DATA-DRIVEN INCENTIVE ALIGNMENT IN CAPITATION SCHEMES

Mark Braverman
Sylvain Chassang

Data-Driven Incentive Alignment in Capitation Schemes
Mark Braverman and Sylvain Chassang
NBER Working Paper No. 28429
February 2021
JEL No. C55,D82,H51,I11,I13

ABSTRACT

This paper explores whether Big Data, taking the form of extensive high dimensional records, can reduce the cost of adverse selection by private service providers in government-run capitation schemes, such as Medicare Advantage. We argue that using data to improve the ex ante precision of capitation regressions is unlikely to be helpful. Even if types become essentially observable, the high dimensionality of covariates makes it infeasible to precisely estimate the cost of serving a given type: Big Data makes types observable, but not necessarily interpretable. This gives an informed private operator scope to select types that are relatively cheap to serve. Instead, we argue that data can be used to align incentives by forming unbiased and non-manipulable ex post estimates of a private operator's gains from selection.

Mark Braverman
Princeton University
mbraverm@cs.princeton.edu

Sylvain Chassang
Department of Economics
Princeton University
Julis Romo Rabinowitz Building
Princeton, NJ 08544
and NBER
chassang@princeton.edu

Data-Driven Incentive Alignment in Capitation Schemes

Mark Braverman   Sylvain Chassang[*†]

Princeton University

October 19, 2020

Abstract

This paper explores whether Big Data, taking the form of extensive high dimensional records, can reduce the cost of adverse selection by private service providers in government-run capitation schemes, such as Medicare Advantage. We argue that using data to improve the ex ante precision of capitation regressions is unlikely to be helpful. Even if types become essentially observable, the high dimensionality of covariates makes it infeasible to precisely estimate the cost of serving a given type: *Big Data makes types observable, but not necessarily interpretable*. This gives an informed private operator scope to select types that are relatively cheap to serve. Instead, we argue that data can be used to align incentives by forming unbiased and non-manipulable ex post estimates of a private operator's gains from selection.

KEYWORDS: adverse selection, big data, capitation, health-care regulation, detail-free mechanism design, delegated model selection.

1   Introduction

This paper explores the value of Big Data in reducing the cost of adverse selection by private service providers in government-run capitation or voucher schemes. We emphasize

an application to healthcare insurance. School voucher programs are an other example.

In the context of public health-insurance, a capitation scheme pays private insurance plans an estimate of the cost of insurance for individuals they enroll. An example is Medicare Advantage, a program which lets US Medicare recipients switch to private health insurance plans. Capitation payments can be conditioned on agreed upon user characteristics (then, they are said to be risk-adjusted). While capitation programs are a popular way to outsource government mandated services to the private sector, they are often plagued by adverse selection. Private plans have strong incentives to select types that are cheaper to serve than their capitation payment, which increases the cost of serving the overall population. In the context of Medicare Advantage, Batata (2004) and Brown et al. (2014) report yearly overpayments in the thousands dollars for patients selected by private plans.

A natural strategy to reduce adverse selection is to increase the precision of risk-adjusted capitation schemes by reimbursing private plans for the expected cost of taking care of the specific patients they select. This suggests that Big Data — i.e., the availability of high-dimensional patient records — could be used to condition capitation payments on precise individual characteristics, and eliminate the scope for adverse selection. We take a different view and argue that in realistic Big Data environments, this naïve use of high-dimensional co-variates is likely to be of limited value. Instead, we suggest that data may be more successfully used to form unbiased ex post estimates of strategic selection by private plans. Correcting capitation formulas with these ex post estimates aligns the public and private plans' incentives.

Our model considers a single public plan p_0 seeking to outsource the provision of healthcare services to a single private plan p_1.[1] The private plan may have a genuine comparative advantage in insuring certain types of patients so that some selection into plans may be welfare enhancing. For instance, plan p_1 may have expertise in tobacco cessation, inducing a comparative advantage insuring current smokers. However, the private plan also has in-

---

[1]In the case of Medicare Advantage, the private plan would correspond to a preferred provider organization (PPO) or a health management organization (HMO).

centives to select patients whose cost of care is overestimated; for instance, patients that are more healthy or less likely to use costly services than their official risk-rating suggests. This leads us to distinguish legitimate selection characteristics, which predict comparative advantage, from illegitimate selection characteristics, which predict costs but not comparative advantage. Efficient selection strategies need only depend on legitimate selection characteristics. We assume that comparative advantage can be captured using a small set of legitimate characteristics, while the underlying set of types predicting cost of care is large. In addition, we assume that public plan p_0 has access to a hold-out sample of cost realizations that is not observed by private plan p_1, such as cost realizations for the patients it serves.

Our modeling choices reflect both the opportunities and limitations presented by Big Data. We assume that high-dimensional records isomorphic to patients' types — i.e. sufficient statistics for patients' cost of care — are observable. However, we also recognize that the number of such possible types need not be small relative to the sample size of available cost data, thereby limiting their use for prediction.[2] This leads us to study mechanism design for environments where both the sample size and the number of relevant covariates are large. In such Big Data environments, sufficient statistics of types are *observable but not interpretable*: even conditional on type and cost data, it is not possible to form a precise estimate of a patient's cost of care. This creates a trade-off when setting capitation rates: "sparse" cost estimates, conditioned on a few patient characteristics, have low standard errors but high bias; in contrast "rich" cost estimates, conditioned on an exhaustive set of patient characteristics, have low bias, but large standard errors.

Concerns over estimation error are reflected in the capitation scheme employed by Medicare Advantage, as well as in the risk-adjustment formula used to calculate transfers between plans under the Affordable Care Act (ACA). The Medicare Advantage risk adjustment model, rolled out in 2004 by the Centers for Medicare and Medicaid Services (CMS), uses Hierarchical Condition Categories (HCCs) (Pope et al., 2004). The HCCs are groups

---

[2]The statistics literature concerned with Big Data makes similar assumptions. See Belloni et al. (2013, 2014) for recent examples in econometrics.

of conditions that can be inferred from the patient diagnosis data. The number of HCCs in the model varies between editions, but generally is under 100. HCCs are used in conjunction with condition severity modifiers and demographic factors to estimate individual patients' expected expenditures in the subsequent year. Thus the model falls under the "sparse capitation" category discussed above: there are relatively few categories, and a reasonably precise estimator can be formed for each category (Evans et al., 2011). The desire for "adequate sample sizes to permit accurate and stable estimates of expenditures" has been a design principle for the risk adjustment scheme, and a factor in keeping the number of patient types in the model relatively low (Pope et al., 2004).

The model used for risk-adjustment transfers under the ACA uses a somewhat expanded set of HCCs (114 in 2014), reflecting the fact that the ACA transfer model is a general-population model, while the Medicare Advantage model is primarily for patients above the age of 65 (Kautter et al., 2014). As in the case of the Medicare Advantage model, the need for statistical power to get good ex ante estimates is one of the design principles limiting the number of categories used (Centers for Medicare and Medicaid Services, 2016). We discuss implications of our analysis to the ACA setting, where no public option is available, in Section 5.

Our first set of results considers traditional capitation schemes, which, as emphasized by Brown et al. (2014), seek to reimburse private plans for the expected cost of treating patients given ex ante observables. Sparse capitation schemes condition cost estimates on a small set of patient characteristics, while rich capitation schemes condition cost estimates on the full set of characteristics made available by Big Data. We show that such schemes induce efficient selection when the private plan is constrained to select only on the basis of legitimate characteristics (determinants of comparative advantage). However, if private plans can engage in illegitimate selection, then patient assignment is bounded away from efficiency in Big Data environments. Indeed, even though types are observable, cost-estimates conditional on types remain noisy even for large samples. As a result it is possible for the

private plan to maintain an informational advantage which induces inefficient selection and increases the average cost of care.

In spite of these limitations, if legitimate selection characteristics are common knowledge, we can construct a prior-free mechanism that achieves efficient patient assignment at no excess cost for the public plan. Big Data is not used to obtain a more precise ex ante capitation formula. Instead, we augment the baseline capitation formula (based on legitimate characteristics) with a single additional term measuring ex post selection by the private plan. This additional term takes the form of an appropriately weighted covariance between the distribution of types selected by the private plan, and the residuals from the sparse capitation regression evaluated on out-of-sample costs. More concretely, it uses Big Data to form an unbiased estimate of the cost savings obtained by the private plan from selecting a non-representative sample of patients. This "strategic capitation scheme" induces efficient selection, and, importantly, does not give the public plan any incentive to bias its report of out-of-sample costs. This last property allows us to extend our approach to health exchanges for which out-of-sample cost realizations would be reported by competing healthcare plans (see Section 5).

Strategic capitation can be extended to environments where legitimate selection characteristics are not common knowledge. It is still possible to achieve a meaningful share of first-best efficiency by using generalized strategic capitation schemes that let private plans specify the characteristics they wish to select on. This flexibility comes at a cost related to the complexity of the class of models the private plan can use to select patients. We show that the performance guarantees of this indirect prior-free mechanism are essentially unimprovable by studying the unrestricted direct mechanism design problem in a specific environment.

The paper contributes to the theoretical literature on adverse selection in insurance markets.[3] Our work is particularly related to Glazer and McGuire (2000), who study optimal

---

[3]See for instance Rothschild and Stiglitz (1976), Bisin and Gottardi (1999, 2006), Dubey and Geanakoplos (2002).

risk-adjustment in a Bayesian setting. They show that when selection is possible, optimal ex ante reimbursement schemes should deviate from simply reimbursing private plans the expected cost of taking care of patients. In particular, capitation schemes should adjust reimbursement rates to dull the effect of "cream-skimming" by private plans. We show how to induce efficient selection by using information about patient types and ex post cost data.

Our mechanism is closely related to that of Mezzetti (2004), which also uses noisy ex post information to provide accurate ex ante incentives. Also related is the work of Riordan and Sappington (1988) who show how to exploit noisy ex post signals to screen agents at no cost to the principal. As we clarify in greater detail in Section 6, our work differs for two main reasons. First, we are interested in prior-free mechanisms and do not make the identification assumptions required in Riordan and Sappington (1988). Second, ex post signals (here the public plan's hold-out cost data) need not be publicly observed and we must ensure that the relevant party has correct incentives for reporting. Third, unlike Mezzetti (2004), we require exact budget-balance.

Our work is motivated by a growing empirical literature which documents cream-skimming in health insurance markets, and studies the efficiency of various risk-adjustment schemes (Frank et al., 2000, Mello et al., 2003, Batata, 2004, Newhouse et al., 2012, Brown et al., 2014). Our analysis is inspired by Brown et al. (2014) which shows that increasing the number of covariates used in Medicare Advantage's capitation formulas has in fact led to an increase in the cost of adverse selection to the state.[4] We complement their result by showing that naïve uses of data are unlikely to resolve adverse selection, but progress can be made by using data to detect selection ex post.

The paper is structured as follows. Section 2 describes our framework, and in particular our approach to Big Data. Section 3 uses a simple example in which legitimate selection characteristics are common knowledge to delineate the mechanics of adverse selection under various capitation schemes. Section 4 generalizes the analysis to settings in which the private

---

[4]Newhouse et al. (2012) argues that the cost of adverse selection may be overstated.

plan's comparative advantage is not common knowledge. Section 5 uses a stylized model to show how strategic capitation can help reduce adverse selection in healthcare exchanges. Section 6 and Appendix A present several extensions addressing potential risk inflation, dynamic selection, and reduced quality provision by private plans. Proofs are collected in Appendix B unless mentioned otherwise.

2 Framework

Our model seeks to capture three features of healthcare capitation. The first is selection by private healthcare plans, such as HMOs or PPOs, which we model as a reduced form cost for attracting different populations. Selection may be achieved through targeted advertisement and marketing (consistent with Starc (2014)), heterogenity in the quality of customer service during enrollment procedures, as well as targeted service bundles.

Second, public and private plans may have heterogeneous comparative advantages in treating patients. While there is controversy about the real value added of private plans over Medicare, there is evidence that insurance plans are more than mere financial intermediaries. Plans play an important role in selecting, monitoring and generally resolving agency problems vis à vis doctors and hospitals, as well as encouraging preventive care and healthy habit formation. Data from Bundorf et al. (2012) provides evidence for such comparative advantage across different plans. In their sample, HMOs have a comparative advantage over PPOs in insuring high risk patients. In our model, the possibility of comparative advantage creates a reason for both public and private plans to be active, and raises the question of efficient patient allocation.

Third, we seek to correctly capture the forces that make Big Data attractive but challenging: we assume that high dimensional records make patients' types observable, but that as a result, even with a large sample of patients, it is not possible to form precise estimates of expected cost of treatment conditional on individual types (the concern for power is explicit in discussions of CMS formulas by Pope et al. (2004), Evans et al. (2011), Kautter et al.

(2014)). In short, types are observable but not interpretable.

The lead example for our work is Medicare Advantage, a program that lets US Medicare recipients switch to private insurance plans such as HMOs and PPOs. Medicare Advantage is a large and growing program. It covers approximately 15 million patients, out of the roughly 50 million enrolled in Medicare, and its size was multiplied by three from 2005 to 2015. Selection by private plans is an ongoing concern threatening the financial sustainability of the program (Batata, 2004, Brown et al., 2014).

2.1   Players, Actions, Payoffs

We study the relationship between a public health care plan p_0, responsible for the health expenses of a set $I = \{1, \cdots, N\}$ of patients, and an independent private plan p_1.

Treatment costs.   Each patient $i \in I$ has a type $\tau_i \in T \subset \mathbb{R}^n$ where the set of types T is potentially very large, but finite. Type τ is a sufficient statistic for a patient's cost of care, and is observable to both plans. For any sample J of patients, we denote by $\mu_J \in \Delta(T)$ the empirical distribution of types τ in sample J, defined by $\mu_J(\tau) \equiv \frac{|J^\tau|}{|J|}$, where $J^\tau \equiv \{j \in J | \tau_j = \tau\}$, and $|J|$ denotes the cardinal of J.

The realized cost of care for a patient i of type τ insured by plan p is positive, bounded above, and denoted by $\widehat{c}_i(p) \in [0, c_{\max}]$. The distribution of potential costs $(\widehat{c}_i(p_0), \widehat{c}_i(p_1))$ conditional on type τ is distributed according to a c.d.f. $F(\tau)$. Treatment costs across patients i are exchangeable conditional on patient type τ, cost distribution F, and plan p. We allow for aggregate uncertainty so that c.d.f. F is itself uncertain.

Selection.   The key difficulty is that private plan p_1 may strategically seek to attract a targeted subset of patients. Specifically, private plan p_1 can choose an expected selection policy $\lambda : T \to [0, 1]$ at a cost $K(\lambda) \geq 0$. A patient i enrolls with private plan p_1 with probability $\lambda(\tau_i)$. Consistent with observations in Starc (2014), this reduced-form cost of

selection may be thought of as a cost of advertisement.[5] Realized selection $\Lambda \subset I$ is a mean preserving spread of intended selection λ defined by

$$\mathbf{1}_{i\in\Lambda} = \lambda(\tau_i) + \varphi_i.$$

Error terms $(\varphi_i)_{i\in I}$ have expectation equal to zero, and may be correlated across different types $\tau \in T$, but are otherwise independent of cost realization $\widehat{c}_i(p)$. For instance, a recruitment ad may unexpectedly attract a population different from the targeted one. Patients $i \in I \setminus \Lambda$ that do not enroll with private plan p_1 are insured by public plan p_0.

Realized payoffs and preferences. Transfers $\Pi \in \mathbb{R}$ from public plan p_0 to private plan p_1 are feasible. Given a selection decision λ by private plan p_1, a realized selection Λ, and a transfer Π, the realized surpluses U_0 and U_1 accruing to the public and private plans are

$$U_0 = -\Pi + \sum_{i\in\Lambda} \widehat{c}_i(p_0) \quad \text{and} \quad U_1 = \Pi - \sum_{i\in\Lambda} \widehat{c}_i(p_1) - K(\lambda).$$

Conditional on selection rule λ and a distribution of costs F, surplus takes the form

$$S(\lambda) = -K(\lambda) + \mathbb{E}_F \left[\sum_{i\in I} \lambda(\tau_i) \left(\widehat{c}_i(p_0) - \widehat{c}_i(p_1)\right) \right].$$

Public plan p_0 has lexicographic preferences: its main objective is to maximize surplus S in expectation; however, taking as given selection behavior λ and surplus S, its secondary objective is to maximize cost savings U_0 in expectation. Private plan p_1 seeks to maximize its profits U_1 in expectation.

Note that although we let the public plan care about surplus, we will seek mechanisms such that the expectation of cost savings U_0 is approximately non-negative in equilibrium.

---

[5]Cost K implicitly includes any information acquisition costs needed to implement λ. Under a more standard model of selection along the lines of Rothschild and Stiglitz (1976), the private plan would screen patients through a menu of discounts and benefits specifically appealing to desirable types. In a dynamic setting, plans may be able to select patients by offering better customer service to types they wish to retain.

This can be thought of as a financing constraint.

2.2 Data

We model explicitly the role that data plays in the contracting problem. We assume that in Big Data environments, the richness of covariates makes patients' types essentially observable, but prevents the public plan from forming a precise estimate of expected treatment costs conditional on types. A consequence illustrated in Section 3 is that even imprecise additional signals of costs may allow the private plan to profit from selecting patients.

Samples. Both plans p_0 and p_1 observe a public dataset of types and cost realizations $D_0 = \{(i, \tau_i, \widehat{c}_i(p_0)) | i \in D_0\}$ for plan p_0, where $i \in D_0$ denotes a patient i whose record is included in D_0. In addition, we denote by $D_0^\tau = \{(i, \tau_i, \widehat{c}_i(p_0)) | \tau_i = \tau, i \in D_0\}$ the cost data relating to patients of type τ. We assume that for every $\tau \in T$, the set D_0^τ is non-empty, which implies $|T| \leq |D_0|$: the sample size of dataset D_0 is at least as large as the type space.

Plan p_1 privately observes a dataset $D_1 = \{(i, x_i, \widehat{c}_i(p_1)) | i \in D_1\}$ reporting both her own costs, and side-signals x_i for a sample of patients $i \in D_1$. Side signal x_i (which can include type τ_i) captures other signals beyond cost realizations that the plan may be able to use in order to select patients.

Finally, we assume that plan p_0 has access to a hold-out sample $H = \{(i, \tau_i, \widehat{c}_i(p_0)) | i \in H\}$ of its own costs, independent of data D_1 conditional on the realization of cost distribution F. Hold-out sample H may consist of ex post cost realizations for the current set of patients enrolled by the public plan. Alternatively, H may correspond to past cost data, securely encrypted, and verifiably released only after patient selection has occurred.[6] Contractual transfers Π will be allowed to depend on hold-out sample H, but we will take seriously the public plan's incentive to reveal correct information. Specifically, we will address the public plan's incentives to bias its records in order to reduce payments to the private plan.

---

[6]For instance, an encrypted version of the data can be released before selection occurs, with a decryption key publicized after patient enrollment has occurred.

For instance the public plan could down-code interventions happening to its own patients. Access to a hold-out sample is essential. It allows the public plan to obtain estimates of its own costs whose errors are uncorrelated to the private plan's information. We motivate the asymmetric treatment of data-sets D_0 (publicly observed) and H (privately observed by the public plan) when we discuss the timing of actions.

Big Data. Recall that $\mu_I(\tau) \equiv \frac{|I^\tau|}{|I|}$ denotes the sample distribution of types τ in patient population I. Our model of Big Data consists of two assumptions:

(i) types $\tau \in T$ are publicly observed;

(ii) even if sample data D_0 is large, type space T is of comparable size; i.e. there exists a fixed constant $a > 0$ such that

$$\mathbb{E}_{\mu_I}\left[\frac{1}{\sqrt{|D_0^\tau|}}\right] > a.$$

Points (i) and (ii) summarize what we think are the opportunities and limitations of Big Data. Point (i) captures the idea that high dimensional records make types observable. Point (ii) implies that even if sample D_0 is very large, the size $|D_0^\tau|$ of many subgroups D_0^τ remains bounded above. As a result the public plan's estimates of type-specific costs on the basis of data D_0 remains noisy.

Note that all our results are non-asymptotic: we provide efficiency bounds that depend explicitly on $\mathbb{E}_{\mu_I}\left[\frac{1}{\sqrt{|D_0^\tau|}}\right]$. Point (ii) clarifies the terms we consider negligible and non-negligible when we provide performance bounds. Our large type-space model of Big Data contrasts with small type-space environments in which the sample size maybe considered large compared to the set of types.

2.3 Beliefs, Contracts, and Equilibrium

Beliefs. The plans' environment is described by: healthcare-cost distribution F; data-sets D_0, D_1 and H; cost of selection K. All these objects are in principle random variables.

For simplicity, we assume that plans p_0 and p_1 share a common prior ν over the tuple (F, D_0, D_1, H, K).

Note that the capitation mechanisms we study do not rely on the common prior assumption. Our performance bounds remain valid in a non-common prior setting if expectations are taken under the private plan's prior.[7]

Reporting of hold-out data. We allow mechanisms to depend on reports of hold out sample H made by public plan p_0. We denote by $\beta : H \mapsto H_R$ the public plan's strategy, mapping hold-out data H to (potentially biased) reported hold-out data H_R. We note that β does not depend on realized selection Λ. This can be thought of as a restriction on timing (the hold-out sample must be encrypted and shared before selection occurs), or a restriction on strategies (the public plan cannot tailor misreporting to simultaneous selection by the private plan). We denote by $\beta^*(H) \equiv H$ the truthful reporting strategy, and break indifferences in favor of truth-telling, reflecting small costs in misreporting.

Contracts. For any set of patients $J \subset I$, let $\tau_J \equiv (\tau_i)_{i \in J}$ and $\widehat{c}_J(p) \equiv (\widehat{c}_i(p))_{i \in J}$ denote profiles of types and costs. We denote by $H_R = \{(i, \tau_i, \widehat{c}_{\tau_i}^R(p_0)), i \in H\}$ the hold-out data reported ex post by public plan p_0. We emphasize that these are reports of privately observed costs, and that the public plan must be given appropriate incentives in order to report truthfully. A capitation contract between the public and private plan is a mapping $\Pi(\Lambda, H_R) \in \mathbb{R}$, specifying the aggregate transfer to private plan p_1 as a function of realized selection Λ, and reported hold-out sample data H_R. Note that contract Π implicitly depends on data D_0, but we suppress this dependency in the notation since D_0 is common-knowledge between plans.

---

[7]For recent work emphasizing prior-free approaches to mechanism design, see Segal (2003), Bergemann and Schlag (2008), Hartline and Roughgarden (2008), Chassang (2013), Carroll (2015), Madarász and Prat (2014), Brooks (2014), Antic (2014).

Timing and equilibrium. Given public data D_0 and a common knowledge contract Π, the timing of the interaction between public and private plans is as follows:

1. the public and private plans simultaneously pick a cost reporting strategy β, and a selection strategy λ;

2. potentially biased hold-out data H_R is reported, and actual selection $\Lambda \subset I$ by the private plan is realized, resulting in a contractual transfer $\Pi(\Lambda, H_R)$.

We assume away potential incentive issues related to the revelation of data D_0 by public plan p_0 for simplicity. It is sufficient for our analysis that the contract Π be common knowledge in stage 1. We only care about data D_0 to the extent that it lets us specify contract Π.

Note that since the public plan's primary objective is to maximize surplus, there would be little incentive to misreport D_0: biased costs would distort the selection of patients λ taking place at a later stage, and reduce surplus S. In contrast, sending a biased report H_R of hold-out data H cannot influence the selection of patients λ, which takes place simultaneously. As a result, biasing hold-out data H does not affect surplus S; it only affects the allocation of surplus across the public plan's cost savings U_0, and the private plan's profits U_1. This is why we take seriously strategic issues related to the reporting of hold-out data H, but simplify away strategic considerations pertaining to the release of initial data D_0.

Given a capitation contract Π, a selection strategy λ, and a reporting strategy β, the public and private plans' expected payoffs under common prior ν are

$$
\mathbb{E}_\nu U_0 = \mathbb{E}_\nu \left[-\Pi + \sum_{i \in \Lambda} \widehat{c}_i(p_0) \Big| \lambda, \beta \right],
$$
$$
\mathbb{E}_\nu U_1 = \mathbb{E}_\nu \left[\Pi - \sum_{i \in \Lambda} \widehat{c}_i(p_1) \Big| \lambda, \beta \right] - K(\lambda).
$$

Given a contract Π, abstractly denoting by $\mathcal{I}_0$ and $\mathcal{I}_1$ the information available to plans p_0

and p_1, a strategy profile (β, λ) is an equilibrium if and only if β and λ respectively solve

$$\max_{\beta} \mathbb{E}_\nu[-\Pi | \mathcal{I}_0, \beta, \lambda] \quad \text{and} \quad \max_{\lambda} \mathbb{E}_\nu \left[\Pi - \sum_{i \in \Lambda} \widehat{c}_i(p_1) \Big| \mathcal{I}_1, \beta, \lambda \right] - K(\lambda).$$

In words, taking selection λ as given, the public plan's reporting strategy β must minimize expected transfers to the private plan. Taking reporting β as given, the private plan picks the selection strategy that maximizes its expected profit.

Design objectives. We seek contracts Π such that for all priors ν, data D_0, D_1, and all equilibria (λ, β), surplus S and public and private plan payoffs U_0, U_1 satisfy:

$$\mathbb{E}_\nu[S|\lambda] = \mathbb{E}_\nu \left[\max_\lambda \mathbb{E}_\nu[S|\lambda, D_0, D_1] \right] - o(|I|); \tag{1}$$

$$\mathbb{E}_\nu \left[U_0 \Big| \lambda, \beta, D_0 \right] \geq -o(|I|); \tag{2}$$

$$\mathbb{E}_\nu \left[U_1 \Big| \lambda, \beta, D_1 \right] \geq 0, \tag{3}$$

where $o(|I|)$ is the usual little-o notation for "negligible relative to $|I|$."

In words, we seek ex post budget-balanced prior-free mechanisms that: (i) maximize efficiency given available information up to a term negligible compared to the size $|I|$ of the patient population; (ii) satisfy approximate interim individual rationality for the public plan (reflecting some willingness to subsidize the market); (iii) satisfy exact interim individual rationality for the private plan.

3 Strategic Capitation

In this section we clarify why existing ex ante capitation schemes may work under idealized conditions, but likely fail in a realistic Big Data setting. Still, we are able to exhibit a prior-free mechanism that achieves first-best allocation by using an ex post measure of selection to correct capitation payments. Since this indirect prior-free mechanism achieves the first-best,

there are no performance benefits from using a prior-dependent Bayesian mechanism.

To make the key forces more transparent we make a few simplifying assumptions that we relax in Section 4.

Legitimate and illegitimate selection. We assume for now that there exists a common knowledge partition E of type space T, with typical element $\eta \in E$, such that treatment costs can be decomposed as

$$\widehat{c}_i(p) = \kappa(\eta_i, p) + e_{i,\tau_i} \tag{4}$$

where error terms e_{i,τ_i} do not depend on provider p, and have mean zero conditional on η_i in the patient population:

$$\mathbb{E}_{\nu,\mu_I}[e_{i,\tau_i}|\eta_i, D_0] = 0. \tag{5}$$

In the example of a private plan with expertise running smoking cessation programs, E would partition patients into smokers, and non-smokers; η would be a patient's smoking status.

Cost decomposition (4) implies that the comparative advantages of plans p_0 and p_1, described by $\kappa(\eta, \cdot)$, depend only on characteristics $\eta \in E$. We think of E as a small set compared to type space T, so that it is possible for each plan to form accurate estimates of its costs conditional on $\eta \in E$. For simplicity, we assume that the costs of the public plan $\kappa(\cdot, p_0)$ are known by both plans, and that private plan p_1 knows its own costs $\kappa(\cdot, p_1)$. Error term $e_{i,\tau}$ captures residual errors in cost estimates.

We assume (for this section) that the private plan can engage in arbitrary costless selection: $\forall \lambda \in [0,1]^T$, $K(\lambda) = 0$. Let $\mathcal{M}(E)$ denote the set of selection rules measurable with respect to E.

Remark 1. *First-best surplus, $S_{\max} \equiv \max_\lambda \mathbb{E}_\nu \left[\sum_{i\in\Lambda} \widehat{c}_i(p_0) - \widehat{c}_i(p_1) \right]$ is attained by a selection policy $\lambda^*(\tau) = \mathbf{1}_{\kappa(\eta,p_0)>\kappa(\eta,p_1)}$ that is measurable with respect to partition E.*

Accordingly, a selection rule is said to be *legitimate* if and only if it is measurable with respect to E (i.e. $\lambda \in \mathcal{M}(E)$). Selection rules that are not measurable with respect to

type-space partition E depend on features of types τ that do not matter for efficiency. They are referred to as *illegitimate*.

A stylized example. The following very stylized example makes for simple computations in the proofs that follow. We denote by ν_0 the corresponding prior. We assume that $|I| = |T| = |D_0| = |D_1|$. Types $\tau \in T$ take the form $\tau = (\eta, n, \delta)$, with η in some finite set E, $n \in \{1, \cdots, \overline{n}\}$ an integer, and $\delta \in \{-1, 1\}$. Sample I includes one patient per type $\tau \in T$. Data D_0 (resp. D_1) consists of a cost realization $\widehat{c}_i(p_0)$ (resp. $\widehat{c}_i(p_1)$). Thus $|I| = 2 \times \overline{n} \times |E|$.

Comparative advantage is determined by $\kappa(\eta, p) = \kappa_0 + \Delta \times \varepsilon_\eta \mathbf{1}_{p=p_1}$, where $\Delta \in [0, \kappa_0/4]$ parameterizes the magnitude of comparative advantages, and $\varepsilon_\eta \sim U\{-1, 1\}$ (uniformly distributed over $\{-1, 1\}$) indicates which plan has a comparative advantage. Error term $e_{i,\tau}$ is distributed $e_{i,\tau} \sim U\{0, \delta\varepsilon_{\eta,n}\kappa_0\}$, with $\varepsilon_{\eta,n} \sim U\{-1, 1\}$. Every n is associated with two types (corresponding to δ equal to 1 or -1), one whose idiosyncratic cost shocks has a positive average, and one with a negative average. Random variables ε_η and $\varepsilon_{\eta,n}$ are drawn independently. Shocks $e_{i,\tau}$ are drawn independently across individuals i conditional on $\varepsilon_{\eta,n}$.

It is immediate that $\mathbb{E}_{\nu, \tau_i \sim \mu_I}[e_{i,\tau_i} | D_0, \eta] = 0$. Cost data for types $(\eta, n, 1)$ and $(\eta, n, -1)$ is either fully revealing of $\varepsilon_{\eta,n}$ or not at all, leading to straightforward conditional expectations.

3.1 Why Ex Ante Capitation Schemes Fail

Existing capitation mechanisms attempt to align incentives through capitation rates that are fixed ex ante. This removes concerns that the public plan may misreport its hold-out cost data H to reduce payments. We clarify when such schemes are potentially effective: when the private plan p_1 is unable to engage in illegitimate selection, or when type-specific cost estimates are arbitrarily precise. We then show that in Big Data environments, if plan p_1 is able to engage in illegitimate selection, such schemes are either bounded away from efficiency, or generate large losses for the public plan.

We consider sparse and rich capitation contracts that differ in the sophistication of the

information used to predict treatment costs. For simplicity, and since we mostly establish negative results, we allow these ex ante capitation schemes to depend on the common prior ν.[8] Transfers take one of the following forms:

$$\Pi^{\mathsf{sparse}}(\Lambda) = \sum_{i \in \Lambda} \mathbb{E}_{\nu,\mu_I}[\widehat{c}_i(p_0)|\eta_i, D_0] = \sum_{i \in \Lambda} \kappa(\eta_i, p_0) \tag{6}$$

$$\Pi^{\mathsf{rich}}(\Lambda) = \sum_{i \in \Lambda} \mathbb{E}_{\nu}[\widehat{c}_i(p_0)|\tau_i, D_0] = \sum_{i \in \Lambda} \kappa(\eta_i, p_0) + \mathbb{E}_{\nu}[e_{i,\tau_i}|\tau_i, D_0]. \tag{7}$$

In both schemes the private plan is paid the public plan's expected cost of treating selected patients, conditional on some set of ex ante observables. We denote by S^{sparse} and S^{rich} the associated surpluses in equilibrium.

Sparse capitation, like the existing CMS formulas, estimates patients' costs conditional on legitimate characteristics η alone. Rich capitation estimates patients' costs conditional on the full set of observables τ — i.e. it exploits Big Data to form targeted estimates.

Under sparse capitation, private plan p_1's expected payoffs for engaging in selection λ can be decomposed as the sum of social surplus, and a rent equal to the covariance between selection profile λ and the bias in cost estimates conditional on type τ:

$$\underbrace{\sum_{i \in I} \lambda(\tau_i) \left[\kappa(\eta_i, p_0) - \kappa(\eta_i, p_1)\right]}_{\text{social surplus}} + \underbrace{\sum_{i \in I} \lambda(\tau_i)\mathbb{E}_{\nu}[e_{i,\tau_i}|D_0, D_1, \tau_i]}_{\text{rents from type selection}}.^{[9]} \tag{8}$$

Proposition 1 (sparse capitation)**.** *Consider capitation scheme Π^{sparse}.*

(i) *If the private plan is constrained to use legitimate selection ($\lambda \in \mathcal{M}(E)$), then efficient selection and truthful reporting (λ^*, β^*) is the unique equilibrium.*

(ii) *Under prior ν_0, if the private plan is not constrained to use legitimate selec-*

---

tion rules, then $\mathbb{E}_{\nu_0}\left[S_{\max} - S^{sparse}\right] = \frac{3}{8}\Delta|I|$.

If $\Delta = 0$ *so that there is no comparative advantage, the public plan makes expected losses* $\mathbb{E}_{\nu_0}[U_0] = -\frac{3}{16}\kappa_0|I|$.

Since $\mathbb{E}_{\nu,\mu_I}[e_{i,\tau_i}|D_0, D_1, \eta_i] = 0$, rents from type selection in (8) are equal to zero whenever selection strategy λ is measurable with respect to legitimate selection characteristics $\eta \in E$. Hence, under the constraint that $\lambda \in \mathcal{M}(E)$, sparse capitation leads to efficient selection.

When the private plan can select on the basis of illegitimate characteristics τ, bias $\mathbb{E}_{\nu}[e_{i,\tau_i}|D_0, D_1, \tau_i]$ in cost estimates can be exploited. The private plan will seek to avoid under-reimbursed patients and recruit over-reimbursed patients, thereby deviating from efficient selection. In addition, rents from such selection come at the expense of the public provider, potentially resulting in large losses to p_0 if the gains from trade are small. The inefficiency loss due to selection does not vanish even for large samples D_0.

In order to correct bias $\mathbb{E}_{\nu}[e_{i,\tau_i}|D_0, D_1, \tau_i]$ in cost estimates, rich capitation Π^{rich} conditions capitation rates on the full set of observables τ. Unfortunately, this does not solve the problem in Big Data settings since conditional cost estimates remain noisy, allowing the private plan to exploit estimation error, instead of estimation bias.

Proposition 2 (rich capitation). *(i) Under any prior* ν, *efficiency loss* $S_{\max} - S^{rich}$ *satisfies* $\mathbb{E}_{\nu}\left[S_{\max} - S^{rich}\right] \leq 2c_{\max}|I|\mathbb{E}_{\mu_I}\left[\frac{1}{\sqrt{|D_0^{\tau}|}}\right]$

(ii) Under prior ν_0, $\mathbb{E}_{\nu}\left[S_{\max} - S^{rich}\right] = \frac{3}{32}\Delta|I|$.

If $\Delta = 0$, *the public plan makes expected losses* $\mathbb{E}_{\nu}[U_0] = -\frac{3}{64}\kappa_0|I|$.

While sparse capitation schemes do not achieve efficiency, regardless of data D_0, rich capitation schemes may achieve near efficiency provided that $\mathbb{E}_{\mu_I}\left[\frac{1}{\sqrt{|D_0^{\tau}|}}\right]$ is small, i.e. for almost every type τ, subsample D_0^{τ} is large. This is ruled out in Big Data environments. Cost estimates $\mathbb{E}_{\nu}[e_{\tau,i}|D_0, \tau_i]$ remain imprecise for a non-vanishing mass of types τ (under the sample distribution of types μ_I). As a result, the private plan's additional data D_1 has a non-vanishing impact on cost estimate $\mathbb{E}_{\nu}[e_{\tau,i}|D_0, D_1, \tau_i]$. This result would arise even if the

private plan did not observe data D_0 but only observed the capitation rates $\mathbb{E}_\nu[\widehat{c}_i(p_0)|\tau_i, D_0]$ offered by the public plan for enrolling patients with type τ_i.

3.2 Aligning Plan Incentives via Strategic Capitation

We now describe a capitation scheme that eliminates incentives for strategic selection by p_1 and strategic reporting by p_0. Payments take the form

$$\Pi^{\mathsf{strat}}(\Lambda, H_R) \equiv \sum_{i \in \Lambda} \pi(\eta_i) + \Delta\pi(\eta_i, H_R, \Lambda), \tag{9}$$

where $\pi(\eta) \equiv \kappa(\eta, p_0)$ is the baseline capitation rate conditional on legitimate characteristics used in sparse capitation. Term $\Delta\pi(\eta, H_R, \Lambda)$ is a correction dependent on reported hold-out data H_R and selected sample Λ. It correlates a measure of the private plan's deviation from legitimate selection with hold-out estimates of cost-prediction bias. It is equal to zero in expectation when the plan does not engage in illegitimate selection. Formally, it takes the form:

$$\Delta\pi(\eta_i, H_R, \Lambda) \equiv \mathrm{cov}_I(s_{\tau_i}, r_{\tau_i}|\eta_i = \eta) = \frac{1}{|I^\eta|} \sum_{i \in I^\eta} s_{\tau_i} r_{\tau_i}, \quad \text{where}$$

- $s_{\tau_i} \equiv \frac{\mu_\Lambda(\tau_i|\eta_i)}{\mu_I(\tau_i|\eta_i)} - 1$ measures selected sample Λ's deviation from legitimate selection;[10]

- $r_{\tau_i} \equiv \frac{1}{|H_R^{\tau_i}|} \sum_{j \in H_R^{\tau_i}} \left[\widehat{c}_j^R(p_0) - \kappa(\eta, p_0)\right]$ is the average residual of costs for type τ_i in the reported hold-out sample $H_R^{\tau_i} \equiv \{(j, \tau_j, \widehat{c}_j^R)|j \in H_R, \tau_j = \tau_i\}$.

Strategic capitation satisfies the following key properties

$$\forall \lambda, \quad \mathbb{E}_\nu[\Delta\pi(\eta, H_R, \Lambda)|D_0, D_1, \beta^*, \lambda] = \mathbb{E}_\nu\left[\sum_{\tau \in \eta}(\mu_\Lambda(\tau|\eta) - \mu_I(\tau|\eta))\mathbb{E}_\nu[e_{\tau,i}|D_1, D_0, \tau]\right], \tag{10}$$

$$\forall \lambda \in \mathcal{M}(E), \forall \beta, \forall \eta \in E, \quad \mathbb{E}_\nu\left[|\Lambda^\eta|\Delta\pi(\eta, H_R, \Lambda)|D_0, D_1, \beta, \lambda\right] = 0. \tag{11}$$

---

[10]Recall that for any sample J, $\mu_J(\tau|\eta) \equiv \frac{|J^\tau|}{|J^\eta|}$ denotes the distribution of types τ conditional on characteristic $\eta \subset T$ in sample J.

Condition (10) implies that under truthful reporting β^*, the strategic capitation adjustment is an unbiased estimate of the excess profits plan p_1 may have obtained through illegitimate selection (the adjustment is negative if private plan p_1 overselects types that are comparatively cheaper to treat). This noisy ex post estimate provides an accurate ex ante correction and dissuades inefficient selection. Condition (11) ensures that regardless of the public plan's reporting strategy β, the private plan can guarantee itself expected capitation payments $\pi(\eta) = \kappa(\eta, p_0)$, provided it uses a legitimate selection strategy $\lambda \in \mathcal{M}(E)$.[11] Strategic capitation essentially turns illegitimate selection into a zero-sum game in which both the public and private plans can guarantee themselves a value of zero.

Proposition 3. *Strategic capitation contract Π^{strat} induces a unique equilibrium (λ^*, β^*). Private plan p_1 selects patients efficiently, and public plan p_0 truthfully reports hold-out sample H. Both plans get positive expected payoffs:*

$$\mathbb{E}_\nu[U_0|D_0, D_1, \lambda^*, \beta^*] \geq 0 \quad and \quad \mathbb{E}_\nu[U_1|D_0, D_1, \lambda^*, \beta^*] \geq 0.$$

Section 6 discusses possible alternative mechanisms and clarifies the importance of different moving parts needed to make strategic capitation work. In a nutshell, Big Data (i.e. observable types τ) is needed to measure the private plan's deviation from legitimate selection, and ensure that in equilibrium, public plan p_0 does not benefit from misreporting hold-out sample costs. The hold-out sample is needed to ensure that residuals r_{τ_i} are uncorrelated to plan p_1's information.

4   General Analysis

The strategic capitation scheme presented in Section 3 relies on cost decomposition (4): the surplus maximizing policy depends on a small number of commonly known characteristics $\eta \in E$. This is not realistic. A private plan's dimensions of comparative advantage are likely

---

[11]This point plays a key role when studying incentives for truthful revelation in exchanges.

to be its private information. For instance, a private plan may be able to innovate and develop comparative advantages along new dimensions. In addition, it need not be the case that the optimal selection policy be measurable with respect to a small set of characteristics. Finally, in practice, the public plan's expected cost of treatment conditional on a characteristic η will have to be estimated from data. This creates additional room for selection by the private plan. This section extends strategic capitation to such environments.

We denote by e_i the idiosyncratic cost shocks of a patient enrolled in public plan p_0, conditional on cost distribution F: $e_i \equiv \widehat{c}_i(p_0) - \mathbb{E}_F[\widehat{c}_i(p_0)|\tau_i]$. By construction, it must be that $e_i \in [-c_{\max}, c_{\max}]$. Finally, let

$$S(\lambda|D_0, D_1) \equiv \mathbb{E}_{\nu, \lambda}\left[\sum_{i \in \Lambda} \widehat{c}_i(p_0) - \widehat{c}_i(p_1)\Big| D_0, D_1\right] - K(\lambda), \quad \text{and}$$

$$S_{E|D_0, D_1} \equiv \max_{\lambda \in \mathcal{M}(E)} S(\lambda|D_0, D_1)$$

respectively denote the surplus achieved by selection rule λ given data D_0, D_1, and the maximum surplus achievable using selection rules measurable with respect to a partition E of types.

4.1 Generalized Strategic Capitation

For any given collection $\mathcal{E}$ of partitions $E \in \mathcal{E}$ – i.e. for any family of possible definitions of legitimate and non-legitimate characteristics – our goal is to approach the maximum achievable efficiency $S_{E|D_0, D_1}$ with respect to partitions $E \in \mathcal{E}$. We take on two difficulties: first, the relevant partition E is no longer common knowledge, but must be elicited from the private plan; second the public plan's expected costs conditional on characteristics $\eta \in E$ must estimated from data. We define the generalized strategic capitation scheme $G_{\mathcal{E}}^{\text{strat}}$ as follows:

1. data D_0 is shared with plan p_1;

2. plan p_1 picks a partition $E \in \mathcal{E}$ according to which it will be allowed to select patients; we continue to refer to characteristics $\eta \in E$ as legitimate selection characteristics;

3. plan p_1 is rewarded using the strategic capitation scheme Π^{strat} defined by

$$\Pi^{\text{strat}}(\Lambda, H_R) \equiv \sum_{i \in \Lambda} \pi(\eta_i) + \Delta\pi(\eta_i, H_R, \Lambda)$$

where $\pi(\eta) = \widehat{\kappa}(\eta, p_0) \equiv \sum_{\tau \in \eta} \mu_I(\tau|\eta) \frac{1}{|D_0^\tau|} \sum_{i \in D_0^\tau} \widehat{c}_i(p_0)$ is the empirical estimate $\widehat{\kappa}(\eta, p_0)$ (in sample data D_0) of the public plan's expected treatment costs conditional on characteristic $\eta \in E$. As in Section 3, $\Delta\pi(\eta, H_R, \Lambda)$ takes the form:

$$\Delta\pi(\eta, H_R, \Lambda) \equiv \text{cov}_I(s_{\tau_i}, r_{\tau_i}|\eta_i = \eta) = \frac{1}{|I^\eta|} \sum_{i \in I^\eta} s_{\tau_i} r_{\tau_i},$$

with

$$s_{\tau_i} \equiv \frac{\mu_\Lambda(\tau_i|\eta_i)}{\mu_I(\tau_i|\eta_i)} - 1 \quad \text{and} \quad r_{\tau_i} \equiv \frac{1}{|H_R^{\tau_i}|} \sum_{j \in H_R^{\tau_i}} \left[\widehat{c}_j^R(p_0) - \widehat{\kappa}(\eta, p_0)\right].$$

An equilibrium of mechanism $G_{\mathcal{E}}^{\text{strat}}$ is a triplet (E, λ, β) where $E \in \mathcal{E}$ is private plan p_1's choice of the characteristics it can use for selection.

Mechanism $G_{\mathcal{E}}^{\text{strat}}$ expands on strategic capitation by letting the private plan specify the set of characteristics it wishes to use for selection. As we show below, this additional degree of freedom results in unavoidable losses related to the complexity of the class of models $\mathcal{E}$ the private plan is allowed to pick from. These losses are related to penalties encountered in the model selection literature (Vapnik, 1998, Massart and Picard, 2007), and indeed one can think of our problem as one of delegated model selection.

Definition 1. *For any class of partitions $\mathcal{E}$ and idiosyncratic errors $e = (e_i)_{i \in D_0}$, let $\Psi(\mathcal{E}, e)$ denote the random variable*

$$\Psi(\mathcal{E}, e) \equiv \max_{E \in \mathcal{E}} \left(\sum_{\eta \in E} |I^\eta| \left[\sum_{\tau \in \eta} \mu_I(\tau|\eta) \frac{1}{|D_0^\tau|} \sum_{i \in D_0^\tau} e_i \right]^+ \right). \tag{12}$$

22

Variable $\Psi(\mathcal{E}, e)$ is an upper-bound to the gains a perfectly informed private plan could obtain from selecting the partition E that lets it optimally target over-reimbursed types. The scope for selection comes from the fact that generalized capitation uses slightly noisy sample averages $\widehat{\kappa}(\eta, p_0)$ to estimate the public plan's cost of service $\mathbb{E}_F[\widehat{c}_i(p_0)|\eta]$ conditional on legitimate characteristics $\eta \in E$. In Section 3 we simply assumed that the public plan knew its own cost $\kappa(\eta, p_0)$ conditional on η.

Generalized capitation extends the performance bounds described in Proposition 3 up to a penalty of order $\mathbb{E}_\nu[\Psi(\mathcal{E}, e)]$.

Proposition 4 (efficiency bounds). *Consider a collection $\mathcal{E}$ of partitions $E \in \mathcal{E}$. In any equilibrium (E, λ, β) of mechanism $G_{\mathcal{E}}^{strat}$ we have that*

$$S(\lambda) \geq \mathbb{E}_\nu \left[\max_{E \in \mathcal{E}} S_{E|D_0, D_1} \right] - 2\mathbb{E}_\nu \left[\Psi(\mathcal{E}, e) \right]; \tag{13}$$

$$\mathbb{E}_\nu \left[-\Pi + \sum_{i \in \Lambda} \widehat{c}_i(p_0) \middle| D_0 \right] \geq -\mathbb{E}_\nu \left[\Psi(\mathcal{E}, e) \right]; \tag{14}$$

$$\mathbb{E}_\nu \left[\Pi - \sum_{i \in \Lambda} \widehat{c}_i(p_1) \middle| D_0, D_1 \right] \geq 0. \tag{15}$$

Note that Proposition 4 takes as given the class of models $\mathcal{E}$. Corollary 2 will show how to endogenize the class of models $\mathcal{E}$, provided $\mathcal{E}$ can be chosen before data D_0 is shared with the private provider.

To operationalize the performance bounds of Proposition 4 we need to evaluate term $\mathbb{E}_\nu[\Psi(\mathcal{E}, e)]$, which depends depends on prior ν through error term e. Lemma 1 provides prior-free bounds for $\mathbb{E}_\nu[\Psi(\mathcal{E}, e)]$. Let $\alpha \equiv \mathbb{E}_{\mu_I} \left[\frac{|I^\tau|}{|D_0^\tau|} \frac{|D_0|}{|I|} \right] \geq 1$ denote the average representativeness of data D_0 for patients in I.[12] Let $M \equiv \sum_{E \in \mathcal{E}} \left(2^{|E|} - 1 \right)$.

Lemma 1 (bounds on penalties). *(i) Let $(e_i')_{i \in I}$ denote i.i.d. random variables uniformly distributed over $\{-c_{max}, c_{max}\}$. For any class $\mathcal{E}$ and any centered error*

---

[12] The fact that $\alpha \geq 1$ follows from the observation that $\alpha = \mathbb{E}_{\mu_I} \left[\mu_I(\tau)/\mu_{D_0}(\tau) \right] \geq 1/\mathbb{E}_{\mu_I} \left[\mu_{D_0}(\tau)/\mu_I(\tau) \right] = 1$.

terms $(e_i)_{i \in I}$ arbitrarily distributed over $[-c_{\max}, c_{\max}]$, we have that

$$\mathbb{E}_\nu \left[\Psi(\mathcal{E}, e) \right] \leq \mathbb{E} \left[\Psi(\mathcal{E}, e') \right] \equiv \overline{\Psi}_{\mathcal{E}}. \tag{16}$$

(ii) Regardless of the distribution of error terms $(e_i)_{i \in I}$,

$$\mathbb{E}_\nu [\Psi(\mathcal{E}, e)] \leq |I| c_{\max} \sqrt{\frac{2\alpha}{|D_0|}} \left(1 + \sqrt{\log M} \right). \tag{17}$$

Bound (16) is a tight bound that can be computed through numerical simulation. We denote by $\overline{\Psi}_{\mathcal{E}}$ the corresponding upper-bound, which does not depend on prior ν. Bound (17) is a more explicit but coarser bound. While it may seem overly conservative, we now show that it delivers non-trivial results in a setting of interest: the class of sparse linear classifiers.

Example: sparse linear classifiers. We now evaluate bound (17) for a natural class of partitions $\mathcal{E}$: those generated by sparse linear classifiers. This is consistent with the use of LASSO and other penalized regression techniques for prediction in Big Data settings. We assume that type space T is a subset of $\mathbb{R}^f$ (we will use the inequality $f \leq |T| \leq |D_0|$). For $d \in \{1, \cdots, f\}$, a d-sparse vector $v = (v_k)_{k \in \{1, \cdots, f\}} \in \mathbb{R}^f$ is a vector with at most d non-zero coordinates. The family of partitions $\mathcal{E}$ induced by d-sparse classifiers is defined as

$$\mathcal{E} \equiv \{E_v \equiv \{\eta_v^+, \eta_v^-\} | v \in \mathbb{R}^f, \ v \text{ is } d\text{-sparse}\}$$

$$\text{where } \eta_v^+ = \{\tau \in T \text{ s.t. } \langle \tau, v \rangle \geq 0\} \text{ and } \eta_v^- = \{\tau \in T \text{ s.t. } \langle \tau, v \rangle < 0\}.$$

For any d-sparse vector $v \in \mathbb{R}^f$, E_v is the partition of $\mathbb{R}^f$ in the two half-spaces η_v^+ and η_v^- defined by direction v.[13]

The private plan is allowed to use any d-sparse linear classifier to decide whether or not

---

[13]Note that affine hyperplanes can be generated by including a constant in the type vector.

to select a particular set of types or not.

Corollary 1. *When the class of possible partitions $\mathcal{E}$ consists of partitions induced by d-sparse classifiers, penalty $\mathbb{E}_\nu\left[\Psi(\mathcal{E}, e)\right]$ satisfies*

$$\mathbb{E}_\nu\left[\Psi(\mathcal{E}, e)\right] \leq 4c_{\max}|I|\sqrt{\frac{\alpha d \log|D_0|}{|D_0|}}. \tag{18}$$

Corollary 1 follows from a direct application of Lemma 1(ii) and the fact that $|T| \leq |D_0|$. Indeed, the number of possible partitions of $|T|$ points generated by d-sparse linear classifiers is bounded by $2^d \times \binom{f}{d} \times \binom{|T|}{d} < \frac{1}{4}|T|^{3d}$, where $\binom{m}{n} = \frac{m!}{(m-n)!n!}$.[14] Since each $E \in \mathcal{E}$ contains two elements, we obtain that $M \leq K^{2d}$. Applying bound (17) implies (18).

Note that for all practical purposes, term $\sqrt{\log|D_0|}$ may be treated as a constant between 4 and 5. For $|D_0| = 48 \times 10^6$, approximately the size of the US Medicare population, $\sqrt{\log|D_0|} \simeq 4.2$. For $|D_0| = 7 \times 10^9$, roughly the current world population, $\sqrt{\log|D_0|} \simeq 4.8$.

Endogenizing $\mathcal{E}$. While Proposition 4 takes as given the class $\mathcal{E}$ of type partitions, it is possible to endogenize $\mathcal{E}$. Consider the following extension of generalized strategic capitation:

1. private plan p_1 picks a class of partitions $\mathcal{E}$, and pays a price $p_\mathcal{E} = \overline{\Psi}_\mathcal{E}$ (where $\overline{\Psi}_\mathcal{E}$ is defined within bound (16));

2. the private and public plan play mechanism $G_\mathcal{E}^{\text{strat}}$.

Note that the private plan must choose $\mathcal{E}$ before information D_0 is shared. The proof of Proposition 4 implies the following corollary.

---

[14]To obtain this bound, observe that there are $\binom{f}{d}$ ways to choose the d non-zero coordinates in the d-sparse classifier. For each such choice, the classifier can be written in the form $a_1x_1 + \ldots + a_dx_d < 1$, where $x_1, \ldots, x_d$ are the relevant coordinates, and $a_1, \ldots, a_d \in \mathbb{R}$ are appropriately chosen coefficients. The set of appropriate d-tuples $(a_1, \ldots, a_d)$ forms a polytope $\mathcal{A}$ in $\mathbb{R}^d$, with each of the $|T|$ points representing a linear constraint on the possible values of $(a_1, \ldots, a_d)$. A node of such a polytope is an intersection of d constraints, and thus $\mathcal{A}$ can be identified using d points from T along with the signs of the d constraints. This gives at most $2^d \times \binom{|T|}{d}$ choices.

Corollary 2. *Let $\mathcal{P}(T)$ denote the set of partitions of type space T. In any equilibrium $(\mathcal{E}, E, \lambda, \beta)$ of the extended generalized strategic capitation mechanism,*

$$\mathbb{E}_\nu\left[S(\lambda|D_0, D_1)\right] \geq \max_{\mathcal{E}' \subset \mathcal{P}(T)} \mathbb{E}_\nu\left[\max_{E \in \mathcal{E}'} S_{E|D_0, D_1}\right] - 2\overline{\Psi}_{\mathcal{E}'}. \tag{19}$$

By charging the private plan a price taking the form of a complexity penalty, it is possible to let the private plan choose the class of models it can use to select patients. Note that if the upfront price $p_\mathcal{E}$ is transferred to the public provider, then the public provider makes no losses in expectation.

4.2 Unimprovability of Strategic Capitation

In the spirit of Hartline and Roughgarden (2008), we now provide a lower-bound for the minimal efficiency losses that any mechanism can guarantee, even if the mechanism is allowed to depend on prior ν. The proof will construct a specific prior ν over the tuple (F, D_0, D_1, H, K) for which we establish minimal efficiency losses under any direct mechanism, along the lines of Myerson and Satterthwaite (1983). Sample size $|D_0|$, as well as the distributions of types $\mu_I \in \Delta(T)$ in the patient population I, and $\mu_{D_0} \in \Delta(T)$ in public data D_0, are known.

We consider the problem of Bayes-Nash implementation using budget-balanced direct mechanisms g of the following form:

- data D_0 is publicly observable;
- plan p_1 sends a message $m_1 = (D_1^m, K^m)$, reporting its data and selection costs;
- the mechanism suggests a selection $\lambda_g(D_0, m_1) \in [0, 1]^T$ by private plan p_1;
- plan p_1 makes a selection decision $\lambda \in [0, 1]^T$, with realized selection $\Lambda \subset I$;
- plan p_0 sends a message $m_0 = H_R$ corresponding to a reported hold-out sample;
- transfers $\Pi(D_0, m_1, m_0, \Lambda)$ from p_0 to p_1 are implemented.

We denote by $\mathcal{G}_\nu$ the set of incentive compatible direct revelation mechanisms under prior

ν. For any mechanism g, we denote by λ_g the equilibrium selection by private plan p_1 that maximizes social surplus.

Proposition 5. *For any class of partitions $\mathcal{E}$, there exists a prior ν such that*

$$\max_{g \in \mathcal{G}_\nu} \mathbb{E}_\nu S(\lambda_g) \leq \mathbb{E}_\nu \left[\max_{E \in \mathcal{E}} S_{E|D_0,D_1} \right] - \hbar|I|c_{max} \max_{E \in \mathcal{E}} \mathbb{E}_{\mu_I} \left[\frac{1}{\sqrt{|D_0^\eta|}} \right]. \tag{20}$$

where $\hbar$ is a fixed constant independent of $\mathcal{E}$ and sample size $|D_0|$.

In particular, the efficiency loss achieved by strategic capitation for linear classifiers (Corollary 1) is tight up to an order $\sqrt{\log|D_0|}$, which, for all plausible values of $|D_0|$, can be treated as a constant less than 5.

5   Extension: Adverse Selection in Exchanges

So far, we have studied adverse selection in a setting with a single private plan and a single public plan. Adverse selection is also a significant concern in insurance markets consisting of multiple private plans, such as the ones organized by the American Healthcare Act. If regulation constrains prices to depend only on a subset of observables (as is the case with community rating), plans have incentives to select patients that are cheaper to serve given characteristics excluded from legal pricing formulas. This increases the cost of serving patients and can result in limited entry. A simple example suggests that strategic capitation may help improve market outcomes in such environments.

A stylized model.   As in Section 2, a set I of patients with types $\tau \in T$ has inelastic unit demand for insurance, where insurance corresponds to a single standardized insurance contract. Plan p_0 is now an incumbent private plan, while p_1 is a potential entrant. For simplicity, we assume that each plan's cost technology is the same: $\forall i \in I, \widehat{c}_i(p_0) \sim \widehat{c}_i(p_1)$. Here the objective is not to improve the allocation of patients to plans, but rather to increase

competition so that insurance is priced at marginal cost. By law, each plan p is constrained to offer insurance at prices $\pi_p(\eta)$ that depend only on a coarse set of patient characteristics $\eta \in E$, where E is a partition of type-space T. Prices are bounded above by $\overline{\pi}$.[15]

We assume that the private plans both know their common expected cost of treatment $\kappa(\tau) \equiv \mathbb{E}[\widehat{c}_i | \tau_i = \tau]$ conditional on type τ. Let $\kappa(\eta) \equiv \mathbb{E}_{\mu_I}[\kappa(\tau)|\eta]$ denote the expected cost of treatment conditional on legitimate characteristics. Each plan p has access to a hold-out sample of its own costs H_p. We assume that both plans have lexicographic preferences over maximizing their own revenue and minimizing that of their competitor.[16] The timing of decisions is as follows:

1. potential entrant p_1 decides to enter the market or not;

2. each plan p active in the market submits a price formula $\pi_p : \eta \mapsto \pi_p(\eta)$;

3. each plan p active in the market chooses a selection $\lambda_p \in [-1/2, 1/2]^T$ of patients;

4. if $\pi_{p_0}(\eta) \neq \pi_{p_1}(\eta)$, patients of type η purchase insurance from the cheapest plan; if $\pi_{p_0}(\eta) = \pi_{p_1}(\eta)$, plan p serves a patient of type $\tau \in \eta$ with probability

$$\mu_I(\tau) \times \left(\frac{1}{2} + \lambda_p(\tau) - \lambda_{\neg p}(\tau) \right),$$

   where $\neg p$ denotes the other plan.

   We assume that the cost of selection $K(\lambda_p)$ is positive, strictly convex, continuously differentiable, and equal to 0 when λ_p is constant and equal to 0: $K(0) = 0$. In addition, $K(\lambda)$ is sufficiently steep around $\lambda = 0$ that $1/2 + \lambda_p - \lambda_{\neg p} \geq 0$ for all individually rational selection policies.

The cross-price elasticity of patient demand is infinite, so that patients always go to the cheapest plan. As a result an entrant will make at most zero profits when entering. We assume that whenever the entrant can guarantee itself zero profits it enters.[17] We denote the realized selected sample of patients purchasing from plan p by Λ_p.

---

[15]Parameter $\overline{\pi}$ may be viewed as the patients' (common) value for insurance.
[16]This could be because reducing a competitor's profits reduces its ability to expand and compete in the future.
[17]This could be due to small subsidies for entry, or high but finite cross-price elasticities.

The following result holds.

Proposition 6. *The market entry game described above has a unique subgame perfect equilibrium. The potential entrant does not enter, and the incumbent charges price $\pi_{p_0}(\eta) = \overline{\pi}$.*

In the off-equilibrium subgame following entry, both the entrant and the incumbent make equilibrium losses $-K(\lambda^*) < 0$ where λ^* solves $\max_{\lambda \in [0,1]^T} \mathbb{E}_\lambda \left[\sum_{i \in \Lambda} \kappa(\eta_i) - \kappa(\tau_i) \right] - K(\lambda)$.

Indeed, because cross-price elasticities are infinite, in equilibrium, both plans price at marginal cost conditional on η: $\pi_p(\eta) = \kappa(\eta)$. Furthermore, since the marginal cost of selection at $\lambda_p = 0$ is zero, both players find it profitable to engage in non-zero selection. In aggregate however, selection efforts by the two plans cancel one another and merely destroy surplus.

Strategic capitation. Consider now the following extension of the strategic capitation scheme introduced in Section 3. The market entry game is modified in two ways:

- at stage 2 each active plan p submits a pricing formulas π_p, and a report $H_{R,p}$ of its hold-out sample;

- after selection has occurred, for each type η it serves, plan p receives price $\pi_p(\eta)$ and capitation adjustment $\Delta\pi(\eta, H_{R,\neg p}, \Lambda_p)$ taking the form:

$$\Delta\pi(\eta_i, H_{R,\neg p}, \Lambda_p) \equiv \text{cov}_I(s_{\tau_i,p}, r_{\tau_i,p} | \eta_i = \eta) = \frac{1}{|I^\eta|} \sum_{i \in I^\eta} s_{\tau_i,p} r_{\tau_i,p},$$

with

$$s_{\tau_i,p} \equiv \frac{\mu_{\Lambda_p}(\tau_i | \eta_i)}{\mu_I(\tau_i | \eta_i)} - 1 \quad \text{and} \quad r_{\tau_i,p} \equiv \frac{1}{|H_{R,\neg p}^{\tau_i}|} \sum_{j \in H_{R,\neg p}^{\tau_i}} \left[\widehat{c}_j^{R,\neg p}(p_0) - \pi_p(\eta) \right].$$

Proposition 7. *The market game with strategic capitation described above has an efficient truthful equilibrium in which: the potential entrant enters; both plans submit prices $\pi_p(\eta) = \kappa(\eta)$; both plans select a representative population in expectation ($\lambda_p = 0$); both plans submit their hold-out sample costs truthfully ($H_{R,p} = H_p$); and expected ex post adjustments are equal to 0 ($\mathbb{E}\Delta\pi_p = 0$).*

The intuition for Proposition 7 result is identical to that of Proposition 3. Given ex ante representative selection, a plan's expected capitation adjustment is equal to zero regardless of messages sent by the other plan. Given truthful revelation of costs, representative selection is a best-response. This reduces the costs of adverse selection and encourages entry.

6   Discussion

This paper explores the value of Big Data in reducing the extent of adverse selection in government-run capitation schemes. We argue that in realistic Big Data environments, including a large number of covariates in an ex ante capitation formula is unlikely to succeed. Instead, we suggest that Big Data may be used to align incentives by using ex post capitation adjustments that interact an unbiased estimate of counterfactual costs to the public plan, with the private plan's deviation from legitimate selection.

This section discusses alternative mechanisms, as well as extensions dealing with dynamic selection, risk-inflation, and heterogeneity in the quality of care.

6.1   Alternative Mechanisms

Generalized strategic capitation deals with three different difficulties: common values, exact budget balance, and prior-free implementation. As Proposition 5 shows, using prior-dependent mechanisms does not imply large performance improvements. Still, in order to clarify the economic forces at work in our analysis it is useful to delineate the mechanics of other relevant mechanisms.

Mechanisms from the literature.   Other work has emphasized the value of ex post noisy signals in environments with quasi-linear preferences. Riordan and Sappington (1988) show that it is possible to efficiently regulate a monopoly with unknown costs by exploiting public signals correlated to the monopoly's type. Using a construction related to that of Cremer and McLean (1988), they show how to extract all the surplus by offering the monopoly

appropriately chosen screening contracts. Strategic capitation also exploits the fact that noisy ex post signals (here, hold-out cost realizations) can be used to construct accurate ex ante incentives, but our environment differs in key ways. First, signals are not public, and we need to take care of the public plan's incentives to reveal its hold-out cost data H. Second, the identification condition at the heart of Riordan and Sappington (1988) is not satisfied: neither the distribution of the public plan's cost, nor the private plan's beliefs thereover, are sufficient statistics of the private plans' costs.

Mezzetti (2004) shows that it is possible to obtain efficiency in common value environments using ex post reports of the players' realized payoffs. In our application the mechanism proposed by Mezzetti (2004) would proceed by making the private plan a negative ex post transfer equal to the public plan's realized cost, and a positive ex ante transfer to cover expected costs. This mechanism does not satisfy budget balance (the public plan does not receive the private plan's transfer – if it did truthful reporting of hold-out data H by the public plan would not be incentive compatible) and relies on priors to set ex ante transfers.[18]

The differences between our environment and that of Mezzetti (2004) help clarify the role played by the Big Data assumption, i.e. the assumption that types are observable but not interpretable. We align incentives under budget balance by forming a measure of the private plan's deviation from legitimate selection, and interacting this measure with an unbiased estimate of the public plan's counterfactual costs. This ensures that in equilibrium, neither the private nor the public plan can affect their expected payoffs by deviating from legitimate selection and truthful reporting. The observability of types is used to compute the private plan's deviation from legitimate selection, as well as correctly reweight the distribution of types in the hold-out sample H to obtain estimates of counterfactual costs in the sample Λ of patients selected by the private plan.

---

[18]Budget balanced expected externality mechanisms along the lines of d'Aspremont and Gérard-Varet (1979) or Athey and Segal (2013) require private values. They are also prior-dependent.

Intuitive alternative mechanisms. A key step in strategic capitation is to use hold-out data to form estimates of counterfactual costs for the public plan. The assumption that types are observable is needed to reweight the distribution of types in the hold-out sample to match that of the selected sample. There may be other ways to form an unbiased estimate of counterfactual costs to public plan p_0. For instance, if it were possible to assign patients selected by the private plan back to the public plan with a fixed uniform probability, one could form an estimate of counterfactual costs without observing types. Beyond feasibility issues (patients would likely object), a difficulty with this approach is that it does not take care of the public plan's incentives to bias its own cost reports.

Strategic capitation dissuades illegitimate selection by forming unbiased estimates of the private plan's excess profits. An alternative way to dissuade illegitimate selection is to impose sufficiently large penalties, say proportional to $\left| \frac{\mu_\Lambda(\tau_i|\eta_i)}{\mu_I(\tau_i|\eta_i)} - 1 \right|$, when the sample selected by the private plan deviates from legitimate selection. This scheme requires the observability of types but does not require the availability of a hold-out sample. If the private plan can perfectly select patients at no cost ($K(\lambda) = 0$ for all λ), this schemes induces efficient legitimate selection. However this scheme induces an efficiency loss if it is costly for private provider p_1 to ensure that realized selection Λ is perfectly representative of population I. Strategic capitation avoids the issue by using hold-out data to form an unbiased estimate of the profits from selection.

6.2 Extensions and Implementation Concerns

We now briefly discuss various realistic challenges with capitation programs like Medicare Advantage, and how they can be addressed within our framework. Details are provided in Appendix A.

Dynamic selection and risk-inflation. The process of selection is dynamic. In the context of Medicare Advantage, patients have the opportunity to switch back and forth

between public and private plans once a year. This implies that costs of care need to be evaluated over time. Plans with low short-term cost of care may end up generating greater longer term costs if they skimp on quality, and encourage patients to disenroll once they get sick enough (Ellis, 1998). Appendix A shows how to adjust strategic capitation to address this issue. It becomes important to keep track of the counterfactual distribution of types over time, had the patient remained with the public plan.

A key insight from Appendix A is that correct dynamic capitation fees remove incentives for risk-inflation by private plans. Indeed, if a patient with legitimate characteristic η_t enrolls in the private plan at time t, then baseline repayments π_{t+s} to the private plan at all times $t + s$ where the patient remains with the private plan take the form

$$\pi_{t+s} = \pi(t + s, \eta_t) \equiv \mathbb{E}[\widehat{c}_{i,t+s}(p_0)|\eta_t].$$

In other words, target repayments depend only on the type η_t of the patient when she enrolls with the private plan, and on elapsed time $t + s$. *Target repayments do not depend on the patient's type η_{t+s} after enrollment time t.* As a result, the private plan has no incentives to exaggerate the medical condition of patients it enrolls (for instance by running a battery of tests detecting mild conditions). Plans may have incentives to exaggerate the medical condition of patients it enrolls when repayments π_{t+s} depend on types η_{t+s} at time $t + s$.

Quality. Throughout the paper we assume that the quality of actual healthcare delivery is homogeneous across plans. In practice, insurance plans may differ in the quality of care they deliver to their enrollees. It may be important to take into account this quality dimension when designing capitation schemes. Otherwise, costs could be kept low at the expense of quality. Appendix A describes an extension of strategic capitation that correctly reflects differences in the quality of care. An important limitation is that it requires health outcomes (including death) to be observable and assigned monetary values.

Surplus Extraction. The paper focuses on the efficient allocation of patients across public and private plans. However, if there is a deadweight loss of public funds, it may be welfare improving for the public plan to extract some of the surplus. Since the private plan has private information over its costs conditional on patient types, this is a difficult multidimensional screening problem. Two observations are helpful to make progress on this issue. First, given that we consider prior-free mechanisms, the argument of Carroll (2017) suggests there may not be much value in complex multidimensional screening. It may be near-optimal to focus on separable one-dimensional screening mechanisms that associate a discounted baseline capitation rate $\rho(\eta)\kappa(\eta, p_0)$ to each patient with characteristics η, with $\rho(\eta) \in [0,1]$ a discount factor. A second useful observation is that strategic capitation adjustments used to prevent selection of mispriced types can be applied to any baseline repayment scheme. This suggests using capitation schemes of the form

$$\Pi(\Lambda, H_R) \equiv \sum_{i \in \Lambda} \rho(\eta)\kappa(\eta, p_0) + \Delta\pi(\eta_i, H_R)$$

where $\Delta\pi(\eta_i, H_R, \Lambda) = \frac{1}{|I^\eta|} \sum_{i \in I^\eta} s_{\tau_i} r_{\tau_i}$, with

$$s_{\tau_i} \equiv \frac{\mu_\Lambda(\tau_i|\eta_i)}{\mu_I(\tau_i|\eta_i)} - 1 \quad \text{and} \quad r_{\tau_i} \equiv \frac{1}{|H_R^{\tau_i}|} \rho(\eta_i) \sum_{j \in H_R^{\tau_i}} \left[\hat{c}_j^R(p_0) - \kappa(\eta, p_0) \right].$$

This separates the problems of extracting revenue and preventing illegitimate selection.

Ethical concerns. Regulators frequently ban indexing ex ante capitation rates on certain observables, such as ethnicity or income. One rationale for such bans is that the law has expressive content that affects social norms, and it is desirable to reinforce the norm that all citizens deserve equal treatment. We believe that these ethical concerns admit an important refinement: differential treatment on the equilibrium path weakens the norm of equality; differential treatment off of the equilibrium path does not. The latter is the case for strategic capitation: it punishes plans off of the equilibrium path for non-representative selection of

types.

Volatility of revenues and profits. One concern with strategic capitation is that capitation payments to the private plan are uncertain at the interim stage. If correlated noise φ_i in selection causes the private plan to enroll types that are relatively cheap (resp. expensive) to treat for the public plan, it receives lower (resp. higher) payments than anticipated. This increases the volatility of revenues, but may in fact reduce the volatility of profits. Indeed, types that are relatively cheap (resp. expensive) to treat to the public plan are also likely to be cheap (resp. expensive) to treat for the private plan. If noise in selection causes a plan to overselect types that are relatively cheap to treat for the public plan, it receives a negative capitation adjustment. However, it is likely that the cost of treating these types is also relatively cheap for the private plan, keeping net profits stable. A similar reasoning applies if the plan overselects types that are relatively expensive to treat for the public plan. Positive capitation adjustments may well compensate a corresponding increase in the private plan's cost of care. In other words, strategic capitation can serve as insurance against selection shocks.

Appendix

A   Extensions

A.1   Dynamic Selection and Risk-Inflation

In dynamic settings, capitation schemes need to control for differential transitions in health status across plans. For simplicity, as in Section 3, we assume that comparative advantage depends on a commonly known set E of legitimate selection characteristics $\eta \in E$, and that expected costs conditional on legitimate characteristics are known. We denote by $\tau_{i,t}$ the type of patient i at date $t \in \mathbb{N}$, by $\eta_{i,t}$ her legitimate selection characteristics at date t, and

by $\widehat{c}_i(t, p)$ her realized cost of care if treated by plan p at time t. Types $(\tau_t)_{t \in \{0, \cdots, \overline{T}\}}$ and characteristics $(\eta_t)_{t \in \{0, \cdots, \overline{T}\}}$ follow Markov chains, whose process ϕ_p depends on the plan p in which the patient is enrolled. Future costs are discounted using discount factor $\delta \in (0, 1]$, and $\overline{T}$ denotes an upper bound to the duration of patients' lives in the system.

For a patient i of type τ_i enrolled with the public plan from time t to time $\overline{T}$, we define

$$\widehat{C}_i(t, p_0) \equiv \sum_{s=t}^{\overline{T}} \delta^{s-t} \widehat{c}_i(s, p_0) \quad \text{and} \quad C(t, \eta, p_0) \equiv \mathbb{E}_\nu \left[\widehat{C}_i(t, p_0) \Big| \eta_t = \eta \right].$$

In dynamic environments, strategic capitation must accommodate the possible reenrollment of patients with the public plan. As a result, transfers must occur at the reentry of patients into the public system. Let us denote by Λ_t the selection of patients enrolled with the private plan at time t, and by Λ_t^{re} the selection of patients disenrolling from the private plan and re-enrolling with the public plan at time t. The following scheme extends strategic capitation. At initial time of enrollment $t = 0$, the public plan commits to the following baseline payments conditional on legitimate characteristics $\eta \in E$:

- A capitation payment $\pi(t, \eta_0) = \mathbb{E}_\nu[\widehat{c}_i(t, p_0)|\eta_{i,0} = \eta_0]$ whenever patient i with initial type η_0 is enrolled with the private plan at time t;

  Note that because transfers depend only on type η_0, there is no incentive for risk-inflation by the plan, i.e. an incentive to requalify the patient as a different type;

- A signed transfer π_i^{re} (with positive transfers being made from the public plan to the private plan) at every time T patient i returns to the public plan: $\pi_i^{re} = \mathbb{E}_\nu[\widehat{C}_i(T, p_0)|\eta_{i,0}] - C(T, \eta_{i,T}, p_0)$.

Provided that the private plan does not engage in illegitimate selection, this scheme induces efficient dynamic behavior by the private plan. To dissuade illegitimate selection, dynamic strategic capitation makes payment adjustments $\Delta\pi(t, \eta_0)$ and $\Delta\pi^{re}(T, \eta_0)$ using reported dynamic hold-out data H_R as follows:

- $\Delta\pi(t, \eta_0) = \frac{1}{|I^{\eta_0}|} \sum_{i \in I^{\eta_0}} s_{i,t} r_{i,t}$, with

$$s_{i,t} \equiv \frac{\mu_{\Lambda_t}(\tau_{i,0}|\eta_{i,0} = \eta_0)}{\mu_I(\tau_{i,0}|\eta_{i,0} = \eta_0)} - 1, \quad \text{and} \quad r_{i,t} \equiv \frac{1}{|H_R^{\tau_{i,0}}|} \sum_{j \in H_R^{\tau_{i,0}}} \widehat{c}_j^R(t, p_0) - \pi(t, \eta_0).$$

- $\Delta\pi^{re}(t, \eta_t) = \frac{1}{|I^{\eta_t}|} \sum_{i \in I^{\eta_t}} s_{i,t}^{re} r_{i,t}^{re}$, with

$$s_{i,t}^{re} \equiv \frac{\mu_{\Lambda_t^{re}}(\tau_{i,t}|\eta_{i,t} = \eta_t)}{\mu_I(\tau_{i,t}|\eta_{i,t} = \eta_t)} - 1, \quad \text{and} \quad r_{i,t}^{re} \equiv \frac{1}{|H_R^{\tau_{i,t}}|} \sum_{j \in H_R^{\tau_{i,t}}} \left[C(T, \eta_{i,T}, p_0) - \widehat{C}_j^R(t, p_0) \right].$$

A.2 Quality

If the private and public plan differ in the quality of health outcomes they deliver to patients, the value associated with different health outcomes needs to be reflected in capitation transfers. We assume that health outcomes (including death) for each patient $i \in I$ treated by plan p are observable and associated with realized monetary values $\widehat{v}_i(p)$. As in the case of costs, we assume that the private plan's comparative advantage is measurable with respect to a relatively small set of characteristics η_i. Given selection rule λ and transfers Π, the surpluses accruing to the public and private plans take the form

$$\mathbb{E}_\nu U_0 = \mathbb{E}_\nu \left[-\Pi + \sum_{i \in \Lambda} \widehat{c}_i(p_0) + \widehat{v}_i(p_1) - \widehat{v}_i(p_0) \Big| \lambda \right],$$

$$\mathbb{E}_\nu U_1 = \mathbb{E}_\nu \left[\Pi - \sum_{i \in \Lambda} \widehat{c}_i(p_1) \Big| \lambda \right] - K(\lambda).$$

Differences in quality of care are isomorphic to a change in the public plan's cost of care. Since we assume that health outcomes are observable, data D_0 should now include values $\widehat{v}_i(p_0)$ of patients in D_0. Health outcomes $\widehat{v}_i(p_1)$ of patients in Λ (selected by private plan p_1) should be visible to the public plan. Strategic capitation can be extended by setting

transfers:

$$\Pi(\Lambda, H_R) \equiv \sum_{i \in \Lambda} \widehat{v}_i(p_1) + \pi(\eta_i) + \Delta\pi(\eta_i, H_R)$$

where

$$\pi(\eta) \equiv \sum_{\tau \in \eta} \mu_I(\tau|\eta) \left[\frac{1}{|D_0^\tau|} \sum_{i \in D_0^\tau} \widehat{c}_i(p_0) - \widehat{v}_i(p_0) \right]$$

and $\Delta\pi(\eta, H_R)$ takes the form:

$$\Delta\pi(\eta_i, H_R, \Lambda) \equiv \frac{1}{|I^\eta|} \sum_{i \in I^\eta} s_{\tau_i} r_{\tau_i},$$

with $s_{\tau_i} \equiv \frac{\mu_\Lambda(\tau_i|\eta_i)}{\mu_I(\tau_i|\eta_i)} - 1$ and

$$r_{\tau_i} \equiv \frac{1}{|H_R^{\tau_i}|} \left[\sum_{j \in H_R^{\tau_i}} \widehat{c}_j^R(p_0) - \widehat{v}_j(p_0) \right] - \pi(\eta).$$

Online Appendix, Not for Publication

B Proofs

B.1 Proofs for Section 3

Proof of Proposition 1: We begin with point (i). Reports from plan p_0 do not affect reimbursements so that truth-telling strategy β^* is dominant. In turn, for any selection λ measurable with respect to characteristics $\eta \in E$, the private plan's expected payoffs from selection take the form

$$\mathbb{E}_\nu \left[\sum_{i \in I} \mathbf{1}_{i \in \Lambda}(\kappa(\eta_i, p_0) - \kappa(\eta_i, p_1) + e_{i,\tau_i}) \right] = \mathbb{E}_\nu \left[\sum_{i \in I} \mathbf{1}_{i \in \Lambda}(\kappa(\eta_i, p_0) - \kappa(\eta_i, p_1)) \right]$$

where we use the fact that $\mathbb{E}_{\nu, \mu_I}[e_{i,\tau_i}|\eta_i] = 0$. It follows that the optimal selection rule is indeed $\lambda(\tau) = \lambda^*(\tau) = \mathbf{1}_{\kappa(\eta,p_0) > \kappa(\eta,p_1)}$.

Let us turn to point (ii). Under environment ν_0, the public plan sets a fixed capitation payment κ_0 for every patient. For $\tau = (\eta, n, \delta)$, the private plan's conditional expectation about error shock e_{i,τ_i} is $\mathbb{E}_\nu[e_{i,\tau}|D_1, \tau] = 0$ if both types $(\eta, n, 1)$ and $(\eta, n, -1)$ have realized costs $\kappa(\eta, p_1)$ in data D_1 – this happens with probability $1/4$. In the complementary event, the private plan is informed, and $\mathbb{E}_\nu[e_{i,\tau}|D_1, \tau] = \delta \varepsilon_{\eta,n} \kappa_0 / 2$. Since the magnitude Δ of comparative advantage is less than $\kappa_0/4$, the optimal selection scheme for the private plan consists of: selecting only the relatively cheap type with characteristics (η, n) when the private plan is informed about $\varepsilon_{\eta,n}$ (proba $3/4$); and selecting type τ on the basis of comparative advantage when the private plan is uninformed about ε_n (proba $1/4$). This implies that in expectation, the efficiency loss is equal to

$$\mathbb{E}_{\nu_0}\left[S_{\max} - S^{\mathsf{sparse}} \right] = \frac{3}{8}\Delta|I|.$$

When $\Delta = 0$, the private plan only profits from selecting the relatively cheap type with characteristics (η, n). This yields an aggregate expected profit equal to $\frac{3}{16}\kappa_0|I|$. Since there is no comparative advantage, this profit comes entirely at the expense of public plan p_0. $\square$

Proof of Proposition 2: We first establish point (i). Given selected set Λ, and data D_0, D_1, private plan p_1 gets payoff

$$\mathbb{E}[U_1|\Lambda, D_0, D_1] = \sum_{i \in \Lambda} \kappa(p_0, \eta_i) - \kappa(p_1, \eta_i) + \mathbb{E}_\nu[e_{\tau_i,i}|\tau_i, D_0] - \mathbb{E}_\nu[e_{\tau_i,i}|\tau_i, D_0, D_1].$$

Using the fact that $a - |b| \le a + b \le a + |b|$, we have that

$$-\Phi + \sum_{i \in \Lambda} \kappa(p_0, \eta_i) - \kappa(p_1, \eta_i) \le \mathbb{E}[U_1|\Lambda, D_0, D_1] \le \Phi + \sum_{i \in \Lambda} \kappa(p_0, \eta_i) - \kappa(p_1, \eta_i)$$

with

$$\Phi = \mathbb{E}_\nu \left[\sum_{i \in I} |\mathbb{E}\left[e_{\tau_i,i}|D_0, D_1\right] - \mathbb{E}\left[e_{\tau_i,i}|D_0\right]| \right].$$

The private plan's optimal selection Λ must yield a higher payoff than the surplus-maximizing selection $\Lambda^* = \mathbf{1}_{\kappa(\eta, p_0) > \kappa(\eta, p_1)}$. Hence it follows that

$$\sum_{i \in \Lambda} \kappa(p_0, \eta_i) - \kappa(p_1, \eta_i) \ge -2\Phi + \sum_{i \in \Lambda^*} \kappa(p_0, \eta_i) - \kappa(p_1, \eta_i).$$

Since $|e_{i,\tau}| < c_{\max}$, it follows that

$$\mathbb{E}_\nu |\mathbb{E}\left[e_{\tau_i,i}|D_0, D_1\right] - \mathbb{E}\left[e_{\tau_i,i}|D_0\right]| \le \left(\mathbb{E}_\nu \left(\mathbb{E}\left[e_{\tau_i,i}|D_0, D_1\right] - \mathbb{E}\left[e_{\tau_i,i}|D_0\right]\right)^2\right)^{1/2} \le$$

$$\left(\mathbb{E}_\nu \left(e_{\tau_i,i} - \mathbb{E}\left[e_{\tau_i,i}|D_0\right]\right)^2\right)^{1/2} \le \left(\frac{c_{max}^2}{|D_0^{\tau_i}|}\right)^{1/2} = \frac{c_{max}}{\sqrt{|D_0^{\tau_i}|}}.$$

Hence $\Phi \le c_{\max}|I|\mathbb{E}_{\mu_I}\left[\frac{1}{\sqrt{|D_0^{\tau_i}|}}\right]$. This establishes point (i).

Point (ii) follows directly from Proposition 1(ii) and the observation that: with probability 3/4 the public plan is fully informed of $\varepsilon_{\eta,n}$, in which case the capitation rate is set precisely, and the private plan selects types efficiently; with probability 1/4, the public plan is uninformed about $\varepsilon_{\eta,n}$ and sets capitation rate $\kappa(\eta, p_0)$. In this last event, the analysis of Proposition 1(ii) applies as is. □

Proof of Proposition 3: We begin by establishing (10) and (11). We begin with (10). We have that

$$
\mathbb{E}_\nu \left[\Delta\pi(\eta, H_R, \Lambda)|D_0, D_1, \beta^*, \lambda \right] = \mathbb{E}_\nu \left[\frac{1}{|I^\eta|} \sum_{i \in I^\eta} \left(\frac{\mu_\Lambda(\tau_i|\eta)}{\mu_I(\tau_i|\eta)} - 1 \right) \mathbb{E}_\nu[e_{\tau_i,i}|D_0, D_1, \tau] \right]
$$

$$
= \mathbb{E}_\nu \left[\sum_{\tau \in \eta} \mu_I(\tau|\eta) \left(\frac{\mu_\Lambda(\tau|\eta)}{\mu_I(\tau|\eta)} - 1 \right) \mathbb{E}_\nu[e_{\tau,i}|D_0, D_1, \tau] \right]
$$

$$
= \mathbb{E}_\nu \left[\sum_{\tau \in \eta} (\mu_\Lambda(\tau|\eta) - \mu_I(\tau|\eta)) \mathbb{E}_\nu[e_{\tau,i}|D_0, D_1, \tau] \right].
$$

We turn to (11). Using the fact that λ is measurable with respect to E (allowing us to define $\lambda(\eta) = \lambda(\tau)$), and the fact that selection error φ_i is orthogonal to r_{τ_i} have that

$$
\mathbb{E}_\nu \left[|\Lambda^\eta| \Delta\pi(\eta, H_R, \Lambda)|D_0, D_1, \beta^*, \lambda \right] = \mathbb{E}_\nu \left[\sum_{\tau \in \eta} \left(|\Lambda^\tau| - \frac{|I^\tau|}{|I^\eta|}|\Lambda^\eta| \right) r_\tau \right]
$$

$$
= \mathbb{E}_\nu \left[\sum_{\tau \in \eta} \left(\lambda(\tau)|I^\tau| - \frac{|I^\tau|}{|I^\eta|}\lambda(\eta)|I^\eta| \right) r_\tau \right] = 0.
$$

We now return to the proof of Proposition 3. In equilibrium, the public plan will get a payoff at most equal to 0. By condition (10), it follows that the public plan can guarantee itself a payoff of 0 by reporting costs truthfully: $\beta = \beta^*$. Hence, it must be that in equilibrium the public plan gets a payoff of 0. Since we break indifferences towards truthtelling, it must be that the public plan reports truthfully in equilibrium. This implies that the private plan gets a payoff equal to $\mathbb{E}_\nu \left[\sum_{i \in \Lambda} \kappa(p_0, \eta_i) - \kappa(p_1, \eta_i) \right]$, which is maximized by using optimal

selection policy $\lambda = \lambda^*$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

B.2 Proofs for Section 4

Proof of Proposition 4: Given a cost distribution F, let $\kappa(\eta, p) \equiv \mathbb{E}_{\mu_I, F}[\widehat{c}_i(p)|\eta_i = \eta]$ denote the expected cost of service for plan p conditional on legitimate selection characteristic η in patient population I. We also denote by $\kappa(\tau, p) \equiv \mathbb{E}_F[\widehat{c}_i(p)|\tau_i = \tau]$ the expected cost of treatment by plan p conditional on type τ.

Given a partition E and a selection rule λ, plan p_1's expected returns are

$$
\begin{aligned}
\mathbb{E}_\nu[U_1|D_0, D_1] =& \mathbb{E}_\nu\left[\sum_{i \in I} \lambda(\tau_i)\left[\widehat{\kappa}(\eta_i, p_0) + \Delta\pi(\eta_i, H_R, \Lambda) - \kappa(\tau_i, p_1)\right] \Big| D_0, D_1\right] - K(\lambda) \\
=& \mathbb{E}_\nu\left[\sum_{i \in I} \lambda(\tau_i)\left[\kappa(\tau_i, p_0) - \kappa(\tau_i, p_1)\right] \Big| D_0, D_1\right] - K(\lambda) \\
&+ \mathbb{E}_\nu\left[\sum_{i \in I} \lambda(\tau_i)\left[\kappa(\eta_i, p_0) + \Delta\pi(\eta_i, H_R, \Lambda) - \kappa(\tau_i, p_0)\right] \Big| D_0, D_1\right] \\
&+ \mathbb{E}_\nu\left[\sum_{i \in I} \lambda(\tau_i)\left[\widehat{\kappa}(\eta_i, p_0) - \kappa(\eta_i, p_0)\right] \Big| D_0, D_1\right] \\
=& U_1^A + U_1^B + U_1^C.
\end{aligned}
$$

where U_1^A, U_1^B and U_1^C are defined as the three respective terms in the expression above. Note that $U_1^A = S(\lambda|D_0, D_1)$. The key steps of the proof are the following,

(i)   in any equilibrium (E, λ, β), $U_1^B \leq 0$;

(ii) for any reporting strategy β, if λ is measurable with respect to E, then $U_1^B = 0$;

(iii) for any E and λ,
$$
\left|\mathbb{E}_\nu[U_1^C|D_0]\right| \leq \mathbb{E}_\nu[\Psi(\mathcal{E}, e)].
$$

Let us first show that points (i), (ii) and (iii) imply properties (13), (14) and (15). We have

that under equilibrium strategies (E, λ, β),

$$\mathbb{E}_\nu[U_1|D_0, D_1] \leq S(\lambda|D_0, D_1) + \mathbb{E}_\nu[U_1^B|D_0, D_1, \lambda, \beta] + \mathbb{E}_\nu[U_1^C|D_0, D_1, \lambda, \beta]$$

$$\leq S(\lambda|D_0, D_1) + \mathbb{E}_\nu[U_1^C|D_0, D_1, \lambda, \beta].$$

In addition, from the fact that the private plan is weakly better off using (E, λ) over any strategy (E', λ') where λ' is measurable with respect to E', it follows that

$$\mathbb{E}_\nu[U_1] \geq \mathbb{E}_\nu\left[\max_{E' \in \mathcal{E}} S_{E'|D_0, D_1}\right] - \mathbb{E}_\nu[\Psi(\mathcal{E}, e)].$$

Altogether, this implies that $S(\lambda|D_0, D_1) \geq \mathbb{E}_\nu\left[\max_{E \in \mathcal{E}} S_{E|D_0, D_1}\right] - 2\mathbb{E}_\nu[\Psi(\mathcal{E}, e)]$. Condition (14) follows from the fact that truthful reporting $\beta^*(c, \tau)$ guarantees that

$$-\mathbb{E}_\nu\left[\sum_{i \in I} \lambda(\tau_i)\left[\widehat{\kappa}(\eta_i, p_0) + \Delta\pi(\eta_i, H_R, \Lambda) - \kappa(\tau_i, p_0)\right]\right] \geq \underbrace{-\mathbb{E}_\nu\left[\sum_{i \in I} \lambda(\tau_i)\left[\kappa(\eta_i, p_0) + \Delta\pi(\eta_i, H_R, \Lambda) - \kappa(\tau_i, p_0)\right]\right]}_{=0}$$

$$\underbrace{-\mathbb{E}_\nu\left[\sum_{i \in I} \lambda(\tau_i)\left[\widehat{\kappa}(\eta_i, p_0) - \kappa(\eta_i, p_0)\right]\right]}_{\geq -\mathbb{E}_\nu[\Psi(\mathcal{E}, e)]}.$$

Finally, condition (15) follows from the fact that plan p_1 can choose a selection strategy measurable with respect to E, which guaranteed p_1 positive expected payoffs.

Let us return to the proofs of points (i), (ii) and (iii) above. Point (i) follows from the fact that in equilibrium the expected transfers of p_0 to plan p_1 under equilibrium reporting strategy β must be weakly lower than under truthful reporting strategy β^*, i.e. $\mathbb{E}_\nu[\Pi|\beta] \leq \mathbb{E}_\nu[\Pi|\beta^*]$. This implies that

$$\mathbb{E}_\nu\left[\sum_{i \in I} \lambda(\tau_i)\left[\widehat{\kappa}(\eta_i, p_0) + \Delta\pi(\eta_i, H_R, \Lambda)\right]\Big|\beta\right] \leq \mathbb{E}_\nu\left[\sum_{i \in I} \lambda(\tau_i)\left[\widehat{\kappa}(\eta_i, p_0) + \Delta\pi(\eta_i, H_R, \Lambda)\right]\Big|\beta^*\right], \text{ so that}$$

$$\mathbb{E}_\nu\left[\sum_{i \in I} \lambda(\tau_i)\left[\kappa(\eta_i, p_0) + \Delta\pi(\eta_i, H_R, \Lambda) - \kappa(\tau_i, p_0)\right]\Big|\beta\right] \leq \mathbb{E}_\nu\left[\sum_{i \in I} \lambda(\tau_i)\left[\kappa(\eta_i, p_0) + \Delta\pi(\eta_i, H_R, \Lambda) - \kappa(\tau_i, p_0)\right]\Big|\beta^*\right].$$

43

Using

$$\mathbb{E}_{\nu}[\Delta\pi(\eta_i, H_R, \Lambda)|i \in \Lambda, \beta^*] = \mathbb{E}_{\nu}\left[\sum_{\tau \in \eta_i}(\mu_{\Lambda}(\tau|\eta_i) - \mu_I(\tau|\eta_i))(\kappa(\tau, p_0) - \widehat{\kappa}(\eta_i, p_0))\Big| i \in \Lambda, \beta^*\right]$$

$$= \mathbb{E}_{\nu}\left[\sum_{\tau \in \eta_i}(\mu_{\Lambda}(\tau|\eta_i) - \mu_I(\tau|\eta_i))(\kappa(\tau, p_0) - \kappa(\eta_i, p_0))\Big| i \in \Lambda, \beta^*\right]$$

$$= \mathbb{E}_{\nu}\left[\sum_{\tau \in \eta_i}\mu_{\Lambda}(\tau|\eta_i)(\kappa(\tau, p_0) - \kappa(\eta_i, p_0))\Big| i \in \Lambda, \beta^*\right]$$

and

$$\mathbb{E}_{\nu}\left[\sum_{i \in \Lambda}\kappa(\eta_i, p_0) - \kappa(\tau_i, p_0) + \Delta\pi(\eta_i, H_R, \Lambda)\right] = \mathbb{E}_{\nu}\left[\sum_{\eta \in E}|\Lambda^{\eta}|\left[\Delta\pi(\eta, H_R) + \sum_{\tau \in \eta}\mu_{\Lambda}(\tau|\eta)[\kappa(\eta, p_0) - \kappa(\tau, p_0)]\right]\right]$$

we obtain that indeed,

$$\mathbb{E}_{\nu}\left[\sum_{i \in I}\lambda(\tau_i)\left[\kappa(\eta_i, p_0) + \Delta\pi(\eta_i, H_R, \Lambda) - \kappa(\tau_i, p_0)\right]\Big|\beta^*\right] = 0,$$

and hence, for any reporting strategy β, $U_1^B \leq 0$, which yields point (i).

Point (ii) follows from the fact that whenever λ is measurable with respect to E, then for all reporting strategies β

$$\mathbb{E}_{\nu}\left[\sum_{i \in I}\lambda(\tau_i)(\kappa(\eta_i, p_0) - \kappa(\tau_i, p_0))\right] = 0$$

and

$$\mathbb{E}_{\nu}\left[\Delta\pi(\eta, H_R, \Lambda)\right] = \mathbb{E}_{\nu}\left[\sum_{\tau \in \eta}[\mu_{\Lambda}(\tau|\eta) - \mu_I(\tau|\eta)]r_{\tau}\right] = 0,$$

where $r_{\tau} = \frac{1}{|H^{\tau}|}\sum_{i \in H^{\tau}}\widehat{c}_i^R(p_0) - \widehat{\kappa}(\eta, p_0)$ denotes the mean residual of the baseline capitation formula computed in the reported hold-out sample.

Finally,

$$
\begin{aligned}
U_1^C &\leq \max_{\lambda \in [0,1]^T, E \in \mathcal{E}} \mathbb{E}_\nu \left[\sum_{i \in I} \lambda(\tau_i)(\widehat{\kappa}(\eta_i, p_0) - \kappa(\eta_i, p_0)) \right] \\
&\leq \max_{\lambda \in \mathcal{M}(E), E \in \mathcal{E}} \mathbb{E}_\nu \left[\sum_{i \in I} \lambda(\tau_i)(\widehat{\kappa}(\eta_i, p_0) - \kappa(\eta_i, p_0)) \right] \\
&\leq \max_{E \in \mathcal{E}} \mathbb{E}_\nu \left[\sum_{\eta \in E} \left[\sum_{i \in I^\eta} \widehat{\kappa}(\eta_i, p_0) - \kappa(\eta_i, p_0) \right]^+ \right],
\end{aligned}
$$

which yields point (iii). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Proof of Lemma 1: We begin with point (i) and show that $\mathbb{E}_\nu [\Psi(\mathcal{E}, e)] \leq \mathbb{E}[\Psi(\mathcal{E}, e')]$ using a coupling argument, i.e. by carefully jointly sampling original errors e and Rademacher errors e'.

Consider the following process for generating errors e and e'. Errors e are generated according to the original distribution of e_i (where the different e_i's are independent of one another). In turn, each error term e_i' is generated from e_i as follows: conditional on e_i, $e_i' \in \{-c_{max}, c_{max}\}$ is chosen so that $\mathbb{E}_\nu[e_i'|e_i] = e_i$. This is possible since $e_i \in [-c_{max}, c_{max}]$, and there is a unique such distribution. Since error terms $(e_i)_{i \in D_0}$ are independent, so are error terms $(e_i')_{i \in I}$. In addition,

$$
\mathbb{E}_\nu[e_i'] = \mathbb{E}_{e_i} \mathbb{E}_\nu[e_i'|e_i] = \mathbb{E}_{e_i} e_i = 0,
$$

which implies $e_i' \sim U\{-c_{max}, c_{max}\}$.

We now show that $\mathbb{E}_\nu [\Psi(\mathcal{E}, e)] \leq \mathbb{E}_\nu[\Psi(\mathcal{E}, e')]$. Note that $\Psi(\mathcal{E}, e)$ can be viewed as the maximum value for $S \subset E \in \mathcal{E}$ of

$$
\Sigma_S \equiv \sum_{\eta \in S} |I^\eta| \left[\sum_{\tau \in \eta} \mu_I(\tau|\eta) \frac{1}{|D_0^\tau|} \sum_{i \in D_0^\tau} e_i \right]. \text{[19]} \tag{21}
$$

45

Fix e, and assume that $\Psi(\mathcal{E}, e)$ is realized by Σ_S for some set S of η's. We have by linearity of expectation that

$$\Psi(\mathcal{E}, e) = \sum_{\eta \in S} |I^\eta| \left[\sum_{\tau \in \eta} \frac{|I^\tau|}{|I^\eta||D_0^\tau|} \sum_{i \in D_0^\tau} e_i \right] = \mathbb{E}_\nu \left[\sum_{\eta \in S} |I^\eta| \left[\sum_{\tau \in \eta} \frac{|I^\tau|}{|I^\eta||D_0^\tau|} \sum_{i \in D_0^\tau} e_i' \right] \Big| e \right]$$

$$\leq \mathbb{E}_\nu \left[\Psi(\mathcal{E}, e') | e \right].$$

Using the law of iterated expectations, this completes the proof of point (i).

We now turn to point (ii). Let $E \in \mathcal{E}$ be any partition, and let $S \subset E$ be a selection of elements in partition E. We first show that for all t,

$$\text{prob}(\Sigma_S > t) \leq \exp\left(-\frac{t^2 |D_0|}{2 c_{max}^2 \alpha^2 |I|^2} \right) \tag{22}$$

where Σ_S is defined by (21). Using Hoeffding's inequality (see Hoeffding (1963) or Cesa-Bianchi and Lugosi (2006), Lemma 2.2) we have

$$\text{prob}(\Sigma_S > t) = \exp\left[-\frac{2t^2}{\sum_{\eta \in S, \tau \in \eta} \sum_{i \in D_0^\tau} 4 c_{max}^2 \frac{|I^\tau|^2}{|D_0^\tau|^2}} \right]$$

$$\leq \exp\left[-\frac{t^2}{2 c_{max}^2 \sum_{\tau \in T} \frac{|I^\tau|^2}{|D_0^\tau|}} \right] = \exp\left[-\frac{t^2}{2 c_{max}^2 \frac{|I|^2}{|D_0|} \sum_{\tau \in T} \frac{|I^\tau|}{|D_0^\tau|} \frac{|D_0|}{|I|} \frac{|I^\tau|}{|I|}} \right]$$

$$\leq \exp\left[-\frac{t^2}{2 c_{max}^2 \frac{|I|^2}{|D_0|} \alpha} \right].$$

Since there are at most $M = \sum_{E \in \mathcal{E}} 2^{|E|} - 1$ possible non-empty sets S, this implies

$$\text{prob}(\Psi(\mathcal{E}, e) > t) \leq M \exp\left[-\frac{t^2}{2 c_{max}^2 \frac{|I|^2}{|D_0|} \alpha} \right].$$

To complete the proof, we use the fact that $\mathbb{E}_\nu[\Psi(\mathcal{E}, e)] = \int_0^{+\infty} \text{prob}(\Psi(\mathcal{E}, e) > t) dt$. Pick t_0

---

[19] Indeed, the corresponding set S will only select ηs such that $\sum_{\tau \in \eta} \mu_I(\tau | \eta) \frac{1}{|D_0^\tau|} \sum_{i \in D_0^\tau} e_i > 0$.

such that $M \exp\left[-\frac{t_0^2}{2c_{\max}^2 \frac{|I|^2}{|D_0|}\alpha}\right] = 1$, i.e. $t_0 = |I|c_{\max}\sqrt{\frac{2\alpha \log M}{|D_0|}}$. We have

$$\mathbb{E}_\nu[\Psi(\mathcal{E},e)] \leq \int_0^{t_0} \mathrm{prob}(\Psi(\mathcal{E},e) > t)\mathrm{d}t + \int_{t_0}^{+\infty} \mathrm{prob}(\Psi(\mathcal{E},e) > t)\mathrm{d}t$$

$$\leq t_0 + \int_{t_0}^{+\infty} M \exp\left[-\frac{t^2}{2c_{\max}^2 \frac{|I|^2}{|D_0|}\alpha}\right]$$

$$\leq |I|c_{\max}\sqrt{\frac{2\alpha \log M}{|D_0|}} + \frac{\sqrt{2\pi}}{2}|I|c_{\max}\sqrt{\frac{\alpha}{|D_0|}} M \exp\left[-\frac{t^2}{2c_{\max}^2 \frac{|I|^2}{|D_0|}\alpha}\right]$$

$$\leq |I|c_{\max}\sqrt{\frac{2\alpha}{|D_0|}}\left(\sqrt{\log M} + 1\right).$$

$\square$

Proof of Corollary 2: Let $\mathcal{E}^*$ be defined as

$$\mathcal{E}^* \in \arg\max_{\mathcal{E} \in \mathcal{P}(T)} \mathbb{E}_\nu[\max_{E \in \mathcal{E}} S_{E|D_0,D_1}] - 2\overline{\Psi}_{\mathcal{E}}.$$

It follows from the proof of Proposition 4 that, net of payments $p_{\mathcal{E}}$, the payoff U_1 to the private plan in equilibrium $(\mathcal{E}, E, \lambda, \beta)$ must satisfy:

$$\mathbb{E}_\nu[U_1] \leq \mathbb{E}_\nu S(\lambda|D_0, D_1), \quad \text{and}$$

$$\forall \mathcal{E}' \subset \mathcal{P}(T), \quad \mathbb{E}_\nu[U_1] \geq \mathbb{E}_\nu\left[\max_{E \in \mathcal{E}'} S_{E|D_0,D_1}\right] - 2\overline{\Psi}_{\mathcal{E}'}.$$

This implies that

$$\mathbb{E}_\nu S(\lambda|D_0, D_1) \geq \max_{\mathcal{E}' \subset \mathcal{P}(T)} \mathbb{E}_\nu\left[\max_{E \in \mathcal{E}'} S_{E|D_0,D_1}\right] - 2\overline{\Psi}_{\mathcal{E}'}.$$

$\square$

47

Proof of Proposition 5: Let E be the partition maximizing $\sum_{\eta \in E} \frac{1}{1+\sqrt{|D_0^\eta|}} \mu_I(\eta)$. We start with the following simple claim:

Claim 1 (hard to distinguish distributions). *For each integer $d \geq 0$, there exists a pair of distributions ϕ_l, ϕ_h taking finitely many values $\widehat{c} \in [0, c_{max}]$ and such that $\mathbb{E}_{\phi_l} \widehat{c} = \kappa_l^d$, $\mathbb{E}_{\phi_h} \widehat{c} = \kappa_h^d$, with $\kappa_h^d, \kappa_l^d \in [c_{max}/4, 3c_{max}/4]$, $\kappa_h^d - \kappa_l^d \geq k' c_{max}/(1 + \sqrt{d})$, and ϕ_l^d is hard to distinguish from ϕ_h^d, in the sense that their statistical distance is less than 1/4:*

$$\sup_{S \subset [0, c_{\max}]^d} \phi_l^d(S) - \phi_h^d(S) \leq 1/4, \tag{23}$$

for some universal constant $k' > 0$, where ϕ_l^d and ϕ_h^d denote the d-product measures.

We defer the proof of Claim 1 until after the proof of the proposition. The following class of environments ν lets us establish our results.

- Selection cost K is identically equal to 0.

- Cost distributions F for the public and private plans are determined as follows. Let $(b_\eta)_{\eta \in E}$ be independent Bernoulli draws over $\{0, 1\}$ such that $\text{prob}(b_\eta = 1) = 1/2$. For patients i with characteristic $\eta \in E$ costs $\widehat{c}_i(p_0)$ are independent and identically distributed according to the distribution $\phi_{b_\eta}^{|D_0^\eta|}$ described in Claim 1. Its expected value is denoted by $\kappa_0^\eta \in \left\{ \kappa_l^{|D_0^\eta|}, \kappa_h^{|D_0^\eta|} \right\}$. Note that illegitimate selection characteristics τ are independent of costs, and play no role in this construction.

- For patients with characteristic $\eta \in E$, costs $\widehat{c}_i(p_1)$ are deterministic and equal to $\kappa_1^\eta \equiv \left(\kappa_0^\eta + \kappa_l^{|D_0^\eta|} + \kappa_h^{|D_0^\eta|} \right)/3$.

- Holdout set H contains sufficient information to identify $(b_\eta)_{\eta \in E}$. Hence, we can let messages m_0 be reports of $(b_\eta)_{\eta \in E}$.

- Private plan p_1 knows $(b_\eta)_{\eta \in E}$. Hence, we can also let messages m_1 be reports of $(b_\eta)_{\eta \in E}$.

Since only legitimate characteristics η are correlated to costs, making illegitimate types τ irrelevant, we can restrict attention to suggested selection strategies $\lambda(D_0, m_1)$ taking values in $[0, 1]^E$.

Step 1: exploiting incentive compatibility to obtain monotonicity constraints.
Fix a characteristic η (there are $\mu_I(\eta)|I|$ such patients), and let $d \equiv |D_0^\eta|$. We establish minimum performance losses that any direct mechanism must accumulate with respect to patients with characteristic η. Our first step is to exploit incentive compatibility constraints to derive a monotonicity condition.

Let m_0^* and m_1^* denote the truthful reporting strategies by plans p_0 and p_1. Let $\overline{m}_0$, $\overline{m}_1$ (resp. $\underline{m}_0$, $\underline{m}_1$) denote the reporting strategies in which the two plans correctly report $(b_{\eta'})_{\eta' \neq \eta}$ but always report $b_\eta = 1$ (resp. report $b_\eta = 0$).

Fix a realization of D_0. In principle, D_0 is an argument of transfer function Π, and of suggested selection λ. We suppress this dependency for now since D_0 is held constant. We denote by $\lambda_{m_1}(\eta)$ the selection intensity of patients with characteristic η given message m_1.

We define plans p_0 and p_1's conditional expected values U_0 and U_1 given messages and a value of b_η as

$$
U_0(m_0, b_\eta | D_0) \equiv \mathbb{E}_\nu \left[-\Pi(m_0, m_1^*, \lambda_{m_1^*}) + \sum_{\eta' \in E} |I^{\eta'}| \times \kappa_0^{\eta'} \times \lambda_{m_1^*}(\eta') \Big| D_0, b_\eta \right]
$$
$$
U_1(m_1, b_\eta | D_0) \equiv \mathbb{E}_\nu \left[\Pi(m_0^*, m_1, \lambda_{m_1}) - \sum_{\eta' \in E} |I^{\eta'}| \times \kappa_1^{\eta'} \times \lambda_{m_1}(\eta') \Big| D_0, b_\eta \right].
$$

By incentive compatibility of plan p_1's messages if $b_\eta = 1$, we must have

$$
U_1(\overline{m}_1, b_\eta = 1 | D_0) \geq U_1(\underline{m}_1, b_\eta = 1 | D_0). \tag{24}
$$

Incentive compatibility of plan p_0's message when $b_\eta = 0$ implies that

$$
U_0(\underline{m}_0, b_\eta = 0 | D_0) \geq U_0(\overline{m}_0, b_\eta = 0 | D_0).
$$

Since the true value of b_η does not affect expected transfers, this inequality simplifies to

$$\mathbb{E}_\nu \left[\Pi(\underline{m}_0, \underline{m}_1, \lambda_{\underline{m}_1}) \,\big|\, D_0 \right] \leq \mathbb{E}_\nu \left[\Pi(\overline{m}_0, \underline{m}_1, \lambda_{\underline{m}_1}) \,\big|\, D_0 \right]. \tag{25}$$

Combining (24) and (25) implies

$$\mathbb{E}_\nu \left[\Pi(\overline{m}_0, \overline{m}_1, \lambda_{\overline{m}_1}) - \Pi(\underline{m}_0, \underline{m}_1, \lambda_{\underline{m}_1}) \,\big|\, D_0 \right] \tag{26}$$
$$\geq \mathbb{E}_\nu \left[\sum_{\eta' \in E} |I^{\eta'}| \times \kappa_1^{\eta'} \times \left(\lambda_{\overline{m}_1}(\eta') - \lambda_{\underline{m}_1}(\eta') \right) \,\bigg|\, D_0, b_\eta = 1 \right].$$

A symmetric argument implies

$$\mathbb{E}_\nu \left[\Pi(\overline{m}_0, \overline{m}_1, \lambda_{\overline{m}_1}) - \Pi(\underline{m}_0, \underline{m}_1, \lambda_{\underline{m}_1}) \,\big|\, D_0 \right] \tag{27}$$
$$\leq \mathbb{E}_\nu \left[\sum_{\eta' \in E} |I^{\eta'}| \times \kappa_1^{\eta'} \times \left(\lambda_{\overline{m}_1}(\eta') - \lambda_{\underline{m}_1}(\eta') \right) \,\bigg|\, D_0, b_\eta = 0 \right].$$

By construction, b_η affects the value κ_1^η but not that of $\kappa_1^{\eta'}$ for $\eta' \neq \eta$. Hence, (26) and (27) imply

$$\mathbb{E}_\nu \left[|I^\eta| \left(\frac{2\kappa_l^d + \kappa_h^d}{3} - \frac{2\kappa_h^d + \kappa_l^d}{3} \right) \times \left(\lambda_{\overline{m}_1}(\eta) - \lambda_{\underline{m}_1}(\eta) \right) \,\bigg|\, D_0 \right] \geq 0.$$

This yields the following monotonicity condition:

$$\mathbb{E}_\nu \left[\lambda_{\overline{m}_1}(\eta) \big| D_0 \right] \leq \mathbb{E}_\nu \left[\lambda_{\underline{m}_1}(\eta) \big| D_0 \right]. \tag{28}$$

Step 2: establishing a lower bound on performance losses. Under first-best patient selection, the unconditional expected cost of treatment for a patient with legitimate characteristic η is

$$\mathbb{E}_\nu \left[\widehat{c}(p_0) + \lambda^*(\eta)(\widehat{c}(p_1) - \widehat{c}(p_0)) \right] = \frac{1}{2}\kappa_l^d + \frac{1}{2} \left(\frac{1}{3}\kappa_l^d + \frac{2}{3}\kappa_h^d \right) = \frac{2\kappa_l^d + \kappa_h^d}{3}. \tag{29}$$

50

Monotonicity condition (28) implies that under any selection λ induced by a direct mechanism, the expected cost of treatment for a patient with characteristic η satisfies:

$$\mathbb{E}_\nu\left[\widehat{c}(p_0) + \lambda(\eta)(\widehat{c}(p_1) - \widehat{c}(p_0))\right] = \mathbb{E}_\nu\left[\mathbb{E}_\nu\left[\widehat{c}(p_0) + \lambda(\eta)(\widehat{c}(p_1) - \widehat{c}(p_0))|D_0\right]\right]$$

$$= \mathbb{E}_\nu\left[\text{prob}(b_\eta = 0|D_0) \times \mathbb{E}_\nu\left[\lambda_{\underline{m}_1}(\eta) \times \frac{2\kappa_l^d + \kappa_h^d}{3} + (1 - \lambda_{\underline{m}_1}(\eta)) \times \kappa_l^d \Big| D_0\right]\right.$$

$$\left. + \text{prob}(b_\eta = 1|D_0) \times \mathbb{E}_\nu\left[\lambda_{\overline{m}_1}(\eta) \times \frac{2\kappa_h^d + \kappa_l^d}{3} + (1 - \lambda_{\overline{m}_1}(\eta)) \times \kappa_h^d \Big| D_0\right]\right]$$

$$= \frac{2\kappa_l^d + \kappa_h^d}{3} + \frac{\kappa_h^d - \kappa_l^d}{3} \times \mathbb{E}_\nu\Big[\text{prob}(b_\eta = 0|D_0)\mathbb{E}_\nu[\lambda_{\underline{m}_1}(\eta)|D_0]$$

$$+ \text{prob}(b_\eta = 1|D_0)\mathbb{E}_\nu[1 - \lambda_{\overline{m}_1}(\eta)|D_0]\Big]$$

$$\geq \frac{2\kappa_l^d + \kappa_h^d}{3} + \frac{\kappa_h^d - \kappa_l^d}{3} \times \mathbb{E}_\nu\Big[\text{prob}(b_\eta = 0|D_0)\mathbb{E}_\nu[\lambda_{\overline{m}_1}(\eta)|D_0]$$

$$+ \text{prob}(b_\eta = 1|D_0)\mathbb{E}_\nu[1 - \lambda_{\overline{m}_1}(\eta)|D_0]\Big]$$

$$\geq \frac{2\kappa_l^d + \kappa_h^d}{3} + \frac{\kappa_h^d - \kappa_l^d}{3} \times \mathbb{E}_\nu\left[\min\left\{\text{prob}(b_\eta = 0|D_0), \text{prob}(b_\eta = 1|D_0)\right\}\right]$$

Step 3: quantify the lower bound using Claim 1. Since $\text{prob}(b_\eta|D_0)$ only depends on D_0^η, it suffices to bound $\mathbb{E}_\nu\left[\min\left\{\text{prob}(b_\eta = 0|D_0^\eta), \text{prob}(b_\eta = 1|D_0^\eta)\right\}\right]$.

Recalling that $d = |D_0^\eta|$, Bayes rule and Claim 1 imply that

$$\mathbb{E}_\nu\left[\min\left\{\text{prob}(b_\eta = 0|D_0^\eta), \text{prob}(b_\eta = 1|D_0^\eta)\right\}\right] =$$

$$\mathbb{E}_\nu\min\left(\frac{\phi_l^d(D_0^\eta)}{(\phi_l^d + \phi_h^d)(D_0^\eta)}, \frac{\phi_h^d(D_0^\eta)}{(\phi_l^d + \phi_h^d)(D_0^\eta)}\right) =$$

$$\mathbb{E}_\nu\left[\frac{1}{2} - \frac{|\phi_l^d(D_0^\eta) - \phi_h^d(D_0^\eta)|}{2(\phi_l^d + \phi_h^d)(D_0^\eta)}\right] = \frac{1}{2} - \sum_{D_0^\eta}\frac{|\phi_l^d(D_0^\eta) - \phi_h^d(D_0^\eta)|}{4} > \frac{1}{4}$$

where we used the fact that the statistical distance satisfies $\frac{1}{2}\sum_{D_0^\eta}|\phi_l^d(D_0^\eta) - \phi_h^d(D_0^\eta)| = \sup_{S \subset [0,c_{\max}]^d}\phi_l^d(S) - \phi_h^d(S)$.

Altogether, it follows that for patients with characteristic η, the per-patient efficiency

loss is least

$$\frac{1}{4} \frac{\kappa_h^d - \kappa_l^d}{3} \geq \frac{k'}{12} \times \frac{c_{max}}{1 + \sqrt{|D_0^{\eta}|}}.$$

Setting $k = k'/12$ completes the proof. $\qquad \square$

We now prove Claim 1.

Proof of Claim 1: Given $d \geq 1$, let $\phi_l \sim c_{\max} B_{1/2-\epsilon}$, and $\phi_h \sim c_{max} B_{1/2+\epsilon}$, where B_q denotes Bernoulli variables taking values in $\{0,1\}$, and equal to 1 with probability q. Set ϵ so that $0 < \epsilon < 1/4$ (with the relationship between ϵ and d to be specified below). The Kullback-Leibler divergence $D(\phi_l^d \| \phi_h^d)$ between ϕ_l^d and ϕ_h^d satisfies (see Cover and Thomas (2012) for a reference)

$$D(\phi_l^d \| \phi_h^d) = d \times D(\phi_l \| \phi_h) = d \times D(B_{1/2-\epsilon} \| B_{1/2+\epsilon}) = d \times O(\epsilon^2),$$

By Pinsker's Inequality (e.g. Csiszar and Körner (2011), Section 3) the statistical distance between ϕ_l^d and ϕ_h^d satisfies

$$2 \times \sup_{S \subset [0, c_{\max}]^d} \phi_l^d(S) - \phi_h^d(S) = \| \phi_l^d - \phi_h^d \|_1 \leq \sqrt{2 \times D(\phi_l^d \| \phi_h^d)} = \sqrt{d \times O(\epsilon^2)} < k_1 \times \epsilon \sqrt{d}, \quad (30)$$

where $k_1 \geq 2$ is a constant.

Choose $\epsilon = 1/(2k_1\sqrt{d}) \leq 1/4$. Claim 1 holds with $\kappa_h^d - \kappa_l^d = 2\epsilon = 1/(k_1\sqrt{d})$. Setting $k' \leq 1/k_1$ completes the proof. $\qquad \square$

B.3 Proofs for Section 6

Proof of Proposition 6: Consider the subgame following entry. For any continuation pricing equilibrium (π_{p_0}, π_{p_1}), the usual Bertrand competition argument implies that price

formulas must satisfy

$$\forall \eta, \pi_0(\eta) = \pi_1(\eta) = \kappa(\eta).$$

Given these prices, profits are determined by the plans' selection behavior $(\lambda_{p_0}, \lambda_{p_1})$. Given the selection rule $\lambda_{\neg p}$ of its competitor, plan p chooses

$$\lambda_p \in \underset{\lambda \in [0,1]^T}{\arg \max} \sum_{\tau \in T} (\kappa(\eta) - \kappa(\tau)) \mu_I(\tau) \left(\frac{1}{2} + \lambda(\tau) - \lambda_{\neg p}(\tau) \right) - K(\lambda)$$

$$= \underset{\lambda \in [0,1]^T}{\arg \max} \sum_{\tau \in T} \mu_I(\tau) \lambda(\tau) (\kappa(\eta) - \kappa(\tau)) - K(\lambda).$$

Since K is strictly convex, minimized at 0, and smooth, it follows that its gradient $\nabla K_{|0}$ at 0 is equal to 0. As a result both plans engage in the same non-zero amount of selection λ^*, and in aggregate selection has no effect on each plan's treated sample. Strict convexity of K implies that $K(\lambda^*) > 0$. The entrant gets strictly negative expected profits following entry.

It follows that the unique equilibrium involves no entry, allowing the incumbent to charge prices equal to $\bar{\pi}$. □

Proof of Proposition 7: Consider the subgame following entry. For any continuation pricing equilibrium (π_{p_0}, π_{p_1}), the usual Bertrand competition argument implies that price formulas must satisfy

$$\forall \eta, \pi_0(\eta) = \pi_1(\eta) = \kappa(\eta).$$

Assuming truthful reporting by plan $\neg p$, strategic capitation ensures that plan p does not benefit from selecting a non representative sample of types. Hence plan p's payoff is equal to

$$\sum_{\tau \in T} (\kappa(\eta) - \kappa(\tau)) \mu_I(\tau) \left(\frac{1}{2} - \lambda_{\neg p}(\tau) \right) - K(\lambda_p).$$

It is therefore optimal for plan p to set $\lambda_p = 0$ and minimize selection cost. Given this choice, it is indeed optimal for plan $\neg p$ to report its hold-out sample truthfully. □

References

ANTIC, N. (2014): "Contracting with Unknown Technologies," *unpublished paper, Princeton University.*

ATHEY, S. AND I. SEGAL (2013): "An efficient dynamic mechanism," *Econometrica,* 81, 2463–2485.

BATATA, A. (2004): "The effect of HMOs on fee-for-service health care expenditures: evidence from Medicare revisited," *Journal of health economics,* 23, 951–963.

BELLONI, A., V. CHERNOZHUKOV, I. FERNÁNDEZ-VAL, AND C. HANSEN (2013): "Program evaluation with high-dimensional data," *arXiv preprint arXiv:1311.2645.*

BELLONI, A., V. CHERNOZHUKOV, AND C. HANSEN (2014): "Inference on Treatment Effects after Selection among High-Dimensional Controls," *The Review of Economic Studies,* 81, 608–650.

BERGEMANN, D. AND K. H. SCHLAG (2008): "Pricing without priors," *Journal of the European Economic Association,* 6, 560–569.

BISIN, A. AND P. GOTTARDI (1999): "Competitive equilibria with asymmetric information," *Journal of Economic Theory,* 87, 1–48.

——— (2006): "Efficient competitive equilibria with adverse selection," *Journal of political Economy,* 114, 485–516.

BROOKS, B. (2014): "Surveying and selling: Belief and surplus extraction in auctions," .

BROWN, J., M. DUGGAN, I. KUZIEMKO, AND W. WOOLSTON (2014): "How Does Risk Selection Respond to Risk Adjustment? New Evidence from the Medicare Advantage Program," *American Economic Review,* 104, 3335–3364.

BUNDORF, M. K., J. LEVIN, AND N. MAHONEY (2012): "Pricing and Welfare in Health Plan Choice," *American Economic Review*, 102, 3214–48.

CARROLL, G. (2015): "Robustness and linear contracts," *American Economic Review*, 105, 536–63.

——— (2017): "Robustness and separation in multidimensional screening," *Econometrica*, 85, 453–488.

CENTERS FOR MEDICARE AND MEDICAID SERVICES (2016): "HHS-operated risk adjustment methodology meeting: discussion paper," Tech. rep.

CESA-BIANCHI, N. AND G. LUGOSI (2006): *Prediction, Learning, and Games*, Cambridge University Press.

CHASSANG, S. (2013): "Calibrated incentive contracts," *Econometrica*, 81, 1935–1971.

COVER, T. M. AND J. A. THOMAS (2012): *Elements of information theory*, John Wiley & Sons.

CREMER, J. AND R. P. MCLEAN (1988): "Full extraction of the surplus in Bayesian and dominant strategy auctions," *Econometrica: Journal of the Econometric Society*, 1247–1257.

CSISZAR, I. AND J. KÖRNER (2011): *Information theory: coding theorems for discrete memoryless systems*, Cambridge University Press.

D'ASPREMONT, C. AND L.-A. GÉRARD-VARET (1979): "Incentives and incomplete information," *Journal of Public economics*, 11, 25–45.

DUBEY, P. AND J. GEANAKOPLOS (2002): "Competitive pooling: Rothschild-stiglitz reconsidered," *Quarterly Journal of Economics*, 1529–1570.

ELLIS, R. P. (1998): "Creaming, skimping and dumping: provider competition on the intensive and extensive margins," *Journal of health economics*, 17, 537–555.

EVANS, M. A., G. C. POPE, J. KAUTTER, M. J. INGBER, S. FREEMAN, R. SEKAR, AND C. NEWHART (2011): "Evaluation of the CMS-HCC Risk Adjustment Model," *CfMM Services, Editor*.

FRANK, R. G., J. GLAZER, AND T. G. MCGUIRE (2000): "Measuring adverse selection in managed health care," *Journal of Health Economics*, 19, 829–854.

GLAZER, J. AND T. G. MCGUIRE (2000): "Optimal risk adjustment in markets with adverse selection: an application to managed care," *American Economic Review*, 1055–1071.

HARTLINE, J. D. AND T. ROUGHGARDEN (2008): "Optimal Mechanism Design and Money Burning," in *Symposium on Theory Of Computing (STOC)*, 75–84.

HOEFFDING, W. (1963): "Probability Inequalities for Sums of Bounded Random Variables," *Journal of the American Statistical Association*, 58, 13–30.

KAUTTER, J., G. C. POPE, M. INGBER, S. FREEMAN, L. PATTERSON, M. COHEN, AND P. KEENAN (2014): "The HHS-HCC risk adjustment model for individual and small group markets under the Affordable Care Act," *Medicare & Medicaid research review*, 4.

MADARÁSZ, K. AND A. PRAT (2014): "Screening with an Approximate Type Space," *Working Paper, London School of Economics*.

MASSART, P. AND J. PICARD (2007): *Concentration inequalities and model selection*, vol. 1896, Springer.

MELLO, M. M., S. C. STEARNS, E. C. NORTON, AND T. C. RICKETTS (2003): "Understanding biased selection in Medicare HMOs," *Health services research*, 38, 961–992.

MEZZETTI, C. (2004): "Mechanism design with interdependent valuations: Efficiency," *Econometrica*, 72, 1617–1626.

MYERSON, R. B. AND M. A. SATTERTHWAITE (1983): "Efficient mechanisms for bilateral trading," *Journal of economic theory*, 29, 265–281.

NEWHOUSE, J. P., M. PRICE, J. HUANG, J. M. MCWILLIAMS, AND J. HSU (2012): "Steps to reduce favorable risk selection in Medicare advantage largely succeeded, boding well for health insurance exchanges," *Health Affairs*, 31, 2618–2628.

POPE, G. C., J. KAUTTER, R. P. ELLIS, A. S. ASH, J. Z. AYANIAN, M. J. INGBER, J. M. LEVY, AND J. ROBST (2004): "Risk Adjustment of Medicare Capitation Payments Using the CMS-HCC Model," *Healthcare Financing Review*, 25, 119.

RIORDAN, M. H. AND D. E. SAPPINGTON (1988): "Optimal contracts with public ex post information," *Journal of Economic Theory*, 45, 189–199.

ROTHSCHILD, M. AND J. STIGLITZ (1976): "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," *The Quarterly Journal of Economics*, 90, 629–649.

SEGAL, I. (2003): "Optimal pricing mechanisms with unknown demand," *The American economic review*, 93, 509–529.

STARC, A. (2014): "Insurer pricing and consumer welfare: Evidence from medigap," *The RAND Journal of Economics*, 45, 198–220.

VAPNIK, V. (1998): *Statistical learning theory*, Wiley New York.