{ijt} captures (possibly persistent) intrinsic preferences of the patient for the hospital. I am interested in first order state dependence. The loyalty state of the patient is determined by the immediately preceding episode: if the patient visited hospital j in episode $t-1$, then $s_{it} = j$. The parameters of the model $\theta_{it} = (\alpha_{it}, \beta_{it}, \gamma_{it})$ might depend on characteristics of the patient, some of which could change across episodes and/or be unobserved by the researcher.

The representation of the patient choice process largely follows the prior literature on hospital choice: the main determinants of hospital choices of patients are convenience (captured by the distance from the patient's home to the hospital) and hospital quality. However, the evaluation of an alternative also depends on whether the patient attended the hospital on her previous visit. In particular, a consumer receives a utility premium γ (which I refer to as the state dependence parameter) from visiting the same hospital as in the previous episode[5]. This effect should be interpreted in a causal sense: if a patient was exogenously

[4] I assume that there is no outside option.

[5] An equivalent interpretation is that the patient has to pay an incremental cost conditional on switching

assigned to a certain hospital in episode $t-1$, then she is more likely to choose that same facility in episode t than an otherwise similar patient. The formulation of utility therefore means that if health shocks or other events made otherwise similar consumers (in terms of their characteristics in the current episode) gravitate towards different hospitals in the past, then their current choices will be different.

Although this formulation of utility is quite general and is similar to models used in previous work, it is restrictive in many dimensions. First, it assumes that patients are myopic: the evaluation of different alternatives only depends on the characteristics of the current episode. In particular, patients do not take into account that current choices will impact future decisions due to lock-in. Second, the loyalty state of the patient could be a more complicated function of past choices, not just of the choice made in the last episode. There is a trade-off between capturing all of the dynamics, and keeping the analysis transparent and tractable. While switching costs in my setting might have both learning and transactional aspects (Farrell and Klemperer, 2007), the model above only captures the latter: a patient who switches from hospital A to hospital B would have to pay the switching cost if she later decides to go back to A. I focus on first-order state dependence because: i) Many of the factors that drive of state dependence operate through the choice made in the immediately preceding episode, and; ii) I expect that the effect of the immediately preceding episode is stronger than the effects of more distant episodes. Third, some of the drivers of state dependence discussed in the introduction are associated with channels that do not operate through the utility function *per se*. For example, state dependence could arise as the hospital used in the previous episode is more likely to enter the consideration set of the patient in the current episode, but with no utility premium from choosing this hospital over other alternatives in the consideration set. In this case, Equation 1 is a reduced form representation of the decision process of the patient. Fourth, the switching cost faced by a patient is symmetric across alternatives: the loyalty premium that a patient gets from sticking to hospital A is equal to the premium from staying with any other hospital B (in other words, γ is independent of j). Fifth, there are no complementarities between different hospitals. However, it might be the case that the cost of switching from hospital A to hospital B is different than the cost of switching from hospital C to hospital B. In some parts of the empirical analysis, I will be able to relax some of these assumptions.

Suppose that I estimate this model from data on patients' actual choices. The identification problem arises from the potential endogeneity of the loyalty state variable. The fact that patient i chooses hospital j in episode t implies that $E[\epsilon_{ijt}|d_{it}=j] > E[\epsilon_{ijt}]$ [6]. If the random

hospitals.

[6]More precisely, it implies that $\delta_{ijt}+\epsilon_{ijt} \geq \max_{k\neq j}\delta_{ikt}+\epsilon_{ikt}$, where $\delta_{ijt}=\alpha_{it}D_{ijt}+\beta_{it}Z_{jt}+\gamma_{it}\mathbb{I}(s_{it}=j)$. The expression in the main text better illustrates the empirical challenge faced by the researcher.

component of utility is correlated across episodes, then $d_{it} = j$ implies that the error term associated to alternative j for patient i will likely be high in episode $t+1$. As a consequence, the estimated state dependence coefficient captures the underlying unobserved propensity of the consumer to choose alternative j, and not just the structural effect of the previous choice on current utility. This is a standard selection problem. Note that the serial correlation of the error term can arise from several sources, such as misspecification of the distribution of the taste coefficients α_{it} and β_{it}, omitted variables, and measurement error. This concern seems particularly well founded in my setting, where there are many attributes of patients and hospitals that I do not observe (religious affiliation of the patient, details about amenities of the hospital, etc). Therefore, I cannot take a positive value of the estimated γ as conclusive evidence of the presence of state dependence.

To deal with this issue, the ideal design would randomly assign patients to hospitals in episode $t-1$ and analyze their choices in episode t; in this case, the loyalty state variable in episode t is uncorrelated with the preferences of the patient[7]. Given that most studies rely on observational data for the analysis, identification has typically relied on both functional form assumptions about the nature of heterogeneity and choice set variation across choice occasions. As Torgovitsky (2016) points out, the first strategy addresses the issue of identification, at least from a mathematical point of view. We could make distributional assumptions about the parameters and the error term in the utility function[8] and estimate the resulting model via maximum likelihood. The problem with this approach is that its validity depends on correctly specifying the distribution of unobserved heterogeneity. Any persistent preference heterogeneity not captured by the model will be loaded onto the econometric error term, leading us to conclude incorrectly that consumer choices exhibit structural state dependence. For example, Dubé et al. (2010) show that allowing for a flexible pattern of heterogeneity can lead to different conclusions than more traditional approaches. Most studies rely (at least in part) on parametric assumptions to separately identify the sources of persistence. Exceptions are Torgovitsky (2016) and Illanes (2016), who recover the values of the parameters that are consistent with the identifying restrictions under different distributions of preference heterogeneity.

Exploiting choice set variation across episodes provides more transparent and credible evidence on the determinants of persistence in choice. The idea is to break the link between the previous choice and unobserved preferences of the consumer, so the selection problem

[7]This allows us to deal with the selection problem and therefore identify state dependence given the structure imposed by Equation 1. However, random assignment is not enough to point identify state dependence in a more general sense. See Subsection 5.4 for a discussion about this issue.

[8]For example, if we assume that ϵ_{ijt} has an extreme value type I distribution, and that $\theta_{it} \sim F(\xi; X_{it})$, where X_{it} are observable characteristics of the patient and F is a c.d.f. indexed by ξ, then we have a random coefficients logit model.

is eliminated or at least attenuated. Heckman (1981) argues that to truly disentangle state dependence from latent heterogeneity I need a sufficiently large variation in the choice set to induce purchases that would not have been made otherwise. Dubé et al. (2010) exploit temporary price changes to identify state dependence in the context of choice of branded products: consumers switch away from their preferred products as a consequence of price variation. The detection of state dependence relies on spells during which the consumer purchases the less-preferred alternatives on successive visits, even after prices return to "normal" levels. Goldfarb (2006b) studies consumers' website choices and exploits product unavailability (caused by Internet denial of service attacks) to identify lock-in. Sudhir and Yang (2014) exploit the mismatch between previous choice and previous consumption created by free upgrades (which are mainly driven by inventory shortages) in the context of car rentals to disentangle state dependence from unobserved heterogeneity. Israel (2005) points out that I can compare two individuals that face the same decision today and have identical loyalty states: one that was forced to use a certain alternative j in the previous episode by an exogenous shock, and another one who chose that same product voluntarily; under selection, the exogenous shock would produce a relatively high number of suboptimal matches among the affected population, which will make consumers depart from alternative j once we return to the usual choice environment.

In the next two sections, I follow this strategy to provide credible evidence about the sources of persistence in hospital choices of patients. In particular, I exploit quasi-exogenous shocks that shift the loyalty state of a patient: emergency hospitalizations (Section 4) and temporary hospital closures (Section 5).

4. Emergencies

In this section, I exploit emergencies as a quasi-exogenous source of variation in the loyalty state of patients. The strategy is to analyze the hospital choice of a patient in episode t following an emergency hospitalization in episode $t-1$. By emergency, I mean an episode in which the patient needs immediate medical care and there is little scope for choosing a particular hospital. I consider emergencies that induce the patient to try a hospital other than the one she had been using. I analyze whether the patient continues using this "new" hospital in the future. As the loyalty state of the patient is initiated by an emergency hospitalization, and to the extent that the new hospital choice is responsive to her preferences, repurchase behavior reflects the extent of state dependence.

The first step is to define emergencies. The identifying assumption is that the locus of treatment for an emergency hospitalization is determined by factors other than the preferences of the patient: the ambulance transport decision and the location of the patient at the time of the health shock (Doyle et al., 2015). This assumption would be violated if: 1) The patient or any of her surrogates requests transportation to a particular hospital during the emergency episode; 2) Patients choose where to live based on health status, so the locus of treatment during an emergency was "chosen" prior to that episode[9,10]. Since the assumption that the hospital used in an emergency is not determined by the preferences of the patient is not verifiable, I take several steps to make the assumption more plausible.

Ideally, I would define emergencies as episodes where the patient suffers a severe and unexpected health shock and arrives to the hospital by ambulance. Unfortunately, I cannot distinguish in my data whether a patient arrived to the hospital by ambulance or self-transport. Given this limitation, I define emergencies as episodes in which the patient is admitted to the hospital through the emergency department (ED) for a severe medical condition that requires immediate care. These non-deferrable conditions correspond to admitting diagnoses (ICD-9-CM) with similar admission rates through the ED on weekdays and weekends (Card et al., 2009). These conditions represent 6% of all ED admissions, and are extremely acute and often life-threatening. Table 1 shows the most common non-deferrable conditions in the full dataset and characteristics of these episodes. Finally, I exclude emergencies in which the patient is admitted to the hospital from a health care facility (e.g. a skilled nursing facility). As the patient is likely to have chosen a health care facility close to her preferred hospital, the use of a hospital during the emergency reflects strong preferences for the facility. In summary, I am confident that the emergencies that I consider are episodes with limited scope for the patient to choose the hospital.

Once I identify an emergency according to the criteria outlined above, I analyze the hospital choice of the patient in the first episode following the emergency (I refer to this episode as the current episode)[11]. In this episode, the loyalty state of the patient is determined by the hospital used for the emergency hospitalization. For example, if the patient used

[9]For example, patients with heart disease might decide to locate close to their preferred hospital, so in the event of a heart attack they are likely to be taken to that facility.

[10]Even if the patient is not involved in the choice process, the emergency hospital could reflect her preferences. Consider two patients who live on opposite sides of the same zip code. I only observe the zip code of residence of a patient, but not her exact address. There are two hospitals, one on each side of the zip code. If patients are taken to the closest facility in an emergency, then the two patients go to different hospitals in that episode. Then, the locus of treatment in an emergency indicates which facility is more convenient for the patient. If the disutility of travel is high, repurchase behavior could reflect unobserved preferences for the emergency hospital (actual distance to the emergency hospital is smaller than in the data). This should be less of a concern the smaller the zip code. I control for this consideration in the analysis below.

[11]As explained in Section 2, I drop readmissions from the working sample. Therefore, I analyze whether the patient continues going to the emergency hospital for episodes not directly related to the emergency itself. In the main specification, I use a 90 days window to identify readmissions. To ensure robustness, I also performed the analysis using other time windows (30, 60, and 120 days), without any substantial change in the nature of the results. These results are available upon request.

hospital A in the emergency episode, then the patient is loyal to hospital A at the time of the next episode. In the analysis, I consider two situations:

1. The current loyalty state was initiated by the emergency. Moreover, the emergency hospital had never been used by the patient before[12]. Therefore, I exclude emergencies in which the patient goes to a hospital different from the last one she used before the emergency but that she used at some point in the past. As a result, the emergency produces a strong shift in the loyalty state of the patient: her choices before the emergency reveal that she does not have strong preferences for the facility used in that episode. Therefore, the repurchase behavior of the patient in the current episode reflects the extent of state dependence.

2. The current loyalty state was initiated before the emergency: the emergency hospital is the same facility that the patient had been using before. In this case, there is a selection issue: the choices of the patient before the emergency reveal that she has strong preferences for that facility. Therefore, the repurchase behavior of the patient in the current episode captures both state dependence and persistent preference heterogeneity.

The repurchase rate is the fraction of patients who in the current episode choose the same hospital used for the emergency hospitalization. Given that the distribution of choice probabilities for a patient might change across episodes, the raw repurchase probability is not very informative about the impact of previous choices on current behavior. For example, in cases where the previous hospital does not offer the type of surgery required in the current episode, the repurchase rate would be zero even in the presence of state dependence. If the hospital used during the emergency is the best hospital for treating the current medical condition of the patient, then a high repurchase rate reflects both state dependence and the quality of the match between the patient and the facility. Therefore, I compare the repurchase rate with the marginal probability of choosing the emergency hospital based on characteristics of the current episode. This way, I measure the likelihood that a patient who used a given hospital j in episode $t-1$ chooses that hospital in episode t relative to an observationally similar patient. To calculate the patient's marginal choice probability, I assume that it is proportional to the market share of the hospital within the group of similar patients. More precisely, I use the following process:

1. Using the full dataset, I define cells of equivalent episodes based on zip code, diagnosis, type of visit, and admission year. Note that a patient might transition across different

[12] More precisely, the patient did not visit the hospital between 1995 (the first year for which I have patient data) and the day of the emergency. Because I only consider emergencies that take place on or after 2005, this restriction means that the patient had not used the facility for at least 10 years before the emergency.

cells over time (for example, if she moves or if she seeks hospital care for different medical conditions). I define these cells to be as precise as possible while maintaining sufficient sample sizes to determine hospital market shares within them. In the main specification, I keep cells that have at least 20 observations. Denote the market share of hospital j within cell k by s_{jk}.

2. I assign each episode following an emergency hospitalization to the corresponding cell. Denote the hospital used in episode i by h_i and the hospital used by the same patient in her previous episode (the emergency hospitalization) by h_i^b.

3. For each episode, the marginal probability of choosing the emergency hospital is the market share of this facility within the episode's cell. Denote the marginal probability of choosing the emergency hospital in episode i by $p(i) = s_{wk}$, where $i \in k$ and $w = h_i^b$.

4. For each $p \in [0, 1]$, the corresponding excess repurchase probability is given by the mean of $x_i = \mathbb{I}\{h_i = h_i^b\} - p(i)$ over episodes with $p(i) = p$.

The idea is to look at patients with different loyalty states (they used different hospitals for the emergency episode) but with the same probability p of choosing the emergency hospital in the current episode. Consider episodes where the loyalty state was initiated by the emergency. If previous choices do not have an impact on current decisions, then we should expect the repurchase rate to be p (so the excess repurchase probability is zero). If the excess repurchase probability is positive, I take that as evidence of state dependence. For episodes where the loyalty state was initiated before the emergency, the excess repurchase probability captures both state dependence and unobserved heterogeneity. Doing the same exercise for all possible values of p, I recover the excess repurchase probability schedule. In practice, the latter is obtained by a locally weighted regression of x_i on $p(i)$. In the main specification, I only consider cases where the marginal choice probability is higher than 0.01. If the marginal probability is lower than this value, it most likely corresponds to a case where the patient will not consider the facility for hospital care.

Table 2 shows summary statistics of the current episodes used in the analysis. I distinguish episodes depending on whether the emergency hospital had never been used before the emergency or was the usual hospital of the patient. There are two differences between these groups to point out. First, patients in the second group had on average more episodes before the emergency hospitalization than patients in the first group. Second, patients who experience a shift in their loyalty state as a result of the emergency episode have lower probability of returning to the emergency hospital based on their observable characteristics.

I construct the excess repurchase probability schedule pooling across all emergencies, hospitals, and current episodes. The results are shown in Figure 1. There are separate schedules for episodes following emergencies that shifted (blue) and did not shift (orange) the loyalty state of the patient. There are two main points to be noted.

First, for cases where the loyalty state of the patient was initiated by the emergency hospitalization, the excess repurchase probability is positive over the unit interval. This indicates that a patient who, due to an emergency, visited a certain hospital j in episode $t-1$ for the first time is more likely to choose hospital j in episode t than a patient with the same characteristics. Moreover, the impact of the previous choice on the current decision is large: for example, if based on current characteristics there is a 20% probability that the patient chooses hospital j, then the choice probability increases to more than 40% for patients who used that same hospital in the previous episode. This is evidence of the presence of state dependence in hospital choices of patients. In general, a positive excess repurchase probability could also signal the presence of unobserved heterogeneity; however, as explained above, I address this concern by focusing on loyalty states initiated by emergencies.

Second, the excess repurchase probability is higher for episodes following an emergency that did not shift the loyalty state of the patient. As explained above, in this case repurchase behavior not only reflects state dependence, but it also captures the latent propensity of the patient to choose the facility used in previous episodes. Therefore, the higher repurchase rate for these cases reflects the presence of substantial unobserved preference heterogeneity. As a result, a naive analysis that does not take into account the endogeneity of previous choices will overstate the magnitude of state dependence.

To assess the robustness of my findings, I construct the excess repurchase probability schedule for episodes following different types of emergencies. The results are shown in Figure 2. The qualitative results are the same as in the main analysis.

The previous results could mask heterogeneity in persistence across different types of hospitals. In particular, there might be differences in loyalty towards high and low quality hospitals. The pattern of heterogeneity might provide insights about the determinants of state dependence. I restrict attention to episodes following an emergency that induced a shift in the loyalty state of the patient, so repurchase behavior reflects state dependence. I distinguish cases based on the quality of the emergency hospital. Quality is not observable, so I use teaching status as a proxy for high quality[13]. Figure 3 shows that the excess repurchase probability schedules of teaching and non-teaching hospitals are similar. Although not conclusive evidence, this suggests that learning is not the primary driver of the observed patterns of persistence: patients are equally loyal to high and low quality hospitals.

[13] Teaching status is obtained from the AHA Annual Survey Database and refers to hospital membership in the Council of Teaching Hospitals (COTH).

5. Hurricane Sandy

5.1. Setting

Hurricane Sandy hit the New York Metropolitan area at the end of October 2012. Damage from the storm led to the temporary closure of three hospitals in New York City: NYU Langone Medical Center, Bellevue Hospital Center, and Coney Island Hospital. Bellevue and NYU Langone are located in Manhattan next to each other, while Coney Island Hospital is located in Brooklyn. Table 3 provides a basic description of the affected hospitals. The facilities differ along various dimensions, so there is heterogeneity in the settings that I analyze. NYU Langone is an academic medical center with a high proportion of privately insured patients. The other two hospitals are part of the city's Health and Hospitals Corporation and attract mostly Medicare, Medicaid, and uninsured patients. NYU Langone and Bellevue are large hospitals (more than 900 beds), while Coney Island Hospital is a medium size facility (371 beds). In terms of service offerings and designations, NYU Langone is the most sophisticated, followed by Bellevue. Although Coney Island Hospital offers a wide range of services, it does not provide the most complex services.

The affected hospitals remained closed for repairs and renovations during several weeks, forcing patients usually served by these facilities to find an alternative hospital for their medical care needs. Consider, for example, a patient who receives medical care at Coney Island Hospital and had never gone elsewhere before. If the patient needed medical care during the time this facility was closed, she would have had to go to another hospital, such as Maimonides Medical Center. In this section, I study whether the temporary unavailability of the affected hospitals had a long-lived impact on patients' preferences: Does the patient in the example continue going to Maimonides Medical Center once Coney Island Hospital reopens or does she return to her usual hospital? Do affected patients become long-term patients of the new facilities? Anecdotal evidence suggests that the possibility of permanently losing patients was a concern for administrators at the shuttered hospitals[14].

This natural experiment is particularly well suited to study the dynamics of hospital choice. First, the type of choice set variation produced by the storm is ideal given how competition takes place in the hospital industry. Second, the hospital closures were unexpected and unrelated to patients' preferences for different facilities, thus providing a quasi-exogenous source of variation in hospital choice. Third, the data allows me to identify the patients most likely to have been affected by the temporary hospital closures: patients with strong preferences for an affected facility who needed hospital care during the unavailability window. Therefore, I can identify patients who would have chosen one of the affected hospitals had it

[14] http://www.nytimes.com/2012/12/04/nyregion/with-some-hospitals-closed-after-hurricane-sandy-others-overflow.html?mcubz=1

been available, but were forced to choose a second-best option. There are two main limitations of my analysis. First, I only observe hospital choices of patients for less than three years after the affected facilities reopened. Second, I analyze a specific empirical context, which places limitations on the external validity of my conclusions. In particular, the analysis is not designed to provide estimates of the extent of state dependence in other settings.

Raval et al. (2017) exploit unexpected hospital closures in different markets following a natural disaster to analyze the substitution patterns predicted by different models of hospital choice. One of the natural disasters that they consider is Hurricane Sandy. In one of their specifications, the authors identify patients who used the affected hospitals in the pre-storm period as those most likely to have experienced the closure of their preferred hospital. However, whether their continued preference for the shuttered facilities is due to switching costs or unobserved heterogeneity is irrelevant for their analysis, so they do not attempt to separate the channels. The main objective of my analysis is to separately identify state dependence from persistent unobserved heterogeneity, while the nature of substitution patterns is not a primary concern.

The analysis proceeds in two steps. First, I show that patients who needed hospital care while a hospital was closed (treatment group) are less likely to visit that facility in the future than patients who did not have hospital visits during the unavailability window (control group). Moreover, non-returning patients in the treatment group favor the facility used during the unavailability window. I show this in a clear way, without relying on complex estimators. Then, I cast the setting into a nonparametric framework. By imposing an independence restriction between preferences and the timing of hospital visits, I can reject the hypothesis of no state dependence under minimal assumptions about the nature of unobserved preference heterogeneity.

5.2. Sample construction

For each case study[15], I construct a panel of patients' hospital choices. As explained in Section 2, readmissions and visits for follow-up care are excluded from the analysis, so hospital visits correspond to different episodes of care. I divide episodes into three periods based on the date of the patient's admission to the hospital: period 0 corresponds to the pre-storm period, period 1 is the time window during which the affected hospital remained closed for repairs, and period 2 goes from the reopening of the shuttered facility trough the end of the sample period. For each patient, I only consider episodes that took place while the patient was living in the service area of the affected hospital: I want to use information on those episodes where

[15] In what follows, I emphasize the case of Coney Island Hospital (CIH, henceforth), because the definitions of service area and the competitive set are more transparent for CIH than for Bellevue and NYU Langone. However, the methodology outlined below applies to all three cases.

the patient is likely to consider this facility for hospital care. In the case of CIH, this step would remove, for example, episodes that take place while the patient lives in Buffalo or Manhattan - so CIH is not viewed as a practical alternative - but would keep hospital visits by that same patient while she lives in southern Brooklyn.

To construct the service area of a hospital, I identify the smallest collection of zip codes that accounted for 90% of inpatient discharges from the facility in the year prior to the storm. The resulting area might span zip codes where the hospital is competitively insignificant. Therefore, I exclude zip codes where the facility had a market share below 4% in the year before the storm. Finally, I make some adjustments to ensure the geographic contiguity of the resulting service area. Figures 4 - 6 show the service area of each of the affected hospitals. I repeated the analysis using alternative thresholds to determine the inclusion of zip codes in the service area and the qualitative results do not change[16].

I construct the final sample to leverage the features of the data that allow me to identify the effects of interest. I keep patients from the full sample who: 1) Had at least one hospital visit in both periods 0 and 2[17], and; 2) Did not have any hospital visits while the affected facility was closed or visited only one hospital during that time window. For each individual in the resulting sample, I keep the last episode of period 0, all episodes (if any) in period 1, and the first episode of period 2. Therefore, for each patient I know: 1) The hospital chosen in the last episode before the storm; 2) Whether she needed hospital care while the affected facility was closed and, if so, which hospital she used, and; 3) The hospital chosen in the first episode of period 2. I distinguish patients based on whether they visited a hospital in period 1. I refer to the set of patients who needed hospital care while the affected hospital was closed as the treatment group, while the other patients constitute the control group.

For the analysis, I focus on the hospital choices of patients in the first episode of period 2. I exclude a small number of cases where the affected hospital does not seem to be in the market for the type of medical care required by the patient in that episode. I infer provision of the required services by aggregating all records in the original data into cells defined by unique combinations of claim type, diagnosis category and semester, and observing the number of patients with those characteristics receiving care at the corresponding facility. I then use a three visit threshold to determine if the hospital is a feasible alternative for the patient. This step removes approximately 5% of patients.

I refer to the sample that results from the selection steps outlined above as sample #1. I define two subsamples that I use in the subsequent analysis. Sample #2 contains those patients who were loyal to the affected hospital at the moment of the storm (they visited that

[16]These results are not shown here, but they are available upon request.

[17]The restriction that patients have at least one episode in period 0 has the benefit that I can use information on actual hospital choices to infer the strength of their preferences for the affected hospital.

facility in the last episode of period 0), while sample #3 contains those patients who chose either the shuttered hospital or one of its main competitors in all the three episodes considered. To identify the main competitors of an affected hospital, I compute the diversion ratio from that facility to other hospitals in the year before the storm. I define the six hospitals with the highest diversion ratios as the main competitors of the affected facility. Tables 4 through 6 present information on the demographic profile and other characteristics of patients in the three samples. There are two main differences between control and treatment groups. First, the treatment group is older and sicker: it has a higher proportion of inpatient episodes and Medicare/Medicaid patients than the control group. Second, patients in the treatment group used the affected hospital in the pre-storm period less frequently than patients in the control group. Although I control for these differences in the analysis below, the imbalance along the latter dimension points to potential threats to identification.

5.3. Reduced form evidence

I now discuss patterns in the data that help identify the presence of state dependence. Consider two patients with similar underlying preferences for CIH (idiosyncratic tastes or unobserved characteristics - separate from state dependence - that make them gravitate towards this hospital) and who were loyal to that hospital at the moment of the storm. The first patient needs hospital care in period 1 and therefore goes to a hospital other than CIH (given that this facility is closed), while the second patient does not need hospital care in period 1. I compare the hospital choices that these patients make in the first episode of period 2. In this episode, the patients have different loyalty states: the patient who got sick during period 1 is not loyal to CIH, while the patient who did not seek hospital care in period 1 remains loyal to CIH. If the temporary unavailability of CIH did not change the underlying preferences of patients for hospitals, then differences in choice probabilities between the two patients capture the impact of state dependence.

The identifying assumption is that the timing of hospital visits is exogenous. In other words, whether a patient seeks hospital care in period 1 or not is independent of her tastes for different facilities. In this case, the distribution of preferences is the same for patients in the treatment and control groups. The assumption seems a priori reasonable in my setting, given that patients most likely visit a hospital due to a health shock. However, there are ways in which the assumption could be violated. I discuss the potential threats to the validity of my approach at the end of the section.

I start by showing that patients who needed hospital care while an affected hospital was closed were less likely to return to that hospital in the first episode of period 2 than similar patients who did not have hospital visits in period 1. I use sample #2 for the analysis, so all the patients considered were loyal to the affected hospital at the moment of the storm.

Figure 7 shows, for both the treatment and control groups, the probability of choosing CIH in the first episode of period 2 as a function of the prior propensity to use the facility. This propensity is defined as the proportion of the patient's hospital visits prior to the storm that were to this facility[18]. There are two things to note. First, the probability of choosing CIH in period 2 increases with the prior propensity to use that facility, for both the treatment and control groups. This is not surprising, given that the pre-storm propensity to use the affected facility captures the strength of the patient's preferences for that hospital. Second, patients in the treatment group are less likely to return to CIH after its reopening than patients in the control group with similar prior propensity. The gap in choice probabilities reflects a lasting effect of the temporary unavailability of the affected hospital on patients' choices. Figures 8 and 9 show the results for Bellevue and NYU Langone, respectively. For NYU Langone, the difference between treatment and control groups is not as clear as in the other two cases; however, the analysis below shows that the difference in patient behavior is significant in this case.

To control for preference heterogeneity related to observables, I estimate a probit of choosing the affected hospital in the first episode of period 2 on the prior propensity to use the facility, zip code fixed effects, diagnosis fixed effects, a spline of the number of days between the reopening of the shuttered facility and the episode, and an indicator for treatment status. The results are displayed in Table 7. In the case of NYU Langone, having a hospital visit in period 1 is associated with a decrease of 6.5 percentage points in the probability of using the facility in period 2; the effect is statistically different from zero at usual significance levels. For CIH and Bellevue, the marginal effect of treatment status is -9.1 and -10.5 percentage points, respectively[19].

Thus far, I have shown that for patients in the treatment group the probability of choosing the affected hospital after its reopening decreased relative to patients in the control group. I now explore to what extent non-returning patients substitute the affected hospital in period 2 with the same facility they used in period 1. Even if patients are less likely to return to the affected hospital once it reopens, they might gravitate towards other hospitals in a proportional way (conditional on heterogeneity); however, if they continue using the hospital they visited in period 1 once the affected facility reopens, then this points to state dependence as the source of the difference in patient behavior. In order to capture the effect of interest, I estimate a conditional logit model of hospital choice using the first episode of

[18]Consider, for example, a patient who had five hospital visits in the pre-storm period. If three of these visits were to the affected hospital, then the prior propensity to use this facility is 0.6 (3/5).

[19]Two possible explanations for the lower effect found in the case of NYU Langone are the following. First, NYU Langone might derive more loyalty from patients given its reputation and higher quality. Second, the hospital might have actively engaged in regaining patients who visited other hospitals during the time it was closed (for example, by contacting them).

period 2. Explanatory variables include hospital-type of visit fixed effects, the distance from the centroid of the patient's zip code of residence to the facility, interactions between the type of insurance of the patient and an indicator for public hospital, and an indicator for whether the hospital was used in the previous episode of care; in addition, I allow the utility from choosing the affected hospital to vary with the prior propensity to use that facility. The estimates from the model are reported in Table 8. Previous use of a facility is a strong predictor of the hospital choice of a patient: having used a hospital in period 1 increases the probability of choosing that same facility in period 2 more than four times above the baseline probability (the magnitude of this effect is similar across the three case studies).

In summary, I have provided evidence that patients who experienced a forced shift in their loyalty state are less likely to return to their usual hospital than similar unaffected patients. Moreover, patients who do not return to the affected hospital favor the new facility they used in period 1. These patterns indicate the presence of state dependence in hospital choices of patients. I now discuss in more detail the assumptions underlying my empirical strategy in order to identify potential problems.

First, I assume that the timing of hospital visits is independent of patients' preferences for hospitals. In other words, I assume that whether a patient seeks hospital care in period 1 or not is independent of her preferences for different hospitals. If patients with very strong preferences for the affected hospital postponed elective medical care until this facility reopened, then there is a selection problem. In particular, the control group would include a higher proportion of patients with strong preferences for the affected facility, which might explain the gap in choice probabilities between the two groups.

Second, the affected hospitals restored services gradually and sometimes at reduced capacity. If capacity constraints were important, I might be artificially inflating persistence due to state dependence: some patients might have wanted to receive care at CIH in period 2 but were unable to do so due to capacity constraints. For this to bias the results, it must be the case that patients in the treatment group were more affected by capacity constraints than patients in the control group.

To address these concerns, I perform the same analysis excluding episodes that took place close to the start of period 2. Capacity constraints likely were more of an issue during the first days after affected hospitals resumed operations, so inference using episodes that occurred later in period 2 should be less prone to bias from this channel. In addition, using episodes further away from the reopening date should lower the impact of selection into treatment status by patients who delay care. I estimate a probit of choosing the affected hospital in period 2 as described above, but excluding cases that took place less than 30, 60, 90, 120, and 150 days after the reopening date. The results are shown in columns 3-7 of Table 7. The decrease in the probability of choosing CIH associated with treatment group status is between

7.2 and 8.4 percentage points, depending on the time window considered. Although the effect is smaller than in the main specification, it remains large and significant. For Bellevue and NYU Langone, the effect of treatment status also decreases as I focus on episodes further away from the date of reopening, although the magnitude of the change is smaller than for CIH.

Third, I assume that the closure and unavailability of an affected hospital did not have any impact on patients' underlying preferences for the facility. In particular, this means that the quality of the hospital did not change and that patients did not update their beliefs about the quality of the hospital. If patients in both the control and treatment groups were equally affected along this dimension, this would not be a problem because identification is based on differences in choice probabilities. However, if patients in the treatment group are more likely to adjust their beliefs about the quality of the affected hospital (or if they feel that some aspects about the facility that they liked changed after its reopening), then this would be a confounding factor in the analysis. Although I cannot control for this effect, I believe that I can largely discount it: the closure was directly related to the damage produced by the storm, which was particularly destructive and affected many areas of the city.

In conclusion, while deviations from the identifying assumptions are certainly possible, the robustness of my results suggests that their potential effects are of second-order importance. However, there are several assumptions implicit in the analysis whose impact on the results is not easily ascertained. First, I assumed a specific distribution for the unobservable portion of utility. The selection of probit or logit to estimate the effects of interest responds more to convention than to economic rationale. One could wonder whether I would still find that there is state dependence in hospital choices of patients under other (possibly "non-standard") distributional assumptions. Second, I have assumed that switching costs are symmetric: the utility premium from sticking to the previous choice is independent of the hospital visited in period 1. However, it is possible that some hospitals induce more loyalty from patients than others. Third, the results previously discussed are informative about the mean effect of state dependence, but they provide limited information about heterogeneity in the impact of previous choices on current behavior. However, it is not clear which modeling assumptions are appropriate in this setting. Fourth, I have assumed that utility is linear in its different components. I may need to explore the effect of non-separability of preferences on measured state dependence. Fifth, I have assumed that patients are myopic when choosing hospitals: they do not consider the implications of their current choices for future episodes of care. The question remains whether conclusions change once I allow for forward looking behavior. In summary, although the main source of identification in my setting is evident, there are several modeling choices that could affect the results. In the next section, I address this concern and show evidence of state dependence using a nonparametric model of dynamic discrete choice.

5.4. Nonparametric analysis

In this section, I quantify the extent of state dependence in hospital choices of patients using the nonparametric approach proposed by Torgovitsky (2016), adapted to the particulars of my case study. Here, I outline the basic framework and present the main results; additional details can be found in Appendix B.

There are N patients and three periods ($t = 0, 1, 2$). All patients choose a hospital in periods 0 and 2, and some patients also have to choose a hospital in period 1. There are J hospitals; however, hospital 1 is not available in period 1 and therefore cannot be chosen by patients who need hospital care in that period. If $\mathcal{J}_t$ denotes the choice set in period t, then $\mathcal{J}_0 = \mathcal{J}_2 = \{1, ..., J\}$ and $\mathcal{J}_1 = \{2, ..., J\}$. This is the setting described in the previous subsection: hospital 1 is the hospital shuttered by Hurricane Sandy and the same definitions of time periods apply. Each patient has a type determined by:

1. Her choice in period $t = 0$, denoted h_0.

2. Her potential choices in each period $t \geq 1$ in which she has to choose a hospital: $h_t(k)$ denotes the alternative the patient would choose in period t had her previous choice counterfactually been k. For a patient who has to choose a hospital in period 1, $h_1(k)$ is the alternative the patient would choose in period 1 had her period 0 choice counterfactually been k, while $h_2(k)$ is the alternative the patient would choose in period 2 had her period 1 choice counterfactually been k. For a patient who does not have to choose a hospital in period 1, $h_2(k)$ is the alternative the patient would choose in period 2 had her period 0 choice counterfactually been k. There are J counterfactual choices $\{h_t(k)\}_{k=1}^{J}$ in period t.

3. Whether she needed hospital care in $t = 1$, indicated by $z \in \{0, 1\}$.

4. Other variables that indicate time-varying or time-invariant characteristics of the patient, captured by vector $x \in \mathcal{X}$. I restrict attention to discrete covariates, so the set $\mathcal{X}$ is discrete.

In what follows, $\mathcal{H}$ denotes the set of all possible types, h denotes a generic type in $\mathcal{H}$ and $h_t = [h_t(1), ..., h_t(J)]$ denotes the corresponding set of potential choices in period t.

Let y_t denote the hospital chosen by a given patient in period t. A patient in my data is associated with a vector of observables $y = \{y_0, y_1, y_2, z, x\}$. I observe the actual hospital choices of patients and their characteristics, but not their types. However, types and observables are related. In particular, a patient with type h makes the following choices:

$$y_0 = h_0 \tag{2a}$$

$$\text{For } z = 1 : y_1 = h_1(y_0) \text{ and } y_2 = h_2(y_1) \tag{2b}$$

$$\text{For } z = 0 : y_2 = h_2(y_0) \tag{2c}$$

Going forward, $\mathcal{Y}$ is the set of all possible vectors of observables, y denotes a generic element of $\mathcal{Y}$, and $y(h) \in \mathcal{Y}$ is the vector of observables generated by a patient with type h through relationships (2a)-(2c). Note that while type h generates a unique vector of observables $y(h) = \bar{y}$, the latter could have been generated by more than one type (see Appendix B for an example).

The data is given by $P : \mathcal{Y} \to [0, 1]$, where P is a probability mass function (p.m.f.) with support in $\mathcal{Y}$. In other words, for each possible vector $y \in \mathcal{Y}$, I know the proportion of patients with those observables. For now, I assume that the distribution of observables is known by the researcher.

For any patient, only one element of the vector of potential choices in a given period is known to the researcher. The remaining potential choices are counterfactuals. This framework is essentially the potential outcomes framework used in policy evaluation applied to a dynamic discrete choice setting. It postulates a set of potential choices that could be observed in alternative states of the world, with the state variable defined by the alternative chosen in the previous choice occasion.

The notion of state dependence arises naturally in this setting, as a patient might make different choices in different counterfactual scenarios. A patient with type h exhibits state dependence in period $t > 0$ if there are two previous counterfactual choices $k, j \in \{1, ..., J\}$ for which the corresponding period t potential choices are different: $h_t(k) \neq h_t(j)$. Let $sd_t(h)$ be an indicator of whether a patient with type h exhibits state dependence in period t. Identifying state dependence at the individual level requires information on at least two potential choices in a period. As explained above, the data reveals only one component of the vector of potential choices in a period. Therefore, whether there is state dependence at the individual level is untestable. However, we might be able to learn something about the extent of state dependence in the population.

The primitive of the model is the distribution of types in the population. From knowledge of this object, one can construct measures of state dependence. In particular, I want to learn about the proportion of patients who exhibit state dependence in period 2 (which I refer to generically as θ). Then, I view this feature of the distribution of types in the population as the object of interest. If the distribution of types is given by the probability mass function $\mathbb{P}$, then the corresponding parameter is:

$$\theta(\mathbb{P}) = \sum_{h \in \mathcal{H}} \mathbb{P}(h) sd_2(h) \tag{3}$$

which is linear in the probabilities of types.

The problem is that the distribution of types is not known by the researcher. Let W be the set of admissible distributions. For the moment, we take W as given. This set contains all the distributions that the researcher considers could have generated the data, based on the assumptions about the data generating process that s/he is willing to maintain. We only consider distributions in W and any such distributions could in principle be the true one. In general, there are many admissible distributions and therefore the parameter of interest is partially identified. The identified set for the parameter of interest is given by:

$$\Theta(W) = \{\theta(\mathbb{P}) : \mathbb{P} \in W\} = [\theta_l(W), \theta_u(W)] \tag{4}$$

In other words, to recover the identified set we need to compute the proportion of patients with state dependence in period 2 under each admissible distribution using Equation (3). The last equality in Equation (4) states that the identified set is a closed interval with extreme points $\theta_l(W)$ and $\theta_u(W)$. This characterization of the identified set only holds under certain conditions on $\theta(\mathbb{P})$ and W, which are satisfied in my setting. Therefore, to characterize the identified set I only need to find the admissible distributions with the lowest and highest state dependence:

$$\theta_l(W) = \min_{\mathbb{P} \in W} \sum_{h \in \mathcal{H}} \mathbb{P}(h) sd_2(h) \tag{5}$$

$$\theta_u(W) = \max_{\mathbb{P} \in W} \sum_{h \in \mathcal{H}} \mathbb{P}(h) sd_2(h) \tag{6}$$

Before discussing the restrictions that define the set of admissible distributions and the computational approach to recover the identified set, let me briefly discuss the two main identification issues we need to deal with.

First, from ordinary (non-experimental) data on patients' hospital choices, I can recover $P[y_t = v | y_{t-1} = w] = \mathbb{P}[h_t(w) = v | y_{t-1} = w]$ for $v \in \mathcal{J}_t, w \in \mathcal{J}_{t-1}$. However, this object differs from $\mathbb{P}[h_t(w) = v]$ due to persistent unobserved heterogeneity. Therefore, the marginal distributions of potential choices in a given period are generally not identified from observational data.

Even if I deal with selection bias (for example, by using experimental data), there is a second identification issue: it is generally not possible to recover the joint distribution of potential choices from the marginal distributions (Heckman et al., 1997). The problem is then to bound an unknown joint distribution from known marginal distributions. This can be illustrated for the case of two alternatives (which I denote 1 and 2) using a 2x2 contingency table, following Heckman et al. (1997). For this example, I focus on the potential choices in period t. In Table 9, the columns refer to the choices in the state in which the patient chose hospital 2 in $t-1$, while the rows refer to the choices in the state in which the patient chose hospital 1 in $t-1$. If I observed each patient in both states of the world, I could fill in the table and recover the full distribution. From the full distribution, I could determine the proportion of patients with state dependence in period t[20]. With experimental data, I would be able to estimate the totals for each row and column but I cannot recover the values of particular cells without further assumptions. This issue stresses the fact that it is not possible to point identify the extent of state dependence even with experimental data[21]. This notion seems to have been overlooked by many empirical studies of persistence in consumer choice that focus almost exclusively on dealing with the selection bias issue[22].

The set of admissible distributions consists of those p.m.f. over $\mathcal{H}$ that are compatible with the observed data and the institutional setting, and that satisfy the additional identifying assumptions that I impose. Let $\mathbb{P}$ denote a generic distribution of types[23]. Then, $\mathbb{P}$ is an admissible distribution if the following conditions are satisfied:

Assumption 1: $\forall \bar{y} \in \mathcal{Y}$, $\sum_{h:y(h)=\bar{y}} \mathbb{P}(h) = P(\bar{y})$

This restriction states that the probability mass assigned to types that generate certain observables $\bar{y}$ must be equal to the proportion of patients in the data with those observables.

Assumption 2: $\forall j \in \{1,..,J\}, h_1(j) \neq 1$

This restriction states that hospital 1 cannot be chosen in period 1 by patients who require hospital care during that period. Then, $\mathbb{P}$ assigns probability 0 to any type that does not satisfy the condition above.

[20]In the example, patients with state dependence in t are those with potential choices $[h_t(1) = 1, h_t(2) = 2]$ or $[h_t(1) = 2, h_t(2) = 1]$. These patients make different choices in different counterfactual scenarios (which are defined by the choice made in the previous episode).

[21]Although knowledge of the marginals is not enough to point identify state dependence, it might be enough to put non-trivial bounds on it. In the example, using the Frechet-Hoeffding bounds, we have:

$$\max\{\mathbb{P}_{1*} + \mathbb{P}_{*2} - 1, 0\} \leq \mathbb{P}_{12} \leq \min\{\mathbb{P}_{1*}, \mathbb{P}_{*2}\}$$

.

[22]It should be noted that dealing with selection bias is in general quite complicated.

[23]In particular, $\mathbb{P}$ satisfies the following conditions: 1) $\mathbb{P}(h) \in [0,1], \forall h \in \mathcal{H}$; 2) $\sum_{h\in\mathcal{H}} \mathbb{P}(h) = 1$

Assumption 3: $\forall j, k, m \in \{1, .., J\}$, $h_t(k) = j \Rightarrow h_t(j) = j$ and $h_t(k) = j \Rightarrow h_t(m) \in \{m, j\}$

I restrict the set of admissible distributions to those p.m.f. that assign probability 0 to any type that does not satisfy the conditions above. There are two assumptions implicit in Assumption 3: 1) Monotone treatment response: if a patient would choose hospital A had she chosen another hospital B in the previous episode, then she would also choose hospital A had she chosen hospital A in the previous choice occasion; 2) No partial compatibility: purchases made in different episodes are compatible only if the same alternative is chosen in both occasions. Assumption 3 would hold, for example, if choices are determined by utility maximization and the utility a patient receives from choosing alternative j in period t depends on: 1) Characteristics of the patient and the hospital in that period; 2) Whether the patient chose alternative j in the previous choice occasion, in which case she receives a non-negative utility premium (but there is no utility premium from switching hospitals). Note that the standard logit model of demand (see, for example, Equation 1 in Section 3) satisfies this assumption. A more detailed discussion of Assumption 3 can be found in Appendix B.

Under Assumptions 1-3[24], the identified set for the proportion of patients with state dependence in period 2 is the interval $[0, 1]$: I can find two distributions of types that attribute all the observed persistence to unobserved heterogeneity and state dependence, respectively. In order to get more informative results, I need to impose additional restrictions. In particular, I translate the identifying assumption discussed in the previous subsection into the following restriction on the set of admissible distributions[25]:

Assumption 4: $\forall j, k \in \{1, .., J\}$ $\forall x \in \mathcal{X}$, $\mathbb{P}[h_2(k) = j | z = 0, x] = \mathbb{P}[h_2(k) = j | z = 1, x]$

This is a conditional independence assumption. It states that the potential choices in period 2 are independent of the treatment status indicator z, conditional on the other covariates x. In other words, the distribution of preferences in period 2 is the same for patients who required hospital care in period 1 and for patients who did not, conditional on other covariates.

This was the main identifying assumption for the analysis in Section 5.3. Here, the meaning of this assumption is more clear. If patients in the treatment and control groups have different loyalty states in period 2, then their choices reveal different elements of the

[24]In principle, the only assumptions that need to be always considered are Assumptions 1 and 2. As I discuss later, I need to impose Assumption 3 ex-ante to accommodate certain practical limitations. However, whenever possible, we might want to treat Assumption 3 as a regular identifying assumption and evaluate its power independently of the other restrictions.

[25]Apart from the ones discussed here, there are other restrictions that could be imposed on the set of admissible distributions. See Appendix B for details.

vector of potential choices in that period. If the distribution of potential choices is the same for both groups, then I can combine the information obtained from each group to learn about the joint distribution of potential choices.

Assumptions 1 through 4 determine the set of distributions W that I need to consider for identification. As discussed above, I could recover the identified set by computing the parameter of interest for each admissible distribution. However, this approach is not very practical. While an analytical characterization seems too complicated, we can exploit the structure of the problem to characterize the identified set in a computationally simple way. As mentioned before, the identified set is a closed interval. Therefore, I only need to recover the lower and upper bounds on the proportion of patients with state dependence in period 2 to fully characterize the identified set. Consider the problem of recovering the lower bound (the same reasoning applies to the problem of recovering the upper bound). As Equation (5) shows, the problem is to find the admissible distribution (probabilities of types) that minimizes the parameter of interest. The latter is linear in the probabilities of types. The set of admissible distributions is defined by Assumptions 1 through 4, which impose linear restrictions on the probabilities of types. Therefore, the lower and upper bounds on the parameter of interest can be found by solving a set of linear programming problems. The full description of the optimization problem can be found in Appendix B. Note that I can recover the identified set under different characterizations of W in a straightforward way: I just need to re-solve the optimization problem under different combinations of assumptions.

To implement this approach, I take certain steps to accommodate practical limitations. The main issue is the size of the set of types: the number of possible potential choices in any period $t \geq 1$ when the cardinality of the choice set is J is J^J, so the number of possible types (without any covariates) is J^{JT+1} if $t = 0, .., T$. To keep the dimensions of the problem to manageable proportions, I make the following modeling choices: 1) I only consider three periods ($T = 2$); 2) I only consider patients who visit the affected hospital or any of its main competitors in any period in which they need hospital care (then, I use sample #3 for the analysis); 3) I impose Assumption 3 ex-ante. Without doing this, the dimensions of the optimization problem would be prohibitive in my application. However, I stress that Assumption 3 has economic content beyond its usefulness to reduce the dimensionality of the problem.

For each case study, I compute the identified set under Assumptions 1 through 4. The results are shown in Table 10. The data and the restrictions that I impose on the distribution of preferences allow me to conclude that at least 9.1% of patients in the sample for Coney Island Hospital exhibit state dependence in period 2. In other words, there is no distribution of patient types consistent with Assumptions 1 through 4, such that the proportion of patients with state dependence in period 2 is lower than 9.1%. In the case of Bellevue and NYU

Langone, the lower bounds on the proportion of patients with state dependence are 9.2% and 12.9%, respectively. In all three cases, the upper bound is equal to 1: I cannot reject the hypothesis that all patients exhibit state dependence in period 2 under Assumptions 1 through 4.

The interpretation of the results is the following. I consider the case of Coney Island Hospital to fix ideas. If I knew the distribution of types in the population, then I would be able to compute the proportion of patients who exhibit state dependence in period 2. However, I do not know what the true distribution is. Therefore, I need to consider all distributions that are consistent with the data. I cannot get informative results from the data alone, so I need to impose additional structure. One possibility would be to specify a fully parametric model of patient behavior (for example, a random utility model), map the full structure of the model into the potential choices framework, estimate the parameters of the model, and use these estimates to compute the proportion of patients with state dependence in period 2. However, it is difficult to assess the sources of identification with this approach and therefore it is not clear how measured state dependence would change under different modeling assumptions. The results obtained using the nonparametric approach indicate that no model with a richer structure will produce less than 9.1% of patients with state dependence in period 2, as long as it is consistent with the identifying assumptions imposed. Therefore, it is possible to rule out the hypothesis that there is no state dependence under transparent and easily interpretable assumptions. The finding of state dependence is not an artifact of particular parametric assumptions but is supported by a more fundamental source of identification. I gain transparency in results at the cost of set identification: there are many values of the parameter of interest that are compatible with the data and the identifying assumptions that I impose.

In summary, this approach allows me to provide robust evidence of the presence of state dependence in hospital choices of patients. I can rule out that persistence in patient behavior is only due to unobserved heterogeneity under minimal assumptions about the data generating process. On the other hand, the identified set is quite wide, so we need to impose additional structure if we want to get more informative results. Therefore, I see the approach discussed in this section as complementary to parametric models of consumer choice: we use these richer models to obtain more meaningful conclusions, but we do this having already shown that the finding of state dependence is not an artifact of the parametric assumptions we impose.

In the previous analysis, I recovered the identified set assuming that the distribution of observables was known, without accounting for sampling error. To construct a confidence region for the identified set, I follow the strategy proposed by Torgovitsky (2016). The idea is to express the potential choices model as a moment inequalities model. The characterization of the identified set is given by a criterion function that penalizes deviations from the identifying assumptions. Then, we define a sample analog of this criterion function and use it as the basis for statistical inference. Specifically, we use it as a test statistic for the null hypothesis that the parameter of interest takes a specific value. To construct confidence regions, we collect all the values of the parameter of interest for which the corresponding null hypothesis is not rejected (test inversion). To recover the distribution of the test statistic, I use the subsampling approach of Romano and Shaikh (2008). This approach requires solving a linearly-constrained quadratic program for many hypothetical values of the parameter of interest and for many subsamples. This is computationally more expensive than solving the linear programming problem used to obtain the point estimates, but the cost is not prohibitive in my application. I am working on implementing this approach to obtain a full characterization of the confidence region; preliminary results show that the hypothesis that the proportion of patients with state dependence in period 2 is equal to 0 (zero) is strongly rejected at usual significance levels.

6. The impact of state dependence on health outcomes

In previous sections, I presented evidence that past hospital choices of a patient influence her current choice. If health or other shocks made a patient gravitate towards a certain hospital in the past, then the patient is more likely to choose that same facility for her current medical needs than an otherwise similar patient. Therefore, absent state dependence, demand patterns would be different.

In this section, I analyze the impact of state dependence on health outcomes. Lock-in might prevent patients from re-optimizing after a change in the environment. In particular, lock-in might prevent a patient from switching to hospitals that are more suitable than her previous choice at treating her current medical condition. Then, absent state dependence, patients would choose hospitals that produce better health outcomes. However, this is an empirical matter. It is possible that, absent state dependence, patients would switch to lower quality facilities, in which case the impact of state dependence on health outcomes is reversed.

I study this in the context of hospital choice for heart surgery (Coronary artery bypass grafting, CABG). I estimate a model of hospital choice that quantifies patients' preferences for hospital attributes and incorporates state dependence. I use the estimates of the model to compare expected mortality in the actual state of the world and in a counterfactual scenario in which there is no state dependence. The main question I want to answer is: Absent state dependence, how many more CABG patients would have survived?

There are two reasons why I focus on a particular medical condition rather than doing a more general analysis. First, risk adjustment and estimation of hospital quality are better defined when they are disease specific. Second, quality measures are difficult to compare across different procedures; therefore, it is hard to find a single measure of health outcomes that is appropriate for a wide range of medical conditions. I focus on CABG surgery for various reasons. First, it is a commonly performed procedure. Second, it is mostly performed on an elective, as opposed to an emergency, basis. Third, risk-adjusted mortality is a well accepted quality metric for CABG surgery. Fourth, there are many studies that analyze hospital choices of patients for CABG surgery, which provide me with a benchmark for assessing any effects I find.

In the analysis, I hold hospital quality fixed as observed in the data: I ignore any feedback between patients' demand patterns and quality choices of hospitals. In order to account for feedback, I would need a formal model of hospital behavior, which is beyond the scope of this study. Therefore, in assessing the impact of state dependence on health outcomes, I only consider the direct effect that takes into account how it affects the allocation of patients to hospitals.

Ideally, I would use the natural experiment discussed in Section 5 for this exercise. However, this is not feasible due to the size of the resulting sample: the number of patients in the treatment group who are hospitalized for CABG is very small[26]. As a result, I must rely on observational data to identify the parameters of interest, at the cost of less clean identification than in the previous section. However, I showed in previous sections that past hospital choices of patients have a causal impact on current behavior. Moreover, the identification of state dependence in those settings did not rely on arbitrary parametric assumptions. This lends credibility to the conclusions obtained from models with a richer structure like the one used in this section.

6.1. Framework

A patient has $T \geq 1$ episodes of care, where episode T corresponds to CABG surgery. Even though patients seek treatment for the same condition in the last episode, they might have received hospital care for different medical conditions in the past. The utility that a patient obtains from choosing a particular hospital for CABG surgery is a function of observed patient and hospital characteristics (which capture the quality of the match between them), the loyalty state of the patient (which summarizes the history of her past hospital choices), and factors unobserved by the researcher. The patient chooses the hospital in the choice set $\mathcal{J} = \{1, .., J\}$ that gives her the highest utility. In what follows, $\mathbb{P}^{SD}_{ij}$ is the ex-ante probability that patient i chooses hospital j in the actual state of the world (taking the loyalty state of the patient as given), and $\mathbb{P}^{NSD}_{ij}$ is the choice probability in the counterfactual scenario where there is no state dependence.

[26] Sample sizes are also small for other medical conditions studied in the literature.

The ultimate goal of the analysis is to evaluate the impact of state dependence on the ex-ante probability of death following CABG surgery. Let $Y_i = 1$ denote that patient i dies within 30 days of CABG surgery. For each hospital, there is a potential outcome Y_{ij} that denotes the outcome that would have been observed had patient i counterfactually chosen hospital j for CABG surgery. Potential outcomes are related to observed variables as follows:

$$Y_i = \sum_{j=1}^{J} Y_{ij} h_{ij} \tag{7}$$

where $h_{ij} = 1$ indicates that patient i chose hospital j.

Following the notation introduced above, the ex-ante expected mortality of patient i in scenario $K \in \{SD, NSD\}$ is given by:

$$EM_i^K = \sum_{j=1}^{J} \mathbb{P}_{ij}^K E(Y_{ij}) \tag{8}$$

This is the expected probability of death taking as given the loyalty state of the patient and before the choice and mortality errors are realized. The impact of state dependence on mortality across all patients who undergo CABG surgery is:

$$\Delta EM = \sum_i \left[EM_i^{SD} - EM_i^{NSD}\right] = \sum_i \sum_j \left[\mathbb{P}_{ij}^{SD} - \mathbb{P}_{ij}^{NSD}\right] E(Y_{ij}) \tag{9}$$

As discussed before, hospital-specific mortality is the same in both scenarios and therefore the difference in health outcomes is only due to different sorting of patients across hospitals. Then, the elements that I need to recover are: 1) The choice probabilities that govern the data generating process; 2) The choice probabilities in the scenario where there is no state dependence, and; 3) Hospital-specific mortality. The choice probabilities are computed from the estimates of a hospital demand model, while I compute hospital quality using patient discharge and mortality data[27]. After describing the data used for the analysis, I discuss these parts of the model in turn.

6.2. Data

I construct the sample for the analysis as follows. First, I identify all patients who had CABG surgery in a New York hospital during the period 2013-2015. I identify CABG cases based on the procedures listed in the discharge record of the patient. I then exclude patients who had valve surgery or another major cardiac surgical procedure during the same hospital stay, and patients who were transferred to the hospital from another health care facility.

[27]Appendix C shows how to cast this question within the nonparametric approach used in Section 5.4.

Second, I exclude CABG surgeries performed during an emergency episode. This way, I only study hospital choices of patients who are able to evaluate the different alternatives available. Third, I exclude a small number of patients who are more than 85 years old and patients with insurance other than Medicare, Medicaid, or private insurance. Fourth, I assume that the choice set of a patient consists of all hospitals in the market for CABG surgery (see below) that are within 100 miles of her home (25 miles for patients in New York City). Then, I drop patients who went to a hospital outside this area. The final sample consists of 7,509 patients.

For each patient, I construct the history of hospital visits before the CABG episode. To avoid picking-up past hospital use directly related to the latter, I consider all visits for cardiac care to the hospital chosen for CABG during the 60 days prior to the surgery as part of the CABG episode. The main loyalty variable is an indicator for whether the hospital was used in the previous episode of care.

The set of hospitals that offer CABG surgery is limited: 39 hospitals out of a total of 264 facilities in the State of New York offered this service during the period 2013-2015[28]. Table 11 shows summary statistics for these hospitals.

Table 12 reports summary statistics of the CABG episodes in the working sample. Almost 80% of patients are male. The average patient is around 66 years old, and more than half of the patients are covered by Medicare. On average, a patient had 4.6 episodes before the CABG surgery. The observed 30-day mortality rate is 0.99%. The average distance to the chosen hospital is 17.5 miles and 41% of patients go the closest hospital. There are 14 hospitals in the average patient's choice set.

In 2,905 cases (39% of the total), the hospital used in the previous episode is in the choice set of the patient. On average, this hospital was used in 59% of previous episodes and in 82% of previous visits to hospitals in the choice set; in 63% of the cases, it is the only hospital in the choice set that was used before. The average time elapsed since the last episode is 2.2 years. Past hospital use is a strong predictor of the current hospital choice of the patient: conditional on having used a hospital in the previous episode, the repurchase probability is 0.62.

6.3. Hospital quality

To compute hospital-specific mortality, I use patient level data of all hospitalizations for CABG surgery during the period 2013-2015. I regress an indicator for mortality within 30 days of discharge on characteristics of the patient (age/race/sex interactions), indicators

[28] Given my sample restrictions, there are two hospitals that were not chosen by any patient in the resulting dataset. Therefore, in the analysis, the set of hospitals in the market for CABG surgery contains 37 hospitals.

for being hospitalized for selected conditions in the year previous to CABG surgery[29], and hospital fixed effects. The hospital fixed effects are the risk-adjusted mortality rate estimates for hospitals and constitute the primary measure of hospital quality in the analysis. A regression of the estimated fixed effects on hospital attributes (not reported) suggests that mortality is negatively associated with provision of sophisticated services such as pediatric cardiac surgery, but there are no systematic differences between teaching and non-teaching hospitals.

The State of New York computes risk-adjusted mortality rates for CABG surgery by hospital. However, as pointed out by Chandra et al. (2016), these rates are computed as the ratio between observed and expected mortality, so they are not equivalent to the hospital fixed effects that I estimate. However, I find that these rates and my estimates are highly correlated (coefficient of correlation is 0.81).

6.4. Hospital demand

The key factors that affect hospital choices of patients are distance from home to the hospital, quality of care, and previous experience with the facility. I assume that the utility that patient i obtains from choosing hospital j for CABG surgery is:

$$u_{ij} = -\alpha_i d_{ij} + \beta_i q_j + \gamma_i \mathbb{I}(L_i = j) + \sum_k \delta_i w_{jk} + \xi_j + \epsilon_{ij} \tag{10}$$

where d_{ij} is the travel distance from patient i's home to hospital j, q_j is the risk-adjusted mortality rate of hospital j, L_i is the loyalty state of the patient, w_{jk} is the value of attribute k (for example, an indicator for public hospital) for hospital j, ξ_j captures unobserved features of the facility, and ϵ_{ij} reflects idiosyncratic preferences of the patient for the hospital. I assume that the latter is iid distributed according to a Type 1 extreme value distribution.

I assume that there is no outside option. I define the alternative specific constants relative to New York-Presbyterian Columbia University Medical Center, which consistently ranks as one of the top cardiology and heart surgery hospitals in the country according to U.S. News. I refer to the model's parameters other than γ as θ. I allow for observed and unobserved heterogeneity in preferences for hospital characteristics. In particular, the coefficients on distance, mortality, and previous use are a function of patient characteristics (demographics and severity) and unobserved factors. More precisely, I assume that:

[29] These conditions are: acute myocardial infarction, diabetes, diabetes with complications, hemiplegia/paraplegia, renal disease, cancer, metastatic cancer, mild liver disease, moderate/severe liver disease, AIDS, congestive heart failure, peripheral vascular disease, cerebrovascular disease, dementia, COPD, rheumatoid disease, and peptic ulcer.

$$\alpha_i = \bar{\alpha} + \sum_r \alpha_r^o z_{ir} + \alpha^u \mu_i \tag{11}$$

$$\beta_i = \bar{\beta} + \sum_r \beta_r^o z_{ir} + \beta^u \nu_i \tag{12}$$

$$\gamma_i = \bar{\gamma} + \sum_r \gamma_r^o z_{ir} + \gamma^u \eta_i \tag{13}$$

where z_{ir} is the value of (observed) attribute r for patient i, and $(\mu_i, \nu_i, \eta_i) \sim N(0, I)$ represent idiosyncratic preferences of consumer i for distance, quality, and loyalty, respectively.

This specification of patient utility is similar to past work, although it allows for richer heterogeneity. The third term in Equation 10 captures the impact of loyalty on current utility. Specifically, the patient gets a utility premium γ from choosing the hospital to which she is loyal. It can also be interpreted as an implied utility cost conditional on switching hospitals. This implies that for patient i to switch hospitals, she must prefer an alternative option by γ_i more than the hospital used in her previous episode.

One possibility is to model the hospital choice of a patient for all her episodes using Equation 10[30]. Then, the likelihood function at the patient level would be computed for the *sequence* of hospital choices. This way, I can exploit changes in choice set across episodes for a given patient to learn about her preferences. However, I would need to assume that preferences are stable across episodes. This would imply, for example, that the disutility of travel of the patient is the same for CABG and for other types of episodes. That assumption seems too strong. Instead, I model the choice of hospital in episode T_i (CABG) alone.

The main identification concern is the endogeneity of the previous choice variable. The fact that a patient chose a given hospital in the past indicates that she might have strong unobserved preferences for that facility; if these preferences are persistent across episodes and are not properly accounted for, then the previous use variable is positively correlated with the econometric error term. Then, a positive value of γ might be capturing both the effect of state dependence and persistent unobserved heterogeneity.

Given that the set of hospitals that offer CABG surgery is small, there are many consumers who have not developed an attachment to any particular hospital in the market for this procedure. For these consumers, state dependence does not affect preferences between hospitals. Then, the parameters in θ are identified from the choices of these "new" patients. This identification strategy has been used in the literature (see, for example, Handel (2013) and Luco (2016)). The state dependence parameter is then identified by the repurchase behavior of patients who are loyal to a particular facility.

[30] The different covariates would need to be indexed by t, as characteristics of patients and hospitals might change across episodes (for example, the patient might need hospital care for different diagnoses over time).

I estimate the model by Maximum Simulated Likelihood (Train, 2009). From the estimates of the model, I recover two sets of choice probabilities: (1) $P_{ij}(\theta, \gamma)$ is the ex-ante probability that patient i chooses hospital j in the actual state of the world: these are the predicted probabilities based on imposing the estimated coefficients of the utility function on the actual data; (2) $P_{ij}(\theta, 0)$ is the choice probability when the state dependence parameter is set equal to zero.

6.5. Results

The estimates from the demand model (see Table 13) indicate that patients dislike traveling for hospital care and value hospital quality. On average, a patient is willing to travel 12 additional miles for getting access to a hospital with one percentage point lower mortality. The impact of state dependence on patients' choices is large: if a hospital was used by the patient in her previous episode, then the probability of choosing that facility is about 3.4 times higher than what would be expected based on other covariates.

The impact of state dependence on health outcomes is a function of the spatial and quality configuration of local markets (the geographic distribution and composition of the patient population in the local area, and the locations and qualities of hospitals). Therefore, although the effects of state dependence on choices are substantial, the effects on health outcomes might not be as large. Moreover, it is not clear that eliminating the frictions driving state dependence will make patients switch to better hospitals in terms of quality.

I consider the impact on health outcomes of a number of thought experiments. I compare the ex-ante expected mortality under different counterfactual scenarios, relative to the actual state of the world. In each scenario, the state dependence parameter is set to a fraction of the baseline estimate: the lower the fraction, the less contaminated by state dependence the hospital choices of patients. We can think of these counterfactuals as illustrating the impact of policies that are partially effective in dealing with the sources of state dependence, or as capturing the possibility that some residual unobserved heterogeneity is picked up by the state dependence parameter. The results are reported in Panel A of Table 14. To provide a reference point to analyze the magnitude of these effects, I consider changes in mortality relative to the baseline scenario from reducing patients' disutility from travel. The results are reported in Panel B of Table 14.

There are two points to be noted. First, I find that reducing the magnitude of the state dependence coefficient leads to reductions in ex-ante expected mortality. In a counterfactual world where there is no state dependence, one extra person is expected to survive, which implies a 3% reduction in mortality relative to the baseline scenario (there were 33 actual deaths). In other words, state dependence prevents a stronger allocation of patients to higher-quality hospitals. To provide a benchmark for assessing the magnitude of this reduction in mortality, consider the findings of Gaynor et al. (2016). They study the impact of a reform in the English National Health Service that removed constraints on patient choice. They find that CABG patients became more responsive to clinical quality: the reallocation to higher quality hospitals led to a 3% reduction in expected mortality. Second, the impact of eliminating state dependence is similar to the effect of reducing patients' preferences for proximity by 40%. Therefore, the allocative role of state dependence is significant, considering that distance is one of the main determinants of patients' hospital choices.

It is important to note that the impact of state dependence on health outcomes that I find is specific to the setting considered. The direction or magnitude of the impact could be reversed for other patient populations (with different medical conditions) or for the same population in different markets.

7. Conclusion

When choosing a hospital, patients favor facilities they have used in the past. While the idea of patient loyalty seems widely accepted, there is no strong prior on whether unobserved heterogeneity or state dependence drive this behavior. These channels have different implications for hospital behavior, the long-run welfare effect of excluding a hospital from an insurer's network, and the design of policies to influence patient demand. To identify the sources of choice persistence in a credible way, I exploit quasi-exogenous shocks that induce a patient to try a new hospital: emergency hospitalizations and temporary hospital closures.

In the first case, I find that patients who visit a new hospital during an emergency hospitalization are more likely to continue using that same facility in subsequent episodes than observationally similar patients. This provides evidence of the presence of state dependence in hospital choices of patients. In cases where the emergency did not shift the loyalty state of the patient, repurchase rates are higher, which indicates that unobserved heterogeneity is also empirically relevant.

In the second case, I exploit the unexpected closures of three hospitals in New York City following Hurricane Sandy. I find that patients who needed hospital care during the time an affected hospital was closed for repairs were less likely to use the facility after its reopening than similar patients who did not have hospital visits during the unavailability window. Moreover, patients continued using the same hospital they visited during the time their usual facility was unavailable. The difference in behavior of these patients points to the presence of state dependence. To provide more credible evidence, I analyze patients' hospital choices using the nonparametric framework of Torgovitsky (2016). I am able to bound the proportion of patients who exhibit state dependence away from zero without making parametric assumptions about the nature of preference heterogeneity.

After showing that state dependence affects patients' choices, I look at its implications for health outcomes. In the context of hospital choice for heart surgery, I find that, absent state dependence, patients would switch to hospitals that are better at producing health outcomes. In particular, in this counterfactual scenario, ex-ante expected mortality would be 3% lower than in the actual state of the world.

There are interesting avenues for future research that originate from the results in this paper. The first is to investigate the microeconomic fundamentals that drive state dependence in this setting. Quantifying switching costs created by hospital policies regarding the sharing of patient data would be particularly interesting given the diffusion of Electronic Health Records (EHR) during the last years and policy efforts to achieve interoperability of different systems.

It would also be interesting to quantify the contribution of state dependence in hospital choices of patients to inertia in consumers' choices of health insurance plan. This is particularly relevant given the diffusion of narrow networks after implementation of the Affordable Care Act.

Another interesting question is how state dependence distortions hospitals' incentives to invest in quality. For example, hospitals might display bargain-then-ripoff behavior using quality as the adjustment variable. In addition, the possibility of developing long-term relationships with patients might lead hospitals to overinvest in services that patients use early in their lives.

References

Andrews, R. L. and T. Srinivasan (1995). "Studying consideration effects in empirical choice models using scanner panel data". *Journal of Marketing Research*, pp. 30–41.

Card, D., C. Dobkin, and N. Maestas (2009). "Does Medicare save lives?" *The Quarterly Journal of Economics* 124 (2), pp. 597–636.

Chandra, A., A. Finkelstein, A. Sacarny, and C. Syverson (2016). "Health Care Exceptionalism? Performance and Allocation in the US Health Care Sector". *American Economic Review* 106 (8), pp. 2110–44.

Desai, S. (2016). "The Role of Interoperability in Hospital Digitization: Network Effects and Information Blocking".

Doyle, J., J. Graves, J. Gruber, and S. Kleiner (2015). "Measuring returns to hospital care: Evidence from ambulance referral patterns". *The journal of political economy* 123 (1), p. 170.

Dubé, J.-P., G. J. Hitsch, and P. E. Rossi (2010). "State dependence and alternative explanations for consumer inertia". *The RAND Journal of Economics* 41 (3), pp. 417–445.

Ericson, K. M. M. (2014). "Consumer inertia and firm pricing in the Medicare Part D prescription drug insurance exchange". *American Economic Journal: Economic Policy* 6 (1), pp. 38–64.

Farrell, J. and P. Klemperer (2007). "Coordination and lock-in: Competition with switching costs and network effects". *Handbook of industrial organization* 3, pp. 1967–2072.

Gaynor, M., C. Propper, and S. Seiler (2016). "Free to Choose? Reform, Choice, and Consideration Sets in the English National Health Service". *American Economic Review* 106 (11), pp. 3521–57.

Goldfarb, A. (2006a). "State dependence at internet portals". *Journal of Economics & Management Strategy* 15 (2), pp. 317–352.

— (2006b). "The medium-term effects of unavailability". *Quantitative Marketing and Economics* 4 (2), pp. 143–171.

Handel, B. R. (2013). "Adverse selection and inertia in health insurance markets: When nudging hurts". *The American Economic Review* 103 (7), pp. 2643–2682.

Heckman, J. J. (1981). "Heterogeneity and state dependence". *Studies in labor markets*. University of Chicago Press, pp. 91–140.

Heckman, J. J., J. Smith, and N. Clements (1997). "Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts". *The Review of Economic Studies* 64 (4), pp. 487–535.

HITPC (2015). *Challenges and Barriers to Interoperability, Report to Congress*. Tech. rep.

Ho, K., J. Hogan, and F. S. Morton (2017). "The Impact of Consumer Inattention on Insurer Pricing in the Medicare Part D Program".

Illanes, G. (2016). *Switching Costs in Pension Plan Choice*. Tech. rep. Working Paper.

Israel, M. (2005). "Tenure dependence in consumer-firm relationships: An empirical analysis of consumer departures from automobile insurance firms". *RAND Journal of Economics*, pp. 165–192.

Jung, K., R. Feldman, and D. Scanlon (2011). "Where would you go for your next hospitalization?" *Journal of health economics* 30 (4), pp. 832–841.

Luco, F. (2016). “Switching Costs and Competition in Retirement Investment”.

Miller, A. R. and C. Tucker (2009). “Privacy protection and technology diffusion: The case of electronic medical records”. *Management Science* 55 (7), pp. 1077–1093.

— (2014). “Health information exchange, system size and information silos”. *Journal of health economics* 33, pp. 28–42.

Nosal, K. (2012). “Estimating switching costs for medicare advantage plans”. *Unpublished manuscript, University of Mannheim*.

Polyakova, M. (2015). *Regulation of Insurance with Adverse Selection and Switching Costs: Evidence from Medicare Part D*. Tech. rep. National Bureau of Economic Research.

Raval, D. and T. Rosenbaum (2017). “Why Do Previous Choices Matter for Hospital Demand? Decomposing Switching Costs from Unobserved Preferences”.

Raval, D., T. Rosenbaum, and N. Wilson (2017). “Using Disaster Induced Closures to Evaluate Discrete Choice Models of Hospital Demand”. *Working paper*.

Romano, J. P. and A. M. Shaikh (2008). “Inference for identifiable parameters in partially identified econometric models”. *Journal of Statistical Planning and Inference* 138 (9), pp. 2786–2807.

Samuelson, W. and R. Zeckhauser (1988). “Status quo bias in decision making”. *Journal of risk and uncertainty* 1 (1), pp. 7–59.

Shepard, M. (2016). *Hospital network competition and adverse selection: Evidence from the Massachusetts health insurance exchange*. Tech. rep. National Bureau of Economic Research.

Stigler, G. J. and G. S. Becker (1977). “De gustibus non est disputandum”. *The american economic review* 67 (2), pp. 76–90.

Sudhir, K and N. Yang (2014). “Exploiting the choice-consumption mismatch: A new approach to disentangle state dependence and heterogeneity”.

Torgovitsky, A. (2016). “Nonparametric Inference on State Dependence with Applications to Employment Dynamics”.

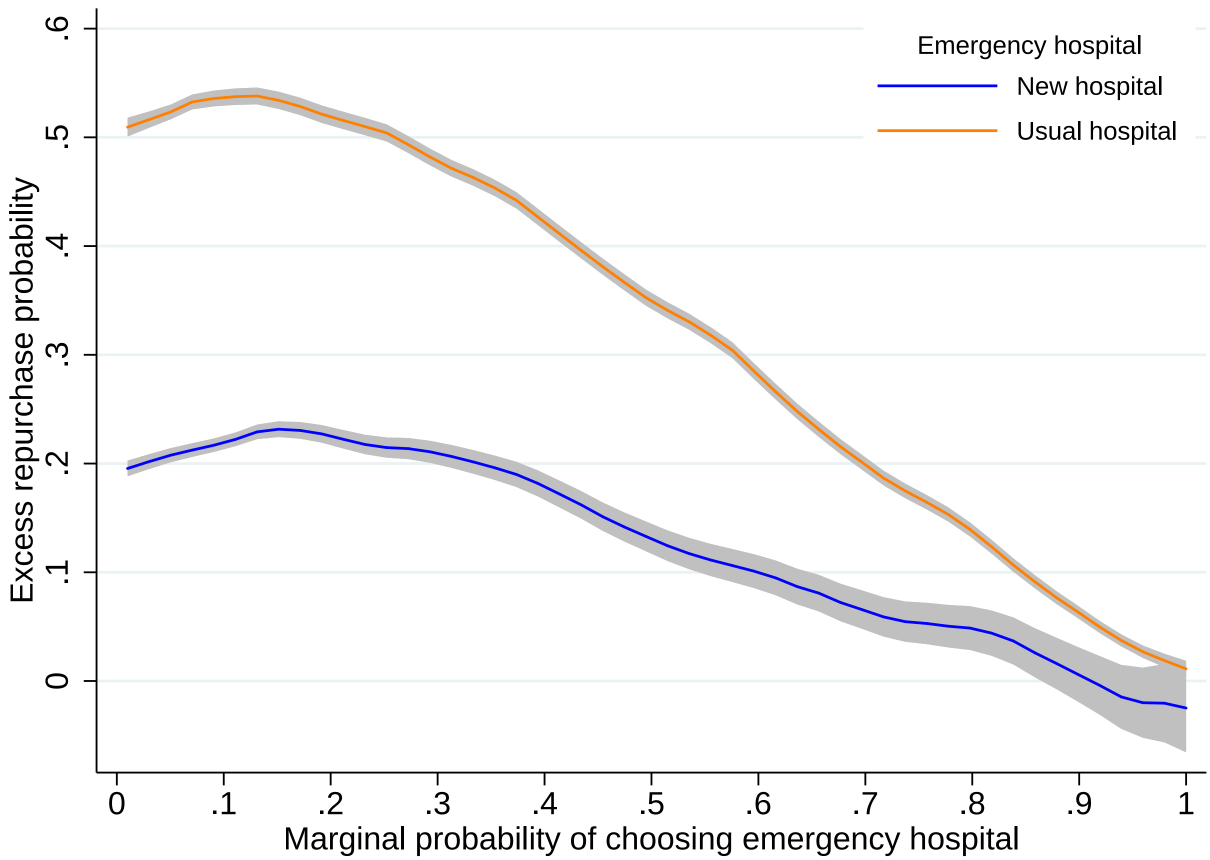
Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.

Table 1: Most common non-deferrable conditions

Description	ICD-9-CM	Weekend	Intensive care	Emergency (SPARCS)	30 days mortality	Length of stay	Admission hour		
							p25	p50	p75
Intertrochanteric fx-cl	820.21	0.28	0.15	0.96	7.9%	7.1	10	15	19
Cellulitis of face	682.0	0.28	0.04	0.98	0.9%	3.7	11	15	18
Cardiac arrest	427.5	0.29	0.81	0.97	75.2%	8.1	9	14	18
Poison-medicinal agt NOS	977.9	0.29	0.45	0.97	1.5%	3.4	7	14	19
Rhabdomyolysis	728.88	0.28	0.15	0.98	4.3%	6.2	10	15	19
Epistaxis	784.7	0.29	0.17	0.98	5.1%	4.1	8	13	19
Ac alcohol intox-contin	303.01	0.29	0.11	0.98	0.5%	3.5	7	15	19
Closed fracture of pubis	808.2	0.29	0.08	0.96	4.2%	5.3	10	15	19
Subarachnoid hemorrhage	430	0.28	0.74	0.98	20.0%	11.8	8	14	19

Notes: The table shows the most common non-deferrable conditions among emergency department (ED) admissions in the full inpatient data. To determine if a diagnosis is considered non-deferrable, I test whether the fraction of ED admissions that occur during the weekend is statistically different from 2/7. The reported statistics correspond to episodes in which the patient is admitted to the hospital through the ED. Intensive care refers to whether the patient spent at least one day in an intensive care unit during the hospital stay. The emergency variable from SPARCS indicates an episode in which "the patient requires immediate medical intervention as a result of severe, life threatening, or potentially disabling conditions." The other two major categories used by SPARCS to categorize hospital admissions are: 1) Urgent: The patient requires immediate attention for the care and treatment of a physical or mental disorder. Generally the patient is admitted to the first available and suitable accommodation; 2) Elective: The patient's condition permits adequate time to schedule the admission based on the availability of a suitable accommodation.

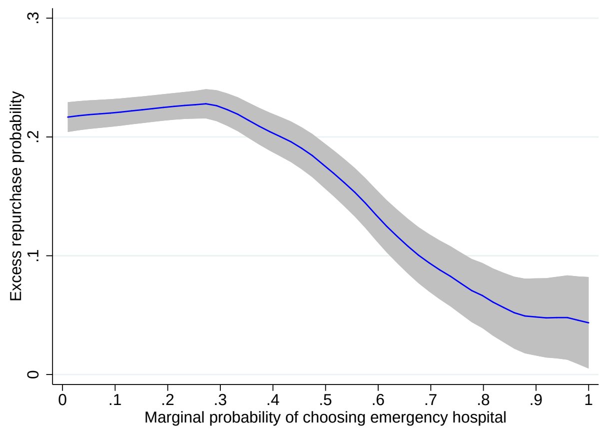
Figure 1: Excess repurchase probability, all cases



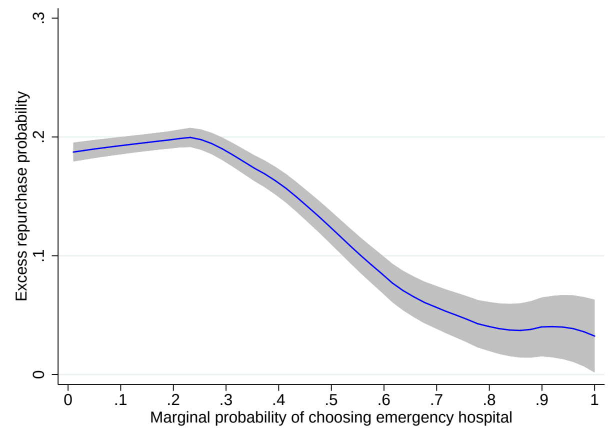
Notes: The figure shows the excess repurchase probability as a function of the marginal probability of choosing the emergency hospital, as described in Section 4. Included: episodes following an emergency hospitalization. There are separate schedules for episodes following emergencies that shifted and did not shift the loyalty state of the patient. The shaded areas represent 95% confidence intervals.

Figure 2: Excess repurchase probability, by type of emergency

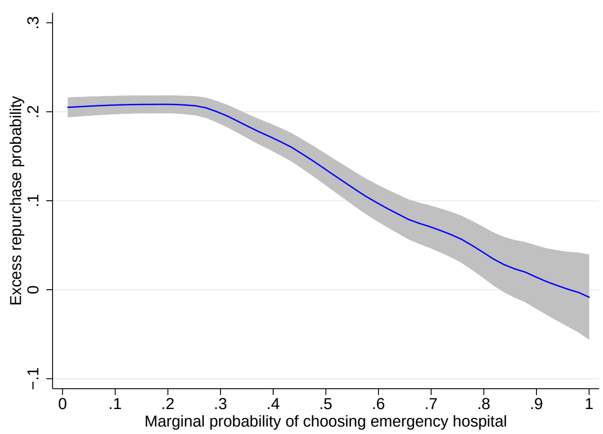
(a) Intensive care



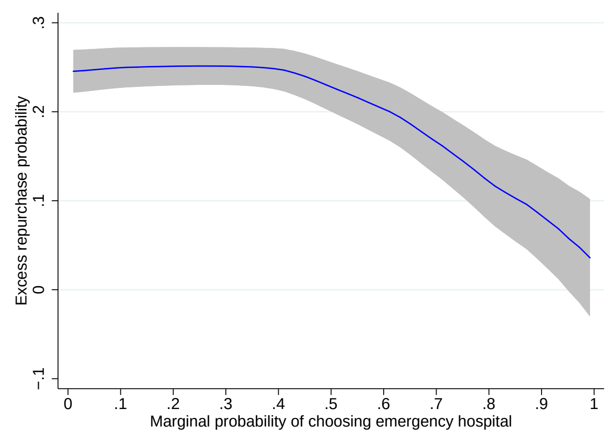
(b) Injury or poisoning



(c) Night admission

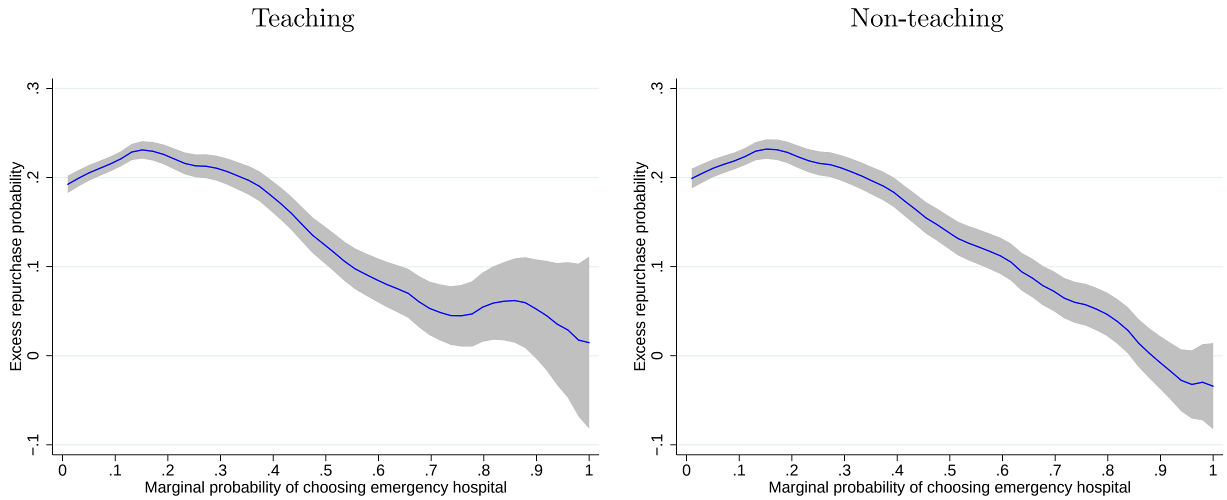


(d) Heart attack and stroke



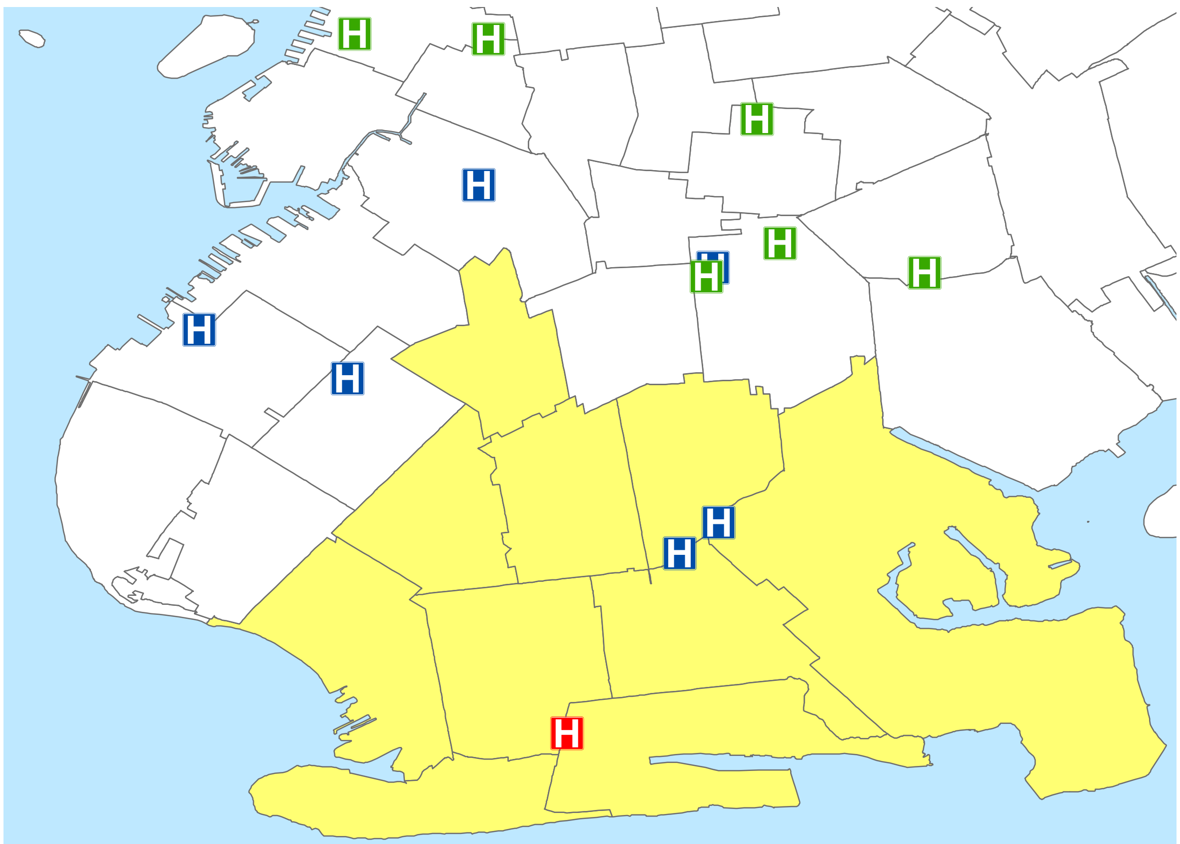
Notes: The figures show the excess repurchase probability as a function of the marginal probability of choosing the emergency hospital, as described in Section 4. Included: episodes following an emergency hospitalization that shifted the loyalty state of the patient. Different figures correspond to different types of emergencies: a) During the emergency hospitalization, the patient spent time in an intensive care unit; b) The emergency was coded as injury or poisoning; c) The emergency admission took place during the night (between 9PM and 5AM), and; d) The emergency was coded as heart attack or cerebrovascular accident. The shaded areas represent 95% confidence intervals.

Figure 3: Excess repurchase probability, by type of hospital



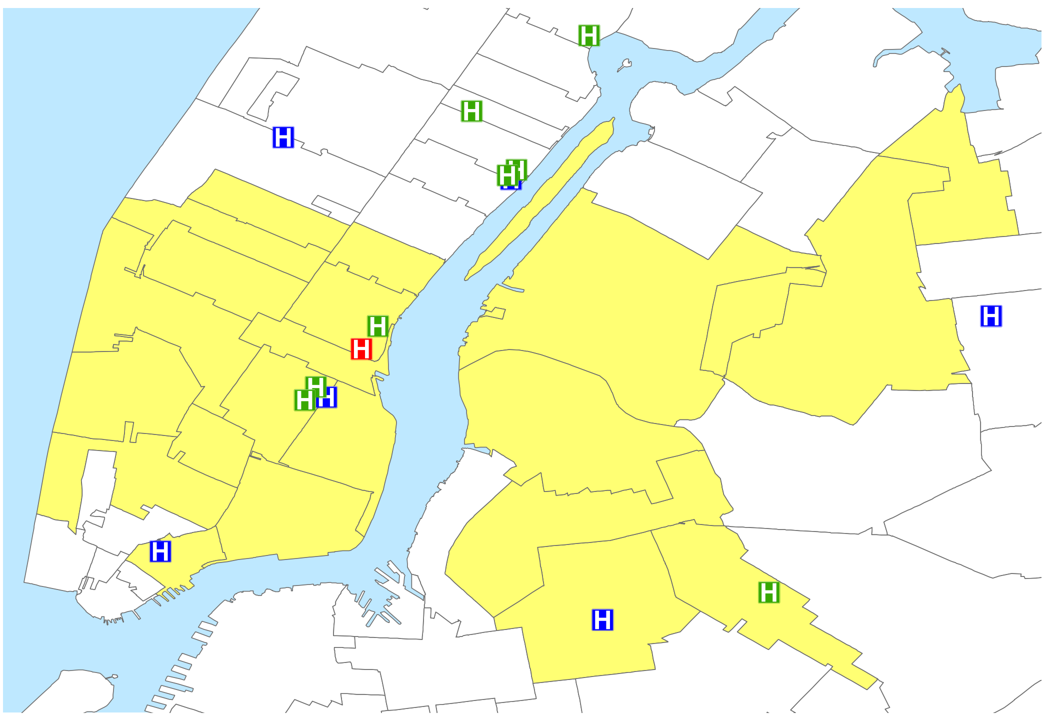
Notes: The figures show the excess repurchase probability as a function of the marginal probability of choosing the emergency hospital, as described in Section 4. Included: episodes following an emergency hospitalization that shifted the loyalty state of the patient. Different figures correspond to different types of emergency hospital. The shaded areas represent 95% confidence intervals.

Figure 4: Service area of Coney Island Hospital



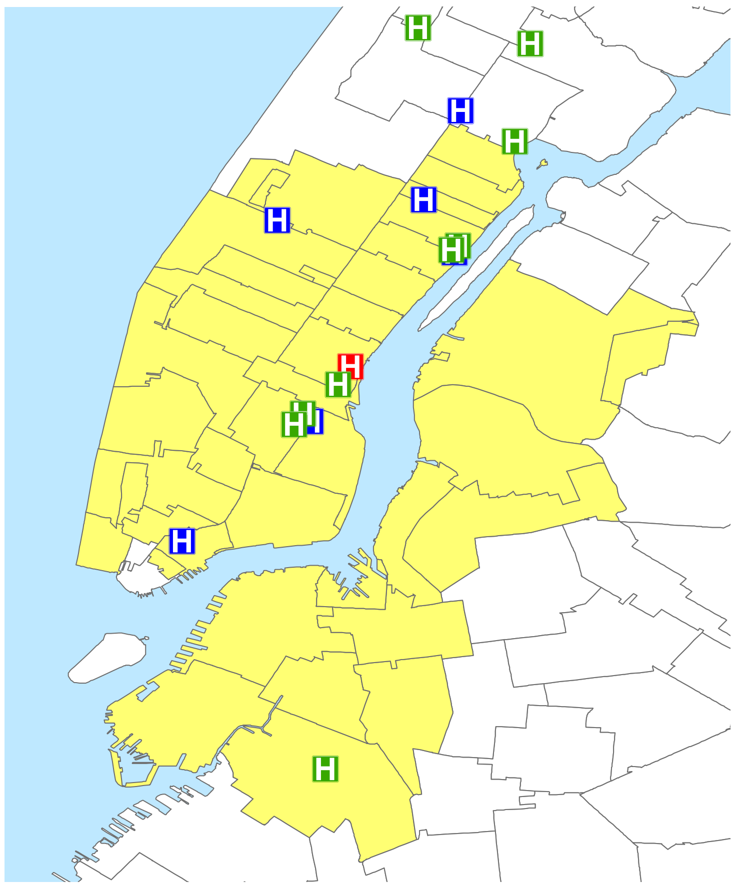
Notes: The figure shows the service area of Coney Island Hospital (zip codes in yellow) constructed as explained in Section 5.2. The red hospital is Coney Island Hospital, the blue hospitals are its main competitors, and the green hospitals are other nearby facilities.

Figure 5: Service area of Bellevue



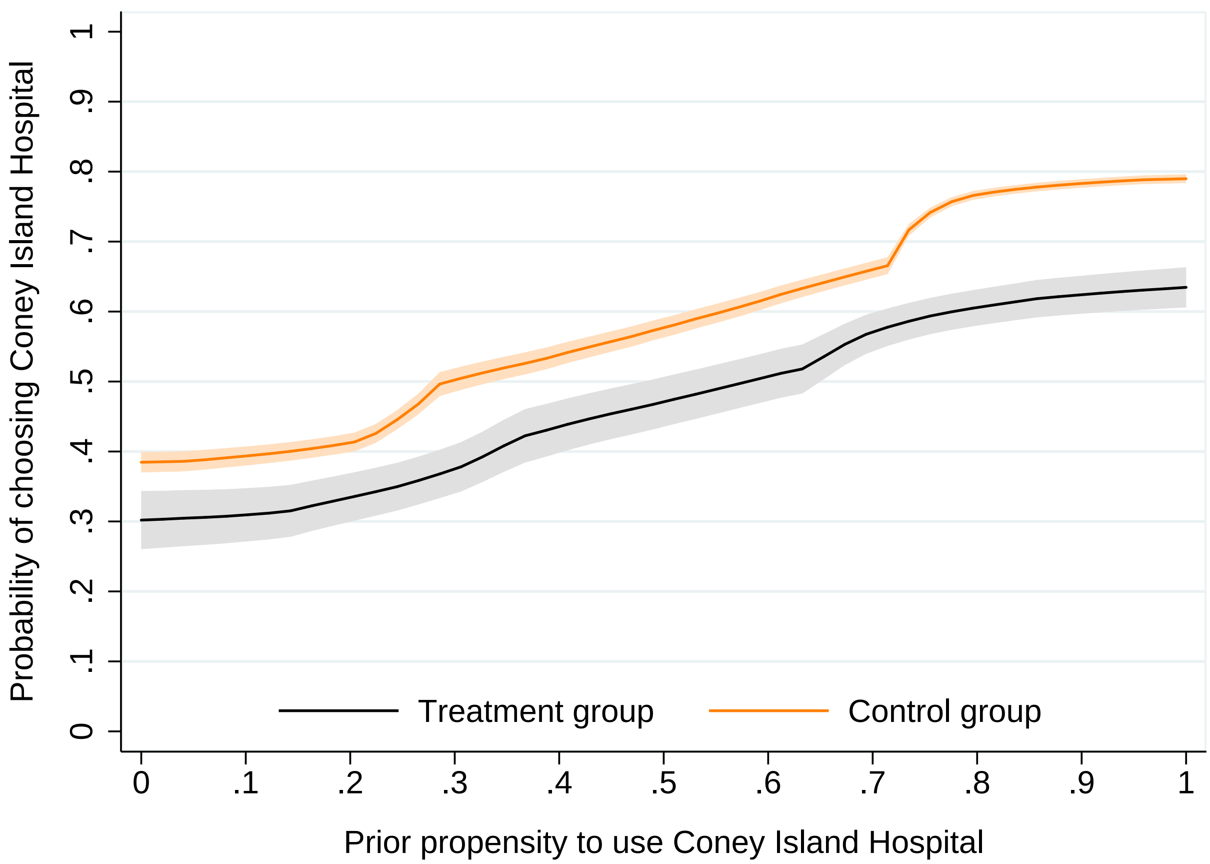
Notes: The figure shows the service area of Bellevue Hospital (zip codes in yellow) constructed as explained in Section 5.2. The red hospital is Bellevue, the blue hospitals are its main competitors, and the green hospitals are other nearby facilities.

Figure 6: Service area of NYU Langone



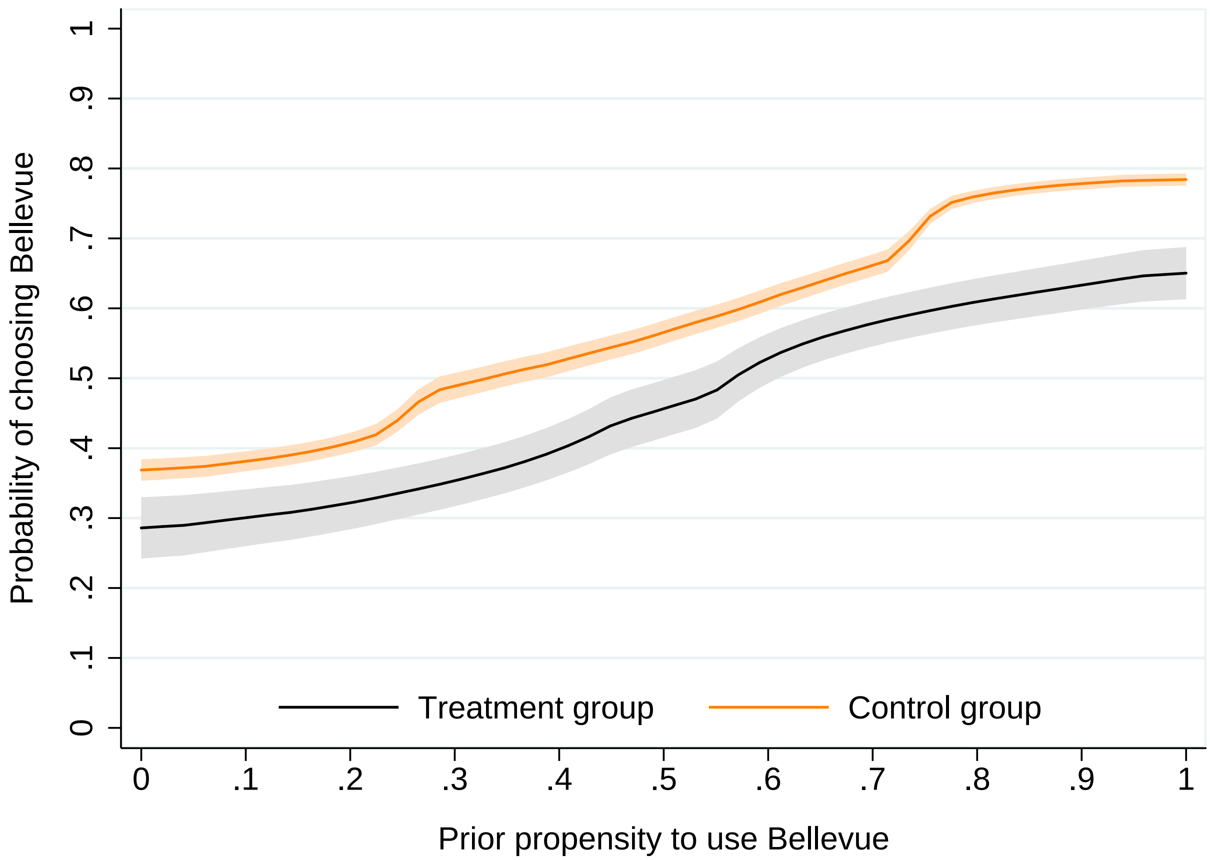
Notes: The figure shows the service area of NYU Langone (zip codes in yellow) constructed as explained in Section 5.2. The red hospital is NYU Langone, the blue hospitals are its main competitors, and the green hospitals are other nearby facilities.

Figure 7: Probability of choosing Coney Island Hospital in period 2



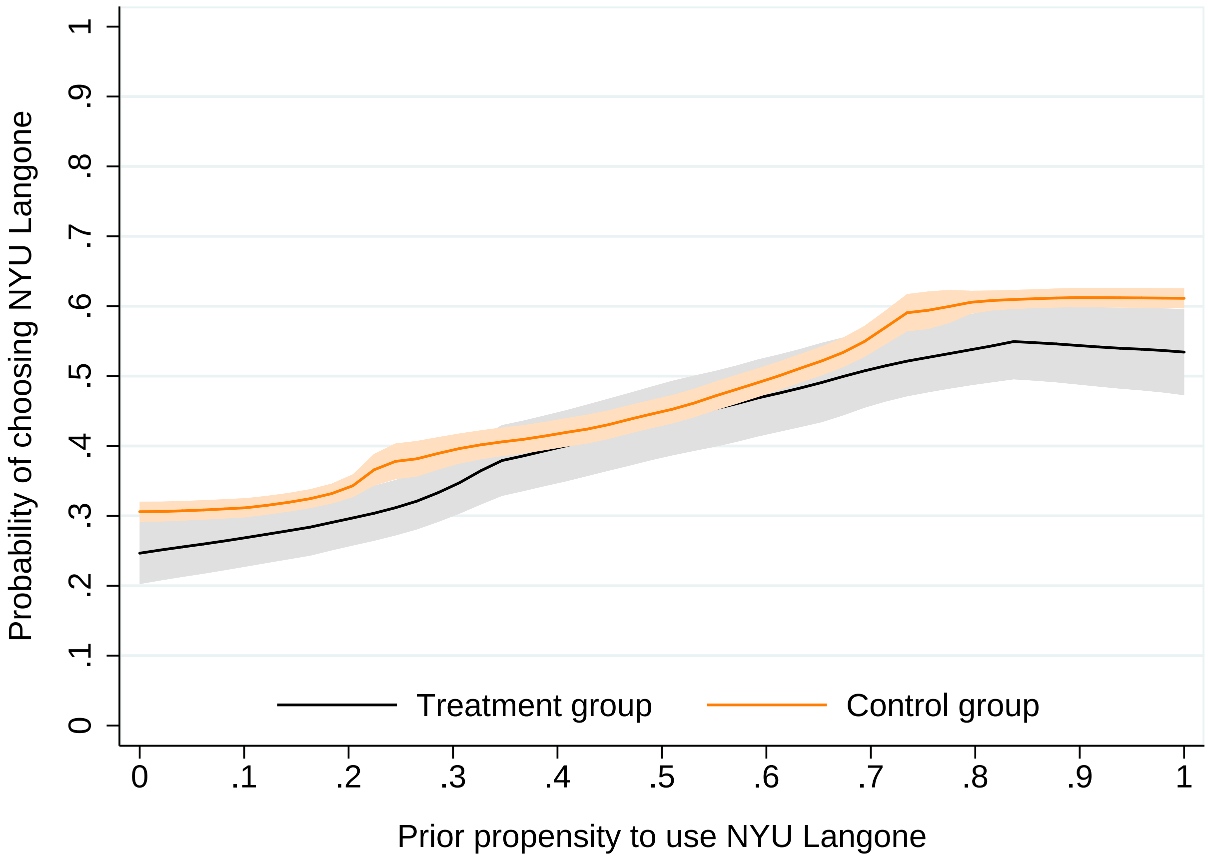
Notes: The figure shows the probability of choosing Coney Island Hospital in the first episode of period 2 as a function of the prior propensity to use the facility. This propensity is defined as the proportion of the patient's hospital visits prior to the storm that were to Coney Island Hospital. Included: patients who were loyal to Coney Island Hospital at the moment of the storm. Kernel-weighted local regression estimates. The shaded areas represent 95% confidence intervals.

Figure 8: Probability of choosing Bellevue in period 2



Notes: The figure shows the probability of choosing Bellevue in the first episode of period 2 as a function of the prior propensity to use the facility. This propensity is defined as the proportion of the patient's hospital visits prior to the storm that were to Bellevue. Included: patients who were loyal to Bellevue at the moment of the storm. Kernel-weighted local regression estimates. The shaded areas represent 95% confidence intervals.

Figure 9: Probability of choosing NYU Langone in period 2



Notes: The figure shows the probability of choosing NYU Langone in the first episode of period 2 as a function of the prior propensity to use the facility. This propensity is defined as the proportion of the patient's hospital visits prior to the storm that were to NYU Langone. Included: patients who were loyal to NYU Langone at the moment of the storm. Kernel-weighted local regression estimates. The shaded areas represent 95% confidence intervals.

Table 2: Episodes following an emergency hospitalization

	Emergency hospital	
	New	Usual
Number of cases	40,549	105,390
Episodes before emergency:		
Number of episodes	1.6	3.8
Propensity to use emergency hospital	0	0.82
Episode after emergency:		
Days since emergency	393	349
Repurchase rate	0.42	0.78
Marginal choice probability	0.24	0.39
Inpatient	0.27	0.31
Emergency Department	0.59	0.57
Ambulatory Surgery	0.14	0.12

Notes: The table shows summary statistics of the episodes following an emergency hospitalization that are used in the analysis of Section 4. I distinguish episodes based on whether the emergency hospital was being used by the patient before the emergency or it is a new hospital. The prior propensity to use the emergency hospital is defined as the proportion of the patient's hospital visits prior to the emergency episode that were to the emergency hospital. The repurchase rate is the proportion of patients who chose the emergency hospital in the episode following the emergency. Inpatient, Ambulatory Surgery and Emergency Department categorize hospital visits. The marginal choice probability is the probability of choosing the emergency hospital based on characteristics of the patient in the current episode (the next episode after the emergency hospitalization).

Table 3: Characteristics of affected hospitals

	Coney Island Hospital	Bellevue	NYU Langone
Visits:			
Ambulatory Surgery	4,487	7,697	17,710
Emergency Department	60,621	93,538	36,475
Inpatient	17,580	26,763	33,095
Demographics, inpatient:			
Age	50.6	42.7	43.6
Female	0.53	0.39	0.58
Medicare	0.35	0.17	0.27
Medicaid	0.48	0.49	0.08
Private	0.09	0.08	0.62
Other insurance	0.08	0.25	0.03
Certified beds:			
Total	371	912	987
Services:			
Perinatal Designation	Level 2	Regional Center	Regional Center
CABG	No	Yes	Yes
Transplant Center	No	No	Yes

Notes: The table presents a summary profile of the hospitals affected by Hurricane Sandy. The number of visits and the demographic profile of patients correspond to the 12 month period preceding the storm (November 2011 - October 2012).

Table 4: Summary statistics, Coney Island Hospital

	Sample 1		Sample 2		Sample 3	
	Control	Treatment	Control	Treatment	Control	Treatment
Type of visit:						
Inpatient	0.21	0.26	0.15	0.25	0.22	0.30
Emergency Department	0.58	0.55	0.73	0.66	0.69	0.62
Ambulatory Surgery	0.21	0.19	0.12	0.10	0.09	0.09
Demographics:						
Age	46.1	49.5	39.6	44.8	43.1	47.6
Female	0.56	0.57	0.53	0.56	0.56	0.58
Medicare	0.24	0.33	0.15	0.27	0.23	0.35
Medicaid	0.31	0.33	0.40	0.44	0.37	0.40
Private insurance	0.28	0.23	0.15	0.11	0.22	0.17
Other insurance	0.17	0.11	0.30	0.17	0.17	0.08
Prior propensity:						
0	0.75	0.77	0.10	0.13	0.63	0.71
0-0.25	0.02	0.04	0.01	0.04	0.02	0.04
0.25-0.50	0.03	0.04	0.04	0.09	0.03	0.05
0.50-0.75	0.04	0.05	0.09	0.15	0.05	0.06
0.75-1	0.18	0.10	0.77	0.59	0.28	0.14
Patients	193,222	19,577	38,615	2,541	102,867	8,882

Notes: The table presents summary statistics of the patients used for the case study of Coney Island Hospital. Treatment status refers to whether the patient had a hospital visit in period 1. The different samples are described in detail in Section 5.2. Sample # 1 contains patients who had at least one hospital visit in both periods 0 and 2. Sample #2 contains those patients in sample #1 who visited Coney Island Hospital in the last episode of period 0. Sample #3 contains those patients in sample #1 who chose either Coney Island Hospital or one of its main competitors in the last episode of period 0, in period 1 (if in the treatment group), and in the first episode of period 2. Characteristics correspond to the first episode of period 2. Prior propensity is defined as the proportion of the patient's hospital visits prior to the storm that were to Coney Island Hospital.

Table 5: Summary statistics, Bellevue

	Sample 1		Sample 2		Sample 3	
	Control	Treatment	Control	Treatment	Control	Treatment
Type of visit:						
Inpatient	0.16	0.21	0.15	0.20	0.15	0.19
Emergency Department	0.65	0.60	0.75	0.72	0.76	0.72
Ambulatory Surgery	0.19	0.19	0.10	0.09	0.09	0.08
Demographics:						
Age	43.7	48.9	42.4	47.5	41.0	45.4
Female	0.54	0.57	0.46	0.49	0.53	0.57
Medicare	0.21	0.32	0.16	0.27	0.19	0.30
Medicaid	0.34	0.36	0.42	0.47	0.41	0.45
Private insurance	0.28	0.21	0.11	0.09	0.20	0.14
Other insurance	0.16	0.11	0.31	0.18	0.20	0.12
Prior propensity:						
0	0.86	0.87	0.14	0.16	0.74	0.80
0-0.25	0.02	0.03	0.03	0.07	0.03	0.05
0.25-0.50	0.02	0.03	0.05	0.08	0.03	0.04
0.50-0.75	0.02	0.03	0.09	0.15	0.03	0.04
0.75-1	0.08	0.05	0.69	0.54	0.17	0.08
Patients	230,040	20,267	24,047	1,510	88,938	6,940

Notes: The table presents summary statistics of the patients used for the case study of Bellevue Hospital. Treatment status refers to whether the patient had a hospital visit in period 1. The different samples are described in detail in Section 5.2. Sample # 1 contains patients who had at least one hospital visit in both periods 0 and 2. Sample #2 contains those patients in sample #1 who visited Bellevue in the last episode of period 0. Sample #3 contains those patients in sample #1 who chose either Bellevue or one of its main competitors in the last episode of period 0, in period 1 (if in the treatment group), and in the first episode of period 2. Characteristics correspond to the first episode of period 2. Prior propensity is defined as the proportion of the patient's hospital visits prior to the storm that were to Bellevue.

Table 6: Summary statistics, NYU Langone

	Sample 1		Sample 2		Sample 3	
	Control	Treatment	Control	Treatment	Control	Treatment
Type of visit:						
Inpatient	0.20	0.22	0.26	0.31	0.28	0.27
Emergency Department	0.53	0.58	0.43	0.44	0.55	0.60
Ambulatory Surgery	0.28	0.21	0.32	0.26	0.18	0.13
Demographics:						
Age	48.2	48.7	50.9	62.5	48.0	50.0
Female	0.55	0.57	0.61	0.59	0.60	0.58
Medicare	0.26	0.31	0.27	0.49	0.27	0.36
Medicaid	0.22	0.32	0.06	0.09	0.17	0.29
Private insurance	0.41	0.27	0.61	0.38	0.50	0.30
Other insurance	0.12	0.11	0.06	0.04	0.06	0.06
Prior propensity:						
0	0.91	0.92	0.19	0.24	0.84	0.89
0-0.25	0.01	0.02	0.02	0.08	0.01	0.03
0.25-0.50	0.01	0.02	0.05	0.12	0.02	0.02
0.50-0.75	0.02	0.02	0.09	0.15	0.02	0.02
0.75-1	0.06	0.03	0.65	0.42	0.11	0.04
Patients	299,009	28,933	20,182	1,118	91,773	8,022

Notes: The table presents summary statistics of the patients used for the case study of NYU Langone. Treatment status refers to whether the patient had a hospital visit in period 1. The different samples are described in detail in Section 5.2. Sample # 1 contains patients who had at least one hospital visit in both periods 0 and 2. Sample #2 contains those patients in sample #1 who visited NYU Langone in the last episode of period 0. Sample #3 contains those patients in sample #1 who chose either NYU Langone or one of its main competitors in the last episode of period 0, in period 1 (if in the treatment group), and in the first episode of period 2. Characteristics correspond to the first episode of period 2. Prior propensity is defined as the proportion of the patient's hospital visits prior to the storm that were to NYU Langone.

Table 7: Probit estimates

Days since reopening	0	30	60	90	120	150
Panel A: Coney Island Hospital						
Treatment status	-0.091	-0.084	-0.080	-0.080	-0.075	-0.072
(Std Error)	(0.009)	(0.010)	(0.010)	(0.011)	(0.011)	(0.012)
Sample size	41,153	38,640	36,318	34,174	32,084	30,095
Panel B: Bellevue						
Treatment status	-0.105	-0.100	-0.104	-0.108	-0.104	-0.099
(Std Error)	(0.012)	(0.013)	(0.014)	(0.015)	(0.016)	(0.017)
Sample size	25,550	23,684	21,891	20,287	18,856	17,560
Panel C: NYU Langone						
Treatment status	-0.065	-0.066	-0.066	-0.064	-0.062	-0.058
(Std Error)	(0.014)	(0.015)	(0.015)	(0.016)	(0.017)	(0.017)
Sample size	21,295	20,155	19,056	17,967	16,914	15,944

Notes: The table shows estimates for the probit model described in Section 5.3. The dependent variable takes the value 1 if the patient visited the affected hospital in period 2 and 0 otherwise. The explanatory variables are an indicator for treatment status, propensity-bin fixed effects, zip code fixed effects, diagnosis group fixed effects, and a spline of the number of days between the reopening of the affected hospital and the episode. The indicator for treatment status takes the value 1 if the patient needed hospital care while the affected hospital was closed for repairs and 0 otherwise. The estimates reported correspond to the average marginal effect of treatment status. Days since reopening refers to the episodes included in the estimation sample: if days since reopening is x, then only episodes that took place at least x days after the reopening of the affected hospital are included in the estimation sample.

Table 8: Logit estimates

	Case study		
	Coney Island Hospital	Bellevue	NYU Langone
Used in previous episode	1.968	2.109	1.890
	(0.068)	(0.101)	(0.091)
Distance	-0.130	-0.476	-0.321
	(0.003)	(0.008)	(0.008)
Affected hospital x Prior propensity			
Prior propensity: 0 - 0.25	0.011	-0.058	-0.036
	(0.104)	(0.092)	(0.111)
Prior propensity: 0.25-0.50	0.338	0.419	0.381
	(0.067)	(0.072)	(0.077)
Prior propensity: 0.50-0.75	0.774	0.871	0.795
	(0.054)	(0.061)	(0.064)
Prior propensity: 0.75-1	1.217	1.442	0.999
	(0.040)	(0.044)	(0.043)
Public hospital x Insurance			
Insurance: Medicare	0.027	0.877	1.022
	(0.045)	(0.058)	(0.099)
Insurance: Medicaid	0.350	1.052	1.724
	(0.037)	(0.050)	(0.122)
Insurance: Other	1.315	1.969	2.290
	(0.044)	(0.058)	(0.115)

Notes: The table shows estimates for the hospital choice model described in Section 5.3. This is a conditional logit model of hospital choice in the first episode of period 2. The previous use variable is an indicator for whether the hospital considered was chosen by the patient in her previous episode of care. Prior propensity is defined as the proportion of the patient's hospital visits prior to the storm that were to the affected hospital. In addition to the variables shown, the model includes hospital - type of visit fixed effects. The effect of previous hospital use is economically significant: having used a hospital in period 1 increases the probability of choosing that same facility in the first episode after the reopening of the affected hospital 4.4, 4.6, and 3.9 times above the baseline probability, respectively.

Table 9: Period t latent choices, J=2

		$h_t(2)$ 1	2	
$h_t(1)$	1	$\mathbb{P}_{11}$	$\mathbb{P}_{12}$	$\mathbb{P}_{1*}$
	2	$\mathbb{P}_{21}$	$\mathbb{P}_{22}$	$\mathbb{P}_{2*}$
		$\mathbb{P}_{*1}$	$\mathbb{P}_{*2}$	

Notes: $\mathbb{P}_{j*}$ denotes the marginal probability $\mathbb{P}[h_t(1) = j]$, $\mathbb{P}_{*j}$ denotes the marginal probability $\mathbb{P}[h_t(2) = j]$, and P_{ij} denotes the probability $\mathbb{P}[h_t(1) = i, h_t(2) = j]$.

Table 10: Proportion of patients who exhibit state dependence in period 2: lower bound

Coney Island Hospital	Bellevue	NYU Langone
0.091	0.092	0.129

Notes: The table shows, for each case study, the lower bound on the proportion of patients who exhibit state dependence in period 2 under Assumptions 1 through 4 discussed in Section 5.4.

Table 11: Hospitals that offer CABG surgery

	Number	Proportion
Public	5	0.14
Teaching	23	0.62
Pediatric surgery	11	0.30
Heart transplant, adult	6	0.16
Heart transplant, pediatric	2	0.05
Total	37	1

Notes: The table shows summary statistics of hospitals in the market for CABG surgery during the period 2013-2015.

Table 12: Summary statistics of CABG patients

Age	66.4
Female	0.22
White	0.70
NYC	0.36
Private insurance	0.32
Medicare	0.55
Medicaid	0.13
Charlson = 0	0.23
Charlson = 1	0.31
Charlson = 2	0.47
Episodes before CABG	4.63
Died (%)	0.99
Distance	17.6
Closest	0.41
Observations	7,509

Notes: The table shows summary statistics of CABG patients in the sample defined in Section 6.2.

Table 13: Estimates of the hospital demand model

		Estimate	Std Error
Used in previous episode	Constant	1.850	0.219
	Female	0.124	0.199
	No white	-0.012	0.183
	Charlson Index = 1	-0.002	0.233
	Charlson Index = 2	0.533	0.215
	Medicare	-0.067	0.184
	Medicaid	0.432	0.283
	Std dev	2.656	0.224
Distance	Constant	0.140	0.007
	NYC	0.095	0.008
	Female	0.000	0.005
	No white	-0.008	0.006
	Charlson Index = 1	0.004	0.006
	Charlson Index = 2	-0.005	0.005
	Medicare	0.000	0.005
	Medicaid	0.014	0.008
	Std dev	0.045	0.004
Mortality	Constant	-1.867	0.172
	Female	0.053	0.043
	No white	0.008	0.040
	Charlson Index = 1	0.041	0.051
	Charlson Index = 2	-0.029	0.049
	Medicare	0.087	0.046
	Medicaid	0.187	0.058
	Std dev	0.130	0.094

Notes: The table shows the estimates from the hospital choice model described in Section 6.4.

Table 14: Ex-ante expected mortality relative to actual state of the world

Panel A: Impact on mortality from reducing state dependence

γ/γ^0	0.8	0.6	0.4	0.2	0
ΔEM	-0.129	-0.291	-0.490	-0.736	-1.007

Panel B: Impact on mortality from reducing disutility from travel

α/α^0	0.8	0.6	0.4	0.2	0
ΔEM	-0.442	-0.950	-1.543	-2.204	-2.776

Notes: The table shows the change in expected mortality (in terms of number of deaths) under a range of counterfactual scenarios, relative to the actual state of the world in which preferences are given by the estimates reported in Table 13. The effects reported correspond to patients for whom the hospital used in the previous episode is in the choice set. In Panel A, different columns correspond to different counterfactual scenarios where the state dependence coefficient has been set equal to a given fraction (0.8, 0.6, 0.4, 0.2, 0) of the baseline estimate. In Panel B, different columns correspond to different counterfactual scenarios where the disutility of distance has been set equal to a given fraction (0.8, 0.6, 0.4, 0.2, 0) of the baseline. γ^0 and α^0 are the baseline coefficients, while γ and α are the counterfactual coefficients.

Appendix A. Hurricane Sandy

Sandy formed as a tropical storm on October 22th, 2012, in the Caribbean and intensified into a hurricane as it moved northward across Jamaica, Cuba and the Bahamas. It then moved northeast of the United States until turning west toward the mid Atlantic coast on the 28th. It transitioned into a post-tropical cyclone just prior to moving onshore near Atlantic City, New Jersey, on the 29th, and dissipated over western Pennsylvania by the 31th. The storm affected many states, but was particularly severe in New York and New Jersey. It is regarded as one of the most destructive storms in U.S. history: according to several sources, the storm caused more than 100 deaths in the U.S., with at least 40 deaths in New York City (NYC), while economic damages are calculated to have exceeded $50 billion.

The storm hit the New York Metropolitan area on October 29th. Preparations in NYC began on October 26th and included, among other measures, mandatory evacuations, flights cancellations, suspension of public transportation, and cancellation of classes. The storm affected the entire city, although zones close to the water suffered the most. Storm surge caused record tide levels in many areas, producing flooding across the city. Strong winds caused loss of power and large parts of the city and surrounding areas remained without electricity for some days. Although there were certainly long lasting effects from the storm, most areas returned to normality within days.

Five city hospitals were forced to evacuate because of the storm: New York Downtown Hospital, Manhattan VA Medical Center, Coney Island Hospital, Bellevue Hospital, and NYU Langone Medical Center. Several hospitals agreed to take some of the affected hospitals' patients. New York Downtown Hospital evacuated its patients before the storm; it suffered minor damage and was able to fully reopen approximately one week after the storm. The other hospitals, on the other hand, suffered extensive damage and the necessary repair work kept the facilities closed for several weeks.

The Manhattan VA Medical Center was evacuated on October 27th, well before the storm hit. It suffered extensive damage to its electrical and mechanical systems as well as clinical equipment. It fully reopened in May 2013.

NYU Langone Medical Center began evacuating around 300 patients on the evening of October 29th after the backup generators at the hospital failed due to flooding. Some limited outpatient services reopened during the first days of November, but the hospital only partially resumed inpatient services on December 27th. The maternity unit and pediatrics reopened on January 14th, 2013. An urgent care center was opened on January 17th, but a true emergency room did not open until April 24th, 2014.

Bellevue Hospital began evacuating about 500 patients on October 31th, when the full extent of the damage became clear: its basement, which housed equipment critical to the hospital's operations, took in millions of gallons of water. The hospital had been operating on backup generators since losing power during the storm. It was the hospital's first evacuation in its 276-year history. It resumed limited outpatient services on November 19th, but it did not fully reopen inpatient services until February 7th, 2013.

Evacuation of about 180 patients at Coney Island Hospital began on the afternoon of October 30th. The hospital had been using generators since the hurricane caused power outages across Southern Brooklyn. An urgent care center was opened by December 3th but patients were not admitted inpatient. It reopened ambulance service and most of its inpatient beds by February 20th, 2013. The labor and delivery unit did not reopen until June 13th, 2013.

The unavailability of these hospitals meant increased patient volume for some nearby facilities. Anecdotal evidence suggests that the latter had difficulties accommodating extra patients and that capacity constraints were severe, with patients facing abnormally long waiting times for getting care in many cases. Some doctors, residents, and nurses at the affected facilities were able to work at other facilities while their hospital remained closed.

Appendix B. Nonparametric analysis

B.1. Patient types

I consider the case of three alternatives ($J = 3$) and no covariates to illustrate the basic framework. In this case, there are three potential choices in both periods 1 and 2. Therefore, the type of a patient is a vector of seven elements: her choice in period 0, her period 1 potential choices (3), and her period 2 potential choices (3). Each row of Table 15 corresponds to a patient type. For each type, I indicate the actual choices made by the patient and whether she exhibits state dependence in period t ($SD_t = 1$). The observed choices are determined by the potential choices that are boxed, but are independent of the potential choices that are not boxed. The presence of state dependence, on the other hand, depends on all the potential choices in a given period.

The first two patients make the same sequence of choices: A-B-C. However, the first patient exhibits state dependence in period 2, while the second patient does not. From the data, the researcher knows the proportion of patients who choose A-B-C. However, we cannot determine the specific types of the patients who make these choices. One possibility is that all these patients are like the first consumer in the example, in which case state dependence is 100%. The other possibility is that all these patients are like the second consumer in the example, in which case state dependence is 0%.

B.2. Optimization problem

Let $\mathbb{U}$ denote the set of possible types in the population. For each type $u \in \mathbb{U}$, let $sd(u) \in \{0, 1\}$ indicate whether a patient with type u exhibits state dependence in period 2. The objective is to recover the identified set for the proportion of patients who exhibit state dependence in period 2. Under certain conditions (which are met in my analysis), the identified set for this parameter is a closed interval. Therefore, I only need to recover the endpoints of the interval. The problem is then to find the admissible distributions under which the extent of state dependence is lowest and highest.

More precisely, I need to find the probabilities $\{p(u)\}_{u\in\mathbb{U}}$ that minimize (maximize) $\sum_{u\in\mathbb{U}} p(u)sd(u)$. Note that this is a linear function of the probabilities. The restrictions on the set of admissible distributions that I consider are also linear in the probabilities. Then, for example, the lower bound on the extent of state dependence is a solution to the following linear problem:

$$
\begin{aligned}
\underset{\{p(u)\}}{\text{minimize}} \quad & \sum_{u \in \mathbb{U}} p(u) sd(u) \\
\text{subject to} \quad & p(u) \geq 0, \forall u \in \mathbb{U} \\
& \sum_{u \in \mathbb{U}} p(u) = 1 \\
& \sum_{u: y(u) = y} p(u) = P(y), \forall y \in Y \\
& \sum_{u \in \mathbb{U}} w(u) p(u) \leq 0
\end{aligned}
\tag{14}
$$

The first two restrictions ensure that I have a proper p.m.f. The third restriction demands that I only consider probability distributions that are consistent with the data. Let Y be the set of all possible sequences of choices. For each $u \in \mathbb{U}$, let $y(u)$ be the sequence of choices made by a patient with type u. The restriction is that the probability assigned to all types that generate a given sequence of choices be equal to the proportion of patients in the data with these observables. The fourth restriction captures other linear constraints on probabilities (such as Assumption 2).

B.3. Ex-ante restrictions

As discussed in Section 5.4, even after limiting the size of the choice set and the number of periods considered, the dimension of the set of types is prohibitive from a computational point of view. To deal with this issue, I impose ex-ante restrictions on the set of admissible types. These restrictions are not only useful to reduce the dimensionality of the problem, but they also have economic content. Here, I discuss Assumption 1 in the context of a standard model of consumer choice.

I assume that choices are determined by utility maximization. The utility that patient i gets from choosing alternative j in period t is given by:

$$
u_{ijt} = g(X_{it}, Z_{jt}) + \sum_{k=1}^{J} \gamma_{kj} A_{ikt} + e_{ijt}
$$

where X_{it} and Z_{jt} are characteristics of the consumer and the alternative in period t, respectively, $y_{i,t-1}$ denotes the alternative chosen by the patient in period $t-1$, and $A_{ikt} = \mathbb{I}[y_{i,t-1} = k]$.

This formulation is very general as it allows for general compatibility patterns[31]: γ_{kj} is the utility premium the consumer gets from choosing alternative j if the alternative chosen

[31]However, it is restrictive along other dimensions. For example, it rules-out forward looking behavior.

in the previous occasion was k.

Consider the additional restriction that the consumer only receives a utility premium from choosing the same alternative as in the previous period. In other words, we have: 1) $k \neq j \Rightarrow \gamma_{kj} = 0$; 2) $\gamma_{jj} \geq 0$. Therefore, utility is given by:

$$u_{ijt} = g(X_{it}, Z_{jt}) + \gamma_{jj}\mathbb{I}[y_{i,t-1} = j] + e_{ijt}$$

Define the portion of utility that depends on current consumer and alternative characteristics as $u^*_{ijt} = g(X_{it}, Z_{jt}) + e_{ijt}$, and let $u^*_{it} = \text{argmax}_j \ u^*_{ijt}$. The assumptions above and transitivity imply that for any counterfactual period $t-1$ choice j, $h_t(j) \in \{j, u^*_t\}$.

Consider a consumer i with $y_{i,t-1} = j \neq k$. Then, we have:

$$u_{ijt} = u^*_{ijt} + \gamma_{jj} \geq u^*_{ijt} \text{ and } u_{ikt} = u^*_{ikt}$$

Then, there are two possibilities:

1. $u^*_t = j \Rightarrow u^*_{ijt} \geq u^*_{ikt} = u_{ikt} \Rightarrow u_{ijt} \geq u_{ikt}$
2. $u^*_t = q \neq j \Rightarrow u^*_{iqt} = u_{iqt} \geq u^*_{ikt} = u_{ikt} \ \forall k \neq j$

Then, a consumer who chooses alternative j in $t-1$ chooses either j or u^*_t in period t.

In my setting, the restrictions imply that, for $j \neq k \neq z$, we have:

1. $h_t(k) = j \Rightarrow h_t(j) = j$
2. $h_t(k) = j \Rightarrow h_t(z) \in \{z, j\}$

In the application, these restrictions are imposed ex-ante because of dimensionality issues, effectively restricting the set of possible types. However, whenever the number of alternatives and time periods is not restrictive we might want to treat them as regular identifying assumptions and evaluate their power independently of other restrictions.

Appendix C. Impact of state dependence on health outcomes

Here, I show how to analyze the impact of state dependence on health outcomes using the nonparametric approach discussed in Section 5.4.

The type of a patient specifies the choices that she would make in a given episode in different states of the world. In the original framework, these counterfactual scenarios are defined by the alternative chosen in the previous episode. However, we could extend the set of counterfactuals to also include the scenario where state dependence is absent.

Let the choice set be $\mathcal{J} = \{1, .., J\}$ and let m_j denote the expected mortality for a patient if she chooses hospital j. As before, for each $k \in \mathcal{J}$, $h_t(k)$ denotes the hospital that the patient would choose in episode t had her previous choice counterfactually been k. In addition, let $h_t(0)$ be the hospital that the patient would choose in episode t absent state dependence. If the patient chose hospital k in episode $T-1$, then the impact of state dependence on expected mortality in episode T is given by:

$$\Delta EM = m_{h_T(k)} - m_{h_T(0)} \tag{15}$$

In the actual state of the world, the consumer chooses alternative $h_T(k)$, so her expected mortality is $m_{h_T(k)}$. In the counterfactual scenario where there is no state dependence, she would choose alternative $h_T(0)$, so her expected mortality is $m_{h_T(0)}$.

Let $y_{h,t}$ be the choice made in episode t by a patient with type h. The aggregate effect of state dependence on health outcomes is given by:

$$\Delta EM = \sum_h \mathbb{P}(h) \left(m_{h_T(y_{h,T-1})} - m_{h_T(0)} \right) \tag{16}$$

Equivalently, we have:

$$\Delta EM = \sum_h \mathbb{P}(h) \left(\sum_k \sum_j \sum_v \mathbb{I}\left(y_{h,T-1} = k, h_T(k) = j, h_T(0) = v \right) (m_j - m_v) \right) \tag{17}$$

The fundamental issue is how to link the latent choice $h_T(0)$ to the data. The problem is that this latent choice is rarely observed. The most common exception is a consumer who is new to the market and therefore has not developed an attachment to any particular product. For consumers with previous experience in the market, on the other hand, choices are affected by state dependence. There are exceptions: for example, cases where inertia has been eliminated as a result of a specific policy. However, in general, we desire to conduct counterfactual analysis based on purely observational data.

Assumption 1 in Section 5.4 has identifying power. If there are consumers who are new to the market (so $T = 1$ for them), an identifying assumption would be:

$$P[h_0 = j|X] = P[h_T(0) = j|T > 1, X] \tag{18}$$

This assumption states that the distribution of first best choices is the same for new and experienced consumers, conditional on observables.

Table 15: Different types of patients

	Latent choices t=1			Latent choices t=2			Observed choices			State dependence	
h_0	$h_1(A)$	$h_1(B)$	$h_1(C)$	$h_2(A)$	$h_2(B)$	$h_2(C)$	y_0	y_1	y_2	SD_1	SD_2
$\boxed{\text{A}}$	$\boxed{\text{B}}$	B	C	A	$\boxed{\text{C}}$	C	A	B	C	1	1
$\boxed{\text{A}}$	$\boxed{\text{B}}$	B	C	C	$\boxed{\text{C}}$	C	A	B	C	1	0
$\boxed{\text{B}}$	C	$\boxed{\text{C}}$	C	C	C	$\boxed{\text{A}}$	B	C	A	0	1
$\boxed{\text{C}}$	B	C	$\boxed{\text{A}}$	$\boxed{\text{C}}$	C	B	C	A	C	1	1