

Hospital Reimbursement in the Presence of Cherry Picking and Upcoding

Nicos Savva

London Business School
nsavva@london.edu

Laurens Debo, Robert A. Shumsky

Tuck School of Business - Dartmouth College
laurens.g.debo@tuck.dartmouth.edu, robert.a.shumsky@tuck.dartmouth.edu

Hospitals throughout the developed world are reimbursed on the basis of diagnosis related groups (DRGs). Under this scheme, patients are divided into clinically meaningful groups and hospitals receive a fixed fee per patient episode that is tied to the patient DRG. The fee is based on the average cost of providing care to patients who belong to the same DRG across all hospitals. This scheme, sometimes referred to as ‘yardstick competition’, provides incentives for cost reduction, as no hospital wants to operate at a higher cost than average, and can be implemented using accounting data alone. Nevertheless, if costs within a DRG are heterogeneous, this scheme may give rise to cherry-picking incentives, where providers ‘drop’ patients who are more expensive to treat than average. To address this problem regulators have tried to reduce within-DRG cost heterogeneity by expanding the number of DRG classes. In this paper we show that, even if cost heterogeneity is eliminated, such expansion will fail to completely eliminate patient cherry picking – in equilibrium, the market will bifurcate into two groups, one of which will continue to cherry pick patients and underinvest in cost reduction, while the other group treats all patients. Furthermore, we show that DRG expansion is particularly problematic if hospitals are also able to ‘upcode’ patients, i.e., intentionally assign patients to a more resource-intensive DRG than needed in order to increase income. Upcoding increases within-DRG cost heterogeneity and amplifies cherry-picking incentives. We examine potential solutions involving yardstick competition based on input statistics.

Key words: Yardstick competition, credence goods, hospital regulation, upcoding, cherry picking.

History: Revised June 7, 2022

1. Introduction

Hospitals throughout the developed world are reimbursed on the basis of diagnosis related groups (DRGs), a patient classification system first developed by the Operations Research Department of Yale University (Fetter 1991). Under the DRG classification system, hospital patients are divided into a small number of clinically meaningful groups, which can be thought of as ‘hospital products,’ such that resources and services required to treat patients within a group are as homogeneous as possible.¹ Hospitals are then reimbursed a fixed fee for every patient episode, which is set at to the

¹ The original goal of the classification system was to allow hospitals to better measure what they “produce” and serve a tool for “budgeting, cost control, and quality control.” In 1990, Professor Fetter, the inventor of the DRG

average cost of treating patients of the same DRG across other similar hospitals, after applying local adjustments, e.g., for differences in labor costs (see for example the Inpatient Prospective Payment System (IPPS) used by Centers of Medicare & Medicaid in the US (CMS.gov 2021a)).

This form of reimbursement was first introduced by CMS in 1983 and similar versions were soon adopted by the private sector and internationally (Mayes 2007, Busse et al. 2013). In contrast to the retrospective cost-based reimbursement system used in the US before 1983, this payment innovation was prospective in the sense that it separated hospital pay from the intensity of services provided, thus generating incentives for cost efficiency. Hospitals treating patients at a cost lower than the DRG fixed fee would be making a profit and inefficient hospitals would have a strong incentive to improve. Furthermore, by linking the payment of each hospital to the cost of treatment of similar patients at other hospitals, the DRG system induced a form of indirect competition between hospitals, sometimes referred to as ‘yardstick competition’ – for any DRG, no hospital would want to operate at a cost higher than the average of all other hospitals and, as a result, the average cost itself would, in equilibrium, be reduced to the efficient level (Shleifer 1985). Importantly, this scheme can be implemented through retrospective cost accounting data alone, placing a relatively small informational burden on the regulator. For these reasons, the DRG payment system has been described as “revolutionary” (Mayes 2007).

One concern with the initial implementation of the DRG system was that it was vulnerable to cherry picking (Newhouse 1996). More specifically, because DRG definitions are based on relatively coarse clinical diagnoses, there may be heterogeneity in the complexity of patient conditions within a DRG and, therefore, in the resources required to effectively treat them. To the extent that providers can observe patient complexity and predict patient profitability, they have an incentive to selectively treat those patients with lower-than-average cost (cherry picking) and avoid treating those with higher-than-average cost (lemon dropping).² Empirical evidence supports this hypothesis. For example Newhouse (1989) finds disproportionately more high-costs patients to be treated in hospitals of last resort that do not have the option to turn patients away and KC and Terwiesch (2011) find evidence that focused cardiology hospitals selectively admit “easy-to-treat” patients. In a similar vein, Shactman (2005) reports that “specialty hospitals [...] concentrate on certain DRGs and treat relatively low-severity cases within them” and Street et al. (2010), Mason et al. (2010) present evidence that UK treatment centers, set up to provide a limited number of high-volume procedures, tend to specialize in treating the less complex elective cases within a DRG compared to public National Health Service (NHS) providers. Alexander (2020) exploits a natural experiment

classification system, was honored with the Franz Edelman Award (Fetter 1991).

² Throughout the paper we refer to the activity of overly selecting low-cost patients as cherry picking and the reciprocal activity of dropping high-cost patients as lemon dropping.

to show that physician financial incentives is one mechanism that gives rise to such cherry picking behaviour.

Over the past 30 years, the DRG system has been refined, in part to alleviate cherry picking incentives. In particular, the number of DRGs has increased to better reflect patient severity (Shactman 2005). In principle, a larger number of DRG classes, with patients of higher severity being allocated to a different DRG class than patients of lower severity, should reduce cost heterogeneity within a DRG, thus reducing providers' ability to cherry pick (or lemon drop) profitable (unprofitable) patients. For example, in 1983 there were two DRGs associated with concussion – 31 for concussion with complications and comorbidities (CCs) and 32 for concussion without CCs, with payment increasing by 33% for treating patients with CC compared to treating those without (Latta and Helbing 1991). In 2020, there were three DRGs associated with concussion – 090 for concussion without CCs, 089 for concussion with CCs, and 088 for concussion with multiple CCs, with payment increasing by 16% and 64% from 090 to 089 and 088, respectively.³ Indeed, when the DRG system was first implemented in the US in 1983 there were 467 DRGs. The number in 2020 was 761 (a relative increase of 63%). A similar trend can also be observed in Europe, where between 2005 and 2011 the number of DRGs increased by 36% in Germany, 127% in the UK, and 239% in France (Busse et al. 2013).

A second issue with the DRG system is 'upcoding,' also referred to as 'DRG creep,' where providers modify patient diagnosis (e.g. by including additional CCs) or even provide unnecessary treatments in order to push the DRG assignment to one that commands a higher fee.⁴ This problem arises due to the fact that care provision is essentially a credence good, where the regulator and/or the patient are not able to assess if the diagnoses and/or treatments provided accurately reflect the needs of the patient (Dulleck and Kerschbamer 2006). The work of Jürges et al. (2013) provides an interesting instance of upcoding in neonatology. Extremely premature babies are expensive to treat and, naturally, are assigned to a higher-paying DRG compared to babies at or close to full term. Due to the approximate nature of estimating gestation term, the assignment to DRGs associated with prematurity in Germany is based on birth weight, with substantial discontinuities at specific weights (e.g. 1000g and 1500g). The authors compare the distribution of reported birth weights before the DRG system was implemented to birth weights after implementation. Under the new DRG system, the number of babies reported to have been born weighing just below the DRG thresholds for prematurity increased substantially, with a similar drop in the number of babies

³ See Table 5 <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/FY2020-FR-Table-5.zip>

⁴ For the purposes of this research we use the term 'upcoding' to refer to both the process of modifying patient diagnosis and the process of providing unnecessary treatment done for reimbursement purposes.

reported just above the threshold. The authors estimate that, in the period 2003–2010, upcoding generated additional reimbursement of 100M Euro (\$ 133M using the average 2010 exchange rate). Beyond neonatology in Germany, Silverman and Skinner (2004) find evidence that, following the implementation of DRG payments in the US, patients with respiratory conditions were more likely to be upcoded to the more generous DRG for pneumonia and more so at for-profit hospitals. Dafny (2005) find that following a change in DRG prices hospitals responded by upcoding patients to DRGs with the largest price increase, and Psaty et al. (1999) estimate the cost of upcoding in heart failure in the US to be as high as \$993M per year.

The focus of this paper is on the interaction between the two phenomena described above: Patient cherry picking and upcoding. In particular, we develop a unifying theoretical framework that extends the classic model of yardstick competition (Shleifer 1985) by explicitly modeling patient heterogeneity within DRGs and allowing for cherry picking and upcoding. The model confirms that, under the standard yardstick competition scheme applied in practice, providers will indeed have every incentive to lemon drop patients whose cost of treatment is higher than the DRG average. Furthermore, dropping patients will distort investment incentives, resulting in higher costs compared to socially optimal levels. More importantly, we show that, in the absence of upcoding, increasing the number of DRGs so that costs within DRG become homogeneous is effective in improving welfare but it does not completely eliminate cherry-picking incentives. In fact, we show that the unique equilibrium outcome of the yardstick competition game is asymmetric – the market bifurcates into two groups of providers: one group that chooses to specialize in treating patients with relatively minor needs (and drop patients with complex needs); and a second generalist group that treats everyone, irrespective of their care needs. Compared to the second group, the first group’s treatment costs are relatively high, but nevertheless both groups make a positive profit – the first loses money on treating patients but this loss is more than offset by reduced investment costs, and the second group makes a profit on treating patients but this profit is partially eroded by higher investment costs. This result may provide a complementary explanation for the recent proliferation of specialist providers, which are often found to cherry-pick less-complex patients (Shactman 2005, KC and Terwiesch 2011) – in addition to the benefits of focus that have been identified by the literature (Freeman et al. 2020), our model suggests that the emergence of such cherry-picking providers, which co-exist with efficient all-purpose providers, may be the rational response of the industry to the expansion in the number of DRG classes within a condition.

Turning to the case where providers can also upcode patients, we show that expanding the number of DRGs is much less effective. Because the practice of upcoding ‘mixes’ some low-cost patients into the high-cost category, costs within a DRG become once again heterogeneous. As a result, providers have an even stronger incentive to drop high-cost patients and, in most cases,

underinvest in cost reduction compared to socially optimal level levels. In fact, we show that in the presence of upcoding, under some conditions, increasing the number of DRGs will actually reduce total welfare.

We conclude by investigating solutions to this problem. One potential solution is to move from prospective payments (DRG-system) to retrospective (or cost-of-service) reimbursement. We show that such a move eliminates the incentives inherent in prospective payment systems to cherry pick and upcode, but at the expense of removing incentives for cost reduction. Instead, we show that there exists an alternative yardstick competition scheme, which can be implemented with currently available information based on ‘input statics,’ that can eliminate upcoding and can also reduce (and sometimes eliminate) the problem of cherry picking. More specifically, the regulator could add an additional payment that is proportional to the difference between the number of patients actually treated and the expected number of patients treated at each provider for every DRG. The expected number of patients could be calculated using appropriately adjusted benchmarks based on patient statistics drawn from all providers, as done in other regulatory schemes used in practice (e.g., the Hospital Readmissions Reduction Program (HRRP) introduced by the Centers of Medicare & Medicaid Services (CMS) in 2012, where each hospital’s observed readmission rates in a number of monitored conditions are compared against a benchmark of expected readmission rates constructed using a national panel of hospital readmissions (Chen and Savva 2018)). This new payment will be positive (a bonus) if the provider treats more patients than expected and negative (a penalty) if the provider treats fewer patients than expected. Therefore, it generates yardstick competition incentives where no provider wants to treat fewer patients than average for any DRG. As we show, it is possible to use information on the costs of other providers to set the penalties in a way that completely eliminates upcoding. It is more difficult to set the penalty sufficiently high in order to eliminate cherry picking, but we show that any positive penalty reduces the aggregate number of patients lemon dropped and, thus, improves welfare.

Finally, in a series of extensions, we show that our findings are robust to alternative model specifications. More specifically, we show that providers will not find it optimal to downcode patients (i.e., assign a lower complexity DRG than warranted based on the patient diagnosis), and examine the case where i) engaging in patient cherry picking or upcoding is itself costly; ii) transfer payments between the payer and the health care providers are not allowed; iii) there are two providers that are asymmetric in the number of patients and the proportion of high complexity patients they treat.

2. Literature Review

The use of relative benchmarks (i.e., yardstick competition) in the context of hospital reimbursement was first analyzed by Shleifer (1985) – see also Holmstrom (1982), Nalebuff and Stiglitz (1983),

Laffont and Tirole (1993) for the use of relative benchmarks in other settings. This early work demonstrates that yardstick competition is effective in inducing regulated firms (e.g., healthcare providers) to exert more effort (e.g., for cost cutting) in a setting where they are privately informed about the cost of effort but the outcomes (e.g., costs) are ex post verifiable. Sobel (1999) examines yardstick competition in a setting where in addition to ex post effort in cost reduction the providers need to also invest in ex ante cost-reducing innovation. The paper shows that yardstick competition is effective in incentivizing effort but may fail to incentivize investment. Lefouili (2015) extends this analysis to the case where there are no transfer payments. A common feature of these studies is that the quality of medical care is assumed to be exogenous and fixed. Tangerås (2009) provides an extension where, in addition to costs, quality is also endogenous and hospitals compete directly based on quality (but not costs). The paper finds that the use of yardstick competition reduces informational rents (compared to individual regulation) but it is fairly complex and thus difficult to implement. Continuing with the theme of exploring the implications of yardstick competition on dimensions other than costs, Savva et al. (2019) examine patient waiting times – a dimension of service quality that involves an externality (see also Naor (1969)). The authors show that the form of yardstick competition implemented in practice fails to incentivize wait-time reduction and propose modifications that better serve this purpose. More recently, the literature has examined yardstick competition as applied in HRRP. The program penalizes hospitals whose (risk-adjusted) readmission rates in a number of targeted conditions is higher than the national average. The research points out shortcomings with the current HRRP implementation that may lead to equilibrium outcomes where some hospitals are not incentivized to improve or an equilibrium might not exist (Zhang et al. 2016, Arifoglu et al. 2021).

In parallel to the work that examines the properties of prospective payments set through yardstick competition, another stream of literature has focused on comparing the performance of exogenously-set prospective payments against cost-based retrospective payments in settings where healthcare providers are local monopolists (Ellis and McGuire 1986, Dranove 1987, Ma 1994) or compete directly for patients on the basis of the quality offered (Pope 1989, Ellis 1998). For example, Dranove (1987) shows that if patient costs are heterogeneous, inefficient hospitals will “dump” high-cost patients, thus raising the cost of care of efficient hospitals. Ma (1994) shows that, if quality is endogenous and hospitals can dump patients, then prospective payments may provide incentives to reduce costs but at the same time, if there is cost heterogeneity, they will induce patient cherry picking which distorts investment incentives. Ellis (1998) finds that, in addition to patient dumping, prospective payments may induce hospitals to overinvest in quality for low cost patients (in order to attract more of them) or choose to underinvest in quality for high-cost patients, phenomena he dubs “creaming” and “skimming”, respectively. A general conclusion from

this literature is that measures that reduce patient cost heterogeneity, such as increasing the number of DRGs, can mitigate the adverse incentives created by prospective payments (e.g., patient dumping) – a conclusion that regulators have implemented in earnest.

In contrast to the stream of literature introduced in the first paragraph, the stream of literature discussed in the second paragraph does not focus on problems of asymmetric information on costs – it either assumes costs are fixed (e.g., Dranove 1987, Ellis 1998) or that the regulator knows the cost technology and is able to set prospective payments exogenously (e.g., Ma 1994). Our work brings together elements from the previous two streams of literature by examining how prospective payments, which due to asymmetric information on costs are set endogenously via yardstick competition, affect patient cherry picking and investment incentives.

More importantly, the literature described above assumes that the patients' medical needs are verifiable at least ex post and, therefore, the provider administers and gets paid for the treatment needed – in other words, a patient without complications or comorbidities (i.e., has minor care needs) will not receive (or be billed as having received) care intended for patients with complications and comorbidities (a major service). In this paper we relax this assumption. In this sense this work borrows ideas from the literature of credence goods, first introduced by Darby and Karni (1973) as an additional category to complement search and experience goods. Unlike search goods, where the customer (or, in our case, the payer) can assess the value of a product before purchase (e.g., a new pair of jeans), or experience goods, where the customer finds out the value of the good after purchase and consumption (e.g., a bottle of wine), in the case of credence goods the customer is not able to tell even after purchase what the value of the good was. A canonical example of a credence good is healthcare, where a customer's problem is diagnosed and treated by an expert who is better informed than the customer (see e.g., Gottschalk et al. 2020). As a result, even after the service is provided the customer is not able to judge if it was appropriate (or indeed in some cases whether it was actually provided). The literature on credence goods is excellently summarized by Dulleck and Kerschbamer (2006). A general finding is that the expert has an incentive to overtreat (provide a higher intensity treatment than needed) and overcharge (charge for a treatment of higher intensity than the one provided). Our work adds to this literature by examining how yardstick competition, which naturally generates incentives for cost reduction, interacts with the credence character of healthcare provision. Importantly, our work shows that once the credence nature of healthcare is accounted for, the conclusion that reducing cost heterogeneity (by increasing the number of DRGs) will ameliorate cherry picking problems is no longer valid. The process of overtreating/overcharging patients reintroduces cost heterogeneity, which amplifies patient cherry picking incentives.

Finally, our work is also related to recent literature in Operations Management that examines the problem of designing payment schemes for healthcare. This includes So and Tang (2000), Lee

and Zenios (2012), Jiang et al. (2012), Gupta and Mehrotra (2015), Adida et al. (2017), Zorc et al. (2017), Guo et al. (2019), Adida and Bravo (2019), Aswani et al. (2019), Jiang et al. (2020), Delana et al. (2021), Nassiri et al. (2021), amongst others. Typically, this literature does not examine the credence-good character of healthcare services or reimbursement via yardstick competition.

3. Model

We consider a setting where a Health Organization (HO) acts as a welfare-maximizing regulator and payer of healthcare provision and is responsible for $N \geq 2$ identical non-competing profit-maximizing service providers (e.g., hospitals) that provide service to a large population of patients as needed (e.g., when they fall ill). We expand on the objectives, decisions, and payments of each of the interacting parties below and present a summary of the main assumptions in Figure 1. Many modeling choices are made to ensure the model (and more importantly the results) are easily comparable to the classic model of yardstick competition by Shleifer (1985).

3.1. Patients

In the catchment area of each service provider there is a large volume of patients λ that may independently visit the service provider in case they experience a service need (e.g., a healthcare problem). The latter happens with a relatively small probability q per time period (e.g., a year), resulting in demand of λq . (In Appendix 3 we examine the case where the demand λq is asymmetric across providers.) The provider will restore the patient’s health and will generate a positive welfare for the patient which we denote by U_0 .

To focus the analysis on the actions of the providers rather than the patients, throughout this work we make two simplifying assumptions. First, that patients co-pays are zero, as they would be in some insurance plans or in national health care systems where care is free at the point of access. Second, patients do not exercise choice and visit only one default provider. While this is clearly a simplification, empirical evidence suggests that this may not be far from reality. For example, Victoor et al. (2016) provide evidence based on semi-structured interviews that “most patients tend to visit the default [hospital] without being concerned about choice.” In a similar vein, de Cruppé and Geraedts (2017) find that patients choose hospitals based on criteria such as distance “with quality information as provided by performance reports [being] of secondary importance” and Schlesinger et al. (2014) note that “health care consumers often make choices that are imperfectly informed and inconsistent with their expressed preferences.”

3.2. Service providers

Provider costs: Once a patient develops a health problem they arrive at one of the $N \geq 2$ providers and they receive service that fully resolves their health problem. For this to happen,

some of the patients (proportion $1 - h$) require a minor intervention which costs the provider on average $c_m = c$ and some (proportion h) require a major intervention which costs on average $c_M = c + \delta$, where $c, \delta > 0$. We will refer to these two groups of patients as low- and high-complexity, respectively. (In Appendix 3 we examine the case where providers are asymmetric in the proportion of high-complexity patients.) The costs assumptions reflect the fact that the service provided to low-complexity patients is a subset of the services provided to high-complexity patients.

The provider starts from historical average cost levels c_0, δ_0 and may decide to invest to reduce the average cost of treatment to some lower level c, δ . The cost of such investment is represented by $R_c(\cdot)$ and $R_\delta(\cdot)$, with $R_c(c_0) = 0$, $\frac{d}{dc}R_c(\cdot) < 0$, $\frac{d^2}{dc^2}R_c(\cdot) > 0$ and similarly for $R_\delta(\cdot)$. This formulation assumes that reducing costs is relatively cheap when costs are high but it becomes increasingly more expensive to continue doing so. It also captures positive spillovers from low- to high-complexity patients, but not vice versa, in the sense that any investment to reduce the cost of the minor component of the treatment c will reduce the cost of both patients needing the minor and major intervention (c_m and c_M , respectively) but any intervention that reduces the cost of the major component of the treatment δ will not have an impact on the costs of the minor treatment. Furthermore, this specification ensures that the average cost of the major treatment is always greater than the average cost of the minor treatment.

Upcoding: The provider, but not the patient or the HO, is able to determine the type of service required by the patient – the service has a credence character. Although the provider may not undertreat patients (i.e., may not offer the minor service to a complex patient), it may upcode a proportion of patients $0 \leq \alpha \leq \bar{\alpha}$ (e.g., as described in the case of coding premature babies in Germany, presented in the introduction). The assumption that undertreatment is not possible is consistent with the fact that providers may be held financially and criminally responsible if patients are systematically discharged without being offered the appropriate level of treatment and it is not consistent with the Hippocratic Oath. In the credence goods literature it is sometimes referred to as the “liability rule” (Dulleck and Kerschbamer 2006). The cost associated with low-complexity patients who are upcoded is on average $c'_m = c_m + \beta(c_M - c_m) = c + \beta\delta$, where $0 < \beta < 1$. This formulation for upcoding nests two special cases: i) overbilling ($\beta \rightarrow 0$), where upcoded patients actually receive the minor treatment or a treatment very close to it in terms of costs ; ii) overtreatment ($0 < \beta < 1$), where at least some elements of the major treatment are offered to low-complexity patients. The fact that the average cost of overtreatment is lower than the average cost of the major treatment (i.e., $\beta < 1$) reflects the fact that upcoded patients are relatively healthier and are offered only a subset of the more expensive major treatment, or may be less time-consuming to treat and experience fewer complications. For the main analysis we assume that providers do

not engage in ‘downcoding’ (i.e., do not code patients that receive the major intervention as having received the minor intervention for reimbursement purposes) – as we show in Appendix A2.1 this is without loss of generality as under reasonable conditions downcoding is not optimal.

Cherry picking and lemon dropping: The provider will treat all low-complexity patients but may select not to treat a fraction $0 \leq \gamma \leq \bar{\gamma}$ of the complex patients. If a high-complexity patient is denied service, they will have to seek service outside the system – for example in taxpayer supported safety-net hospitals or academic medical centers at a higher average cost $c_{out} > c_0 + \delta_0$ covered by the HO. The assumption that providers will not deny service to low-complexity patients is consistent with practice – providers would find it difficult to justify denying service to patients with relatively simple needs. We will refer to the fraction γ of patients dropped by the provider as the lemon-dropping rate. Such lemon dropping could take place through several mechanisms, such as offering physicians direct or indirect financial incentives to select low-cost patients (as empirically documented in Alexander (2020)) or ex ante through the hospital’s investment decisions (e.g., specialist hospitals that choose not to invest in the staff and resources required to treat patients needing a major intervention (Shactman 2005)). The increased average cost c_{out} could be due to the fact that the ‘hospital of last resort’ operates with a less cost-efficient technology, or it could represent the direct reduction in patient welfare associated with the inconvenience incurred by the patient if they do not receive care at their hospital of first choice.⁵

We note that in the above formulation, the costs associated with treating patients (c , $c + \delta$, $c + \beta\delta$, c_{out}) are averages and the realized costs are subject to substantial (ex post) uncertainty (i.e., some patients requiring the major treatment will end up costing more than the average cost $c + \delta$ and others less). Nevertheless, providers are not able to estimate realized cost ex ante and, therefore, cannot lemon drop patients based on realized costs (i.e., they can only lemon drop patients based on complexity). In addition, despite the fact that the average cost of an upcoded patient may be lower than the average cost of providing the major treatment, the residual (ex post) cost uncertainty makes it impossible for the HO to determine with certainty that upcoding is taking place (i.e., the HO cannot rule out that a typical upcoded patient is a high-complexity patient that turned out to be cheaper than average). Furthermore, since the HO and providers are risk neutral, for the rest of the analysis we will ignore this ex post cost uncertainty and work directly with average costs.

⁵ Instead of assuming that lemon dropped patients are treated outside the system it is possible to assume that (at least) some of them receive treatment at another provider within the system. The main results of the paper remain qualitatively unchanged.

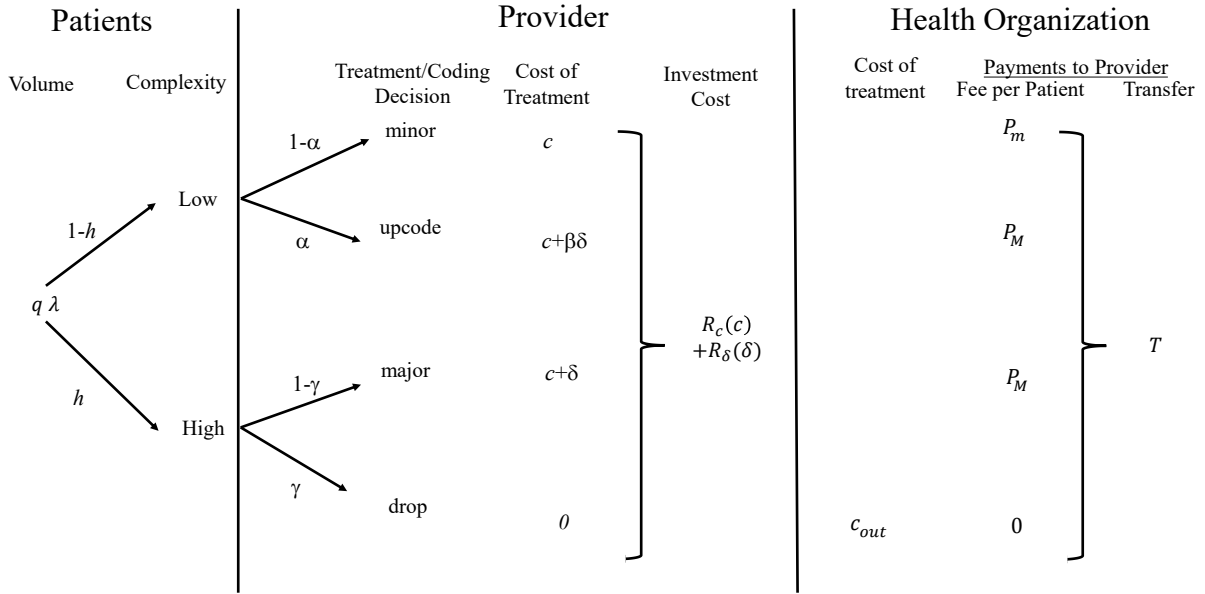


Figure 1 Main modeling assumptions: A large population λ may develop a medical need with probability q and seek care at one of N providers. A proportion h of patients is of high complexity and may be given the major treatment at an average cost $c + \delta$ or may be 'dropped' in which case they will receive treatment outside the system at an average cost c_{out} paid for by the HO. A proportion $1 - h$ of patients is of low complexity and may be given the minor treatment at an average cost of c or may be 'upcoded' to the major treatment at an average cost of $c + \beta\delta$, where $0 < \beta < 1$. Providers need to decide on the cost level c and δ they want to operate at by investing $R(c)$ and $R(\delta)$, respectively, and the rate of patient upcoding α (up to a limit $\bar{\alpha}$) and dropping γ (up to a limit $\bar{\gamma}$). The HO decides on reimbursement rates for minor and major treatments (P_m and P_M , respectively) and transfer payment (T) to be paid to the providers.

Provider Payments and Objective: Providers receive payments p_M and p_m from the HO for each treatment provided in a period (typically a year). The payments may depend on the type of treatment, major or minor respectively, but not on the type of patient (low- or high-complexity) as the former is observable by the HO but not the latter. Providers may also receive a direct transfer payment $T \geq 0$ from the HO. In Appendix A2.4 we extend the model to the case where transfer payments are not allowed. Providers are assumed to be profit maximizers and their profit is given by:

$$\begin{aligned} \pi(c, \delta, \alpha, \gamma) = & T + \lambda q ((h(1 - \gamma) + (1 - h)\alpha) p_M + (1 - h)(1 - \alpha) p_m) \\ & - \lambda q ((h(1 - \gamma) + (1 - h)\alpha\beta)\delta + (1 - h\gamma)c) - R_c(c) - R_\delta(\delta). \end{aligned} \quad (1)$$

The first line represents the provider's revenues, which include a transfer payment and fee for every treatment provided (including upcoding) and the second line represents the provider's costs, including the cost of providing treatments and the investment in cost reduction. We note that, in the preceding discussion, we have effectively assumed that lemon dropping and upcoding are costless up to an exogenous threshold ($\bar{\gamma}$ and $\bar{\alpha}$, respectively) and infinitely expensive after this

limit. In Appendix A2.3 we extend the model by presenting a continuously increasing cost (or, equivalently, risk of detection) in the proportion of upcoding or lemon dropping.

3.3. The Health Organization

The HO, such as the CMS in the US and the NHS in the UK, acts as the regulator of and payer for healthcare provision. We make the usual assumption in this literature that the HO's objective is to maximize total welfare subject to the constraint that providers break even ($\pi \geq 0$). Total welfare is the sum of the patient welfare and the provider profit within the regulated period (e.g., a year). As typical in such models, we assume that all within system payments (i.e., the transfer payments T , the per patient fees p_M and p_m) can be collected frictionlessly and therefore do not affect welfare directly. Of course, they affect welfare indirectly as they may influence provider actions.

The HO's objective function (for each provider) is given by:

$$U(c, \delta, \alpha, \gamma) = \lambda q U_0 - [(h(1 - \gamma) + (1 - h)\alpha\beta)\delta + (1 - h\gamma)c] \lambda q - R_c(c) - R_\delta(\delta) - \gamma h \lambda q c_{out}, \quad (2)$$

where U_0 is the patient welfare generated by having their health problem resolved irrespective of where the treatment is provider or of the complexity of the treatment. The next terms represent the provider cost (including both the variable cost and the investment cost) and the final term represents the additional cost (or the direct patient welfare loss) associated with providing treatment to lemon-dropped patients. We note that more general formulations, where the patient welfare lost due to lemon dropping is not linear in the lemon-dropping rate (γ), or where the patient welfare is a decreasing function of the upcoding rate (α) would generate similar results. The total welfare over all N providers is simply the sum of (2) over all providers.

3.4. HO's (first-best) solution

The first-best solution, which maximizes the HO's objective subject to the constraint that the providers break even, is given by the following equations, which derive from the problem's first order conditions:

$$\gamma^* = \alpha^* = 0, -\frac{d}{dc} R_c(c^*) = \lambda q, -\frac{d}{d\delta} R_\delta(\delta^*) = \lambda q h,$$

and any combination of payments (T, p_M, p_m) that satisfies $T \geq R_c(c^*) + R_\delta(\delta^*) - \lambda q(h(p_M - c^* - \delta^*) + (1 - h)(p_m - c^*))$. The second-order conditions $\frac{d^2}{dc^2} R_c() > 0$, $\frac{d^2}{d\delta^2} R_\delta() > 0$ ensure that the first order conditions are necessary and sufficient to characterize the optimal solution. For the rest of this work we assume that they are both satisfied and we focus on the more interesting case where the initial costs satisfy $c_0 > c^*$ and $\delta_0 > \delta^*$. We also assume that the HO will choose the minimum transfer payment that satisfies the providers' participation constraint ($\pi = 0$).

The solution makes intuitive sense. Since lemon-dropping patients generates additional costs ($c_{out} > c_0 + \delta_0$) the HO finds it optimal not to do it ($\gamma^* = 0$). Similarly, since upcoding patients increases costs ($\beta > 0$) without conferring any benefit to the patient, the HO finds it optimal not to upcode ($\alpha^* = 0$). The last two conditions determine the optimal costs of providing care – the HO finds it optimal to reduce the cost of care c such that the marginal benefit of reducing the cost by Δc for the λq patients requiring treatment is equal to the investment cost associated with such a reduction ($-\frac{d}{dc}R_c(c)$), and similarly for the proportion $h\lambda q$ that require major treatment. Note that, as consequence of the assumption that $\frac{d^2}{di^2}R_i(.) > 0$, investment in cost reduction exhibit economies of scale – the more patients a provider treats the first best solution shifts to lower costs. This is consistent with empirical evidence in this industry (Freeman et al. 2020).

Clearly, for the HO to implement this solution they need to know the cost functions $R_c(.)$, $R_\delta(.)$, the disease incident rate (q) and the proportion of patients that require major treatment (h). Given the complex and ever-changing nature of healthcare costs and medical technology, this is a tall order. For the rest of the paper, we investigate payment schemes that do not place this unrealistic informational burden on the HO.

We note that, when the costs of treating patients are ex ante homogeneous (i.e., if $\delta_0 = 0$), or if it is impossible to lemon drop or upcode patients (i.e., if $\bar{\gamma} = \bar{\alpha} = 0$) then the model reduces to one equivalent to Shleifer (1985) albeit with the difference that we assume that patient demand is exogenous to prices. This is a simplification that is realistic in the context of national health systems (e.g., Medicare in the US, NHS in the UK).

4. Cost-based yardstick competition

Since implementing the first best solution directly places an unreasonable informational burden on the HO, in this section we will investigate the effectiveness of yardstick competition, a regulatory scheme that has been implemented in practice as it requires information more readily available to the HO (Shleifer 1985). Throughout this section we will not assume that the HO knows the provider's cost functions ($R_c(.)$, $R_\delta(.)$) or anything relating to the disease (e.g., q or h). Instead, we will assume that the HO has access to patient-level ex post costing data, which are audited for accuracy, and is able to accurately estimate the realized average cost of treating patients within a DRG at least at the provider level. This is indeed the case in the UK (see NHS Digital (2021)) where the average cost data per DRG are publicly available (e.g., see (Freeman et al. 2020)) and in the US as described in CMS.gov (2021a)). The regulator can use this cost information when setting provider payments.

To make this more exact, we need to amend the notation by adding the subscript i on each provider's decision variable (costs c_i and δ_i , dropping rate γ_i , and upcoding rate α_i). We need to

distinguish two possible cases. In the first, the HO covers the costs of provider i by paying a single fee per patient episode (p_i) irrespective of the treatment provider (i.e., $p_i = p_{Mi} = p_{mi}$) – this is equivalent to assuming that there is only one coarse DRG associated with the condition. In this case, the price p_i is set at the observed average cost per episode at all other providers, which is given by⁶

$$\bar{c}_i := \frac{1}{N-1} \sum_{j \neq i} \left[c_j + \delta_j \frac{h(1-\gamma_j) + (1-h)\alpha_j\beta}{1-h\gamma_j} \right]. \quad (3)$$

In the second case, the HO breaks the condition into two distinct DRGs – one for the minor and another for the major treatment and covers the providers' costs by paying two distinct fees, one for each DRGs. In this case, the fee per patient episode are given by $p_{Mi} = \bar{c}_{Mi}$, $p_{mi} = \bar{c}_{mi}$ where \bar{c}_{Mi} and \bar{c}_{mi} are the average costs of providing the major and minor treatments at all other providers given by

$$\bar{c}_{Mi} := \frac{1}{N-1} \sum_{j \neq i} \left[c_j + \delta_j \frac{h(1-\gamma_j) + (1-h)\alpha_j\beta}{h(1-\gamma_j) + (1-h)\alpha_j} \right] \text{ and } \bar{c}_{mi} := \frac{1}{N-1} \sum_{j \neq i} c_j, \quad (4)$$

respectively. In both cases, the realized average investment cost incurred by all other providers is given by:

$$\bar{R}_i := \frac{1}{N-1} \sum_{j \neq i} [R_c(c_j) + R_\delta(\delta_j)] \quad (5)$$

and the providers are paid a transfer payment equal to $T_i = \bar{R}_i$.

Note that, in order to implement the payments defined above, the HO must credibly commit to setting payments equal to the average costs realized by the other service providers. Furthermore, since providers are identical in our setting the average cost of other providers should be a good proxy of the cost of provider i . If, however, providers are heterogeneous in observable and exogenous dimensions (e.g., labor costs, teaching status, patient demographic characteristics) these costs could be adjusted accordingly (see (Shleifer 1985) for a theoretical treatment and (CMS.gov 2021a) for an description on how this is done in practice).

Given these terms, each provider decides on the cost-reduction investment (which determines the costs c_i and δ_i they will operate at) and whether they want to engage in any patient upcoding α_i or lemon dropping γ_i . Note that due to the relative benchmarking associated with yardstick competition, each provider's payoff depends on the actions of other providers through the fees p_i and transfer payment T_i . Therefore, even though providers do not engage in direct competition, the reimbursement mechanism described above forces them to engage in a simultaneous-move game. We will characterize the equilibrium outcome of this game for different cases.

⁶ To help the reader understand this formula, note that the cost of providing the minor service to patients who are not upcoded at a provider j is $\lambda q(1-h)(1-\alpha_j)c_j$, the cost of providing the major service to patients who are not lemon dropped is $\lambda q h(1-\gamma_j)(c_j + \delta_j)$ and the cost of upcoding α_j minor patients is $\lambda q(1-h)\alpha_j(c_j + \beta\delta_j)$. The total volume of patients treated, is $\lambda q(1-h\gamma)$. The sum of the first three terms divided by the fourth term gives the summand.

4.1. In the absence of upcoding and cherry picking.

We begin the analysis by examining the simpler case where patient upcoding is not possible (e.g., because patient needs are observable and verifiable by the HO thus alleviating any credence goods concerns) and that providers are not able to select patients (e.g., because all hospitals have to treat every patient). This case is equivalent to setting $\bar{\gamma} = \bar{\alpha} = 0$.

Proposition 1 *Under yardstick competition, and irrespective of the number of DRGs used (one or two), in the absence of patient lemon dropping and upcoding ($\bar{\gamma} = \bar{\alpha} = 0$) there exists a unique Nash equilibrium which is symmetric. All providers invest in cost reduction optimally (i.e., c^* , δ^*) and they all break even.*

The proposition above demonstrates the power of yardstick competition as first described by Shleifer (1985). Without knowing the providers' cost structure, just by observing the providers' realized (average) costs, the HO can incentivize first-best investment in cost reduction. The relative benchmark forces the providers to engage in indirect competition – no provider wants to operate at a cost that is higher than the average of the rest of the providers – and, as a result, the average cost of treatment is set at the level the HO would have chosen. Furthermore, the HO can achieve this without surrendering rents to the providers – the prices and transfer payments are set so as to break even. More importantly, for the purposes of this work, if cherry picking and upcoding are not possible, it does not matter if the patient categories are finely or coarsely defined – having one or two DRGs in this case makes no difference. As we shall see in the rest of this section, this is no longer the case if cherry picking and/or upcoding are a possibility.

4.2. Patient cherry picking and category expansion.

To isolate the impact of cherry picking on yardstick competition, we proceed by assuming that upcoding is not possible ($\bar{\alpha} = 0$) but providers may engage in patient cherry picking if it is profitable to do so (i.e., they may choose to drop a proportion of high-complexity patients $0 \leq \gamma_i \leq \bar{\gamma}$). If the HO implements yardstick competition without differentiating between major or minor treatments (i.e., pays the providers on the basis of a single DRG, where the payment is given by $p_{Mi} = p_{mi} = \bar{c}_i$ as defined in Equation (3)), the providers engage in a simultaneous move game. Each provider's strategy is characterized by the tuple $(\gamma_i, c_i, \delta_i)$.

Before we characterize the equilibrium of this game, it is convenient to define the costs c^{e1}, δ^{e1} as the unique solutions to the equations

$$-\frac{d}{dc}R_c(c) = \lambda q(1 - h\bar{\gamma}), \quad -\frac{d}{d\delta}R_\delta(\delta) = \lambda qh(1 - \bar{\gamma}), \quad (6)$$

respectively. These are the cost levels that the HO would have chosen for a provider who engages fully in lemon dropping (i.e., given the provider's decision to drop $\bar{\gamma}$ high-complexity patients,

these cost levels maximize welfare). For this reason, we will refer to them as the *cherry-picking-best* costs.

Note that the cherry-picking-best variable costs are higher than first-best variable costs (i.e., $c^{e1} > c^*$ and $\delta^{e1} > \delta^*$) – this is a consequence of the fact that a provider that engages in lemon dropping will treat fewer patients, and, because investment in cost reduction is characterized by economies of scale, the provider will find it optimal to invest less compared to first best. Although in theory it could be possible for the cost distortions associated with lemon dropping to be so extreme that they render lemon dropping unprofitable (i.e., lemon dropping becomes unprofitable if the average cost of a major treatment at cherry-picking-best costs is greater than the cost of a major treatment at first-best cost, $c^{e1} + \delta^{e1}h \frac{1-\bar{\gamma}}{1-h\bar{\gamma}} \geq c^* + \delta^*$), for ease of exposition we will make the assumption that this is not the case. (We explore the implications of such extreme cherry-picking-best costs in Appendix A2.2.)

Proposition 2 *In the absence of upcoding ($\bar{\alpha} = 0$), if the HO implements yardstick competition based on a single DRG and cherry-picking-best costs are not too extreme (i.e., $c^{e1} + \delta^{e1}h \frac{1-\bar{\gamma}}{1-h\bar{\gamma}} < c^* + \delta^*$), then there exists a unique Nash equilibrium which is symmetric. Providers drop as many patients as possible and invest in cherry-picking-best costs (i.e., all providers choose $(\bar{\gamma}, c^{e1}, \delta^{e1})$, and $c^{e1} > c^*$, $\delta^{e1} > \delta^*$). All providers break even.*

In sharp contrast to the case in which cherry picking was not possible (see Proposition 2), when providers have the ability to select patients based on their cost of treatment, the use of yardstick competition payments with a single coarse DRG is problematic in the sense that it encourages providers to drop high-complexity patients and underinvestment in cost reduction. The crux of the problem is cost heterogeneity within a DRG. Providers are reimbursed with a single fee for treating patients, irrespective of the patients' needs. This single fee is designed to cover the average cost of providing treatment and, as a result, a provider makes a profit when treating low-complexity patients and a loss when treating high-complexity patients. This naturally generates strong cherry-picking and lemon-dropping incentives. Furthermore, the best response of a provider who knows that all other providers will drop the maximum number of patients possible is to also drop as many patients as possible – yardstick competition does nothing to ameliorate such an incentive. Nevertheless, yardstick competition still provides incentives for cost reduction – given the providers' cherry picking behavior the cost levels chosen by the providers (c^{e1}, δ^{e1}) are those that the HO would have chosen. Finally, in the symmetric equilibrium all providers receive a transfer payment (equal to the investment cost of other providers) and, thus, they break even.

We next investigate whether expanding the number of DRGs, as has been done in practice, can improve the equilibrium outcome. More specifically, the HO may use two distinct DRGs: one

associated with the minor and another with the major treatment according to $p_{Mi} = \bar{c}_{Mi}$, $p_{mi} = \bar{c}_{mi}$ as defined in Equation (4). In this case the DRG expansion is able to completely remove ex ante cost heterogeneity within DRGs.

Before we characterize the equilibrium outcome of this competition, we define the following two quantities:

$$\begin{aligned} v_1 &:= \lambda q(h(\delta^{e1} + c^{e1} - \delta^* - c^*) + (1-h)(c^{e1} - c^*)) + R_c(c^{e1}) + R_\delta(\delta^{e1}) - R_c(c^*) - R_\delta(\delta^*) \\ u_1 &:= \lambda q(h(1-\bar{\gamma})(\delta^* + c^* - \delta^{e1} - c^{e1}) + (1-h)(c^* - c^{e1})) + R_c(c^*) + R_\delta(\delta^*) - R_c(c^{e1}) - R_\delta(\delta^{e1}) \end{aligned}$$

The quantity v_1 is the profit of a provider who is paid according to yardstick competition based on two DRGs and chooses $(0, c^*, \delta^*)$ when all other providers choose $(\bar{\gamma}, c^{e1}, \delta^{e1})$ and vice versa for the quantity u_1 . Note that $v_1 > 0$ and $u_1 > 0$ (see proof of Proposition 4).

Proposition 3 *In the absence of upcoding ($\bar{\alpha} = 0$), if the HO implements yardstick competition based on two DRGs, there exists a unique Nash equilibrium which is asymmetric: $N - \theta_1$ providers drop as many patients as possible and choose cherry-picking-best costs (i.e., these providers choose $(\bar{\gamma}, c^{e1}, \delta^{e1})$ and $c^{e1} > c^*$, $\delta^{e1} > \delta^*$) and θ_1 providers do not engage in lemon dropping and invest efficiently (first-best) in cost reduction (i.e., these providers choose $(0, c^*, \delta^*)$). The number of efficient providers θ_1 is the unique integer in the interval $\left[\frac{(N-1)v_1}{v_1+u_1}, \frac{Nv_1+u_1}{v_1+u_1} \right]$. All providers receive a positive rent and total welfare is higher under two DRG compared to one DRG.*

The proposition shows that expanding the number of DRGs associated with a condition has a desirable effect. Compared to the equilibrium outcome when the HO used a single DRG, the symmetric equilibrium where all providers engage in lemon dropping and underinvest in cost reduction (see Proposition 3) disappears. Nevertheless, and quite surprisingly, eliminating cost heterogeneity is not completely effective in eliminating lemon dropping or restoring first-best investment incentives across the whole market – some providers will find it optimal to continue dropping high-complexity patients and underinvest in cost reduction.

First, we will explain why no symmetric equilibrium exists despite the fact that all providers are identical and are subject to a symmetric payment mechanism. The explanation is also a rough sketch of the proof of Proposition 4. In a symmetric equilibrium, had one existed, all providers would have chosen the same costs (c, δ) and to drop the same proportion of high-complexity patients (γ) . In contrast to the case where the HO deployed one coarse DRG definition, in this case the use of two DRGs eliminates cost heterogeneity within DRG and ensures that, in any symmetric equilibrium, providers will be paid a higher fee for treating high-complexity patients compared to low-complexity patients (i.e., $p_M > p_m$), and that these fees will cover the costs of providing

treatment for both types of patients (i.e., $p_m = c$ and $p_M = c + \delta$). The transfer payment would ensure that all providers break even. Note that, in any symmetric equilibrium candidate where providers drop patients (i.e., $\gamma > 0$), there would be underinvestment in cost reduction compared to first best (i.e., $c > c^*$ and $\delta > \delta^*$) – this is a consequence of the fact that investment in cost reduction is characterized by economies of scale. A symmetric equilibrium where providers drop patients (i.e., $\gamma > 0$, $c > c^*$, and $\delta > \delta^*$) can be ruled out, as at least one provider will find it profitable to deviate. Note that deviating to any action will not change the fees the provider is paid (these only depend on the actions of other providers). Furthermore, since treating high-complexity patients is not loss making (i.e., $p_M = c + \delta$), if a deviating provider stopped dropping patients their profit would not change. But this deviating provider is now treating more patients so they would also find it optimal to invest more in cost reduction (to benefit from economies of scale) which would in turn increase their profit. Therefore, the only symmetric equilibrium candidate that survives such deviation is one where all providers take first-best actions (i.e., all providers choose $\gamma = 0$, $c = c^*$, and $\delta = \delta^*$). Nevertheless, this is not an equilibrium outcome either because the best response of a provider that knows that everyone else will not drop patients and invest in first-best cost reduction is to drop some high-complexity patients and underinvest in cost reduction compared to first best. Again, such a deviation will not change the fees the provider is paid (as these depend on the actions of other providers only). Therefore, this provider, who will now operate at a higher cost than everyone else, will experience a loss in treating any patient that they do not drop. Nevertheless, this loss is more than offset by the lower investment cost. Since such a deviation is profitable, first-best actions by all providers cannot be sustained as a symmetric equilibrium outcome and, more generally, no symmetric equilibrium outcome exists.

Instead, what emerges is an asymmetric equilibrium where the market for providing care bifurcates into two distinct groups. The first group of ‘specialist’ providers drops the maximum number of high-complexity patients and focuses in treating low-complexity patients. As a result of treating fewer patients than first best, these providers underinvest in cost reduction compared to first best. The second group of ‘generalist’ providers treat both low- and high-complexity patients (i.e., do not drop any patients) and invest optimally in cost reduction. The source of profit for these two groups are diametrically opposed. Since the generalists have a lower treatment cost than the specialists, and since the fee per patient is equal to the average cost across the market, generalists are making a profit from treating patients and specialists are making a loss. This explains why specialists drop patients and generalists do not. In contrast, since specialists invest less in cost reduction compared to generalists, and the transfer payment is equal to the average investment cost across the market, specialists experience an investment profit and generalists an investment loss. Naturally, the more specialist providers (i.e., the smaller θ_1) the higher the fee-per-patient will be and the lower the

transfer payment – this favours the generalists who treat patients profitably and harms the specialists who draw their profit from the transfer payments. As a result, the equilibrium number of providers is reached when the profit of providers in each of the two groups are roughly equal – or to be more precise, sufficiently close so that none of the providers find it optimal to deviate to the other group.

Despite the fact that expanding the number of DRGs does not restore first-best outcomes, total welfare is clearly higher if the HO uses two DRGs instead of one. This observation follows from the fact that, in contrast to the case of one DRG where all providers engaged in lemon dropping and underinvest in cost reduction, in this case only a fraction $1 - \frac{\theta_1}{N}$ does so.

Interestingly, Proposition 3 may provide an additional and complementary explanation for the relatively recent emergence of specialized providers. Such providers are often found to cherry-pick low-complexity patients (and actively avoid treating patients with comorbidities and other complications that are expensive to treat) compared to generalist providers who treat everyone (Shactman 2005, Greenwald et al. 2006, KC and Terwiesch 2011). While benefits of focus and scope may certainly play a role in the emergence of such providers (e.g., as described in the case of the Shouldice Hospital, a specialist provider of hernia operations, the focus on treating low-complexity patients has allowed the hospital to implement process standardization that generates cost efficiency and better patient outcomes (Heskett 2003), see also Freeman et al. (2020) for more large-scale empirical evidence) the result above suggests that it may be the rational response of the market to the increase in the number of DRGs associated with a given condition. Such specialist providers in our model generate a deadweight loss for the HO (both because they underinvest in cost reduction and because patients who are refused service have to be treated at an higher cost c_{out}) but in our model they do not harm the profit of the generalist providers that do not engage in lemon dropping.

In summary, the preceding analysis only partially confirms the rationale behind the progressive increase in the number of DRGs observed in multiple health systems over the previous years – removing cost heterogeneity within DRGs improves incentives but is not a panacea.

4.3. Cost-based yardstick competition with patient cherry picking and upcoding

We proceed by examining the case where, in addition to lemon dropping, providers may also engage in patient upcoding (i.e., $0 \leq \gamma_i \leq \bar{\gamma}$, $0 \leq \alpha_i \leq \bar{\alpha}$). Clearly, upcoding is meaningless in the case where there is only one coarse DRG definition as there is no higher-paying DRG code to upcode to and the results of Proposition 3 would apply. For this reason, we will focus on the case where the HO uses two distinct DRGs. The presence of two DRGs, one of which is reimbursed at a higher rate, should generate incentives for providers to upcode. What is less clear is how upcoding interacts with cherry-picking incentives.

Each provider's strategy is characterized by the 4-tuple $(\alpha_i, \gamma_i, c_i, \delta_i)$. Before we present results, we first define the costs δ^{e2}, δ^{e3} as the unique solutions to the equations

$$-\frac{d}{d\delta}R_\delta(\delta) = \lambda q(h(1-\bar{\gamma}) + (1-h)\bar{\alpha}\beta), -\frac{d}{d\delta}R_\delta(\delta) = \lambda q(h + (1-h)\bar{\alpha}\beta),$$

respectively. The costs levels (c^{e1}, δ^{e2}) are those that the HO would have chosen (i.e., the costs that maximize welfare) for a provider who engages fully in both upcoding and lemon dropping. For this reason we will refer to these costs as 'cherry-picking-upcoding-best' (CPU-best) costs. Similarly, costs (c^*, δ^{e3}) are the cost levels that the HO would have chosen for a provider who engages fully in upcoding but not lemon dropping – we will refer to these costs as 'upcoding-best' costs. Note that $\delta^{e3} < \delta^*, \delta^{e2} < \delta^{e1}$.

With these definitions in place, we are in a position to examine the basic economics of upcoding under yardstick competition. On the one hand, if a provider chooses to upcode a patient, they will receive a fee equal to the average cost of treating major patients at other providers instead of a fee equal to the average cost of treating minor patients at other providers. Therefore, upcoding will generate an additional revenue which depends on how cost efficient other providers are – the more other providers invest in cost reduction for the major treatment, the lower this additional revenue will be. On the other hand, by upcoding a patient, a provider will incur the additional cost associated with overtreatment ($\beta\delta_i$). This additional cost depends on how cost efficient the upcoding provider is – the less they invest in cost reduction the higher this additional cost will be. It is conceivable that a provider may be so inefficient compared to other providers, so that the net effect of upcoding would be negative. Such a provider would not engage in upcoding. Since we want to focus on upcoding, we would like to rule out such scenarios. The following inequality ensures that this is the case

$$\beta < \frac{h\delta^{e3}}{h\delta^{e1} + (\delta^{e1} - \delta^{e3})(1-h)\bar{\alpha}}. \quad (7)$$

Clearly, this is always satisfied if upcoding is primarily due to overbilling as opposed to overtreatment (i.e., β is sufficiently close to 0). For the rest of this section we assume this inequality holds.

We also define the following two quantities:

$$\begin{aligned} v_2 &:= \lambda q \left((h + (1-h)\bar{\alpha}) \frac{h(1-\bar{\gamma}) + (1-h)\bar{\alpha}\beta}{h(1-\bar{\gamma}) + (1-h)\bar{\alpha}} \delta^{e2} - (h + (1-h)\bar{\alpha}\beta) \delta^{e3} + c^{e1} - c^* \right) \\ &\quad + R_c(c^{e1}) + R_\delta(\delta^{e2}) - R_c(c^*) - R_\delta(\delta^{e3}), \\ u_2 &:= \lambda q \left(\delta^{e3} (h + (1-h)\bar{\alpha}\beta) \frac{h(1-\bar{\gamma}) + (1-h)\bar{\alpha}}{h + (1-h)\bar{\alpha}} - (h(1-\bar{\gamma}) + (1-h)\bar{\alpha}\beta) \delta^{e2} + (1-\bar{\gamma}h)(c^* - c^{e1}) \right) \\ &\quad + R_c(c^*) + R_\delta(\delta^{e3}) - R_c(c^{e1}) - R_\delta(\delta^{e2}). \end{aligned}$$

The quantity v_2 is the profit of a provider who is paid according to yardstick competition with two DRGs and chooses $(\bar{\alpha}, 0, c^*, \delta^{e3})$ when all other providers choose $(\bar{\alpha}, \bar{\gamma}, c^{e1}, \delta^{e2})$ and vice versa for

the quantity u_2 . Note that in this case $u_2 > 0$, but the sign of v_2 will depend on model parameters (see Proof of Proposition 4).

Finally, we will call the CPU-best costs as ‘comparable’ to upcoding-best costs if either of the following two conditions hold:

- The upcoding-best cost is not less than the CPU-best costs (i.e., $c^* + \delta^{e3} > c^{e1} + \delta^{e2} \frac{h(1-\bar{\gamma})+(1-h)\bar{\alpha}\beta}{h(1-\bar{\gamma})+(1-h)\bar{\alpha}}$).
- The profit of a provider who is paid according to yardstick competition based on two DRGs and chooses $(\bar{\alpha}, 0, c^*, \delta^{e3})$ when all other providers choose $(\bar{\alpha}, \bar{\gamma}, c^{e1}, \delta^{e2})$ is negative (i.e., $v_2 < 0$).

Proposition 4 *If both lemon dropping and upcoding are possible, if the HO implements yardstick competition based on two DRGs, then there exists a unique Nash equilibrium:*

A. If CPU-best costs are comparable to upcoding-best costs, then the equilibrium is symmetric. Providers upcode and lemon drop as many patients as possible and invest in upcoding-cherry-picking-best costs (i.e., all providers choose $(\bar{\alpha}, \bar{\gamma}, c^{e1}, \delta^{e2})$). Furthermore, there is underinvestment in cost reduction compared to first best for the minor treatment ($c^{e1} > c^$) and if $\beta < \frac{h}{1-h}\bar{\gamma}\bar{\alpha}$ then there is also underinvestment for the major treatment ($\delta^{e2} > \delta^*$), otherwise there is overinvestment. All providers break even.*

B. Otherwise, the equilibrium is asymmetric: $N - \theta_2$ providers upcode and drop as many patients as possible and choose cherry-picking-upcoding-best costs (i.e., these providers choose $(\bar{\alpha}, \bar{\gamma}, c^{e1}, \delta^{e2})$) and θ_2 providers upcode as many patients as possible but do not engage in lemon dropping and invest in upcoding best costs (i.e., these providers choose $(\bar{\alpha}, 0, c^, \delta^{e3})$). The number θ_2 is the only integer in the interval $\left[\frac{(N-1)v_2}{u_2+v_2}, \frac{Nv_2+u_2}{u_2+v_2} \right]$. All providers receive positive rents.*

The proposition above shows that upcoding is deeply problematic. Not surprisingly, if upcoding is possible, providers will fully engage in it – yardstick competition does nothing to curtail the providers’ incentives to upcode patients. What is more interesting is the finding that, in the presence of upcoding, increasing the number of DRGs does little to restrain providers from lemon dropping. We discuss why below.

Depending on model parameters, the structure of the equilibrium outcome in the presence of upcoding (Proposition 4) is sometimes (Case A) similar to that of the symmetric equilibrium of Proposition 2 and sometimes (Case B) similar to the asymmetric equilibrium of Proposition 3.

In the first case (Case A), which emerges whenever the CPU-best costs are not too extreme, all providers engage in upcoding low-complexity patients and lemon-dropping high-complexity patients, despite the fact that the HO uses two DRGs. This result occurs because upcoding essentially *mixes* low-complexity patients into the high-cost major treatment category. As a result of this mixing, the average cost of patients treated under the major DRG is reduced to a level below the cost of treating

high-complexity patients. And since under yardstick competition the payment is, in equilibrium, equal to the average cost (see Equation (4)), the payment will not be enough to cover the cost of treating high-complexity patients (i.e., since $\beta < 1$, for any $\gamma \geq 0$ the payment for providing the major treatment $c + \delta \frac{h(1-\gamma)+(1-h)\bar{\alpha}\beta}{h(1-\gamma)+(1-h)\bar{\alpha}}$ is lower than the cost of treating high-complexity patients $c + \delta$). Therefore, dropping high-complexity patients is now profitable. In other words, upcoding reintroduces cost heterogeneity within DRGs and, therefore, reinstates the incentive for patient lemon dropping. Furthermore, the practice of lemon dropping and upcoding also distort providers' investment incentives. For the minor component of the treatment the distortion is always towards underinvestment – since in equilibrium $h\bar{\gamma}$ patients are dropped there is a corresponding reduction in the minor component of the treatment administered and, since investment is characterized by economies of scale, there will be less investment in cost reduction. For the major component of the treatment, lemon dropping and upcoding interact in a more complex manner. On the one hand, $h\bar{\gamma}$ fewer major patients will be treated leading to a reduction in activity. On the other hand, $(1-h)\bar{\alpha}$ minor patients will receive a fraction β of the major treatment, leading to an increase in activity. Naturally, if the aggregate impact is a decrease in major activity (i.e., $h\bar{\gamma} > (1-h)\bar{\alpha}\beta$) then there will be underinvestment compared to first best and vice versa. Note that, all things being equal, when upcoding is largely due to overbilling (i.e., β is sufficiently low) the equilibrium outcome will be characterised by underinvestment.

In the second case (case B) of Proposition 4, where provider costs are inflated by lemon dropping to a high degree (i.e., CPU-best costs are no longer comparable to upcoding-best costs), the symmetric outcome described above is no longer sustained in equilibrium. The best response of a provider that knows that all other providers will drop the maximum number of patients and underinvest in cost reduction will be to invest heavily in cost reduction so that they no longer make a loss when treating patients requiring the major treatment. Instead the market bifurcates into two groups of providers, as was the case when upcoding was not possible (see Proposition 3). All providers will continue to upcode low-complexity patients as much as possible, and some $(N - \theta_2)$ will also continue to lemon-drop high-complexity patients while the rest (θ_2) will find it optimal to treat everyone and invest even more than first-best in cost reduction. Compared to the first group of providers, the second group will make a profit out of treating patients, but this profit will be largely eroded by the the inflated investment in cost reduction. As in the case of Proposition 3, the number of providers in each group is set so that their profits are roughly equal.

In contrast to the case where upcoding is not possible, the results of Proposition 4 cast doubt on whether it is useful to expand the number of DRGs in the presence of upcoding. It seems that, in addition to the welfare loss associated with upcoding itself, upcoding reinstates providers' lemon-dropping incentives that the DRG expansion was meant to eliminate. We examine this more formally with the proposition below.

Proposition 5 *If both lemon dropping and upcoding are possible and if cherry-picking-best costs are not too extreme (i.e., $c^{e1} + \delta^{e1}h \frac{1-\bar{\gamma}}{1-h\bar{\gamma}} < c^* + \delta^*$) and CPU-best-costs are comparable to upcoding-best costs, then total welfare is (weakly) higher if the HO implements yardstick competition based on one DRG instead of two DRGs.*

Proposition 5 confirms that, in the presence of upcoding, the HO is better off with coarser DRG definitions, at least for the case where the cost distortions associated with cherry picking and upcoding are not too extreme (i.e., when the equilibria are symmetric). Although coarser DRG definitions do nothing to curtail cherry picking, at least they do not suffer from the problem of upcoding.

5. Numerical Example

In this section we illustrate the inefficiencies associated with lemon dropping and upcoding using a numerical example. We use publicly available data⁷ and we focus on the DRGs associated with concussion, which were briefly mentioned in the introduction. In 2020 there were three DRGs associated with concussion: 088, 089, and 090 with weights 1.3891, 0.9863, and 0.8483, and with 753, 1887, and 605 discharges, respectively. The last two DRGs (089 and 090) refer to patients with at most one CC, and since they have relatively similar weights we will treat them as the minor treatment, with a volume weighted DRG weight of 0.9469 and a total number of discharges of 2640. DRG 088 refers to patients with multiple CCs and, therefore, we will treat it as the major treatment, with 753 discharges. Assuming no cherry picking or upcoding, this suggests that the proportion of high complexity patients is $h = 18.64\%$. There are 62.6 million Medicare enrollees, which suggests that the rate of concussion per year is $q = 0.0052\%$. There are $N = 5,141$ US hospitals which serve on average $\lambda = 62,000$ patients each. Turning to costs, the Medicare base rate is \$6,555, which implies the average cost for the minor treatment is \$6,197 and the additional cost for the major treatment is \$2,894. We take these to be the first best costs under yardstick competition and assume that in the absence of yardstick competition the costs would have been m times higher (i.e., $c_0 = mc^*$ and $\delta_0 = m\delta^*$). The cost-multiplier $m > 1$ is a measure of the efficiency gains that can be achieved via yardstick competition (e.g., when $m = 1$ there is no need to implement yardstick competition). We parameterize the cost functions with $R_c(c) = \zeta_c(c - c_0)^2$ and $R_\delta(\delta) = \zeta_\delta(\delta - \delta_0)^2$ as in Savva et al. (2019). This specification implies that $\zeta_c = \frac{\lambda q}{2(c_0 - c^*)}$ and $\zeta_\delta = \frac{\lambda q h}{2(c_0 - c^*)}$. In the base case, we set $\bar{\gamma} = 0.2$, $\bar{\alpha} = 0.2$, $\beta = 0.2$, and $m = 1.5$, but we vary these parameters in the following intervals $\bar{\gamma} \in [0.01, 0.4]$, $\alpha \in [0.01, 0.25]$, $\beta \in [0.01, 0.22]$, $m \in [1.2, 4]$. The range of parameters was chosen to be as wide as possible around the base-case without violating the condition (7). For all parameters, cherry-picking-best costs are not extreme (see Proposition 2).

⁷ We use Tables 5 and 7 reported here https://www.medpac.gov/wp-content/uploads/2021/11/medpac_payment_basics_21_hospital_final_sec.pdf

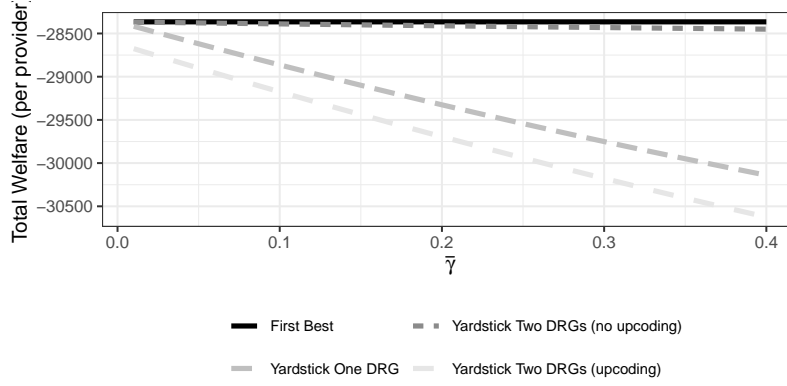


Figure 2 Numerical example: total welfare as a function of the patient maximum lemon-dropping rate $\bar{\gamma}$ under different scenarios (see figure legend). All other parameters are set to base value.

We normalize the utility from receiving treatment to zero ($U_0 = 0$), therefore the total welfare (per provider) depicted in Figures 2 and 3 is equivalent to costs. From Figure 2, it is clear that if lemon dropping is possible (i.e., if $\bar{\gamma} > 0$) using one DRG generates a welfare loss which is increasing in $\bar{\gamma}$. This loss is reduced but not eliminated if the HO uses two DRGs, but only in the absence of upcoding (i.e., if $\bar{\alpha} = 0$). In this case, most providers (95%) are generalists (i.e., treat everyone and invest optimally in cost reduction) while the rest are specialists (i.e., lemon drop high-complexity patients and underinvest in cost reduction). In the presence of upcoding, using two DRGs is counterproductive – the resulting equilibrium is symmetric and all providers upcode and, as a result, lemon drop as much as possible (i.e., $\theta_2 = 0$).

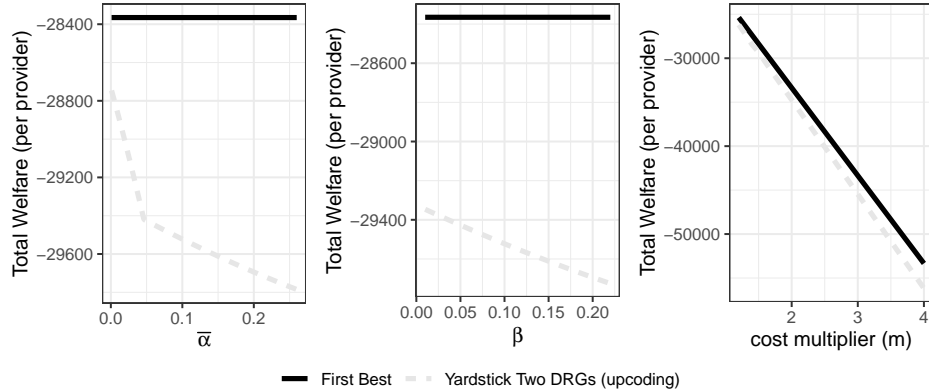


Figure 3 Numerical example: total welfare as a function of the patient maximum upcoding rate ($\bar{\alpha}$), cost of upcoding (β), and cost multiplier (m). All other parameters are set to base value.

Figure 3 confirms that the welfare lost due to upcoding and lemon dropping increases as i) the maximum proportion of patient upcoding ($\bar{\alpha}$) increases; ii) the cost of upcoding patients (β) increases; iii) the cost multiplier (m) increases. Interestingly, the slope of the total welfare with

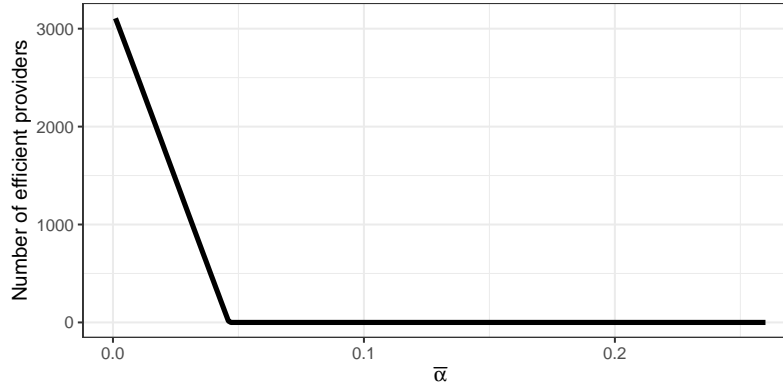


Figure 4 Numerical example: Number of efficient providers (θ_2) as a function of the patient maximum upcoding rate $\bar{\alpha}$. All other parameters are set to base value.

respect to $\bar{\alpha}$ is larger when $\bar{\alpha}$ is small and decreases at larger values of $\bar{\alpha}$. This happens because when $\bar{\alpha}$ is small, the equilibrium is asymmetric – most providers are efficient generalists that do not lemon drop patients and only a fraction of providers are inefficient specialists. As $\bar{\alpha}$ increases, welfare decreases for two reasons. First, the inefficient providers upcode more patients which reduces welfare. Second, the equilibrium number of inefficient providers increases which also reduces welfare. This is depicted in Figure 4, where we show the number of efficient generalist providers (θ_2) as a function of $\bar{\alpha}$. When $\bar{\alpha}$ is sufficiently high (about 0.047 in this example), the equilibrium becomes symmetric and all providers are inefficient. As a result, as $\bar{\alpha}$ increases further welfare reduces due to the first mechanism (i.e., inefficient providers become more inefficient) but not the second (as all providers are inefficient).

6. Potential Solution: Yardstick based on input statistics

Clearly, expanding the number of DRGs so that costs become more homogeneous within a DRG provides only a partial solution to the problem of cherry picking and only in the absence of upcoding. Upcoding reintroduces heterogeneity by mixing low-complexity patients into the DRG for the major treatment. This heterogeneity reduces the average cost and, in a yardstick competition world, also reduces the payment associated with the major treatment to a level that is below the cost of treating high-complexity patients. One obvious solution to this problem is to rigorously audit reimbursement to correct for upcoding ex post and impose penalties to deter it ex ante. This should also reduce (but not eliminate) cherry picking. A second solution is to move away from reimbursement based on average cost benchmarks (yardstick competition). For example, as we demonstrated in the Appendix A2.0, returning to a cost-of-service reimbursement model (similarly to how CMS used to reimburse hospitals before 1983), where providers are paid according to their reported costs, completely removes incentives for patient cherry picking and upcoding. However, cost-of-service reimbursement also removes incentives for cost reduction.

A more promising solution that does not increase the HO's informational burden is to implement yardstick competition based on input statistics – that is, monitor the number of patients treated for each DRG and link the providers' reimbursement to these numbers. If providers are indeed identical (or differ from each other in a number of observable features) then they should, on average, treat a similar composition of majors and minors (perhaps after controlling for observable differences in catchment size and patient composition). Any difference could be a sign of patient cherry picking or upcoding and penalties could be placed to discourage such behavior. We note that yardstick competition based on input statistics is similar in spirit to the HRRP implemented by CMS in 2012. The observed readmission rate of each hospital in a number of monitored conditions is compared to the expected readmission rate, which is estimated using data from all eligible US hospitals. Hospitals whose readmission rate is higher than the expected rate are then penalized (Chen and Savva 2018).

Using this idea, we can show that there exists a relatively simple mechanism that eliminates the problem of upcoding and could potentially also resolve the problem of cherry picking. In the case of two DRGs, this can be achieved by augmenting the cost-based yardstick competition payments ($p_{mi} = \bar{c}_{mi}, p_{Mi} = \bar{c}_{Mi}$ as defined in Equation (4)) with the following payment S_i for provider i :

$$S_i = \kappa_i(M_i - \bar{M}_i) + \phi_i(m_i - \bar{m}_i), \quad (8)$$

where M_i and m_i , are the number of major and minor patients treated by provider i , respectively, and \bar{M}_i, \bar{m}_i is the average number of major and minor patients treated by all other providers, respectively. In other words, the HO awards a bonus (penalty) to any provider that treats more (fewer) patients than expected in each DRG based on a benchmark estimated using input statistics from other providers. The parameters κ_i and ϕ_i need to be chosen such that $\kappa_i \geq 0$ and $\phi_i - \kappa_i = \bar{\delta}_i - \epsilon$, where $0 < \epsilon < \beta\delta^{e3}$ and $\bar{\delta}_i := \bar{c}_{Mi} - \bar{c}_{mi}$.

Before we characterize the equilibrium outcome of the proposed scheme, we define

$$\bar{\kappa} := \min \left\{ c^{e1} + \delta^{e1} - c^* - \delta^*, \frac{u_1}{\lambda q h \bar{\gamma}} \right\}.$$

Proposition 6 *Under the two-DRG payment scheme with input statistics described above, with $\kappa_i = \kappa$ for all i , there exists a unique Nash equilibrium:*

- *If $\kappa > \bar{\kappa}$, the equilibrium is symmetric in which all providers choose first-best actions (i.e., all providers choose $(0, 0, c^*, \delta^*)$).*
- *If $0 \leq \kappa \leq \bar{\kappa}$, the equilibrium is asymmetric. No provider engages in upcoding and $N - \theta_4$ providers drop as many patients as possible and choose cherry-picking-best costs (i.e., these providers choose $(0, \bar{\gamma}, c^{e1}, \delta^{e1})$ and $c^{e1} > c^*, \delta^{e1} > \delta^*$) and θ_4 providers do not engage in cherry*

picking and choose first-best costs (i.e., these providers choose $(0, 0, c^, \delta^*)$). The number of efficient providers is the only integer in the interval $[\frac{(N-1)(v_1 + \lambda q h \bar{\gamma} \kappa)}{v_1 + u_1}, \frac{N v_1 + u_1 + (N-1) \lambda q h \bar{\gamma} \kappa}{v_1 + u_1}]$, satisfies $\theta_4 \geq \theta_1$ and is non-decreasing in κ .*

Intuitively, the additional yardstick competition payment creates indirect competition between providers in one additional dimension – the number of patients treated. More specifically, no provider wants to treat fewer patients than average in each DRG, as doing so will trigger a penalty; conversely, every provider wants to treat more patients than average in each DRG, as doing so will trigger a bonus payment. By choosing the parameters (κ_i, ϕ_i) wisely, the HO can eliminate upcoding and reduce if not completely eliminate cherry picking.

To understand how this works, and how high these parameters need to be set, let's consider an equilibrium outcome where no provider engages in upcoding. If a provider i decided to deviate from this outcome by upcoding one low-complexity patient, then, on the negative side they would incur the additional cost associated with upcoding $(\beta \delta_i)$ and, since they would be providing one fewer minor treatment than anyone else, they would incur a penalty ϕ_i . On the positive side, they would receive the additional reimbursement associated with the major treatment $(\bar{\delta}_i = \bar{c}_{Mi} - \bar{c}_{mi})$ and, since they are providing one additional major treatment, they would also receive a bonus κ_i . So, provided that $\phi_i > \kappa_i + \bar{\delta}_i - \beta \delta_i$, then there is no incentive to upcode. The HO can always ensure that this is the case by setting $\phi_i > \kappa_i + \bar{\delta}_i - \beta \delta^{e3}$ (since any $\delta_i \geq \delta^{e3}$). Therefore the scheme proposed above eliminates incentives to upcode.

As the discussion above makes clear, by using the appropriate parameter ϕ_i the HO can eliminate upcoding. Therefore, by setting $\kappa_i = 0$ for all i the regulator can achieve the results of Proposition 4 (where upcoding was not possible), which are already better from a welfare perspective than the results of Proposition 5 (where upcoding was possible). However, the regulator can do even better by increasing κ_i . To see why, consider that, for every patient dropped by provider i then, on the one hand, they would have to pay a penalty κ_i and would forgo a fee of p_{Mi} . On the other hand, they would not incur the costs $c_i + \delta_i$. Therefore, increasing the parameter κ_i reduces the value of dropping high-complexity patients. As a result, if $\kappa_i = \kappa$ for all providers, then by increasing κ from 0 to $\bar{\kappa}$ for all providers, the asymmetric equilibrium will involve more providers who choose not to lemon drop patients and invest in first-best cost reduction (i.e., θ_4 is non-decreasing in κ). When κ exceeds the threshold $\bar{\kappa}$, all providers choose not to lemon drop and to implement first-best investment in cost reduction and the equilibrium becomes symmetric.

It is important to note that the HO can eliminate upcoding (by setting the parameter ϕ_i appropriately) without using any information not already available – relative benchmarks ensure that no provider wants to treat fewer minor patients than everyone else. Furthermore, in equilibrium,

the solution to the upcoding problem is free. No penalties or rewards would have to be paid – the threat of such penalties (promise of rewards) is enough to discourage upcoding. In contrast, to ensure that the penalty κ_i is set sufficiently high to eliminate lemon dropping, the HO would need to be able to calculate $\bar{\kappa}$, which in turn depends on the providers' cost functions $R_c(\cdot)$ and $R_\delta(\cdot)$ that the HO does not have access too. Therefore, one should treat the parameter κ_i as a lever which could be set by the HO at a tentative level – any $\kappa > 0$ improves the equilibrium outcome in the sense that more providers choose not to lemon drop compared to setting $\kappa = 0$. Furthermore, in the asymmetric equilibrium (i.e., if $\kappa \leq \bar{\kappa}$) there will be penalties applied to inefficient providers and bonuses paid to efficient providers, while in the symmetric equilibrium (i.e., if $\kappa > \bar{\kappa}$) no such penalties (or bonuses) will actually need to be paid.

A potential concern with this scheme is that it may incentivize providers to downcode patients (i.e., provide the major treatment to high-complexity patients and then code them as having received the minor treatment for reimbursement purposes). This would be the case if the difference $\phi_i - \kappa_i > \bar{\delta}_i$, as in this case the bonus associated with treating one additional minor patient is greater than the penalty associated with treating one fewer major patient. Nevertheless, this can be avoided by ensuring that $\phi_i - \kappa_i = \bar{\delta}_i - \epsilon$ where $\epsilon > 0$. If in addition $\epsilon < \beta\delta^{e3}$, then upcoding continues not to be optimal. We prove this more rigorously in Appendix A2.1.

A drawback of yardstick competition based on input statistics is that it adds complexity to an already complex system. Particularly, if hospitals are serving catchment areas of different size and with substantial case mix heterogeneity, taking the average number of patients treated in other hospitals may not be a sensible benchmark. Nevertheless, if these differences are based on observable and exogenous characteristics – for example due to differences in the catchment areas in terms of demographics, size, or prevalence of complications and comorbidities – then appropriate adjustments could be accommodated. More specifically, the benchmarks \bar{m}_i and \bar{M}_i could be estimated using information about the number of patients of all providers as well as other observable covariates that may influence the number of patients and the mixture between minor and major cases that seek care in provider i . To the extent that these covariates successfully account for heterogeneity and cannot be directly influenced by the provider (i.e., they are exogenous), then the reimbursement scheme proposed in the section would provide the right incentives (see Shleifer (1985) and Savva et al. (2019) for more formal arguments). We note that such adjustments are often used in practice in different yardstick competition schemes to account for provider heterogeneity – for example hospital benchmark costs used in IPPS are adjusted for local market conditions (CMS.gov 2021a), the benchmark readmission rate used in HRRP is adjusted for patient risk factors (CMS.gov 2021b).

7. Conclusions

The creation of DRGs is arguably one the most impactful innovations to have come out of an Operations Research group since World War II – the use of DRGs, coupled with payments based on relative benchmarks, has been credited with saving Medicare billions in the US (Fetter 1991) and has been copied extensively in Europe and elsewhere (Busse et al. 2013). Over the past 30 years, as more data and better coding practices became available, the system has been refined to increase the number of patient categories. At least in part, the motivation behind these successive refinements has been a desire to reduce patient cherry picking incentives; with a larger number of DRGs the cost heterogeneity within a DRG is reduced, leaving no ‘cherries’ for providers to pick or ‘lemons’ to drop. This work casts doubt on the effectiveness of these successive expansions in reducing cherry picking incentives. In particular, we show that increasing the number of DRGs leads to the market bifurcating into two groups of providers – one that treats patients efficiently (as proponents of increase in the number of DRGs intended) and another that drops expensive patients and underinvests in cost reduction. In addition, if providers engage in upcoding, where low-complexity patients are coded as belonging to a more expensive (and therefore higher reward) category, they effectively reintroduce cost heterogeneity within a DRG, thus exacerbating patient cherry-picking incentives. Finally, this work has proposed a potential solution based on input statistics. While the solution requires little additional information to that already collected, it will certainly increase the complexity of the reimbursement system and, if implemented, should be carefully monitored to ensure it does not induce any unintended adverse effects.

In an online Appendix we extend the main model to show that under realistic conditions downcoding is not optimal and also to allow for i) continuous increasing costs of upcoding and cherry picking; ii) the absence of a transfer payment; iii) asymmetric providers. The main conclusion continues to hold; expanding the number of DRGs to tackle the problem of cherry picking improves welfare but is less effective in the presence of patient upcoding. Moreover, we show that yardstick competition based on input statistics provides a solution in these cases as well.

References

- Adida, E., F. Bravo. 2019. Contracts for healthcare referral services: Coordination via outcome-based penalty contracts. *Management Science* **65**(3) 1322–1341.
- Adida, E., H. Mamani, S. Nassiri. 2017. Bundled payment vs. fee-for-service: Impact of payment scheme on performance. *Management Science* **63**(5) 1606–1624.
- Alexander, D. 2020. How do doctors respond to incentives? Unintended consequences of paying doctors to reduce costs. *Journal of Political Economy* doi:10.1086/710334.
- Arifoglu, K., H. Ren, T. Tezcan. 2021. Hospital Readmissions Reduction Program does not provide the right incentives: Issues and remedies. *Management Science* .

- Aswani, A., Z.-J. M. Shen, A. Siddiq. 2019. Data-driven incentive design in the medicare shared savings program. *Operations Research* **67**(4) 1002–1026.
- Busse, R., A. Geissler, A. Aaviksoo, et al. 2013. Diagnosis related groups in Europe: moving towards transparency, efficiency, and quality in hospitals? *BMJ* **346**. doi:10.1136/bmj.f3197.
- Chen, C., N. Savva. 2018. Unintended consequences of hospital regulation: The case of the hospital readmissions reduction program. *Available at SSRN 3236983* .
- CMS.gov. 2021a. Acute inpatient pps. <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS>.
- CMS.gov. 2021b. Hrrp readmission measures. <https://qualitynet.cms.gov/inpatient/measures/readmission/methodology>.
- Dafny, L.S. 2005. How do hospitals respond to price changes? *American Economic Review* **95**(5) 1525–1547.
- Darby, M.R., E. Karni. 1973. Free competition and the optimal amount of fraud. *The Journal of Law and Economics* **16**(1) 67–88.
- de Crupé, W., M. Geraedts. 2017. Hospital choice in Germany from the patient’s perspective: a cross-sectional study. *BMC health services research* **17**(1) 720.
- Delana, K., N. Savva, T. Tezcan. 2021. Proactive customer service: Operational benefits and economic frictions. *Manufacturing & Service Operations Management* **23**(1) 70–87.
- Dranove, D. 1987. Rate-setting by diagnosis related groups and hospital specialization. *The RAND Journal of Economics* **18**(3) 417–427.
- Dulleck, U., R. Kerschbamer. 2006. On doctors, mechanics, and computer specialists: The economics of credence goods. *Journal of Economic literature* **44**(1) 5–42.
- Ellis, R. P. 1998. Creaming, skimping and dumping: provider competition on the intensive and extensive margins. *Journal of Health Economics* **17**(5) 537–555.
- Ellis, R. P., T. G. McGuire. 1986. Provider behavior under prospective reimbursement: Cost sharing and supply. *Journal of Health Economics* **5**(2) 129–151.
- Fetter, R. B. 1991. Diagnosis related groups: Understanding hospital performance. *Interfaces* **21**(1) 6–26.
- Freeman, M., N. Savva, S. Scholtes. 2020. Economies of scale and scope in hospitals: An empirical study of volume spillovers. *Management Science* **67**(2) 3147–3167.
- Gottschalk, F., W. Mimra, C. Waibel. 2020. Health services as credence goods: A field experiment. *The Economic Journal* **130**(629) 1346–1383.
- Greenwald, L., J. Cromwell, W. Adamache, S. Bernard, E. Drozd, E. Root, K. Devers. 2006. Specialty versus community hospitals: referrals, quality, and community benefits. *Health Affairs* **25**(1) 106–118.
- Guo, P., C.S. Tang, Y. Wang, M. Zhao. 2019. The impact of reimbursement policy on social welfare, revisit rate, and waiting time in a public healthcare system: Fee-for-service versus bundled payment. *Manufacturing & Service Operations Management* **21**(1) 154–170.

- Gupta, D., M. Mehrotra. 2015. Bundled payments for healthcare services: Proposer selection and information sharing. *Operations Research* **63**(4) 772–788.
- Heskett, J. L. 2003. Shouldice Hospital Limited. HBS Case 9-683-068. Harvard Business School, Boston.
- Holmstrom, B. 1982. Moral hazard in teams. *The Bell Journal of Economics* **13**(2) 324–340.
- Jiang, H., Z. Pang, S. Savin. 2012. Performance-based contracts for outpatient medical services. *Manufacturing & Service Operations Management* **14**(4) 654–669.
- Jiang, H., Z. Pang, S. Savin. 2020. Performance incentives and competition in health care markets. *Production and Operations Management* **29**(5) 1145–1164.
- Jürges, H., J. Köberlein, et al. 2013. DRG upcoding in german neonatology. *Annual Conference 2013 (Duesseldorf): Competition Policy and Regulation in a Global Economic Order*. 79866, Verein für Socialpolitik/German Economic Association.
- KC, D., C. Terwiesch. 2011. The effects of focus on performance: Evidence from california hospitals. *Management Science* **57**(11) 1897–1912.
- Laffont, J. J., J. Tirole. 1993. *A Theory of Incentives in Procurement and Regulation*. MIT Press, Cambridge, MA.
- Latta, V.B., C. Helbing. 1991. Medicare short-stay hospital services by diagnosis-related groups. *Health Care Financing Review* **12**(4) 105.
- Lee, D. K. K., S. A. Zenios. 2012. An evidence-based incentive system for Medicare’s End-Stage Renal Disease program. *Management Science* **58**(6) 1092–1105.
- Lefouili, Y. 2015. Does competition spur innovation? The case of yardstick competition. *Economics Letters* **137** 135–139.
- Ma, C. A. 1994. Health care payment systems: Cost and quality incentives. *Journal of Economics & Management Strategy* **3**(1) 93–112.
- Mason, A., A. Street, R. Verzulli. 2010. Private sector treatment centres are treating less complex patients than the nhs. *Journal of the Royal Society of Medicine* **103**(8) 322–331.
- Mayes, R. 2007. The origins, development, and passage of Medicare’s revolutionary prospective payment system. *Journal of the History of Medicine and Allied Sciences* **62**(1) 21–55.
- Nalebuff, B. J., J. E. Stiglitz. 1983. Information, competition, and markets. *The American Economic Review* **73**(2) 278–283.
- Naor, P. 1969. The regulation of queue size by levying tolls. *Econometrica* **37**(1) 15–24.
- Nassiri, S., E. Adida, H. Mamani. 2021. Reference pricing for healthcare services. *Manufacturing & Service Operations Management* .
- Newhouse, J.P. 1989. Do unprofitable patients face access problems? *Health Care Financing Review* **11**(2) 33.

- Newhouse, J.P. 1996. Reimbursing health plans and health providers: Efficiency in production versus selection. *Journal of Economic Literature* **34**(3) 1236–1263.
- NHS Digital. 2021. Patient level information and costing system (plics) data collections. <https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/patient-level-information-and-costing-system-plics-data-collections>.
- Pope, G. C. 1989. Hospital nonprice competition and Medicare reimbursement policy. *Journal of Health Economics* **8**(2) 147–172.
- Psaty, B.M., R. Boineau, L.H. Kuller, R.V. Luepker. 1999. The potential costs of upcoding for heart failure in the United States. *American Journal of Cardiology* **84**(1) 108–109.
- Savva, N., T. Tezcan, Ö. Yıldız. 2019. Can yardstick competition reduce waiting times? *Management Science* **65**(7) 3196–3215.
- Schlesinger, M., D.E. Kanouse, S.C. Martino, D. Shaller, L. Rybowski. 2014. Complexity, public reporting, and choice of doctors: a look inside the blackest box of consumer behavior. *Medical Care Research and Review* **71**(5S) 38S–64S.
- Shactman, D. 2005. Specialty hospitals, ambulatory surgery centers, and general hospitals: charting a wise public policy course. *Health Affairs* **24**(3) 868–873.
- Shleifer, A. 1985. A theory of yardstick competition. *The RAND Journal of Economics* **16**(3) 319–327.
- Silverman, E., J. Skinner. 2004. Medicare upcoding and hospital ownership. *Journal of Health Economics* **23**(2) 369–389.
- So, K. C., C. S. Tang. 2000. Modeling the impact of an outcome-oriented reimbursement policy on clinic, patients, and pharmaceutical firms. *Management Science* **46**(7) 875–892.
- Sobel, J. 1999. A reexamination of yardstick competition. *Journal of Economics & Management Strategy* **8**(1) 33–60.
- Street, A., P. Sivey, A. Mason, M. Miraldo, L. Siciliani. 2010. Are English treatment centres treating less complex patients? *Health Policy* **94**(2) 150–157.
- Tangerås, T. P. 2009. Yardstick competition and quality. *Journal of Economics & Management Strategy* **18**(2) 589–613.
- Victoor, A., D. Delnoij, R. Friele, J. Rademakers. 2016. Why patients may not exercise their choice when referred for hospital care. an exploratory study based on interviews with patients. *Health Expectations* **19**(3) 667–678.
- Zhang, D.J., I. Gurvich, J.A. Van Mieghem, E. Park, R. S. Young, M. V. Williams. 2016. Hospital readmissions reduction program: An economic and operational analysis. *Management Science* **62**(11) 3351–3371.
- Zorc, S., S.E. Chick, S. Hasija. 2017. Outcomes-based reimbursement policies for chronic care pathways Working Paper No. 2017/35/DSC/TOM, INSEAD, Fontainebleau, France.

Appendix 1: Proofs of Propositions

Proof of Proposition 1: If there is no upcoding or cherry picking the profit of provider i is given by

$$\pi_i(c_i, \delta_i, 0, 0) = \lambda q [h(p_{Mi} - \delta_i - c_i) + (1 - h)(p_{mi} - c_i)] - R_c(c_i) - R_\delta(\delta_i) + T_i$$

(see Equation (1)). Since the provider's choice of c_i and δ_i does not affect the reimbursement received, and this is true irrespective of the number of DRGs used by the HO, the profit-maximizing choice of the provider is given by

$$-\frac{d}{dc}R_c(c_i) = \lambda q, \quad -\frac{d}{d\delta}R_\delta(\delta_i) = \lambda q h.$$

Any values of c_i and δ_i that satisfy these $2 \times N$ conditions are equilibria. Naturally, if all providers choose $c_i = c^*$ and $\delta_i = \delta^*$ the conditions above are identical for all providers, and in fact reduce to the first order conditions of the welfare-maximization problem. In addition, the transfer payment ensures that all providers break even. Therefore, the first-best investment decisions constitute a symmetric Nash equilibrium that achieves first-best investment in cost reduction. Furthermore, since $R_c'' > 0$ and $R_\delta'' > 0$, the symmetric equilibrium is unique and no asymmetric equilibrium exists. \square

Proof of Proposition 2: In the absence of upcoding ($\bar{\alpha} = 0$), under the yardstick competition scheme with a single DRG, the profit of provider i is given by

$$\pi_i(c_i, \delta_i, 0, \gamma_i) = \lambda q [(1 - h\gamma_i)\bar{c}_i - \delta_i h(1 - \gamma_i) - c(1 - h\gamma_i)] - R_c(c_i) - R_\delta(\delta_i) + \bar{R}_i,$$

where $\bar{c}_i = \frac{1}{N-1} \sum_{j \neq i} \left[c_j + \delta_j \frac{h(1-\gamma_j)}{1-h\gamma_j} \right]$ and $\bar{R}_i := \frac{1}{N-1} \sum_{j \neq i} [R_c(c_j) + R_\delta(\delta_j)]$ as defined in §4. The derivatives of the profit function of provider i are given by:

$$\frac{\partial}{\partial \gamma_i} \pi_i = \lambda q h (c_i + \delta_i - \bar{c}_i), \tag{9}$$

$$\frac{\partial}{\partial c_i} \pi_i = -\frac{d}{dc} R_c(c_i) - \lambda q (1 - h\gamma_i), \tag{10}$$

$$\frac{\partial}{\partial \delta_i} \pi_i = -\frac{d}{d\delta} R_\delta(\delta_i) - \lambda q h (1 - \gamma_i). \tag{11}$$

In any equilibrium outcome the last two conditions will be equal to zero for all providers. Otherwise the provider for whom one of these conditions is not zero could increase their profit by changing c_i or δ_i . Furthermore, the conditions above imply that any two providers with $\gamma_i = \gamma_j$ will have the same costs $c_i = c_j$ and $\delta_i = \delta_j$. Furthermore, since $R_c'' > 0$ and $R_\delta'' > 0$, if a provider has $\gamma_i > \gamma_j$ then $c_i > c_j$, $\delta_i > \delta_j$ and the converse is also true – if a provider has costs such that $\delta_i > \delta_j$ (or $c_i > c_j$) then $\gamma_i > \gamma_j$. Furthermore, since $0 \leq \gamma_i \leq \bar{\gamma}$, from (10) and (11), provider costs must satisfy $c^* \leq c_i \leq c^{e1}$, $\delta^* \leq \delta_i \leq \delta^{e1}$.

Consider a symmetric equilibrium where all providers choose (γ, c, δ) . In such a symmetric equilibrium then $\frac{\partial}{\partial \gamma_i} \pi_i = \delta \frac{1-h}{1-h\gamma} > 0$. Therefore, $\gamma = \bar{\gamma}$ which also implies that $c = c^{e1}, \delta = \delta^{e1}$ would constitute a candidate for a symmetric equilibrium. Furthermore, since $R_c'' > 0$ and $R_\delta'' > 0$ the symmetric equilibrium candidate is unique. We note that in this symmetric equilibrium candidate, the transfer payment ensures that all providers break even (i.e., make a profit of zero). For this to be an equilibrium outcome no provider must find it profitable to unilaterally deviate to a different strategy. Consider the payoff of one provider (labeled j) that chooses to deviate to a different strategy $(\gamma_j, c_j, \delta_j)$ when all other providers choose $(\bar{\gamma}, c^{e1}, \delta^{e1})$. The derivative of the profit function of provider j with respect to γ_j is given by $\frac{\partial}{\partial \gamma_j} \pi_j = \lambda q h (c_j + \delta_j - c^{e1} - \delta^{e1} \frac{h(1-\bar{\gamma})}{1-h\bar{\gamma}})$. Since $c_j + \delta_j \geq c^* + \delta^*$ then if $c^* + \delta^* > c^{e1} + \delta^{e1} \frac{h(1-\bar{\gamma})}{1-h\bar{\gamma}}$ (i.e., if cherry-picking-best costs are not too extreme) which we have assumed to be the case, then choosing $\gamma_j < \bar{\gamma}$ cannot be a profitable deviation. Therefore, $(\bar{\gamma}, c^{e1}, \delta^{e1})$ is the unique symmetric equilibrium.

We will next investigate the existence of asymmetric equilibria. If an asymmetric equilibrium exists, then at least one provider (labeled j) would have the highest $(\gamma_j, c_j, \delta_j)$ (i.e., $\gamma_j \geq \gamma_i$ for all i and the inequality is strict for at least one i , and similarly for c_j, δ_j). Therefore, $c_j + \delta_j > \bar{c}_j$ (recall that \bar{c}_j is the average cost of all other providers and at least some of these providers will have lower costs). From (9), this implies that $\gamma_j = \bar{\gamma}$, which also implies that the costs $c_j = c^{e1}, \delta_j = \delta^{e1}$. Therefore, in any asymmetric equilibrium, some providers (at least one) will choose $(\bar{\gamma}, c^{e1}, \delta^{e1})$. The rest of the providers will have $\gamma_k < \bar{\gamma}, c_k < c^{e1}, \delta_k < \delta^{e1}$. For this to be an equilibrium outcome, from (9) it must be the case that $c_k + \delta_k - \bar{c}_k \leq 0$. If it was not the case then $\frac{\partial}{\partial \gamma_k} \pi_k > 0$, implying that the provider's profit could increase by increasing γ_k which is a contradiction. Consider a provider with $c_k + \delta_k - \bar{c}_k = 0$. The profit of this provider can be written as $[\lambda q [-\delta_k h - c_k] - R_c(c_k) - R_\delta(\delta_k)] - \lambda q \gamma_k h (\bar{c}_k - \delta_k - c_k) + C$, where C is an exogenous constant. Note that the first term is independent of γ_i and is maximized at c^* and δ^* . Consider a deviation from $(\gamma_k, c_k, \delta_k)$ to $(0, c^*, \delta^*)$. This deviation does not affect the second term (it is zero under both strategies) and increases the first term (the first term is maximized at c^*, δ^*). Therefore this deviation is profitable. This suggests that no provider with costs $c_k + \delta_k - \bar{c}_k = 0$ can exist, which implies that any provider with cost other than c^{e1}, δ^{e1} must satisfy $c_k + \delta_k - \bar{c}_k < 0$, which implies $\frac{\partial}{\partial \gamma_k} \pi_k < 0$. Therefore, this provider must choose $(0, c^*, \delta^*)$ (any other choice of $\gamma_k > 0$ cannot be an equilibrium outcome as provider k can increase their profit by reducing γ). Therefore the condition $c_k + \delta_k - \bar{c}_k < 0$ becomes $c^* + \delta^* < \bar{c}_k$, and note that $\bar{c}_k > c^{e1} + \delta^{e1} \frac{h(1-\bar{\gamma})}{1-h\bar{\gamma}}$. In words, in any asymmetric equilibrium, providers will divide in two groups, θ_0 providers will not drop any patients and choose to operate at a cost as low as first best $(0, c^*, \delta^*)$, and $N - \theta_0$ providers will drop the maximum number of patients and operate at a higher cost compared to first best $(\bar{\gamma}, c^{e1}, \delta^{e1})$.

Consider one of the θ_0 low-cost providers. The fee per patient treated by this provider will be given by $\bar{c}_k = \frac{\theta_0-1}{N-1}(c^* + h\delta^*) + \frac{N-\theta_0}{N-1}(c^{e1} + h\frac{1-\bar{\gamma}}{1-\bar{\gamma}h}\delta^{e1})$. The condition $c_k + \delta_k - \bar{c}_k < 0$ implies that $c^* + \delta^* < \nu(c^* + h\delta^*) + (1-\nu)(c^{e1} + h\frac{1-\bar{\gamma}}{1-\bar{\gamma}h}\delta^{e1})$, for $\nu = \frac{\theta_0-1}{N-1}$. This is a contradiction as $c^* + \delta^* > (c^* + h\delta^*)$ and since we have assume that cherry-picking-best costs are not extreme, $c^* + \delta^* > c^{e1} + h\frac{1-\bar{\gamma}}{1-\bar{\gamma}h}\delta^{e1}$. Therefore, an asymmetric equilibrium cannot exist. \square

Proof of Proposition 3: In the absence of upcoding ($\bar{\alpha} = 0$), under the yardstick competition scheme with two DRGs, the profit of provider i is given by

$$\pi_i(c_i, \delta_i, 0, \gamma_i) = \lambda q [h(1-\gamma_i)(\bar{c}_{Mi} - \delta_i - c_i) + (1-h)(\bar{c}_{mi} - c_i)] - R_c(c_i) - R_\delta(\delta_i) + \bar{R}_i,$$

where $\bar{c}_{Mi} := \frac{1}{N-1} \sum_{j \neq i} [c_j + \delta_j]$, $\bar{c}_{mi} := \frac{1}{N-1} \sum_{j \neq i} c_j$, and $\bar{R}_i := \frac{1}{N-1} \sum_{j \neq i} [R_c(c_j) + R_\delta(\delta_j)]$ as defined in §4. The derivatives of the profit function of provider i are given by:

$$\frac{\partial}{\partial \gamma_i} \pi_i = \lambda q h (c_i + \delta_i - \bar{c}_{Mi}), \quad (12)$$

$$\frac{\partial}{\partial c_i} \pi_i = -\frac{d}{dc} R_c(c_i) - \lambda q (1-h\gamma_i), \quad (13)$$

$$\frac{\partial}{\partial \delta_i} \pi_i = -\frac{d}{d\delta} R_\delta(\delta_i) - \lambda q h (1-\gamma_i). \quad (14)$$

In any equilibrium outcome, the last two conditions will be equal to zero for all providers. Otherwise the provider for whom one of these conditions is not zero could increase their profit by changing c_i or δ_i . Furthermore, the conditions above imply that any two providers with $\gamma_i = \gamma_j$ will have the same costs $c_i = c_j$ and $\delta_i = \delta_j$. Since $R_c'' > 0$ and $R_\delta'' > 0$, if a provider has $\gamma_i > \gamma_j$ then $c_i > c_j$, $\delta_i > \delta_j$ and the converse is also true – if a provider has costs such that $\delta_i > \delta_j$ (or $c_i > c_j$) then $\gamma_i > \gamma_j$. Furthermore, since $0 \leq \gamma_i \leq \bar{\gamma}$, from (13) and (14), provider costs satisfy $c^* \leq c_i \leq c^{e1}$, $\delta^* \leq \delta_i \leq \delta^{e1}$.

Clearly, in any symmetric equilibrium $\frac{\partial}{\partial \gamma_i} \pi_i = 0$. Therefore, any $\gamma_i = \gamma$ where $0 \leq \gamma \leq \bar{\gamma}$ along with $c_i = c$ and $\delta_i = \delta$ such that $-\frac{d}{dc} R_c(c) = \lambda q (1-h\gamma)$, $-\frac{d}{d\delta} R_\delta(\delta) = \lambda q h (1-\gamma)$ would be a candidate for a symmetric equilibrium outcome. For any such γ , the values of c and δ are unique and are increasing in γ (since $R_c'' > 0$ and $R_\delta'' > 0$). Now consider any such symmetric equilibrium candidate where $\gamma > 0$ and consider the profit of provider i which will be given by $\pi_i = [-\lambda q (h\delta_i + c_i) - R_c(c_i) - R_\delta(\delta_i)] - \lambda q h \gamma_i (\bar{c}_{Mi} - \delta_i - c_i) + C$, where C is an exogenous constant. Note that the first term is independent of γ_i and is maximized at c^* and δ^* . Consider a deviation from the symmetric equilibrium candidate (γ, c, δ) to $(0, c^*, \delta^*)$. This deviation will be profitable for provider i as it would increase the first term and leave the second term unaffected (it is zero under both strategies). Therefore, $\gamma > 0$ cannot be a symmetric equilibrium outcome. Therefore, the only symmetric equilibrium candidate that survives is $(0, c^*, \delta^*)$. For this to be an equilibrium outcome, no provider must find it profitable to unilaterally deviate to a different strategy. Consider the payoff of one provider (labeled j) that chooses to

deviate to a different strategy $(\gamma_j, c_j, \delta_j)$ when all other providers choose $(0, c^*, \delta^*)$. The derivative of the profit function of provider j with respect to γ_j is given by $\frac{\partial}{\partial \gamma_j} \pi_j = \lambda q h(c_j + \delta_j - c^* - \delta^*) > 0$. Therefore, this provider can improve their profit by deviating to a strategy where they drop some patients (i.e., $\gamma_j > 0$) and invest less in cost reduction, (i.e., $c_j > c^*, \delta_j > \delta^*$). Therefore no symmetric equilibrium outcome can exist.

We will now turn to asymmetric equilibria. In any asymmetric equilibrium, at least one provider (labeled j) would have the highest $(\gamma_j, c_j, \delta_j)$ (i.e., $\gamma_j \geq \gamma_i$ for all i and the inequality is strict for at least one i , and similarly for c_j, δ_j). Therefore, $c_j + \delta_j > \bar{c}_{Mj}$ (recall that \bar{c}_{Mj} is the average cost of all other providers and at least some of these providers will have lower costs). From (12), this implies that $\gamma_j = \bar{\gamma}$, which also implies that the costs $c_j = c^{e1}, \delta_j = \delta^{e1}$. Conversely, at least one provider (labeled k) will have the lowest c_k, δ_k, γ_k (i.e., $c_k \leq c_i$ for all i and the inequality is strict for at least one i , and similarly for δ_k, γ_k). Therefore, $c_k + \delta_k < \bar{c}_{Mk}$ (recall that \bar{c}_{Mk} is the average cost of all other providers and at least some of these providers will have higher costs). This implies that this provider will choose $\gamma_k = 0$ and $c_k = c^*, \delta_k = \delta^*$. Furthermore, consider a provider with costs other than c^* or c^{e1} , which we label as provider s . This provider must have costs c_s and δ_s such that $c_s + \delta_s = \bar{c}_{Ms}$ and a corresponding γ_s . Consider the profit of this provider, which can be written as $\pi_s = [-\lambda q(h\delta_s + c_s) - R_c(c_s) - R_\delta(\delta_s)] - \lambda q h \gamma_s (\bar{c}_{Ms} - \delta_s - c_s) + C$, where C is an exogenous constant. Note that the first term is independent of γ_s and is maximized at $c_s = c^*$ and $\delta_s = \delta^*$. Consider a deviation from $(\gamma_s, c_s, \delta_s)$ to $(0, c^*, \delta^*)$. This deviation does not affect the second term (it is zero under both strategies) and increases the first term. Therefore this deviation is profitable. This suggests that no provider with costs γ_s, c_s, δ_s can exist. In words, in any asymmetric equilibrium, providers will divide in two groups: θ_1 providers will not drop any patients and choose to operate at a cost as low as first best $(0, c^*, \delta^*)$ and $N - \theta_1$ providers will drop the maximum number of patients and operate at a higher cost compared to first best $(\bar{\gamma}, c^{e1}, \delta^{e1})$.

For such an asymmetric equilibrium to exist, the profit of the θ_1 low cost providers and the profit of the $N - \theta_1$ high-cost providers need to be non-negative. Consider one of the θ_1 efficient providers. The fee for providing the major treatment is given by $\bar{c}_{Mk} = \frac{N-\theta_1}{N-1}(\delta^{e1} + c^{e1}) + \frac{\theta_1-1}{N-1}(\delta^* + c^*)$, the minor treatment is given by $\bar{c}_{mk} = \frac{N-\theta_1}{N-1}c^{e1} + \frac{\theta_1-1}{N-1}c^*$, and the transfer payment they will receive is given by $\bar{T}_k = \frac{N-\theta_1}{N-1}(R_\delta(\delta^{e1}) + R_c(c^{e1})) + \frac{\theta_1-1}{N-1}(R_\delta(\delta^*) + R_c(c^*))$. After some algebra, the profit of the efficient provider can be written as $\frac{N-\theta_1}{N-1}v_1$, where

$$v_1 := \lambda q(h(\delta^{e1} + c^{e1} - \delta^* - c^*) + (1-h)(c^{e1} - c^*)) + R_c(c^{e1}) + R_\delta(\delta^{e1}) - R_c(c^*) - R_\delta(\delta^*).$$

Note that the expression $-\lambda q(h(\delta + c) + (1-h)c) - R_c(c) - R_\delta(\delta)$ is maximized at $c = c^*$ and $\delta = \delta^*$, therefore $v_1 > 0$. Similarly, the profit of one of the $N - \theta_1$ high cost providers can be written as $\frac{\theta_1}{N-1}u_1$, where

$$u_1 := \lambda q(h(1-\bar{\gamma})(\delta^* + c^* - \delta^{e1} - c^{e1}) + (1-h)(c^* - c^{e1})) + R_c(c^*) + R_\delta(\delta^*) - R_c(c^{e1}) - R_\delta(\delta^{e1}).$$

Note that the expression $-\lambda q(h(1-\bar{\gamma})(\delta+c) + (1-h)c) - R_c(c) - R_\delta(\delta)$ is maximized at $c = c^{e1}$ and $\delta = \delta^{e1}$, therefore $u_1 > 0$.

We will next determine the value of θ_1 . For this to be an equilibrium outcome, it must be the case that the profit one of the low-cost providers makes by being low cost is greater than the profit they would make if they deviated to being a high-cost provider. After some algebra, this condition can be written as

$$\frac{N-\theta_1}{N-1}v_1 \geq \frac{\theta_1-1}{N-1}u_1.$$

Conversely, the profit of one of the high-cost providers must be greater than the payoff they would make if they deviated to being an low-cost provider. After some algebra, this condition reduces to

$$\frac{\theta_1}{N-1}u_1 \geq \frac{N-\theta_1-1}{N-1}v_1.$$

Together the last two inequalities imply that the number of efficient providers must satisfy

$$\frac{(N-1)v_1}{v_1+u_1} \leq \theta_1 \leq \frac{Nv_1+u_1}{v_1+u_1}.$$

Note that this interval contains exactly 1 integer as the difference $\frac{Nv_1+u_1}{v_1+u_1} - \frac{(N-1)v_1}{v_1+u_1} = 1$. Furthermore, since $\frac{Nv_1+u_1}{v_1+u_1} < N$, this integer is always less than N . \square

Proof of Proposition 4: Under the yardstick competition scheme with two DRGs, the profit of provider i is given by

$$\begin{aligned} \pi_i(c_i, \delta_i, \alpha_i, \gamma_i) = & \lambda q\{[h(1-\gamma_i) + (1-h)\alpha_i]\bar{c}_{Mi} + (1-h)(1-\alpha_i)\bar{c}_{mi} \\ & - [(h(1-\gamma_i) + (1-h)\alpha_i\beta)\delta_i + (1-h\gamma_i)c_i]\} \\ & - R_c(c_i) - R_\delta(\delta_i) + \bar{R}_i, \end{aligned}$$

where $\bar{c}_{Mi} := \frac{1}{N-1} \sum_{j \neq i} [c_j + \delta_j \frac{h(1-\gamma_j) + (1-h)\alpha_j\beta}{h(1-\gamma_j) + (1-h)\alpha_j}]$, $\bar{c}_{mi} := \frac{1}{N-1} \sum_{j \neq i} c_j$, and $\bar{R}_i := \frac{1}{N-1} \sum_{j \neq i} [R_c(c_j) + R_\delta(\delta_j)]$ as defined in §4. The derivatives of the profit function of provider i are given by:

$$\frac{\partial}{\partial \alpha_i} \pi_i = \lambda q(1-h)(\bar{c}_{Mi} - \bar{c}_{mi} - \beta \delta_i), \quad (15)$$

$$\frac{\partial}{\partial \gamma_i} \pi_i = \lambda q h(c_i + \delta_i - \bar{c}_{Mi}), \quad (16)$$

$$\frac{\partial}{\partial c_i} \pi_i = -\frac{d}{dc} R_c(c_i) - \lambda q(1-h\gamma_i), \quad (17)$$

$$\frac{\partial}{\partial \delta_i} \pi_i = -\frac{d}{d\delta} R_\delta(\delta_i) - \lambda q[h(1-\gamma_i) + (1-h)\alpha_i\beta]. \quad (18)$$

In any equilibrium outcome, the last two conditions will be equal to zero for all providers. Otherwise the provider for whom one of these conditions is not zero could increase their profit by changing

c_i or δ_i . Furthermore, since $0 \leq \gamma_i \leq \bar{\gamma}$, $0 \leq \alpha_i \leq \bar{\alpha}$, and $R_c'' > 0$ and $R_\delta'' > 0$, from (17) and (18) provider costs satisfy $c^* \leq c_i \leq c^{e1}$, $\delta^{e3} \leq \delta_i \leq \delta^{e1}$, where δ^{e3} is the unique solution to $-\frac{d}{d\delta}R_\delta(\delta_i) = \lambda q[h + (1-h)\bar{\alpha}\beta]$.

Turning to (15), this condition will be positive for provider i if $\frac{1}{N-1} \sum_{j \neq i} \left[\delta_j \frac{h(1-\gamma_j) + (1-h)\alpha_j\beta}{h(1-\gamma_j) + (1-h)\alpha_j} \right] > \beta\delta_i$. Note that the LHS is minimized when all providers other than i choose $\gamma_j = 0$, $\alpha_j = \bar{\alpha}$, $\delta_j = \delta^{e3}$. The RHS is maximized if provider i chooses $\delta_i = \delta^{e1}$. Therefore, a sufficient condition for (15) to be positive for all providers i is

$$\beta < \frac{h\delta^{e3}}{h\delta^{e1} + (\delta^{e1} - \delta^{e3})(1-h)\bar{\alpha}}, \quad (19)$$

which we have assumed holds. Therefore, in any equilibrium outcome, all providers will choose $\alpha_i = \bar{\alpha}$. Furthermore, (17) and (18) imply that any two providers with $\gamma_i = \gamma_j$ will have the same costs $c_i = c_j$ and $\delta_i = \delta_j$. Since $R_c'' > 0$ and $R_\delta'' > 0$, if a provider has $\gamma_i > \gamma_j$ then $c_i > c_j$, $\delta_i > \delta_j$ and the converse is also true – if a provider has costs such that $\delta_i > \delta_j$ (or $c_i > c_j$) then $\gamma_i > \gamma_j$.

Consider a symmetric equilibrium such that $\alpha_i = \alpha$, $\gamma_i = \gamma$, $c_i = c$ and $\delta_i = \delta$ for all i . In any symmetric equilibrium $\bar{c}_{Mi} - \bar{c}_{mi} - \beta\delta = \delta \frac{h(1-\gamma)(1-\beta)}{h(1-\gamma) + (1-h)\alpha} > 0$, which implies $\frac{\partial}{\partial \alpha_i} \pi_i > 0$. Therefore, in any symmetric equilibrium $\alpha_i = \bar{\alpha}$ for all i . This, implies that $c + \delta - \bar{c}_{Mi} = \frac{(1-h)(1-\beta)\delta\bar{\alpha}}{h(1-\gamma) + (1-h)\bar{\alpha}} > 0$, which also implies that $\frac{\partial}{\partial \gamma_i} \pi_i > 0$. Therefore in any symmetric equilibrium $\gamma = \bar{\gamma}$ for all i . The values of $c = c^{e1}$ and $\delta = \delta^{e2}$ are the solution to $-\frac{d}{dc}R_c(c) = \lambda q(1-h\bar{\gamma})$, $-\frac{d}{d\delta}R_\delta(\delta) = \lambda q[h(1-\bar{\gamma}) + (1-h)\bar{\alpha}\beta]$, and they are unique (since $R_c'' > 0$ and $R_\delta'' > 0$). In addition, the transfer payment ensures that all providers break even. Furthermore, since $(1-h\bar{\gamma}) < 1$, this implies that $c^{e1} > c^*$. If $h(1-\bar{\gamma}) + (1-h)\bar{\alpha}\beta < h$ then $\delta^{e2} > \delta^*$, otherwise the opposite holds.

For this to be an equilibrium outcome, no provider must find it profitable to unilaterally deviate to a different strategy. Consider the payoff of one provider (labeled j) that chooses to deviate to a different strategy $(\alpha_j, \gamma_j, c_j, \delta_j)$ when all other providers choose $(\bar{\alpha}, \bar{\gamma}, c^{e1}, \delta^{e2})$. Due to condition (7), it is not profitable to choose any $\alpha_j < \bar{\alpha}$. The derivative of the profit function of provider j with respect to γ_j is given by $\frac{\partial}{\partial \gamma_j} \pi_j = \lambda q h(c_j + \delta_j - c^{e1} - \delta^{e2} \frac{h(1-\bar{\gamma}) + (1-h)\bar{\alpha}\beta}{h(1-\bar{\gamma}) + (1-h)\bar{\alpha}})$. Since $c_j + \delta_j \geq c^* + \delta^{e3}$ then if $c^* + \delta^{e3} \geq c^{e1} + \delta^{e2} \frac{h(1-\bar{\gamma}) + (1-h)\bar{\alpha}\beta}{h(1-\bar{\gamma}) + (1-h)\bar{\alpha}}$ then choosing $\gamma_j < \bar{\gamma}$ cannot be a profitable deviation. Therefore, $(\bar{\alpha}, \bar{\gamma}, c^{e1}, \delta^{e2})$ will constitute a symmetric equilibrium. If however, $c^* + \delta^{e3} \leq c^{e1} + \delta^{e2} \frac{h(1-\bar{\gamma}) + (1-h)\bar{\alpha}\beta}{h(1-\bar{\gamma}) + (1-h)\bar{\alpha}}$ then it may be profitable to deviate to $(\bar{\alpha}, 0, c^*, \delta^{e3})$. For this to be the case, it must be the case that the profit of the provider who deviates to $(\bar{\alpha}, 0, c^*, \delta^{e3})$ when all other providers choose $(\bar{\alpha}, \bar{\gamma}, c^{e1}, \delta^{e2})$ is non-negative (as the profit associated with not deviating is zero). This condition can be written as $v_2 \geq 0$, where

$$v_2 := \lambda q \left((h + (1-h)\bar{\alpha}) \frac{(h(1-\bar{\gamma}) + (1-h)\bar{\alpha}\beta)}{h(1-\bar{\gamma}) + (1-h)\bar{\alpha}} \delta^{e2} - (h + (1-h)\bar{\alpha}\beta) \delta^{e3} + c^{e1} - c^* \right) + R_c(c^{e1}) + R_\delta(\delta^{e2}) - R_c(c^*) - R_\delta(\delta^{e3}).$$

We will then consider asymmetric equilibria. Due to condition (7), in any asymmetric equilibrium, all providers will choose $\alpha_j = \bar{\alpha}$. At least one provider (labeled j) would have the highest $(\gamma_j, c_j, \delta_j)$ (i.e., $\gamma_j \geq \gamma_i$ for all i and the inequality is strict for at least one i , and similarly for c_j, δ_j). Therefore, $c_j + \delta_j > \bar{c}_{Mj}$ (recall that \bar{c}_{Mj} is the average cost of all other providers and at least some of these providers will have lower costs). From (16), this implies that $\gamma_j = \bar{\gamma}$, which also implies that the costs $c_j = c^{e1}, \delta_j = \delta^{e2}$. Conversely, at least one provider (labeled k) will have the lowest $(\gamma_k, c_k, \delta_k)$ (i.e., $\gamma_k \leq \gamma_i$ for all i and the inequality is strict for at least one i , and similarly for c_k, δ_k). Therefore, $c_k + \delta_k < \bar{c}_{Mk}$ (recall that \bar{c}_{Mk} is the average cost of all other providers and at least some of these providers will have higher costs). From (16), this implies that $\gamma_k = 0$ and $c_k = c^*, \delta_k = \delta^{e3}$. For this to be possible, it must be the case that $c^* + \delta^{e3} < c^{e1} + \delta^{e2} \frac{h(1-\bar{\gamma}) + (1-h)\bar{\alpha}\beta}{h(1-\bar{\gamma}) + (1-h)\bar{\alpha}}$.

Furthermore, consider a provider with γ_s other than 0 or $\bar{\gamma}$, which we label as provider s . Due to condition (7), this provider will still have $\alpha_s = \bar{\alpha}$, and from (16) must have costs c_s and δ_s such that $c_s + \delta_s = \bar{c}_{Ms}$. Consider the profit of this provider, which can be written as $\pi_s = [-\lambda q(h\delta_s + c_s) - R_c(c_s) - R_\delta(\delta_s)] - \lambda q h \gamma_s (\bar{c}_{Ms} - \delta_s - c_s) + C$, where C is a constant that does not depend on $(\gamma_s, c_s, \delta_s)$. Note that the first term is independent of γ_s and is maximized at $c_s = c^*$ and $\delta_s = \delta^*$. Consider a deviation from $(\gamma_s, c_s, \delta_s)$ to $(0, c^*, \delta^*)$. This deviation does not affect the second term (it is zero under both strategies) and increases the first term. Therefore this deviation is profitable. This suggests that no provider with costs γ_s, c_s, δ_s can exist. In words, in any asymmetric equilibrium, providers will divide in two groups: θ_2 providers will upcode the maximum number of patients, will not drop any patients and choose to operate at relatively low costs $(\bar{\alpha}, 0, c^*, \delta^{e3})$ and $N - \theta_2$ providers that will upcode the maximum number of patients, will drop the maximum number of patients and operate at a higher cost $(\bar{\alpha}, \bar{\gamma}, c^{e1}, \delta^{e2})$.

For such an asymmetric equilibrium to exist, the profit of the θ_2 low-cost providers and the profit of the $N - \theta_2$ high-cost providers need to be non-negative. Consider one of the θ_2 high-cost providers. The fee they are paid for providing the major treatment is given by $\bar{c}_{Mk} = \frac{N-\theta_2}{N-1} (\delta^{e2} \frac{h(1-\bar{\gamma}) + (1-h)\bar{\alpha}\beta}{h(1-\bar{\gamma}) + (1-h)\bar{\alpha}} + c^{e1}) + \frac{\theta_2-1}{N-1} (\delta^{e3} \frac{h(1-\bar{\gamma}) + (1-h)\bar{\alpha}\beta}{h(1-\bar{\gamma}) + (1-h)\bar{\alpha}} + c^*)$, the fee for the minor treatment is given by $\bar{c}_{mk} = \frac{N-\theta_2}{N-1} c^{e1} + \frac{\theta_2-1}{N-1} c^*$, and the transfer payment they will receive is given by $\bar{T}_k = \frac{N-\theta_2}{N-1} (R_\delta(\delta^{e2}) + R_c(c^{e1})) + \frac{\theta_2-1}{N-1} (R_\delta(\delta^{e3}) + R_c(c^*))$. The profit of this provider will be given by $\frac{N-\theta_2}{N-1} v_2$. Similar algebra shows that the profit of one of the high-cost providers will be given by $\frac{\theta_2}{N-\theta_2} u_2$, where

$$u_2 := \lambda q (\delta^{e3} (h + (1-h)\bar{\alpha}\beta) \frac{h(1-\bar{\gamma}) + (1-h)\bar{\alpha}}{h + (1-h)\bar{\alpha}} - (h(1-\bar{\gamma}) + (1-h)\bar{\alpha}\beta) \delta^{e2} + (1-\bar{\gamma}h)(c^* - c^{e1})) + R_c(c^*) + R_\delta(\delta^{e3}) - R_c(c^{e1}) - R_\delta(\delta^{e2}).$$

Therefore, for the asymmetric equilibrium to exist, it must be the case that $v_2 \geq 0$ and $u_2 \geq 0$. Note that $u_2 > 0$. To see this, note that u_2 can be written as

$$u_2 = [-\lambda q ((1-\bar{\gamma}h)c^{e1} + ((1-\bar{\gamma})h + (1-h)\bar{\alpha}\beta)\delta^{e2}) - R_c(c^{e1}) - R_\delta(\delta^{e2})]$$

$$\begin{aligned}
& + [\lambda q ((1 - \bar{\gamma}h)c^* + ((1 - \bar{\gamma})h + (1 - h)\bar{\alpha}\beta)\delta^{e3}) - R_c(c^*) - R_\delta(\delta^{e3})] \\
& + h(1 - h) \frac{\bar{\alpha}\bar{\gamma}(1 - \beta)}{h + (1 - h)\bar{\alpha}} \delta^*.
\end{aligned}$$

Note that the expression $[-\lambda q [(1 - \bar{\gamma}h)c + ((1 - \bar{\gamma})h + (1 - h)\bar{\alpha}\beta)\delta] - R_c(c) - R_\delta(\delta)]$ is maximized at c^{e1}, δ^{e2} , therefore the sum of the terms in the first two brackets is positive. The third term is also positive, which implies that $u_2 > 0$. The sign of v_2 will depend on model parameters.

We next determine the value of θ_2 . If one of the θ_2 low-cost providers was to deviate and become a high-cost provider then their profit would be given by $\frac{\theta_2 - 1}{N - 1}u_2$ and if one of the $N - \theta_2$ high-cost providers was to deviate and become a low-cost provider it would be $\frac{N - \theta_2 - 1}{N - 1}v_2$. In the symmetric equilibrium it must be the case that these deviations are not profitable. Therefore, $\frac{N - \theta_2}{N - 1}v_2 > \frac{\theta_2 - 1}{N - 1}u_2$ and $\frac{\theta_2}{N - 1}u_2 \geq \frac{N - \theta_2 - 1}{N - 1}v_2$. These conditions imply that θ_2 satisfies $\frac{(N - 1)v_2}{u_2 + v_2} \leq \theta_2 \leq \frac{Nv_2 + u_2}{u_2 + v_2}$. If $v_2 \geq 0$, this interval contains exactly 1 integer as the difference between the RHS and the LHS of the inequalities is $\frac{Nv_2 + u_2}{u_2 + v_2} - \frac{(N - 1)v_2}{u_2 + v_2} = 1$ and this integer is always less than N .

□

Proof of Proposition 5: Under one DRG, if upcoding is possible (i.e., $\bar{\alpha} > 0$) the derivative of the profit of any provider with respect to α_i is given by $\frac{\partial}{\partial \alpha_i}\pi_i = -\lambda q(1 - h)\beta\delta_i < 0$. Therefore, in equilibrium no provider would choose to upcode and the equilibrium outcome is identical to that presented in Proposition 3 – namely, given the condition $c^{e1} + \delta^{e1}h\frac{1 - \bar{\gamma}}{1 - h\bar{\gamma}} < c^* + \delta^*$ the symmetric equilibrium is characterized by all providers choosing $(0, \bar{\gamma}, c^{e1}, \delta^{e1})$. Note that this is equivalent to the solution that maximizes total welfare (i.e., maximizes the objective of the HO as defined in (2)) under the constraint $\gamma = \bar{\gamma}$. Under two DRGs, the equilibrium outcome is given by Proposition 5. Namely, given that CPU-best costs are comparable to upcoding best costs, the equilibrium is symmetric and characterized by all providers choosing $(\bar{\alpha}, \bar{\gamma}, c^{e1}, \delta^{e2})$. Note that this is the solution that maximizes total welfare (i.e., maximizes the objective of the HO as defined in (2)) under the constraints $\gamma = \bar{\gamma}$ and $\alpha = \bar{\alpha}$. Note that the feasible region of this welfare-maximization problem is a subset of the feasible region of the previous welfare-maximization problem. Therefore welfare under two-DRG symmetric equilibrium cannot be greater than the welfare under the one-DRG equilibrium. □

Proof of Proposition 6: Under this yardstick competition scheme, the profit of provider i is given by

$$\begin{aligned}
\pi_i(c_i, \delta_i, \alpha_i, \gamma_i) &= \lambda q ([h(1 - \gamma_i) + (1 - h)\alpha_i]\bar{c}_{Mi} + (1 - h)(1 - \alpha_i)\bar{c}_{mi} \\
&\quad - [(h(1 - \gamma_i) + (1 - h)\alpha_i\beta)\delta_i + (1 - h\gamma_i)c_i]) \\
&\quad - R_c(c) - R_\delta(\delta) + \bar{R}_i + \kappa(M_i - \bar{M}_i) + \phi_i(m_i - \bar{m}_i),
\end{aligned}$$

where $M_i = \lambda q(h(1 - \gamma_i) + (1 - h)\alpha_i)$, $m_i = \lambda q(1 - h)(1 - \alpha_i)$, $\kappa \geq 0$, $\phi_i - \kappa > \bar{\delta}_i - \beta\delta^{e3}$, $\bar{\delta}_i := \bar{c}_{Mi} - \bar{c}_{mi}$, $\bar{c}_{Mi} = \frac{1}{N-1} \sum_{j \neq i} \left[c_j + \delta_j \frac{h(1 - \gamma_j) + (1 - h)\alpha_j \beta}{h(1 - \gamma_j) + (1 - h)\alpha_j} \right]$, $\bar{c}_{mi} = \frac{1}{N-1} \sum_{j \neq i} c_j$, and $\bar{R}_i = \frac{1}{N-1} \sum_{j \neq i} [R_c(c_j) + R_\delta(\delta_j)]$.

The derivatives of the profit function of provider i are given by:

$$\frac{\partial}{\partial \alpha_i} \pi_i = \lambda q(1 - h)(-\beta\delta_i + \bar{\delta}_i + \kappa - \phi_i), \quad (20)$$

$$\frac{\partial}{\partial \gamma_i} \pi_i = \lambda q h(c_i + \delta_i - \bar{c}_{Mi} - \kappa), \quad (21)$$

$$\frac{\partial}{\partial c_i} \pi_i = -\frac{d}{dc} R_c(c_i) - \lambda q(1 - h\gamma_i), \quad (22)$$

$$\frac{\partial}{\partial \delta_i} \pi_i = -\frac{d}{d\delta} R_\delta(\delta_i) - \lambda q[h(1 - \gamma_i) + (1 - h)\alpha_i \beta]. \quad (23)$$

In any equilibrium outcome, the expressions (22) and (23) have to be equal to zero for all providers. Otherwise the provider for whom one of these is not zero could increase their profit by changing c_i or δ_i . Furthermore, since $0 \leq \gamma_i \leq \bar{\gamma}$ and $0 \leq \alpha_i \leq \bar{\alpha}$, from (22) and (23) provider costs satisfy $c^* \leq c_i \leq c^{e1}$, $\delta^{e3} \leq \delta_i \leq \delta^{e1}$. Turning to (20), since $\phi_i - \kappa > \bar{\delta}_i - \beta\delta^{e3}$, in any equilibrium $\frac{\partial}{\partial \alpha_i} \pi_i < 0$ which implies that $\alpha_i = 0$ for all i . This implies that any two providers with $\gamma_i = \gamma_j$ will have the same costs $c_i = c_j$ and $\delta_i = \delta_j$ and since $R'_c > 0$ and $R'_\delta > 0$, if a provider has $\gamma_i > \gamma_j$ then $c_i > c_j$, $\delta_i > \delta_j$ and the converse is also true – if a provider has costs such that $\delta_i > \delta_j$ (or $c_i > c_j$) then $\gamma_i > \gamma_j$. Furthermore, since $\alpha_i = 0$ for all i we can use (23) to narrow down the range of possible costs δ_i to $\delta^* \leq \delta_i \leq \delta^{e1}$.

In any symmetric equilibrium $\frac{\partial}{\partial \gamma_i} \pi_i = -\kappa < 0$, therefore $\gamma_i = 0$, and $\delta_i = \delta^*$ and $c_i = c^*$ for all i . Therefore, the strategy $(0, 0, c^*, \delta^*)$ is the only candidate for a symmetric equilibrium outcome. For this to be an equilibrium outcome no provider must find it profitable to unilaterally deviate to a different strategy. Consider the payoff of one provider (labeled j) that chooses to deviate to a different strategy $(0, \gamma_j, c_j, \delta_j)$ when all other providers choose $(0, 0, c^*, \delta^*)$. The derivative of the profit function of provider j with respect to γ_j is given by $\frac{\partial}{\partial \gamma_j} \pi_j = \lambda q h(c_j + \delta_j - c^* - \delta^* - \kappa)$. If $\kappa > c^{e1} + \delta^{e1} - c^* - \delta^*$ then no profitable deviation can exist, therefore the strategy $(0, 0, c^*, \delta^*)$ is the unique symmetric equilibrium outcome. Otherwise, provider j may find it profitable to deviate to $(0, \bar{\gamma}, c^{e1}, \delta^{e1})$. In this case, the profit of provider j needs to be non-negative

$$\begin{aligned} u_4 &:= \lambda q(h(1 - \bar{\gamma})(\delta^* + c^* - \delta^{e1} - c^{e1}) + (1 - h)(c^* - c^{e1})) + R_c(c^*) + R_\delta(\delta^*) - R_c(c^{e1}) - R_\delta(\delta^{e1}) - \lambda q h \bar{\gamma} \kappa \\ &= u_1 - \lambda q h \bar{\gamma} \kappa. \end{aligned}$$

Note that $u_1 > 0$ (see Proof of Proposition 4). Therefore, if $\kappa < \min\{c^{e1} + \delta^{e1} - c^* - \delta^*, \frac{u_1}{\lambda q h \bar{\gamma}}\}$ then a profitable deviation will exist and no symmetric equilibrium outcome can exist.

We now turn to asymmetric equilibria. As shown above, in any asymmetric equilibrium $\alpha_i = 0$ for all i . Therefore, if $\kappa = 0$ for all i , then the problem of finding asymmetric equilibria reduces to that

of finding equilibria in the case where there is cherry picking but not upcoding (see Proposition 4). We will consider the case where $\kappa > 0$. In this case, the provider with the highest γ_j will also be the provider with the highest costs c_j and δ_j (see similar argument given in the Proof of Proposition 4). Consider one such provider. From (21), $\frac{\partial}{\partial \gamma_j} \pi_j = \lambda q h (c_j + \delta_j - \bar{c}_{Mj} - \kappa_j)$. Note that $c_j + \delta_j > \bar{c}_{Mj}$, nevertheless, if $\kappa \geq c^{e1} + \delta^{e1} - c^* - \delta^*$ then $\frac{\partial}{\partial \gamma_j} \pi_j < 0$, suggesting that lowering γ_j would increase the provider's profit. Therefore, for sufficiently high κ there cannot exist a provider with higher rate γ_i or higher costs than other providers, suggesting that an asymmetric equilibrium does not exist. If $\kappa < c^{e1} + \delta^{e1} - c^* - \delta^*$, then $\gamma_j = \bar{\gamma}$, which also implies that the costs $c_j = c^{e1}$, $\delta_j = \delta^{e1}$. Conversely, at least one provider (labeled k) will have the lowest c_k, δ_k, γ_k (i.e., $c_k \leq c_i$ for all i and the inequality is strict for at least one i , and similarly for δ_k, γ_k). Therefore, $c_k + \delta_k < \bar{c}_{Mk}$ (recall that \bar{c}_{Mk} is the average cost of all other providers and at least some of these providers will have higher costs). This implies that this provider will choose $\gamma_k = 0$ and $c_k = c^*, \delta_k = \delta^*$. Furthermore, consider a provider with costs other than c^* or c^{e1} , which we label as provider s . This provider must have costs c_s and δ_s such that $c_s + \delta_s = \bar{c}_{Ms} + \kappa$ and a corresponding γ_s . Consider the profit of this provider, which can be written as $\pi_s = [-\lambda q (h \delta_s + c_s) - R_c(c_s) - R_\delta(\delta_s)] - \lambda q h \gamma_s (\bar{c}_{Ms} - \delta_s - c_s + \kappa) + C$, where C is an exogenous constant. Note that the first term is independent of γ_s and is maximized at $c_s = c^*$ and $\delta_s = \delta^*$. Consider a deviation from $(0, \gamma_s, c_s, \delta_s)$ to $(0, 0, c^*, \delta^*)$. This deviation does not affect the second term (it is zero under both strategies) and increases the first term. Therefore this deviation is profitable. This suggests that no provider with costs γ_s, c_s, δ_s can exist. In words, in any asymmetric equilibrium providers will divide in two groups: θ_4 providers will not drop any patients and choose to operate at a cost as low as first best $(0, 0, c^*, \delta^*)$ and $N - \theta_4$ providers will drop the maximum number of patients and operate at a higher cost compared to first best $(0, \bar{\gamma}, c^{e1}, \delta^{e1})$.

For such an asymmetric equilibrium to exist, the profit of the θ_4 low-cost providers and the profit of the $N - \theta_4$ high-cost providers need to be non-negative. Consider one of the θ_4 efficient providers. The fee for providing the major treatment is given by $\bar{c}_{Mk} = \frac{N-\theta_1}{N-1}(\delta^{e1} + c^{e1}) + \frac{\theta_1-1}{N-1}(\delta^* + c^*)$, the minor treatment is given by $\bar{c}_{mk} = \frac{N-\theta_1}{N-1}c^{e1} + \frac{\theta_1-1}{N-1}(c^*)$, the number of major treatments provided by others $\bar{M}_i = \lambda q \left(\frac{N-\theta_1}{N-1} h (1 - \bar{\gamma}) + \frac{\theta_1-1}{N-1} h \right)$, the number of minor treatments provided by others $\bar{m}_i = \lambda q \left(\frac{N-\theta_1}{N-1} (1 - h \bar{\gamma}) + \frac{\theta_1-1}{N-1} \right)$, and the transfer payment they will receive is given by $\bar{T}_k = \frac{N-\theta_1}{N-1} (R_\delta(\delta^{e1}) + R_c(c^{e1})) + \frac{\theta_1-1}{N-1} (R_\delta(\delta^*) + R_c(c^*))$. After some algebra, the profit of the efficient provider can be written as $\frac{N-\theta_1}{N-1} v_4$, where

$$\begin{aligned} v_4 &:= \lambda q (h(\delta^{e1} + c^{e1} - \delta^* - c^*) + (1-h)(c^{e1} - c^*) + h \bar{\gamma} \kappa) + R_c(c^{e1}) + R_\delta(\delta^{e1}) - R_c(c^*) - R_\delta(\delta^*) \\ &= v_1 + \lambda q h \bar{\gamma} \kappa. \end{aligned}$$

Note that $v_1 > 0$ (see Proof of Proposition 4), therefore, $v_4 > 0$. Similarly, the profit of one of the $N - \theta_4$ high-cost providers can be written as $\frac{\theta_4}{N-1}u_4$, where $u_4 := u_1 - \lambda q h \bar{\gamma} \kappa$. Note that $u_1 > 0$ (see Proof of Proposition 4). Therefore, for the asymmetric equilibrium to exist, it must be the case that $\kappa < \frac{u_1}{\lambda q h \bar{\gamma}}$.

We will next determine the value of θ_4 . For this to be an equilibrium outcome the profit one of the low-cost providers makes by being low cost must be greater than the profit they would make if they deviated to being a high-cost provider. After some algebra this condition can be written as

$$\frac{N - \theta_4}{N - 1}v_4 \geq \frac{\theta_4 - 1}{N - 1}u_4.$$

Conversely, the profit of one of the high-cost providers must be greater than the payoff they would make if they deviated to being a low-cost provider. After some algebra this condition reduces to

$$\frac{\theta_4}{N - 1}u_4 \geq \frac{N - \theta_4 - 1}{N - 1}v_4.$$

Together the last two inequalities imply that the number of efficient providers must satisfy

$$\frac{(N - 1)(v_1 + \lambda q h \bar{\gamma} \kappa)}{v_1 + u_1} \leq \theta_4 \leq \frac{N v_1 + u_1 + (N - 1)\lambda q h \bar{\gamma} \kappa}{v_1 + u_1}.$$

Note that this interval contains exactly 1 integer as the difference between the RHS and the LHS of the inequalities is 1. Furthermore, θ_4 is non-decreasing in κ . Since $\theta_4 = \theta_1$ when $\kappa = 0$, it follows that $\theta_4 \geq \theta_1$. \square

Appendix 2: Cherry picking and Upcoding under Alternative Assumptions

In this Appendix we examine how patient cherry picking and upcoding affect the DRG design problem under cost-of-service regulation and under alternative assumptions.

A2.0. Cost of service regulation

Under cost-of-service regulation, the HO observes the chosen costs of the providers and sets reimbursement on the basis of these observed costs. This is similar to how hospitals used to be reimbursed by Medicare until 1983 (Dranove 1987). For this section we will assume that the status quo, where costs are c_0, δ_0 and there is no upcoding or lemon dropping ($\alpha = \gamma = 0$), are going to be the chosen equilibrium outcome unless there is an incentive to deviate – in other words, there is an implicit cost of managerial effort in changing the cost structure and/or finding ways to lemon drop some patients and upcode others that breaks ties in favour of the status quo. We need to distinguish two possibilities. In the first, the HO covers the provider's costs by paying a single fee per patient episode (p) irrespective of the treatment provided (i.e., $p_m = p_M = p$) – this is equivalent

to assuming that there is only one coarse DRG associated with the condition. The fee per patient episode will be set equal to the ex post observed average cost of treating patients reduces to:

$$p = c + \delta \frac{h(1 - \gamma) + (1 - h)\alpha\beta}{1 - h\gamma}. \quad (24)$$

Alternately, the HO could break the condition into two distinct DRGs – one for the minor and another one for the major condition – and cover the provider's costs by paying two distinct fees, one for each DRG. This type of cost-of-service payment is more akin to fee-for-service, where the payment depends on the volume and intensity of services the provider offers to patients. In this case, the fee for the minor and the major condition will be set equal to the ex post average cost of treating patients, which are given by

$$p_m = c \text{ and } p_M = c + \delta \frac{h(1 - \gamma) + (1 - h)\alpha\beta}{h(1 - \gamma) + (1 - h)\alpha}, \quad (25)$$

respectively. In order to cover any investment cost incurred by the provider, in addition to the fees described above, the provider also receives a transfer payment equal to the observed investment costs $T = R_c(c) + R_\delta(\delta)$. All payments depend on the cost levels c, δ and/or the degree of upcoding and lemon dropping α, γ , all of which are all chosen by the provider. Furthermore, although the HO can observe the realized costs and the number of patients treated, they are not able to observe if the cost investments made by the providers are optimal (because they do not know the cost functions $R_c(\cdot)$ and $R_\delta(\cdot)$) and are not in a position to ascertain whether patient selection or upcoding are taking place. We characterize the providers' optimal actions with the proposition below.

Proposition 7 *Under cost-of-service regulation, irrespective of the number of DRGs used, i) the provider does not upcode or lemon drop (i.e., chooses $\alpha = \gamma = 0$); ii) the provider makes no investment in cost reduction (i.e., chooses costs c_0, δ_0).*

On the downside, the proposition shows that, because payments are linked to costs, providers have no incentive to invest in any cost reduction as any such cost reduction will result in a lower fee per treatment. This result is well known in the literature (see Shleifer 1985, Ma 1994). On the upside, the proposition shows that cost-of-service regulation is effective in curtailing patient upcoding and cherry picking practices. That it does so irrespective of the number of DRGs specified is a little surprising. For example, if the HO uses one DRG then the payment per patient will, by definition, be greater than the cost of treating low-complexity patients (c) but lower than the cost of treating high-complexity patients ($c + \delta$) – see Equation (24). Nevertheless, the provider has no incentive to cherry pick low-complexity patients or lemon drop high-complexity patients – if they did that, the average cost of treating patients and, therefore, the payment received per patient would be reduced

accordingly. Similarly, if the HO were to use two DRGs, then the payment for the major treatment would be greater than the payment for the minor treatment (see Equation (25)). Nevertheless, the provider has no incentive to upcode low-complexity patients because such upcoding would reduce the average cost of providing the major treatment and, as a result, the payment for the major treatment would be reduced accordingly.

At a high level, the proposition suggests that in situations where the main concern is not to incentivize investment in cost reduction (e.g., because costs are largely fixed and exogenous to the providers' efforts) but instead to incentivize providers to prescribe the right treatment (i.e., solve the credence-goods problem), cost-of-service regulation, where providers are reimbursed for their costs of providing the service, performs well.

A2.1. Provider downcoding

In the main paper we had assume that providers were not able to undertreat patients (because of ethical and liability considerations). We had also assumed that providers do not downcode patients. That is, providers never code a high-complexity patient that was given the major treatment as having received the minor treatment for reimbursement purposes. In this section we relax this assumption and allow providers to downcode a proportion $1 - \zeta$ of the non-dropped high complexity patients. We assume that the rate of downcoding is bounded from above by $1 - \bar{\zeta} \geq 0$. The updated model is shown in Figure 5. We will focus on the case where the regulator is using two DRGs as downcoding is not possible/meaningful otherwise.

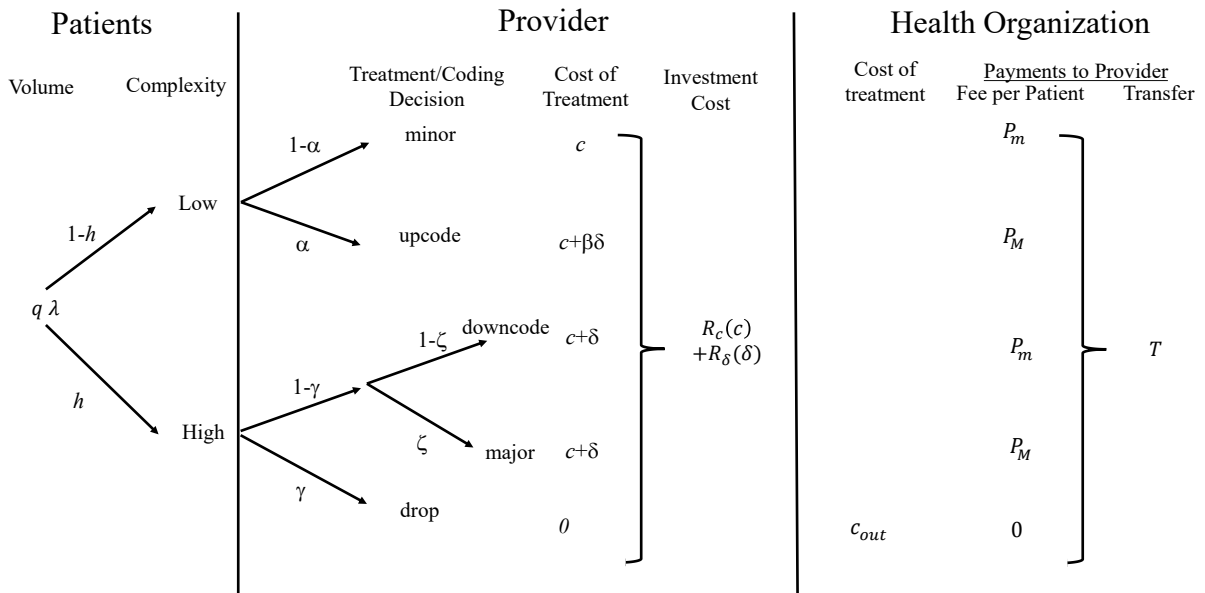


Figure 5 In this Appendix in addition to the assumptions made in the main paper (see Figure 1), providers may decide that a proportion $1 - \zeta$ of the high-complexity patients that they do not lemon-drop will be downcoded. These patients still cost $c + \delta$ to treat (as they receive the major treatment) but will be reimbursed at the minor rate P_m .

We first examine the case of cost-of-service regulation. In this case, provider reimbursement for the minor and the major condition are given by the average cost of providing treatment:

$$p_m = c + \delta \frac{h(1-\gamma)(1-\zeta)}{h(1-\gamma)(1-\zeta) + (1-h)(1-\alpha)}, \text{ and } p_M = c + \delta \frac{h(1-\gamma)\zeta + (1-h)\alpha\beta}{h(1-\gamma)\zeta + (1-h)\alpha}. \quad (26)$$

Note that if $\zeta = 1$ (i.e., providers do not downcode) these equations reduce to those of (25).

Proposition 7d: *Under cost-of-service regulation providers never find it optimal to downcode ($\zeta = 1$), and the optimal outcome is given by Proposition 1.*

Turning to the case where the regulator uses yardstick competition to reimburse providers, the realized average costs for major and minor treatments at all other providers besides provider i , are given by:

$$\begin{aligned} \bar{c}_{Mi} &:= \frac{1}{N-1} \sum_{j \neq i} \left[c_j + \delta_j \frac{h(1-\gamma_j)\zeta_j + (1-h)\alpha_j\beta}{h(1-\gamma_j)\zeta_j + (1-h)\alpha_j} \right], \\ \bar{c}_{mi} &:= \frac{1}{N-1} \sum_{j \neq i} c_j + \delta_j \frac{h(1-\gamma_j)(1-\zeta_j)}{h(1-\gamma_j)(1-\zeta_j) + (1-h)(1-\alpha_j)}, \end{aligned}$$

respectively. Again, note that if $\zeta = 1$ (i.e., providers do not downcode) these equations reduce to those of (4).

We characterize the equilibrium outcome when providers can upcode, downcode, and lemon drop patients with the proposition below, which extends the results of Proposition 4 of the main paper.

Proposition 4d: *If the HO implements yardstick competition based on two DRGs, then in any equilibrium providers do not engage in downcoding if $\bar{c}_{Mi} - \bar{c}_{mi} > 0$ for all i . A sufficient condition for this to be true is $\bar{\zeta} > \bar{\alpha}$. If this condition holds the equilibrium outcome is given by Proposition 4.*

The proposition suggests that under very mild conditions (i.e., as long as the reimbursement for the major treatment is higher than the reimbursement for the minor treatment) then providers find it optimal not to engage in downcoding. A sufficient condition based on model primitives for this to hold is that for every provider the proportion of high complexity patients that are coded correctly (ζ_i) is not lower than the proportion of low complexity patients that are upcoded (α_i).

Next we turn our attention to the case where the HO uses input statistics as described in §6. The following proposition extends the results of Proposition 6 of the main paper to the case where providers can downcode.

Proposition 6d: *Under the two-DRG payment scheme with input statistics, in any equilibrium providers do not engage in downcoding if $\bar{c}_{Mi} - \bar{c}_{mi} > 0$ for all i . A sufficient condition for this to be true is $\bar{\zeta} > \bar{\alpha}$. If this condition holds the equilibrium outcome is given by Proposition 6.*

A2.2. Extreme Cherry-Picking-Best costs

In this section we extend the results of Proposition 2 to the case where cherry-picking-best costs are extreme (i.e., $c^* + \delta^* \leq c^{e1} + \delta^{e1} \frac{h(1-\bar{\gamma})}{1-h\bar{\gamma}}$). We will define the following two quantities:

$$\begin{aligned} v_0 &:= \lambda q \left[c^{e1} - c^* + h \left(\frac{1-\bar{\gamma}}{1-\bar{\gamma}h} \delta^{e1} - \delta^* \right) \right] + R_c(c^{e1}) + R_\delta(\delta^{e1}) - R_c(c^*) - R_\delta(\delta^*), \\ u_0 &:= \lambda q \left[(1-\bar{\gamma}h)(c^* - c^{e1}) + h(\delta^* - \delta^{e1}) + \bar{\gamma}h(\delta^{e1} - h\delta^*) \right] + R_c(c^*) + R_\delta(\delta^*) - R_c(c^{e1}) - R_\delta(\delta^{e1}). \end{aligned}$$

The quantity v_0 is the profit of a provider who is paid according to yardstick competition and chooses $(0, c^*, \delta^*)$ when all other providers choose $(\bar{\gamma}, c^{e1}, \delta^{e1})$, and vice versa for u_0 .

The following proposition describes equilibrium outcomes when cherry-picking-best costs are extreme and complements the results presented in Proposition 2.

Proposition 8 *In the absence of upcoding ($\bar{\alpha} = 0$) if cherry-picking-best costs are extreme (i.e., $c^* + \delta^* \leq c^{e1} + \delta^{e1} \frac{h(1-\bar{\gamma})}{1-h\bar{\gamma}}$), then if the HO implements yardstick competition based on a single DRG, then there exists a unique Nash equilibrium:*

- If $v_0 < 0$ the equilibrium is symmetric and is given by proposition 2.
- Otherwise, the equilibrium is asymmetric, where $N - \theta_0$ providers drop as many patients as possible and choose cherry-picking-best costs (i.e., these providers choose $(\bar{\gamma}, c^{e1}, \delta^{e1})$, and $c^{e1} > c^*$, $\delta^{e1} > \delta^*$) and θ_0 providers do not engage in cherry picking and invest in cost reduction as much as in first best (i.e., these providers choose $(0, c^*, \delta^*)$). The number of efficient providers θ_0 is the only integer in the interval $\left[\frac{(N-1)v_0}{u_0+v_0}, \frac{Nv_0+u_0}{u_0+v_0} \right]$ and $\theta_1 \geq \theta_0$. All providers receive a positive rent.

The proposition shows that even if costs are extreme (i.e., $c^* + \delta^* \leq c^{e1} + \delta^{e1} \frac{h(1-\bar{\gamma})}{1-h\bar{\gamma}}$) the equilibrium might still be the same as that described in Proposition 2 (this is the case if $v_0 < 0$) where all providers drop as many high-complexity patients as possible and underinvest in cost reduction compared to first best. Otherwise, if $v_0 \geq 0$ the equilibrium is asymmetric. Some providers choose to drop high-complexity patients and underinvest in cost reduction while other providers treat all patients and invest optimally in cost reduction. This result is similar to the case where the provider uses two DRGs (see Proposition 3). Nevertheless, the number of efficient providers is lower under one DRG compared to two DRGs, suggesting that, in all cases, the equilibrium outcome with two DRGs dominates in terms of welfare compared to that with one DRG.

A2.3. Continuous increasing cost of patient upcoding and cherry picking

For the main paper we assumed that if providers choose to upcode or drop patients, they can do so without any cost up to the predetermined maximum amount of upcoding rate $\bar{\alpha}$ and patient lemon-dropping rate $\bar{\gamma}$, after which we have implicitly assumed that the cost of additional upcoding or lemon dropping becomes infinite. In this Appendix we investigate the case of continuously

increasing cost in engaging with patient upcoding and cherry picking. More specifically, we assume that when the provider decides to upcode (or drop) a proportion of patients α (γ), the provider incurs a cost $E_\alpha(\alpha)$ ($E_\gamma(\gamma)$). Furthermore, we assume that the cost functions $E_\gamma(\cdot)$ and $E_\alpha(\cdot)$ are positive, increasing, convex, and ‘sufficiently’ well-behaved (see conditions A and B imposed below) with $E_\alpha(0) = E_\gamma(0) = 0$ and $\lim_{x \rightarrow 1} E_\alpha(x) = \lim_{x \rightarrow 1} E_\gamma(x) = \infty$. These assumptions reflect the situation where it is relatively cheap for the provider to upcode (or drop) a small number of patients, but as they engage more in this practice, the costs start to increase (in a convex manner). They also ensure that the chosen rate of upcoding/dropping is in the $[0, 1)$ interval. (Alternately, one could view the cost as the expected penalty that the provider may incur if caught engaging in this practice, and that the size of the penalty and/or the probability of getting caught are increasing in proportion to the amount of upcoding/lemon dropping that takes place.)

In this case, the profit of the provider is as defined in §3.2 minus the cost of engaging in upcoding and overtreatment $E_\alpha(\alpha_i) + E_\gamma(\gamma_i)$. We note that, under these alternative assumptions, the first-best solution remains unchanged – the HO would still find it optimal not to engage in any lemon dropping or upcoding. Similarly, cost-of-service regulation would continue to provide the right incentives for providers not to engage in lemon dropping or upcoding but at the expense of eliminating incentives to invest in cost reduction.

Under yardstick competition, providers would be reimbursed with per-patient fees given by (3) or (4), if the HO uses one or two DRGs, respectively. We also assume that the transfer payment is now going to be equal to $T_i = \bar{R}_i + \bar{E}_{\alpha_i} + \bar{E}_{\gamma_i}$, where the last two terms are the average cost of upcoding and cherry picking at other providers (i.e., $\bar{E}_{\alpha_i} = \sum_{j \neq i} \frac{E_\alpha(\alpha_j)}{N-1}$ and similarly for \bar{E}_{γ_i}). This assumption is consistent with the case where the HO can only measure total investment cost and is not able to distinguish whether the investment was made to reduce costs or to upcode/cherry pick patients. (This assumption is not critical for any of our results, as the transfer payment does not affect incentives other than participation.)

We analyze the case where the HO uses one DRG and upcoding is not possible ($\alpha_i = 0$ for all i). This is the equivalent of Proposition 2.

Proposition 2c: *In the absence of upcoding ($\alpha_i = 0$ for all i), if the HO implements yardstick competition based on a single DRG and if Conditions A hold, then there exists a unique symmetric Nash equilibrium where providers choose $(\gamma^c, c^c, \delta^c)$ with $\gamma^c > 0, c^c > c^*, \delta^c > \delta^*$.*

Conditions A: Let $c(\gamma)$ and $\delta(\gamma)$ be the implicit functions defined by the unique solutions to $-R'_c(c) - \lambda q(1 - h\gamma) = 0$, $R'_\delta(\delta) - \lambda q h(1 - \gamma) = 0$, respectively.

- The following inequality holds for all $0 \leq \gamma < 1$: $(\lambda q h)^2 \left(\frac{1}{R''_\delta(\delta(\gamma))} + \frac{1}{R''_c(c(\gamma))} \right) - E''_\gamma(\gamma) < 0$.
- If γ^c is the unique solution to $\lambda q h \delta(\gamma^c) \frac{1-h}{1-\gamma^c h} - E'_\gamma(\gamma^c) = 0$ then $\lambda q h(c^* + \delta^* - c(\gamma^c) - h\delta(\gamma^c) \frac{1-\gamma^c}{1-\gamma^c h}) - E'_\gamma(0) > 0$.

Clearly, using one DRG continues to be an issue – providers will choose to drop patients and investment in cost reduction will be reduced accordingly. We now turn our attention to the case where the HO uses two DRGs.

Proposition 3c: *In the absence of upcoding ($\alpha_i = 0$ for all i), if the HO implements yardstick competition based on two DRGs and if Conditions A hold, then there exists a unique symmetric Nash equilibrium where providers choose $(0, c^*, \delta^*)$.*

The results confirm that expanding the number of DRGs is helpful, at least in the absence of upcoding. If cherry picking costs are increasing at a sufficiently high rate, there exists a unique symmetric equilibrium where providers choose not to upcode.

Turning to the case where upcoding is possible, the following proposition summarizes the main results.

Proposition 4c: *If the HO implements yardstick competition based on two DRGs and if Conditions B hold, then there exists a unique symmetric Nash equilibrium where providers choose $(\alpha^d, \gamma^d, c^d, \delta^d)$ with $\alpha^d > 0, \gamma^d > 0, c^d > c^*$. If $\beta < \frac{h}{1-h}\gamma^d\alpha^d$ then there is underinvestment for the major treatment ($\delta^d > \delta^*$), otherwise there is overinvestment ($\delta^d < \delta^*$).*

Conditions B: Let $c(\gamma), \delta(\alpha, \gamma)$ be the implicit functions defined by the unique solutions to $-R'_c(c) - \lambda q(1 - h\gamma) = 0$, $R'_\delta(\delta) - \lambda q(h(1 - \gamma) + (1 - h)\alpha) = 0$, respectively. Let

$$\begin{aligned} \pi(\alpha, \gamma) = & -\lambda q((h(1 - \gamma) + (1 - h)\alpha\beta)\delta(\alpha, \gamma) + (1 - h\gamma)c(\gamma)) \\ & - R_c(c(\gamma)) - R_\delta(\delta(\alpha, \gamma)) - E_\gamma(\gamma) - E_\alpha(\alpha). \end{aligned}$$

Also define α^d and γ^d as the unique solutions to the system of two equations: $\lambda q h \delta(\alpha^d, \gamma^d) \frac{h(1-\beta)(1-\gamma^d)}{h(1-\gamma^d)+(1-h)\alpha^d} - E'_\alpha(\alpha^d) = 0$, $\lambda q h \delta(\alpha^d, \gamma^d) \frac{(1-h)(1-\beta)\alpha^d}{h(1-\gamma^d)+(1-h)\alpha^d} - E'_\gamma(\gamma^d) = 0$. The following relationships hold:

- The Hessian of $\pi(\alpha, \gamma)$ is negative definite,
- $\lambda q h (\delta(\alpha^d, \gamma^d) \frac{h(1-\gamma^d)+(1-h)\alpha^d\beta}{h(1-\gamma^d)+(1-h)\alpha^d} - \beta\delta^*) - E'_\alpha(0) > 0$,
- $\lambda q h (c^* + \delta^* - c^d - \delta(\alpha^d, \gamma^d) \frac{h(1-\gamma^d)+(1-h)\alpha^d\beta}{h(1-\gamma^d)+(1-h)\alpha^d}) - E'_\gamma(0) > 0$.

The result described above mirrors the results of Proposition 4 – the presence of upcoding reintroduces incentives for providers to engage in cherry picking. There is, however, one difference worth outlining. In contrast to the case where cherry picking and upcoding were costless, where welfare was best served by not increasing the number of DRGs, in this case increasing the number of DRGs may confer welfare benefits. This is particularly the case if upcoding is relatively expensive compared to dropping patients. In this case, providers will not engage in too much upcoding and, as a result, the amount of heterogeneity in the high-cost DRG will be limited, thus reducing the financial benefit of dropping high-complexity patients. As a result, providers will only drop

a limited number of high-complexity patients, and welfare under two DRGs may be higher than welfare under one DRG.

Turning to the solution with input statistics, the following proposition characterizes the equilibrium outcome.

Proposition 6c *Under the two-DRG payment scheme with input statistics described in §6, if Conditions A are satisfied and if $\kappa_i \geq 0$ there exists a unique symmetric Nash equilibrium where all providers choose first-best actions (i.e., $(0, 0, c^*, \delta^*)$).*

Importantly for the purposes of this research, yardstick competition based on input statistics as described in the main paper would eliminate incentives to cherry pick and upcode patients and restore first-best investments. To see why this is the case, consider that payments based on input statistics remove incentives to upcode, even when doing so is costless. When it is costly to do so, the incentives to upcode cannot be any greater. Once upcoding has been eliminated so has the heterogeneity in costs within DRG, leading to a drastic reduction in incentives to cherry pick. In fact, as Proposition 7c suggests, if the cost of cherry picking is rising sufficiently fast (as per Conditions A), then the cherry picking is eliminated too for any $\kappa_i \geq 0$.

A2.4. No transfer payments

In the main paper we had assumed that the HO was able to pay a fee per patient episode and a transfer payment. The role of the former was to reimburse providers for the variable cost of treating patients and the role of the latter was to reimburse for the investment in cost reduction. In several healthcare systems, for example, Medicare in the US, providers are paid only a fee per patient episode which has to cover both the variable cost of treatment and the fixed investment cost.

We note that in classic yardstick competition literature (Shleifer 1985), where demand was endogenous to prices, the lack of a transfer payments was inherently problematic. Since patients would have to pay inflated prices (i.e., higher than the marginal cost of treatment) some patients whom the HO would have found optimal to treat will decide not to seek treatment. Furthermore, as a result of this inefficient reduction in demand, providers would find it optimal to invest less in cost reduction. In our case, since demand is exogenous, the impact of the absence of a transfer payment, if any, will be through the way it affects upcoding and cherry-picking incentives.

In the absence of transfer payments, if the HO wanted to implement yardstick competition on the basis of a single DRG, then the payment per patient to hospital i will be given by the average total cost of providing care in all other hospitals:

$$p = \frac{1}{N-1} \sum_{j \neq i} \left[c_j + \delta_j \frac{h(1-\gamma_j) + (1-h)\alpha_j\beta}{1-\gamma_j h} + \frac{R_c(c_j) + R_\delta(\delta_j)}{(1-\gamma_j h)\lambda q} \right]. \quad (27)$$

The first term in the sum represents the average cost for the minor treatment (offered to all patients who are not dropped). The second term represents the average cost of the major treatment (offered to high-complexity patients who are not dropped and to low-complexity patients who are upcoded). The third term represents the average investment cost amortized proportionally on all patients receiving treatment. Clearly, since in this case there is only one DRG, the provider has no incentive to upcode. Furthermore, because the per-patient-episode payment is now inflated by the investment cost, the providers' incentives to lemon drop are diluted in the sense that the additional profit associated with lemon dropping a patient is less in the absence of transfer payments compared to the case where transfer payments are allowed. Nevertheless, the results of Proposition 3 remain largely unchanged.

More specifically, define

$$\begin{aligned} v'_0 &:= v_0 + \frac{\bar{\gamma}h}{1-\bar{\gamma}h}(R_c(c^{e1}) + R_\delta(\delta^{e1})), \\ u'_0 &:= u_0 - \bar{\gamma}h(R_c(c^*) + R_\delta(\delta^*)). \end{aligned}$$

The quantity v'_0 is the profit of a provider who is paid according to yardstick competition and chooses $(0, c^*, \delta^*)$ when all other providers choose $(\bar{\gamma}, c^{e1}, \delta^{e1})$ when transfer payments are not allowed, and vice versa for u_0 . We will call the cherry-picking-best costs as comparable to first-best costs if either of the following conditions hold:

- $c^* + \delta^* \geq c^{e1} + \delta^{e1}h \frac{1-\bar{\gamma}}{1-h\bar{\gamma}} + \frac{R_c(c^{e1}) + R_\delta(\delta^{e1})}{\lambda q(1-\bar{\gamma}h)}$
- $v'_0 \leq 0$ or $u'_0 \leq 0$.

The following proposition is the equivalent of Propositions 3 and 8 in the absence of transfer payments.

Proposition 2n *In the absence of upcoding ($\bar{\alpha} = 0$), if the HO implements yardstick competition based on a single DRG, then there exists a unique Nash equilibrium:*

- *If cherry-picking-best costs are ‘comparable’ to first-best costs (as defined above), the equilibrium is symmetric. Providers drop as many patients as possible and invest in cherry-picking-best costs (i.e., all providers choose $(\bar{\gamma}, c^{e1}, \delta^{e1})$, and $c^{e1} > c^*$, $\delta^{e1} > \delta^*$).*
- *Otherwise, the equilibrium is asymmetric, where $N - \theta'_0$ providers drop as many patients as possible and choose cherry-picking-best costs (i.e., these providers choose $(\bar{\gamma}, c^{e1}, \delta^{e1})$, and $c^{e1} > c^*$, $\delta^{e1} > \delta^*$) and θ'_0 providers do not engage in cherry picking and invest in cost reduction as much as in first best (i.e., these providers choose $(0, c^*, \delta^*)$). The number of efficient providers θ'_0 is the only integer in the interval $\left[\frac{(N-1)v'_0}{u'_0 + v'_0}, \frac{Nv'_0 + u'_0}{u'_0 + v'_0} \right]$. Furthermore, $\theta'_0 \geq \theta_0$.*

The proposition suggests that, in the absence of transfer payments, then i) the inefficient symmetric equilibrium (where all providers choose to drop the maximum number of patients and

underinvest in cost reduction) is the outcome for a smaller set of parameter values; and ii) the asymmetric equilibrium is more efficient in the sense that more providers choose not to engage in cherry picking and invest optimally in cost reduction.

Turning to the case where the HO splits the condition into two DRGs for reimbursement purposes, the payments for the major and the minor condition, set at the average (treatment and investment) cost of all other providers, are given by

$$\bar{p}_{Mi} = \frac{1}{N-1} \sum_{j \neq i} \left[c_j + \delta_j \frac{h(1-\gamma_j) + (1-h)\alpha_j\beta}{h(1-\gamma_j) + (1-h)\alpha_j} + \frac{R_\delta(\delta_j) + R_c(c_j)}{\lambda q(1-h\gamma_j)} \right], \quad (28)$$

$$\bar{p}_{mi} = \frac{1}{N-1} \sum_{j \neq i} \left[c_j + \frac{R_\delta(\delta_j) + R_c(c_j)}{\lambda q(1-h\gamma_j)} \right], \quad (29)$$

respectively.

Before we characterize the equilibrium outcome of this competition, we define the following two quantities:

$$\begin{aligned} v'_1 &:= v_1 + \frac{h\bar{\gamma}}{1-h\bar{\gamma}}(R_\delta(\delta^{e1}) + R_c(c^{e1})), \\ u'_1 &:= u_1 - h\bar{\gamma}(R_\delta(\delta^*) + R_c(c^*)). \end{aligned}$$

As in the case with transfer payments, the quantity v'_1 is the profit of a provider who is paid according to yardstick competition based on two DRGs in the absence of transfer payments and who chooses $(0, c^*, \delta^*)$ when all other providers choose $(\bar{\gamma}, c^{e1}, \delta^{e1})$, and vice versa for the quantity u'_1 . Note that in this case $v'_2 > 0$ while the sign of u'_1 depends on model parameters.

In the absence of transfer payments under two DRGs, we will call the cherry-picking-best costs comparable to first-best costs if either of the following two conditions hold:

- $\lambda q(c^{e1} + \delta^{e1} - c^* - \delta^*) \leq R_\delta(\delta^*) + R_c(c^*)$.
- $u'_1 \leq 0$.

Proposition 3n *In the absence of upcoding ($\bar{\alpha} = 0$), if the HO implements yardstick competition based on two DRGs without transfer payments, then there exists a unique Nash equilibrium:*

- *If cherry-picking-best costs are ‘comparable’ to first-best costs (as defined above), the equilibrium is symmetric. Providers will not engage in cherry picking and invest in cost reduction as much as first best (i.e., all providers choose $(0, c^*, \delta^*)$).*
- *Otherwise, the equilibrium is asymmetric, where $N - \theta'_1$ providers drop as many patients as possible and choose cherry-picking-best costs (i.e., these providers choose $(\bar{\gamma}, c^{e1}, \delta^{e1})$, and $c^{e1} > c^*$, $\delta^{e1} > \delta^*$) and θ'_1 providers do not engage in cherry picking and invest in cost reduction as much as in first best (i.e., these providers choose $(0, c^*, \delta^*)$). The number of efficient providers θ_1 is the only integer in the interval $\left[\frac{(N-1)v'_0}{u'_0 + v'_0}, \frac{Nv'_0 + u'_0}{u'_0 + v'_0} \right]$. Furthermore, $\theta'_1 \geq \theta_1$.*

Again, the proposition demonstrates that the absence of transfer payments reduces providers' incentives to lemon drop. In contrast to the case with transfer payments, there may exist a symmetric equilibrium where all providers engage in first-best actions. Even if such an equilibrium does not exist, the number of providers that will choose to be inefficient is smaller in the absence of transfer payments compared to the case with transfers.

Turning to the case where the HO uses two DRGs and upcoding is possible, note that the absence of a transfer payment does not affect the value of upcoding – the difference in payment between a low-complexity patient who is upcoded and that of a low-complexity patient who is not remains the same whether transfer payments are used or not. Therefore, the condition (7)

$$\beta < \frac{h\delta^{e3}}{h\delta^{e1} + (\delta^{e1} - \delta^{e3})(1-h)\bar{\alpha}}$$

continues to ensure that providers will find upcoding profitable and will engage in it fully. Define

$$\begin{aligned} v'_2 &:= v_2 + \frac{h\bar{\gamma}}{1-h\bar{\gamma}}(R_\delta(\delta^{e2}) + R_c(c^{e1})), \\ u'_2 &:= u_2 - h\bar{\gamma}(R_\delta(\delta^{e3}) + R_c(c^*)). \end{aligned}$$

The quantity v'_2 is the profit of a provider who is paid according to yardstick competition and chooses $(\alpha, 0, c^*, \delta^{e3})$ when all other providers choose $(\bar{\alpha}, \bar{\gamma}, c^{e1}, \delta^{e2})$ when transfer payments are not allowed, and vice versa for u'_2 . We will call the cherry-picking-upcoding-best (CPU-best) costs comparable to upcoding-best costs if any of the following conditions hold:

- $\lambda q(c^* + \delta^{e3} - c^{e1} - \delta^{e2} \frac{h(1-\bar{\gamma}) + (1-h)\bar{\alpha}\beta}{h(1-\bar{\gamma}) + (1-h)\bar{\alpha}}) \geq \frac{R_\delta(\delta^{e2}) + R_c(c^{e1})}{\lambda q(1-h\bar{\gamma})}$,
- $v'_2 \leq 0$ or $u'_2 \leq 0$.

Proposition 4n *If both patient cherry picking and upcoding are possible, if the HO implements yardstick competition based on two DRGs without transfer payments, then there exists a unique Nash equilibrium:*

- *If CPU-best costs are ‘comparable’ to upcoding-best costs (as defined above), then the equilibrium is symmetric. Providers upcode and drop as many patients as possible and invest in CPU-best costs (i.e., all providers choose $(\bar{\alpha}, \bar{\gamma}, c^{e1}, \delta^{e2})$). Furthermore, there is underinvestment in cost reduction compared to first best for the minor condition ($c^{e1} > c^*$) and if $\beta < \frac{h}{1-h}\bar{\gamma}\bar{\alpha}$ then there is also underinvestment for the major condition ($\delta^{e2} > \delta^*$), otherwise there is overinvestment.*

- *Otherwise, the equilibrium is asymmetric. $N - \theta'_2$ providers upcode and drop as many patients as possible and choose CPU-best costs (i.e., these providers choose $(\bar{\alpha}, \bar{\gamma}, c^{e1}, \delta^{e2})$) and θ'_2 providers upcode as many patients as possible but do not engage in cherry picking and invest in upcoding-best costs (i.e., these providers choose $(\bar{\alpha}, 0, c^*, \delta^{e3})$). The number of low-cost providers θ_2 is the only integer in the interval $\left[\frac{(N-1)v'_2}{u'_2 + v'_2}, \frac{Nv'_2 + u'_2}{u'_2 + v'_2} \right]$. Furthermore, $\theta'_2 \geq \theta_2$.*

The proposition demonstrates that the presence of upcoding strengthens incentives to drop high-complexity patients, even in the absence of a transfer payment.

Turning to yardstick competition with input statistics, we find it convenient to define

$$\bar{\kappa}' := \min\left\{c^{e1} + \delta^{e1} - c^* - \delta^* - \frac{R_\delta(\delta^*) + R_c(c^*)}{\lambda q}, \frac{u'_1}{\lambda q h \bar{\gamma}}\right\}$$

Proposition 6n *Under the two-DRG payment scheme with input statistics described in §6, with $\kappa_i = \kappa$ for all i then there exists a unique Nash equilibrium:*

- *If $\kappa > \bar{\kappa}'$, the equilibrium is symmetric in which all providers choose first-best actions (i.e., $(0, 0, c^*, \delta^*)$).*
- *If $0 \leq \kappa \leq \bar{\kappa}'$, the equilibrium is asymmetric. No provider engages in upcoding and $N - \theta_4$ providers drop as many patients as possible and choose cherry-picking-best costs (i.e., these providers choose $(0, \bar{\gamma}, c^{e1}, \delta^{e1})$ and $c^{e1} > c^*$, $\delta^{e1} > \delta^*$) and θ'_4 providers do not engage in cherry picking and invest in cost reduction as much as in first best (i.e., these providers choose $(0, 0, c^*, \delta^*)$). The number of efficient providers $\theta'_4 \geq \theta'_1$ and is non-decreasing in κ' . Furthermore, $\bar{\kappa}' \leq \bar{\kappa}$.*

The proposition suggests that, compared to the case where transfer payments are allowed, in their absence yardstick competition based on input statistics continues to be effective in eliminating the problem of upcoding, and is even more effective in alleviating the problem of cherry picking.

Collectively, the results of this appendix suggest that the absence of transfer payments helps reduce the problem of cherry picking but does nothing to alleviate the problem of upcoding. In a setting where investment costs are relatively small compared to variable costs, the results of the main paper persist in the absence of transfer fees. According to CMS, for the fiscal year 2022, “operating base payments,” which are meant to reimburse “labor and supply costs,” were estimated to \$6,122 and “capital base payments,” which are meant to reimburse capital expenditures such as “costs for depreciation, interest, rent, and property related insurance and taxes,” were estimated to \$473.⁸ Therefore, upcoding and patient dropping are likely to remain an issue in the hospital reimbursement setting.

Proofs of results appearing in Appendix 2

Proof of Proposition 7: Under cost-of-service regulation, and irrespective of the number of DRGs used, the provider’s payoff is equal to zero. Therefore, the provider resorts to selecting the status quo – zero upcoding and cherry picking and zero investment in cost reduction (i.e., the provider will choose $\alpha = \gamma = 0$, $c = c_0$, $\delta = \delta_0$). \square

Proof of Proposition 2d: Under cost-of-service regulation the provider’s profit is zero, therefore the provider has no incentive to downcode. The proof proceeds as that of Proposition 7. \square

⁸ See https://www.medpac.gov/wp-content/uploads/2021/11/medpac_payment_basics_21_hospital_final_sec.pdf

Proof of Proposition 4d: The derivative of a provider's profit with respect to ζ_i is given by

$$\frac{\partial}{\partial \zeta_i} \pi_i = \lambda q h (1 - \gamma_i) (\bar{c}_{Mi} - \bar{c}_{mi}).$$

This implies that $\frac{\partial}{\partial \zeta_i} \pi_i > 0$ if $\bar{c}_{Mi} - \bar{c}_{mi} > 0$ which suggests that providers will find it optimal not to downcode ($\zeta_i = 1$). Note that

$$\begin{aligned} \bar{c}_{Mi} - \bar{c}_{mi} &= \frac{1}{N-1} \sum_{j \neq i} \delta_j \left[\frac{h(1-\gamma_j)\zeta_j + (1-h)\alpha_j\beta}{h(1-\gamma_j)\zeta_j + (1-h)\alpha_j} - \frac{h(1-\gamma_j)(1-\zeta_j)}{h(1-\gamma_j)(1-\zeta_j) + (1-h)(1-\alpha_j)} \right] \\ &= \frac{1}{N-1} \sum_{j \neq i} \delta_j \left[h(1-\gamma_j) \left(\frac{\zeta_j}{h(1-\gamma_j)\zeta_j + (1-h)\alpha_j} - \frac{1-\zeta_j}{h(1-\gamma_j)(1-\zeta_j) + (1-h)(1-\alpha_j)} \right) \right. \\ &\quad \left. + \frac{(1-h)\alpha_j\beta}{h(1-\gamma_j)\zeta_j + (1-h)\alpha_j} \right]. \end{aligned}$$

This is positive if $\frac{\zeta_j}{h(1-\gamma_j)\zeta_j + (1-h)\alpha_j} - \frac{1-\zeta_j}{h(1-\gamma_j)(1-\zeta_j) + (1-h)(1-\alpha_j)}$, which is always the case if $\zeta_i > \alpha_i$. Since $\zeta_i \geq \bar{\zeta}$ and $\alpha_i \leq \bar{\alpha}$, then $\bar{c}_{Mi} - \bar{c}_{mi} > 0$ if $\bar{\zeta} > \bar{\alpha}$. The rest of the proof proceeds as Proposition 4. \square

Proof of Proposition 6d: The derivative of a provider's profit with respect to ζ_i is given by

$$\frac{\partial}{\partial \zeta_i} \pi_i = \lambda q h (1 - \gamma_i) (\bar{c}_{Mi} - \bar{c}_{mi} + \kappa - \phi_i) > 0.$$

Since $\phi_i - \kappa < \bar{c}_{Mi} - \bar{c}_{mi}$, this implies that if $\bar{c}_{Mi} - \bar{c}_{mi} > 0$ then $\frac{\partial}{\partial \zeta_i} \pi_i > 0$ which suggests that providers will find it optimal not to downcode ($\zeta_i = 1$). Sufficient conditions for $\bar{c}_{Mi} - \bar{c}_{mi} > 0$ are derived in the proof of Proposition 5d and the rest of the proof proceeds as that of Proposition 6. \square

Proof of Proposition 8: This proof continues from the proof of Proposition 2, where we have shown that if $c^* + \delta^* > c^{e1} + \delta^{e1} \frac{h(1-\bar{\gamma})}{1-h\bar{\gamma}}$ then there exists a symmetric equilibrium where all providers choose (γ, c, δ) . In contrast, if this condition is not satisfied, then a provider that knows all providers choose (γ, c, δ) may find it profitable to deviate to $(0, c^*, \delta^*)$. For this to be the case, the profit of the provider who deviates to $(0, c^*, \delta^*)$ when all other providers choose $(\bar{\gamma}, c^{e1}, \delta^{e1})$ must be positive (as the profit associated with not deviating is zero). After some algebra, this condition can be written as $v_0 \geq 0$. If this is not satisfied (i.e. if $v_0 < 0$), the equilibrium is symmetric and is as given in Proposition 2. Otherwise (i.e. if $v_0 \geq 0$) if no symmetric equilibrium exists.

Turning to the case of asymmetric equilibria, in the Proof of Proposition 2 we have ruled out their existence if $c^* + \delta^* > c^{e1} + \delta^{e1} \frac{h(1-\bar{\gamma})}{1-h\bar{\gamma}}$. If this condition is not satisfied then an asymmetric equilibrium will exist if the profit of the θ_0 low-cost providers and the profit of the $N - \theta_0$ high-cost providers are non-negative. The profit of a low-cost provider will be $\frac{N-\theta_0}{N-1} v_0$. Similarly, the profit of a high-cost provider will be given by $\frac{\theta_0}{N-1} u_0$. Therefore, for the asymmetric equilibrium to exist it

must be the case that $v_0 \geq 0$ and $u_0 \geq 0$. Note that $u_0 > 0$. To see this, note that u_0 can be written as

$$\begin{aligned} u_0 = & [-\lambda q ((1 - \bar{\gamma}h)c^{e1} + (1 - \bar{\gamma})h\delta^{e1}) - R_c(c^{e1}) - R_\delta(\delta^{e1})] \\ & + [\lambda q ((1 - \bar{\gamma}h)c^* + (1 - \bar{\gamma})h\delta^*) - R_c(c^*) - R_\delta(\delta^*)] \\ & + \bar{\gamma}h(1 - h)\delta^*. \end{aligned}$$

Note that the expression $[-\lambda q [(1 - \bar{\gamma}h)c + (1 - \bar{\gamma})h\delta] - R_c(c) - R_\delta(\delta)]$ is maximized at c^{e1}, δ^{e1} , therefore the sum of the terms in the first two brackets is positive. The third term is also positive, which implies that $u_0 > 0$. For $v_0 \geq 0$ it must be the case that

$$\lambda q(c^{e1} - c^* + h[\frac{1 - \bar{\gamma}}{1 - \bar{\gamma}h}\delta^{e1} - \delta^*]) \geq R_c(c^*) + R_\delta(\delta^*) - R_c(c^{e1}) - R_\delta(\delta^{e1}) \quad (30)$$

We will next determine the value of θ_0 . If one of the θ_0 low-cost providers was to deviate and become a high-cost provider, then their profit would be given by $\frac{\theta_0 - 1}{N - 1}u_0$. If one of the $N - \theta_0$ high-cost providers was to deviate and become a low cost provider it would be $\frac{N - \theta_0 - 1}{N - 1}v_0$. In the asymmetric equilibrium it must be the case that these deviations are not profitable. Therefore, $\frac{N - \theta_0}{N - 1}v_0 > \frac{\theta_0 - 1}{N - 1}u_0$ and $\frac{\theta_0}{N - 1}u_0 \geq \frac{N - \theta_0 - 1}{N - 1}v_0$. These conditions imply that θ_0 satisfies $\frac{(N - 1)v_0}{u_0 + v_0} \leq \theta_0 \leq \frac{Nv_0 + u_0}{u_0 + v_0}$. If condition (30) is satisfied, this interval contains exactly 1 integer as the difference $\frac{Nv_0 + u_0}{u_0 + v_0} - \frac{(N - 1)v_0}{u_0 + v_0} = 1$. Furthermore, this integer is always less than N .

We will then compare θ_1 and θ_0 . Note that $v_1 - v_0 = \lambda q h(1 - h)\frac{\bar{\gamma}}{1 - \bar{\gamma}h}\delta^{e1} > 0$, therefore $v_1 > v_0$. Also, $u_1 - u_0 = -\lambda q \bar{\gamma}h(1 - h)\delta^* < 0$, therefore $u_1 < u_0$. It follows that $\frac{v_1}{u_1} > \frac{v_0}{u_0}$, which implies that $\frac{v_1}{v_1 + u_1} > \frac{v_0}{v_0 + u_0}$. Since θ_1 is the first integer higher than $(N - 1)\frac{u_1}{v_1 + u_1}$ and θ_0 is the first integer higher than $(N - 1)\frac{v_0}{v_0 + u_0}$, it follows that $\theta_1 \geq \theta_0$. \square

Proof of Proposition 2c: In the absence of upcoding ($\alpha_i = 0$), under the yardstick competition scheme with a single DRG the profit of provider i is given by

$$\pi_i(c_i, \delta_i, 0, \gamma_i) = \lambda q [(1 - h\gamma_i)\bar{c}_i - \delta_i h(1 - \gamma_i) - c(1 - h\gamma_i)] - R_c(c_i) - R_\delta(\delta_i) - E_\gamma(\gamma_i) + \bar{R}_i,$$

where $\bar{c}_i = \frac{1}{N - 1} \sum_{j \neq i} [c_j + \delta_j \frac{h(1 - \gamma_j)}{1 - h\gamma_j}]$, and $\bar{R}_i := \frac{1}{N - 1} \sum_{j \neq i} [R_c(c_j) + R_\delta(\delta_j)]$ as defined in §4. The derivatives of the profit function of provider i are given by:

$$\frac{\partial}{\partial \gamma_i} \pi_i = \lambda q h(c_i + \delta_i - \bar{c}_i) - E'_\gamma(\gamma_i), \quad (31)$$

$$\frac{\partial}{\partial c_i} \pi_i = -\frac{d}{dc} R_c(c_i) - \lambda q(1 - h\gamma_i), \quad (32)$$

$$\frac{\partial}{\partial \delta_i} \pi_i = -\frac{d}{d\delta} R_\delta(\delta_i) - \lambda q h(1 - \gamma_i). \quad (33)$$

In any equilibrium outcome, the last two conditions, which do not depend on the actions of other providers, will have to be equal to zero for all providers. Otherwise the provider for whom one of these conditions is not zero could increase their profit by changing c_i or δ_i .

Consider a symmetric equilibrium where all providers choose $(\gamma^c, c^c, \delta^c)$. From (31)–(33), these values should satisfy $\lambda q h \delta^c \frac{1-h}{1-\gamma^c h} - E'_\gamma(\gamma^c) = 0$, $-R'_c(c^c) - \lambda q(1-h\gamma^c) = 0$, and $R'_\delta(\delta^c) - \lambda q h(1-\gamma^c) = 0$. Consider a provider that chooses $(\gamma_i, c_i, \delta_i)$ when all other providers choose $(\gamma^c, c^c, \delta^c)$. We will show that this provider will also choose $(\gamma^c, c^c, \delta^c)$. The derivative of the provider's profit is given by $\frac{\partial}{\partial \gamma_i} \pi_i = \lambda q h(c_i + \delta_i - c^c - h \delta^c \frac{1-\gamma^c}{1-\gamma^c h}) - E'_\gamma(\gamma_i)$. Clearly, the derivative is zero if $\gamma_i = \gamma^c$. A sufficient condition for this to be the unique maximizer is $\frac{\partial^2}{\partial \gamma_i^2} \pi_i = (\lambda q h)^2 \left(\frac{1}{R''_\delta(\delta^c)} + \frac{1}{R''_c(c^c)} \right) - E''_\gamma(\gamma^c) < 0$. A necessary condition for the solution to be in the interval $(0, 1]$ is $\lambda q h(c^* + \delta^* - c^c - h \delta^c \frac{1-\gamma^c}{1-\gamma^c h}) - E'_\gamma(0) > 0$. If these conditions hold, $(\gamma^c, c^c, \delta^c)$ is the unique symmetric equilibrium. Note that these conditions are a subset of Conditions A.

The proof above does not rule out asymmetric equilibria, which may exist. \square

Proof of Proposition 3c: In the absence of upcoding ($\alpha_i = 0$), under the yardstick competition scheme with two DRGs the profit of provider i is given by

$$\pi_i(c_i, \delta_i, 0, \gamma_i) = \lambda q [h(1-\gamma_i)(\bar{c}_{Mi} - \delta_i - c_i) + (1-h)(\bar{c}_{mi} - c_i)] - R_c(c_i) - R_\delta(\delta_i) - E_\gamma(\gamma_i) + \bar{R}_i,$$

where $\bar{c}_{Mi} := \frac{1}{N-1} \sum_{j \neq i} [c_j + \delta_j]$, $\bar{c}_{mi} := \frac{1}{N-1} \sum_{j \neq i} c_j$, and $\bar{R}_i := \frac{1}{N-1} \sum_{j \neq i} [R_c(c_j) + R_\delta(\delta_j)]$ as defined in §4. The derivatives of the profit function of provider i are given by:

$$\frac{\partial}{\partial \gamma_i} \pi_i = \lambda q h(c_i + \delta_i - \bar{c}_{Mi}) - E'_\gamma(\gamma_i), \quad (34)$$

$$\frac{\partial}{\partial c_i} \pi_i = -\frac{d}{dc} R_c(c_i) - \lambda q(1-h\gamma_i), \quad (35)$$

$$\frac{\partial}{\partial \delta_i} \pi_i = -\frac{d}{d\delta} R_\delta(\delta_i) - \lambda q h(1-\gamma_i). \quad (36)$$

In any equilibrium outcome, the last two conditions have to be equal to zero for all providers. Otherwise the provider for whom one of these conditions is not zero could increase their profit by changing c_i or δ_i .

Consider a symmetric equilibrium where all providers choose $(\gamma^c, c^c, \delta^c)$. Equation (34) implies that $\frac{\partial}{\partial \gamma_i} \pi_i = -\frac{d}{d\gamma_i} E_\gamma(\gamma_i) < 0$, therefore $\gamma^c = 0$, which also implies $c^c = c^*$, $\delta^c = \delta^*$. Consider a provider that chooses $(\gamma_i, c_i, \delta_i)$ when all other providers choose $(0, c^*, \delta^*)$. We will show that this provider will also choose $(0, c^*, \delta^*)$. The derivative of the provider's profit is given by $\frac{\partial}{\partial \gamma_i} \pi_i = \lambda q h(c_i + \delta_i - c^* - \delta^*) - \frac{d}{d\gamma_i} E_\gamma(\gamma_i)$. At $\gamma_i = 0$ this derivative is negative. Furthermore, if $(\lambda q h)^2 \left(\frac{1}{R''_\delta(\delta^c)} + \frac{1}{R''_c(c^c)} \right) - E''_\gamma(\gamma^c) < 0$ for all γ_c such that $-R'_c(c^c) - \lambda q(1-h\gamma^c) = 0$, $R'_\delta(\delta^c) - \lambda q h(1-\gamma^c) = 0$, then $\frac{\partial}{\partial \gamma_i} \pi_i < 0$ for all γ_c . This implies that no profitable deviation exists and $(0, c^*, \delta^*)$ is the unique symmetric equilibrium. Note that these conditions are a subset of Conditions A.

The proof above does not rule out asymmetric equilibria, which may exist. \square

Proof of Proposition 4c: Under the yardstick competition scheme with two DRGs the profit of provider i is given by

$$\begin{aligned}\pi_i(c_i, \delta_i, \alpha_i, \gamma_i) = & \lambda q \{ [h(1 - \gamma_i) + (1 - h)\alpha_i] \bar{c}_{Mi} + (1 - h)(1 - \alpha_i) \bar{c}_{mi} \\ & - [(h(1 - \gamma_i) + (1 - h)\alpha_i\beta)\delta_i + (1 - h\gamma_i)c_i] \} \\ & - R_c(c) - R_\delta(\delta) - E_\gamma(\gamma_i) - E_\alpha(\alpha_i) + \bar{R}_i,\end{aligned}$$

where $\bar{c}_{Mi} := \frac{1}{N-1} \sum_{j \neq i} \left[c_j + \delta_j \frac{h(1-\gamma_j) + (1-h)\alpha_j\beta}{h(1-\gamma_j) + (1-h)\alpha_j} \right]$, $\bar{c}_{mi} := \frac{1}{N-1} \sum_{j \neq i} c_j$, and $\bar{R}_i := \frac{1}{N-1} \sum_{j \neq i} [R_c(c_j) + R_\delta(\delta_j)]$ as defined in §4. The derivatives of the profit function of provider i are given by:

$$\frac{\partial}{\partial \alpha_i} \pi_i = \lambda q (1 - h) (\bar{c}_{Mi} - \bar{c}_{mi} - \beta \delta_i) - E'_\alpha(\alpha_i), \quad (37)$$

$$\frac{\partial}{\partial \gamma_i} \pi_i = \lambda q h (c_i + \delta_i - \bar{c}_{Mi}) - E'_\gamma(\gamma_i), \quad (38)$$

$$\frac{\partial}{\partial c_i} \pi_i = -\frac{d}{dc} R_c(c_i) - \lambda q (1 - h \gamma_i), \quad (39)$$

$$\frac{\partial}{\partial \delta_i} \pi_i = -\frac{d}{d\delta} R_\delta(\delta_i) - \lambda q [h(1 - \gamma_i) + (1 - h)\alpha_i\beta]. \quad (40)$$

In any equilibrium outcome, the last two conditions have to be equal to zero for all providers. Otherwise the provider for whom one of these conditions is not zero could increase their profit by changing c_i or δ_i .

Consider a symmetric equilibrium where all providers choose $(\alpha^d, \gamma^d, c^d, \delta^d)$. From (37)–(40), these values should satisfy $\lambda q h \delta^d \frac{h(1-\beta)(1-\gamma^d)}{h(1-\gamma^d) + (1-h)\alpha^d} - E'_\alpha(\alpha^d) = 0$, $\lambda q h \delta^d \frac{(1-h)(1-\beta)\alpha^d}{h(1-\gamma^d) + (1-h)\alpha^d} - E'_\gamma(\gamma^d) = 0$, $-R'_d(c^d) - \lambda q (1 - h \gamma^d) = 0$, and $R'_\delta(\delta^d) - \lambda q (h(1 - \gamma^d) + (1 - h)\beta \alpha^d) = 0$. Consider the payoff of a provider that chooses $(\alpha_i, \gamma_i, c_i, \delta_i)$ when all other providers choose $(\alpha^d, \gamma^d, c^d, \delta^d)$. We will show that this provider will also choose $(\alpha^d, \gamma^d, c^d, \delta^d)$. The derivative of the provider's profit is given by $\frac{\partial}{\partial \alpha_i} \pi_i = \lambda q h (\delta^d \frac{h(1-\gamma^d) + (1-h)\alpha^d\beta}{h(1-\gamma^d) + (1-h)\alpha^d} - \beta \delta_i) - \frac{d}{d\alpha_i} E_\alpha(\alpha_i)$ and $\frac{\partial}{\partial \gamma_i} \pi_i = \lambda q h (c_i + \delta_i - c^d - \delta^d \frac{h(1-\gamma^d) + (1-h)\alpha^d\beta}{h(1-\gamma^d) + (1-h)\alpha^d}) - \frac{d}{d\gamma_i} E_\gamma(\gamma_i)$. Clearly, both derivatives are zero if $\alpha_i = \alpha^d, \gamma_i = \gamma^d$. A sufficient condition for this to be the unique maximizer is the Hessian of the profit with respect to α_i and γ_i (when $-R'_d(c_i) - \lambda q (1 - h \gamma_i) = 0$, and $R'_\delta(\delta_i) - \lambda q (h(1 - \gamma_i) + (1 - h)\beta \alpha_i) = 0$) to be negative definite. Necessary conditions for the solution to be in the interval $(0, 1]$ are $\lambda q h (\delta^d \frac{h(1-\gamma^d) + (1-h)\alpha^d\beta}{h(1-\gamma^d) + (1-h)\alpha^d} - \beta \delta^*) - E'_\alpha(0) > 0$ and $\lambda q h (c^* + \delta^* - c^d - \delta^d \frac{h(1-\gamma^d) + (1-h)\alpha^d\beta}{h(1-\gamma^d) + (1-h)\alpha^d}) - E'_\gamma(0) > 0$. If these conditions hold, $(\alpha^d, \gamma^d, c^d, \delta^d)$ is the unique symmetric equilibrium. These conditions are summarized by Conditions B. Furthermore, from (39) and since $\gamma^d > 0$ it follows that $c^d > c^*$. From (40), if $h(1 - \gamma^d) + (1 - h)\alpha^d\beta > h$ there will be overinvestment in the major treatment $\delta^d > \delta^*$, otherwise underinvestment.

The proof above does not rule out asymmetric equilibria, which may exist. \square

Proof of Proposition 6c: Under the yardstick competition scheme with two DRGs with input statistics the profit of provider i is given by

$$\begin{aligned}\pi_i(c_i, \delta_i, \alpha_i, \gamma_i) = & \lambda q \{ [h(1 - \gamma_i) + (1 - h)\alpha_i] \bar{c}_{Mi} + (1 - h)(1 - \alpha_i) \bar{c}_{mi} \\ & - [(h(1 - \gamma_i) + (1 - h)\alpha_i \beta) \delta_i + (1 - h\gamma_i) c_i] \} \\ & - R_c(c) - R_\delta(\delta) - E_\gamma(\gamma_i) - E_\alpha(\alpha_i) + \bar{R}_i + \kappa_i(M_i - \bar{M}_i) + \phi_i(m_i - \bar{m}_i),\end{aligned}$$

where $M_i = \lambda q(h(1 - \gamma_i) + (1 - h)\alpha_i)$, $m_i = \lambda q(1 - h)(1 - \alpha_i)$, $\kappa_i \geq 0$, $\phi_i - \kappa_i > \bar{\delta}_i - \beta \delta^{e3}$, $\bar{\delta}_i := \bar{c}_{Mi} - \bar{c}_{mi}$, $\bar{c}_{Mi} = \frac{1}{N-1} \sum_{j \neq i} \left[c_j + \delta_j \frac{h(1-\gamma_j) + (1-h)\alpha_j \beta}{h(1-\gamma_j) + (1-h)\alpha_j} \right]$, $\bar{c}_{mi} = \frac{1}{N-1} \sum_{j \neq i} c_j$, and $\bar{R}_i = \frac{1}{N-1} \sum_{j \neq i} [R_c(c_j) + R_\delta(\delta_j)]$.

The derivatives of the profit function of provider i are given by:

$$\frac{\partial}{\partial \alpha_i} \pi_i = \lambda q(1 - h)(-\beta \delta_i + \bar{\delta}_i + \kappa_i - \phi_i) - E'_\alpha(\alpha_i), \quad (41)$$

$$\frac{\partial}{\partial \gamma_i} \pi_i = \lambda q h(c_i + \delta_i - \bar{c}_{Mi} - \kappa_i) - E'_\gamma(\gamma_i), \quad (42)$$

$$\frac{\partial}{\partial c_i} \pi_i = -\frac{d}{dc} R_c(c_i) - \lambda q(1 - h\gamma_i), \quad (43)$$

$$\frac{\partial}{\partial \delta_i} \pi_i = -\frac{d}{d\delta} R_\delta(\delta_i) - \lambda q[h(1 - \gamma_i) + (1 - h)\alpha_i \beta]. \quad (44)$$

In any equilibrium, $\frac{\partial}{\partial \alpha_i} \pi_i < 0$, which implies that $\alpha_i = 0$ for all i . In any symmetric equilibrium, $\frac{\partial}{\partial \gamma_i} \pi_i < 0$, therefore a symmetric equilibrium candidate is $(0, 0, c^*, \delta^*)$. The rest of the proof proceeds identically to the Proof of Proposition 4c. \square

Proof of Proposition 2n: The proof proceeds similarly to the proofs of Propositions 2 and 8, with the updated notation for cost being comparable and v'_0 . The result that $\theta'_0 \geq \theta_0$ follows from the observation that $v'_0 > v_0$ and $u'_0 < u_0$. \square

Proof of Proposition 3n: The proof proceeds similarly to the proof of Proposition 3, with the updated notation for cost being comparable. The result that $\theta'_1 \geq \theta_1$ follows from the observation that $v'_1 > v_1$ and $u'_1 < u_1$. \square

Proof of Proposition 4n: The proof proceeds similarly to the proof of Proposition 4, with the updated notation for cost being comparable. The result that $\theta'_2 \geq \theta_2$ follows from the observation that $v'_2 > v_2$ and $u'_2 < u_2$. \square

Proof of Proposition 6n: The proof proceeds similarly to the proof of Proposition 6. The result that $\bar{\kappa}' \leq \bar{\kappa}$ follows from the observation that $u'_1 < u_1$ and $R_\delta(\delta^*) + R_c(c^*) > 0$. \square

Appendix 3: The case of two asymmetric providers

In the main analysis we assume that providers are symmetric and have argued that any scheme that achieves first best in the case of symmetric providers can be modified to account for provider

heterogeneity based on factors that are observable by the HO and exogenous to the provider. In this section we extent the analysis to examine the case where there is heterogeneity that the HO does not account for in the reimbursement scheme (e.g., because it is not observable or because the HO is not sophisticated enough to make the necessary adjustments). Naturally, the results in this case will depend on the exact way in which providers are asymmetric, but to gain insights on the impact of asymmetry on equilibrium outcomes we will focus on the case where there are two providers ($N = 2$) that differ in the number of patients and in the proportion of high-complexity patients they treat. More specifically we assume that $\lambda_1 q_1 = \lambda q$ and $\lambda_2 q_2 = \eta \lambda q$, where $\eta > 1$ and $h_1 = h$, $h_2 = \xi h$ where $\xi > 1$. This specification ensures that Provider 2 treats more patients in total and more high-complexity patients than Provider 1. Naturally, we require that $h\xi < 1$ (i.e., the proportion of high complexity patients of Provider 2 is less than 1). We continue to assume that providers have access to the same cost reduction technology (i.e., $R_c(\cdot)$ and $R_\delta(\cdot)$ are identical) but, given their different size, they may decide to invest differently in cost reduction. Providers may choose to lemon drop a proportion $\gamma_i \leq \bar{\gamma}$ of their high-complexity patients and upcode a proportion $\alpha_i \leq \bar{\alpha}$ of their low-complexity patients.

In this case, the first best actions of Provider 1 are given by the same conditions as the main text:

$$\gamma_1^* = \alpha_1^* = 0, -\frac{d}{dc}R_c(c_1^*) = \lambda q, -\frac{d}{d\delta}R_\delta(\delta_1^*) = \lambda q h,$$

while the first best actions of Provider 2 become

$$\gamma_2^* = \alpha_2^* = 0, -\frac{d}{dc}R_c(c_2^*) = \lambda q \eta, -\frac{d}{d\delta}R_\delta(\delta_2^*) = \lambda q h \eta \xi.$$

Given the properties of $R_c(\cdot)$ and $R_\delta(\cdot)$ and since $\eta > 1, \xi > 1$, it follows that $c_1^* > c_2^*$ and $\delta_1^* > \delta_2^*$ (i.e., since the larger Provider 2 treats more patients, they find it optimal to invest more heavily in cost reduction and operate at lower cost levels compared to the smaller Provider 1).

Under cost-based yardstick competition, if the HO uses one DRG the fees per patient for each provider are given by $p_{M1} = p_{m1} = \bar{c}_1$ and $p_{M2} = p_{m2} = \bar{c}_2$, where \bar{c}_i is the average treatment cost of the other provider:

$$\begin{aligned} \bar{c}_1 &:= c_2 + \delta_2 \frac{h\xi(1 - \gamma_2) + (1 - h\xi)\alpha_2\beta}{1 - h\xi\gamma_2}, \\ \bar{c}_2 &:= c_1 + \delta_1 \frac{h(1 - \gamma_1) + (1 - h)\alpha_1\beta}{1 - h\gamma_1}. \end{aligned} \tag{45}$$

If the HO breaks the condition into two distinct DRGs the fees per patient episode for the major and the minor condition for each provider are given by $p_{Mi} = \bar{c}_{Mi}$ and $p_{mi} = \bar{c}_{mi}$, respectively, where

$$\bar{c}_{M1} := c_2 + \delta_2 \frac{h\xi(1 - \gamma_2) + (1 - h\xi)\alpha_2\beta}{h\xi(1 - \gamma_2) + (1 - h\xi)\alpha_2} \text{ and } \bar{c}_{m1} := c_2, \tag{46}$$

$$\bar{c}_{M2} := c_1 + \delta_1 \frac{h(1-\gamma_1) + (1-h)\alpha_1\beta}{h(1-\gamma_1) + (1-h)\alpha_1} \text{ and } \bar{c}_{m2} := c_1.$$

In both cases, the transfer payment for each provider is given by

$$\begin{aligned} \bar{R}_1 &:= R_c(c_2) + R_\delta(\delta_2), \\ \bar{R}_2 &:= R_c(c_1) + R_\delta(\delta_1). \end{aligned} \tag{47}$$

We will analyse the equilibrium outcome for different cases below. In all the analysis we focus on the provider's actions and how they compare to first-best actions. In doing so, we will not impose participation constraints, i.e., we will allow equilibria where one of the two providers may be receiving a negative rent. (If participation constraints apply then in those cases where at least one of the providers receives a negative rent no equilibrium would exist.)

We start by looking at the case where there is no upcoding or cherry picking. The following proposition summarizes the equilibrium actions of the providers.

Proposition 1A: *Under yardstick competition, and irrespective of the number of DRGs used (one or two), in the absence of patient lemon dropping and upcoding ($\bar{\gamma} = \bar{\alpha} = 0$) providers invest in cost reduction optimally.*

Proposition 1A shows that yardstick competition scheme is effective in providing first-best incentives for cost reduction even if the providers are asymmetric. This is the case even if the HO does not adjust for this asymmetry.

We will next examine the case where providers may lemon drop a proportion $\gamma_i \leq \bar{\gamma}$ patients but cannot upcode. We find it convenient to define cherry-picking-best costs $(c_1^{e1}, \delta_1^{e1})$ for Provider 1 as the solutions to

$$-\frac{d}{dc}R_c(c_1^{e1}) = \lambda q(1 - \bar{\gamma}h), -\frac{d}{d\delta}R_\delta(\delta_1^{e1}) = \lambda qh(1 - \bar{\gamma}),$$

and $(c_2^{e1}, \delta_2^{e1})$ for Provider 2 as the solutions to

$$-\frac{d}{dc}R_c(c_2^{e1}) = \lambda q\eta(1 - \bar{\gamma}\xi h), -\frac{d}{d\delta}R_\delta(\delta_2^{e1}) = \lambda q\eta\xi h(1 - \bar{\gamma}).$$

Due to economies of scale (i.e., $R_c''(.) > 0$, $R_\delta''(.) > 0$), cherry-picking-best costs are higher than first best costs for both providers (i.e., $c_i^{e1} > c_i^*$ and $\delta_i^{e1} > \delta_i^*$).

Proposition 2A: *In the absence of upcoding ($\bar{\alpha} = 0$), if the HO implements yardstick competition based on a single DRG, then*

- *There exists a Nash equilibrium where Provider 1 drops as many patients as possible ($\gamma_1 = \bar{\gamma}$), chooses cherry-picking-best costs $(c_1^{e1}, \delta_1^{e1})$. If $c_1^{e1} + \delta_1^{e1}h\frac{1-\bar{\gamma}}{1-h\bar{\gamma}} < c_2^* + \delta_2^*$, then Provider 2 also drops as many patients as possible ($\gamma_2 = \bar{\gamma}$) and chooses cherry-picking-best costs $(c_2^{e1}, \delta_2^{e1})$. If $c_2^* + \delta_2^* < c_1^{e1} + \delta_1^{e1}h\frac{1-\bar{\gamma}}{1-h\bar{\gamma}}$ then Provider 2 does not drop any patients ($\gamma_2 = 0$) and invest optimally in cost reduction (i.e. chooses costs c_2^*, δ_2^*). Otherwise, Provider 2 drops an intermediate proportion γ_2^v of patients, where $0 < \gamma_2^v < \bar{\gamma}$ and chooses intermediate costs (c_2^v, δ_2^v) , where $c_2^{e1} > c_2^v > c_2^*$, $\delta_2^{e1} > \delta_2^v > \delta_2^*$.*

• If $c_1^* + \delta_1^* < c_2^{e1} + \delta_2^{e1} \frac{h\xi(1-\bar{\gamma})}{1-h\xi\bar{\gamma}}$ then there exists a second Nash equilibrium in which Provider 2 drops as many patients as possible ($\gamma_2 = \bar{\gamma}$), and chooses cherry-picking-best costs (c_2^{e1}, δ_2^{e1}). If $c_1^* + \delta_1^* h < c_2^{e1} + \delta_2^{e1}$ then Provider 1 does not drop any patients ($\gamma_1 = 0$) and invests optimally in cost reduction (i.e. chooses costs c_1^*, δ_1^*), otherwise Provider 1 drops an intermediate proportion γ_1^m of patients, where $0 < \gamma_1^m < \bar{\gamma}$ and chooses intermediate costs (c_1^m, δ_1^m) , where $c_1^{e1} > c_1^m > c_1^*$, $\delta_1^{e1} > \delta_1^m > \delta_1^*$.

This proposition is equivalent to Propositions 2 and 8 of the symmetric case. It shows that if lemon dropping is possible then at least one or both providers will engage in it. As a result of lemon dropping, providers will also underinvest in cost reduction. More specifically, the symmetric equilibrium of Proposition 2, where both providers exercised maximum lemon dropping, still exists and is “symmetric” in the sense that providers still exercise maximum lemon dropping but it is no longer symmetric in the sense that providers choose to operate at different cost levels. Similarly, the asymmetric equilibrium of Proposition 8, where only one provider engaged in lemon dropping, still exists but in this case the provider who engages in lemon dropping could be the larger or, under some conditions, the smaller provider. It is notable that there are no model parameters where the equilibrium does not involve at least some lemon dropping.

We turn to the case where the HO uses two DRGs next.

Proposition 3A: *In the absence of upcoding ($\bar{\alpha} = 0$), if the HO implements yardstick competition based on two DRGs, there exists a Nash equilibrium in which Provider 1 drops as many patients as possible and underinvests in cost reduction compared to first best, Provider 2 does not drop any patients and invests optimally in cost reduction. In addition, if $c_2^{e2} + \delta_2^{e2} - c_1^* - \delta_1^* \geq 0$, then there exists a second Nash equilibrium in which Provider 2 drops as many patients as possible and underinvests in cost reduction compared to first best, Provider 1 does not drop any patients and invests optimally in cost reduction.*

This result echoes that of Proposition 3 of the symmetric case. Expanding the number of DRGs is helpful in the sense that now one of the two providers acts optimally. If the providers are sufficiently asymmetric (e.g., even if the larger Provider 2 drops as many patients as possible it still treats more patients than Provider 1 treats if they do not drop any patient, or, to be more precise, so that the first-best cost of the smaller provider are larger than the cherry-picking-best costs of the larger provider) then the equilibrium is unique and the larger Provider 2 is the one that acts optimally. If however the providers are relatively similar an additional equilibrium will emerge the smaller Provider 1 is the one that acts optimally. Therefore, just like in the case of symmetric providers, DRG expansion is not a panacea in this case either.

We turn to the case where in addition to lemon dropping providers can also upcode. Before we present results, we first define the costs $(\delta_1^{e2}, \delta_1^{e3})$ as the unique solutions to the equations

$$-\frac{d}{d\delta}R_\delta(\delta) = \lambda q(h(1-\bar{\gamma}) + (1-h)\bar{\alpha}\beta), -\frac{d}{d\delta}R_\delta(\delta) = \lambda q(h + (1-h)\bar{\alpha}\beta),$$

respectively, and the costs $(\delta_2^{e2}, \delta_2^{e3})$ as the unique solutions to the equations

$$-\frac{d}{d\delta}R_\delta(\delta) = \lambda q\eta(\xi h(1-\bar{\gamma}) + (1-\xi h)\bar{\alpha}\beta), -\frac{d}{d\delta}R_\delta(\delta) = \lambda q\eta(\xi h + (1-\xi h)\bar{\alpha}\beta),$$

respectively. Note that since $\eta > 1, \xi > 1, h\xi < 1$ it follows that $\delta_2^{e2} < \delta_1^{e2}, \delta_2^{e3} < \delta_1^{e3}, \delta_i^* > \delta_i^{e3}$ and $\delta_i^{e2} > \delta_i^{e3}$. To focus on cases where upcoding is profitable for both providers, we make the following assumptions that are equivalent to condition (7) of the main text (see discussion in the main text for an explanation as to why these conditions ensure that upcoding is profitable):

$$\delta_2^{e3} \frac{\xi h + (1-h\xi)\bar{\alpha}\beta}{\xi h + (1-h\xi)\bar{\alpha}} > \beta \delta_1^{e1}, \delta_1^{e3} \frac{h + (1-h)\bar{\alpha}\beta}{h + (1-h)\bar{\alpha}} > \beta \delta_2^{e1}. \quad (48)$$

For the rest of the analysis we will assume that these two conditions hold.

Proposition 4A: *If both lemon dropping and upcoding are possible, if the HO implements yardstick competition based on two DRG,*

- *There exists a Nash equilibrium where Provider 1 upcodes and drops as many patients as possible ($\alpha_1 = \bar{\alpha}, \gamma_1 = \bar{\gamma}$) and chooses costs $(c_1^{e1}, \delta_1^{e2})$. If $c_2^{e1} + \delta_2^{e2} > c_1^{e1} + \delta_1^{e2} \frac{h(1-\bar{\gamma}) + (1-h)\bar{\alpha}\beta}{h(1-\bar{\gamma}) + (1-h)\bar{\alpha}}$, then Provider 2 also upcodes and drops as many patients as possible ($\alpha_2 = \bar{\alpha}, \gamma_2 = \bar{\gamma}$) and chooses costs $(c_2^{e1}, \delta_2^{e1})$. If $c_2^* + \delta_2^{e3} < c_1^{e1} + \delta_1^{e2} \frac{h(1-\bar{\gamma}) + (1-h)\bar{\alpha}\beta}{h(1-\bar{\gamma}) + (1-h)\bar{\alpha}}$ then Provider 2 upcodes as much as possible does not drop any patients ($\alpha_2 = \bar{\alpha}, \gamma_2 = 0$) and chooses costs (c_2^*, δ_2^{e1}) . Otherwise, Provider 2 upcodes as much as possible ($\alpha_2 = \bar{\alpha}$), drops an intermediate proportion (γ_2^v) of patients, where $0 < \gamma_2^v < \bar{\gamma}$ and chooses intermediate costs (c_2^v, δ_2^v) , where $c_2^{e1} > c_2^v > c_2^*, \delta_2^{e3} > \delta_2^v > \delta_2^*$.*

- *If $c_1^* + \delta_1^{e3} < c_2^{e1} + \delta_2^{e2} \frac{\xi h(1-\bar{\gamma}) + (1-h\xi)\bar{\alpha}\beta}{\xi h(1-\bar{\gamma}) + (1-h\xi)\bar{\alpha}}$ then there exists a second Nash equilibrium in which Provider 2 upcodes as many patients as possible ($\alpha_2 = \bar{\alpha}$) drops as many patients as possible ($\gamma_2 = \bar{\gamma}$), and chooses costs $(c_2^{e1}, \delta_2^{e2})$. If $c_2^{e2} + \delta_2^{e2} > c_1^* + \delta_1^{e3} \frac{h + (1-h)\bar{\alpha}\beta}{h + (1-h)\bar{\alpha}}$ then Provider 1 upcodes as many patients as possible ($\alpha_1 = \bar{\alpha}$) does not drop any patients ($\gamma_1 = 0$) and chooses costs (c_1^*, δ_1^{e3}) , otherwise Provider 1 drops an intermediate proportion (γ_1^m) of patients $0 < \gamma_1^m < \bar{\gamma}$ and chooses intermediate costs (c_1^m, δ_1^m) , where $c_1^{e1} > c_1^m > c_1^*, \delta_1^{e1} > \delta_1^m > \delta_1^{e3}$.*

The results presented in Proposition 4A mirror that of Proposition 4 in the symmetric case. If upcoding is possible providers will engage fully in it and, even worse, upcoding will make it optimal for one or both providers to also engage in lemon dropping.

Collectively, this section demonstrates that in the case of two providers that are asymmetric in the number of patients they treat and/or in the proportion of high complexity patients, the

results of the symmetric case continue to apply. Namely, in the absence of upcoding, expanding the number of DRGs eliminates lemon dropping incentives only for one of the two providers (usually the larger). The other provider will continue to lemon drop patients. If upcoding is possible then this will, in most cases, reinstate lemon dropping incentives for both provider as well.

Proofs of results appearing in Appendix 3

Proof of Proposition 1A: If there is no upcoding or cherry picking the profit of each provider is given by

$$\pi_1(c_1, \delta_1, 0, 0) = \lambda q [h(p_{M1} - \delta_1 - c_1) + (1 - h)(p_{m1} - c_1)] - R_c(c_1) - R_\delta(\delta_1) + R_c(c_2) + R_\delta(\delta_2)$$

$$\pi_2(c_2, \delta_2, 0, 0) = \lambda q \eta [h\xi(p_{M2} - \delta_2 - c_2) + (1 - h\xi)(p_{m2} - c_2)] - R_c(c_2) - R_\delta(\delta_2) + R_c(c_1) + R_\delta(\delta_1)$$

Since the provider's choice of c_i and δ_i does not affect the reimbursement received, and this is true irrespective of the number of DRGs used by the HO, the profit-maximizing choice of the provider is given by

$$-\frac{d}{dc}R_c(c_1) = \lambda q, \quad -\frac{d}{d\delta}R_\delta(\delta_1) = \lambda q h, \quad -\frac{d}{dc}R_c(c_2) = \lambda q \eta, \quad -\frac{d}{d\delta}R_\delta(\delta_2) = \lambda q h \eta \xi.$$

These conditions are identical to the first order conditions of the welfare-maximization problem. Therefore, the first-best investment decisions constitute a Nash equilibrium. Furthermore, since $R'_c > 0$ and $R''_\delta > 0$ these values are unique.

□

Proof of Proposition 2A: In the absence of upcoding and if the HO uses a single DRG, the profit of each provider is given by

$$\pi_1 = \lambda q [(1 - h\gamma_1)\bar{c}_1 - \delta_1 h(1 - \gamma_1) - c_1(1 - h\gamma_1)] - R_c(c_1) - R_\delta(\delta_1) + R_c(c_2) + R_\delta(\delta_2)$$

$$\pi_2 = \lambda q \eta [(1 - h\xi\gamma_2)\bar{c}_2 - \delta_2 h\xi(1 - \gamma_2) - c_2(1 - h\xi\gamma_2)] - R_c(c_2) - R_\delta(\delta_2) + R_c(c_1) + R_\delta(\delta_1),$$

with $\bar{c}_1 = c_2 + \delta_2 \frac{h\xi(1-\gamma_2)}{1-h\xi\gamma_2}$, $\bar{c}_2 = c_1 + \delta_1 \frac{h(1-\gamma_1)}{1-h\gamma_1}$. The derivatives of the profit function of Provider 1 are given by:

$$\frac{\partial}{\partial \gamma_1} \pi_1 = \lambda q h(c_1 + \delta_1 - \bar{c}_1), \tag{49}$$

$$\frac{\partial}{\partial c_1} \pi_1 = -\frac{d}{dc}R_c(c_1) - \lambda q(1 - h\gamma_1), \tag{50}$$

$$\frac{\partial}{\partial \delta_1} \pi_1 = -\frac{d}{d\delta}R_\delta(\delta_1) - \lambda q h(1 - \gamma_1), \tag{51}$$

and for Provider 2

$$\frac{\partial}{\partial \gamma_2} \pi_2 = \lambda q \eta h \xi (c_2 + \delta_2 - \bar{c}_2), \tag{52}$$

$$\frac{\partial}{\partial c_2} \pi_2 = -\frac{d}{dc} R_c(c_2) - \lambda q \eta (1 - h \xi \gamma_2), \quad (53)$$

$$\frac{\partial}{\partial \delta_2} \pi_2 = -\frac{d}{d\delta} R_\delta(\delta_2) - \lambda q \eta h \xi (1 - \gamma_2). \quad (54)$$

In equilibrium the conditions (50), (51), (53), (54) must all be equal to zero, otherwise the provider for whom one of these conditions is not zero could improve their profits.

Turning to (49), we need to check three cases.

- First, if at the optimal solution (49) is positive, i.e., $c_1 + \delta_1 - c_2 - \delta_2 h \xi \frac{1-\gamma_2}{1-h\xi\gamma_2} > 0$, then $\frac{\partial}{\partial \gamma_1} \pi_1 > 0$, therefore Provider 1 must find it optimal to choose $\gamma_1 = \bar{\gamma}$ and $c_1 = c_1^{e1}$, $\delta_1 = \delta_1^{e1}$. Turning to (52) we will examine three subcases:

- First, if at the optimal solution $c_2 + \delta_2 > c_1^{e1} + \delta_1^{e1} \frac{h(1-\bar{\gamma})}{1-h\bar{\gamma}}$ then $\frac{\partial}{\partial \gamma_2} \pi_2 > 0$, therefore Provider 2 must find it optimal to choose $\gamma_2 = \bar{\gamma}$ and $c_2 = c_2^{e1}$, $\delta_2 = \delta_2^{e1}$. At these choices, the condition on (49) is indeed satisfied. However, for the condition on (52) to be satisfied the following condition must hold $c_2^{e1} + \delta_2^{e1} > c_1^{e1} + \delta_1^{e1} \frac{h(1-\bar{\gamma})}{1-h\bar{\gamma}}$.

- Second, if at the optimal solution $c_2 + \delta_2 < c_1^{e1} + \delta_1^{e1} \frac{h(1-\bar{\gamma})}{1-h\bar{\gamma}}$ then $\frac{\partial}{\partial \gamma_2} \pi_2 < 0$ therefore Provider 2 finds it optimal to choose $\gamma_2 = 0$ and $c_2 = c_2^*$, $\delta_2 = \delta_2^*$. At these choices, the condition on (49) is indeed satisfied. However, for the condition on (52) to be satisfied the following condition must hold $c_2^* + \delta_2^* < c_1^{e1} + \delta_1^{e1} \frac{h(1-\bar{\gamma})}{1-h\bar{\gamma}}$.

- Third, if at the optimal solution $c_2 + \delta_2 = c_1^{e1} + \delta_1^{e1} \frac{h(1-\bar{\gamma})}{1-h\bar{\gamma}}$ then $\frac{\partial}{\partial \gamma_2} \pi_2 = 0$. Therefore, Provider 2 will find it optimal to choose $\gamma_2 = \gamma_2^v$, where $0 < \gamma_2^v < \bar{\gamma}$ with c_2^v and δ_2^v such that (53) and (54) are equal to zero. γ_2^v is such that $c_2^v + \delta_2^v = c_1^{e1} + \delta_1^{e1} \frac{h(1-\bar{\gamma})}{1-h\bar{\gamma}}$. At these choices the condition on (49) is satisfied.

- Second, if at the optimal solution (49) is negative, i.e., $c_1 + \delta_1 - c_2 - \delta_2 h \xi \frac{1-\gamma_2}{1-h\xi\gamma_2} < 0$ then $\frac{\partial}{\partial \gamma_1} \pi_1 \leq 0$, suggesting that Provider 1 would find it optimal to choose $\gamma_1 = 0$ and $c_1 = c_1^*$, $\delta_1 = \delta_1^*$. Turning to (52) we will examine three subcases:

- First, if $c_2 + \delta_2 < c_1^* + \delta_1^* h$ then $\frac{\partial}{\partial \gamma_2} \pi_2 < 0$ therefore Provider 2 finds it optimal to choose $\gamma_2 = 0$ and $c_2 = c_2^*$, $\delta_2 = \delta_2^*$. At these choices, the condition on (49) cannot be satisfied, leading to a contradiction. Therefore, this cannot be an equilibrium outcome.

- Second, if $c_2 + \delta_2 > c_1^* + \delta_1^* h$ then $\frac{\partial}{\partial \gamma_2} \pi_2 > 0$ therefore Provider 2 finds it optimal to choose $\gamma_2 = \bar{\gamma}$ and $c_2 = c_2^{e1}$, $\delta_2 = \delta_2^{e1}$. At these choices, the conditions on (49) and (52) are both satisfied if $c_1^* + \delta_1^* < c_2^{e1} + \delta_2^{e1} \frac{h\xi(1-\bar{\gamma})}{1-h\xi\bar{\gamma}}$ and $c_2^{e1} + \delta_2^{e1} > c_1^* + \delta_1^* h$.

- Third, if $c_2 + \delta_2 = c_1^* + \delta_1^* h$ then $\frac{\partial}{\partial \gamma_2} \pi_2 = 0$. Therefore, Provider 2 will find it optimal to choose $\gamma_2 = \gamma_2^{vv}$, where $0 < \gamma_2^{vv} < \bar{\gamma}$ with c_2^{vv} and δ_2^{vv} such that (53) and (54) are equal to zero. γ_2^{vv} is such that $c_1^* + \delta_1^* h = c_2^{vv} + \delta_2^{vv}$. At these choices condition (49) becomes $c_2^{vv} + \delta_2^{vv} \frac{h\xi(1-\gamma_2^{vv})}{1-h\xi\gamma_2^{vv}} > c_1^* + \delta_1^*$. At these choices, adding up (49) and (52) gives $\delta_2^{vv} \frac{1-h\xi}{1-h\xi\gamma_2^{vv}} + \delta_1^*(1-h) < 0$, leading to a contradiction. Therefore, this cannot be an equilibrium outcome.

• Third, if at the optimal solution (49) is equal to zero, i.e., $c_1 + \delta_1 - c_2 - \delta_2 h \xi \frac{1-\gamma_2}{1-h\xi\gamma_2} = 0$ then $\frac{\partial}{\partial \gamma_1} \pi_1 = 0$, suggesting that Provider 1 would find it optimal to choose $\gamma_1 = \gamma_1^m$ and $c_1 = c_1^m$, $\delta_1 = \delta_1^m$ such that conditions (50), (51) are equal to zero. Turning to (52) we will examine three subcases:

— First, if $c_2 + \delta_2 > c_1^m + \delta_1^m h \frac{1-\gamma_1^m}{1-h\gamma_1^m}$ then $\frac{\partial}{\partial \gamma_2} \pi_2 > 0$ therefore Provider 2 finds it optimal to choose $\gamma_2 = \bar{\gamma}$ and $c_2 = c_2^{e1}$, $\delta_2 = \delta_2^{e2}$. At these choices, the condition on (52) implies that $c_2^{e1} + \delta_2^{e1} > c_1^m + \delta_1^m h \frac{1-\gamma_1^m}{1-h\gamma_1^m}$. Given the condition on (49), this can be written as $\delta_1^m \frac{1-h}{1-h\gamma_1^m} + \delta_2^{e1} \frac{1-h\xi}{1-h\xi\bar{\gamma}} > 0$, which is always satisfied. Turning to the condition on (49), there will exist a $\gamma_1^m < \bar{\gamma}$ to satisfy this condition only if $c_1^* + \delta_1^* < c_2^{e1} + \delta_2^{e1} \frac{h\xi(1-\bar{\gamma})}{1-h\xi\bar{\gamma}}$.

— Second, if $c_2 + \delta_2 < c_1^m + \delta_1^m h \frac{1-\gamma_1^m}{1-h\gamma_1^m}$ then $\frac{\partial}{\partial \gamma_2} \pi_2 < 0$ therefore Provider 2 finds it optimal to choose $\gamma_2 = 0$ and $c_2 = c_2^*$, $\delta_2 = \delta_2^*$. At these choices, the condition on (49) becomes $c_1^m + \delta_1^m = c_2^* + \delta_2^* h \xi$. This cannot be true as $c_1^m > c_2^*$, $\delta_1^m > \delta_2^*$ and $h\xi < 1$, leading to a contradiction. Therefore, this cannot be an equilibrium outcome.

— Third, if $c_2 + \delta_2 = c_1^m + \delta_1^m h \frac{1-\gamma_1^m}{1-h\gamma_1^m}$ then $\frac{\partial}{\partial \gamma_2} \pi_2 = 0$. Therefore, Provider 2 will find it optimal to choose $\gamma_2 = \gamma_2^{mm}$, where $0 < \gamma_2^{mm} < \bar{\gamma}$ with c_2^{mm} and δ_2^{mm} such that (53) and (54) are equal to zero. γ_2^{mm} is such that $c_2^{mm} + \delta_2^{mm} = c_1^m + \delta_1^m h \frac{1-\gamma_1^m}{1-h\gamma_1^m}$. At these choices, adding up (49) and (52) gives $\delta_2^{mm} \frac{1-h\xi}{1-h\xi\gamma_2^{mm}} + \delta_1^m \frac{1-h}{1-h\gamma_1^m} = 0$, leading to a contradiction. Therefore, this cannot be an equilibrium outcome.

Collectively, these cases describe the two equilibria presented in the Proposition.

□

Proof of Proposition 3A: In the absence of upcoding ($\bar{\alpha} = 0$), under the yardstick competition scheme with two DRGs, the profit of each provider is given by

$$\begin{aligned} \pi_1 &= \lambda q [h(1-\gamma_1)(\bar{c}_{M1} - \delta_1 - c_1) + (1-h)(\bar{c}_{m1} - c_1)] - R_c(c_1) - R_\delta(\delta_1) + R_c(c_2) + R_\delta(\delta_2), \\ \pi_2 &= \lambda q \eta [h\xi(1-\gamma_2)(\bar{c}_{M2} - \delta_2 - c_2) + (1-h\xi)(\bar{c}_{m2} - c_2)] - R_c(c_2) - R_\delta(\delta_2) + R_c(c_1) + R_\delta(\delta_1), \end{aligned}$$

where $\bar{c}_{M1} := c_2 + \delta_2$, $\bar{c}_{m1} := c_2$, $\bar{c}_{M2} := c_1 + \delta_1$, $\bar{c}_{m2} := c_1$. The derivatives of the profit function of Provider 1 are given by:

$$\frac{\partial}{\partial \gamma_1} \pi_1 = \lambda q h (c_1 + \delta_1 - \bar{c}_{M1}), \quad (55)$$

$$\frac{\partial}{\partial c_1} \pi_1 = -\frac{d}{dc} R_c(c_1) - \lambda q (1-h\gamma_1), \quad (56)$$

$$\frac{\partial}{\partial \delta_1} \pi_1 = -\frac{d}{d\delta} R_\delta(\delta_1) - \lambda q h (1-\gamma_1), \quad (57)$$

and for Provider 2 by

$$\frac{\partial}{\partial \gamma_2} \pi_2 = \lambda q h \eta \xi (c_2 + \delta_2 - \bar{c}_{M2}), \quad (58)$$

$$\frac{\partial}{\partial c_2} \pi_2 = -\frac{d}{dc} R_c(c_2) - \lambda q \eta (1 - h \xi \gamma_2), \quad (59)$$

$$\frac{\partial}{\partial \delta_2} \pi_2 = -\frac{d}{d\delta} R_\delta(\delta_2) - \lambda q h \eta \xi (1 - \gamma_2). \quad (60)$$

In equilibrium the conditions (56), (57), (59), (60) must all be equal to zero, otherwise the provider for whom one of these conditions is not zero could improve their profits.

Turning to (55), if $c_1 + \delta_1 - c_2 - \delta_2 > 0$ then $\frac{\partial}{\partial \gamma_1} \pi_1 > 0$, which suggests that Provider 1 would find it optimal to choose $\gamma_1 = \bar{\gamma}$ and $c_1 = c_1^{e1}$, $\delta_1 = \delta_1^{e1}$ and from (58) then $\frac{\partial}{\partial \gamma_2} \pi_2 < 0$, which suggests that Provider 2 would find it optimal to choose $\gamma_2 = 0$, $c_2 = c_2^*$, $\delta_2 = \delta^*$. Given the optimal choices, the condition $c_1 + \delta_1 - c_2 - \delta_2 > 0$ is indeed satisfied.

If $c_1 + \delta_1 - c_2 - \delta_2 \leq 0$ then $\frac{\partial}{\partial \gamma_1} \pi_1 \leq 0$, which suggests that Provider 1 would find it optimal to choose $\gamma_1 = 0$ and $c_1 = c_1^*$, $\delta_1 = \delta_1^*$ and (58) would imply that $\frac{\partial}{\partial \gamma_2} \pi_2 \geq 0$ which in turn suggests that Provider 2 would find it optimal to choose $\gamma_2 = \bar{\gamma}$, $c_2 = c_2^{e1}$, $\delta_2 = \delta_2^{e1}$. Given the optimal choices, the condition $c_1 + \delta_1 - c_2 - \delta_2 < 0$ can only be satisfied if $c_1^* + \delta_1^* - c_2^{e1} - \delta_2^{e1} < 0$.

□

Proof of Proposition 4A: In the presence of upcoding, under the yardstick competition scheme with two DRGs, the profit of each provider is given by

$$\begin{aligned} \pi_1 &= \lambda q [(h(1 - \gamma_1) + (1 - h)\alpha_1)\bar{c}_{M1} + (1 - h)(1 - \alpha_1)\bar{c}_{m1} - h(1 - \gamma_1)\delta_1 - (1 - \gamma_1 h)c_1] \\ &\quad - R_c(c_1) - R_\delta(\delta_1) + R_c(c_2) + R_\delta(\delta_2), \\ \pi_2 &= \lambda q \eta [(h\xi(1 - \gamma_2) + (1 - h\xi)\alpha_2)\bar{c}_{M2} + (1 - h\xi)(1 - \alpha_2)\bar{c}_{m2} - h\xi(1 - \gamma_2)\delta_2 - (1 - \gamma_2 h\xi)c_2] \\ &\quad - R_c(c_2) - R_\delta(\delta_2) + R_c(c_1) + R_\delta(\delta_1), \end{aligned}$$

where $\bar{c}_{M1} := c_2 + \delta_2 \frac{\xi h(1 - \gamma_2) + (1 - h\xi)\alpha_2 \beta}{\xi h(1 - \gamma_2) + (1 - h\xi)\alpha_2}$, $\bar{c}_{m1} := c_2$, $\bar{c}_{M2} := c_1 + \delta_1 \frac{h(1 - \gamma_1) + (1 - h)\alpha_1 \beta}{h(1 - \gamma_1) + (1 - h)\alpha_1}$, $\bar{c}_{m2} := c_1$. The derivatives of the profit function of Provider 1 are given by:

$$\frac{\partial}{\partial \alpha_1} \pi_1 = \lambda q h (\bar{c}_{M1} - \bar{c}_{m1} - \beta \delta_1), \quad (61)$$

$$\frac{\partial}{\partial \gamma_1} \pi_1 = \lambda q h (c_1 + \delta_1 - \bar{c}_{M1}), \quad (62)$$

$$\frac{\partial}{\partial c_1} \pi_1 = -\frac{d}{dc} R_c(c_1) - \lambda q (1 - h \gamma_1), \quad (63)$$

$$\frac{\partial}{\partial \delta_1} \pi_1 = -\frac{d}{d\delta} R_\delta(\delta_1) - \lambda q [h(1 - \gamma_1) + (1 - h)\alpha_1 \beta], \quad (64)$$

and for Provider 2 by

$$\frac{\partial}{\partial \gamma_2} \pi_2 = \lambda q h \eta \xi (\bar{c}_{M2} - \bar{c}_{m2} - \beta \delta_2), \quad (65)$$

$$\frac{\partial}{\partial \gamma_2} \pi_2 = \lambda q h \eta \xi (c_2 + \delta_2 - \bar{c}_{M2}), \quad (66)$$

$$\frac{\partial}{\partial c_2} \pi_2 = -\frac{d}{dc} R_c(c_2) - \lambda q \eta (1 - h \xi \gamma_2), \quad (67)$$

$$\frac{\partial}{\partial \delta_2} \pi_2 = -\frac{d}{d\delta} R_\delta(\delta_2) - \lambda q \eta [h \xi (1 - \gamma_2) + (1 - h \xi) \alpha_2 \beta]. \quad (68)$$

In equilibrium the conditions (63), (64), (67), (68) must all be equal to zero, otherwise the provider for whom one of these conditions is not zero could improve their profits. Condition (61) can be written as $\frac{\partial}{\partial \alpha_1} \pi_1 = \delta_2 \frac{\xi h(1-\gamma_2) + (1-h\xi)\alpha_2\beta}{\xi h(1-\gamma_2) + (1-h\xi)\alpha_2} - \beta \delta_1$. Note that the first term is minimized at $\gamma_2 = 0$ and the second term is maximized at δ_1^{e2} . Therefore, (61) is positive if (48) holds and this implies that $\alpha_1 = \bar{\alpha}$. A similar argument shows that if (48) holds then $\alpha_2 = \bar{\alpha}$.

Turning to condition (62), this can be written as $\frac{\partial}{\partial \gamma_1} \pi_1 = c_1 + \delta_1 - c_2 - \delta_2 \frac{\xi h(1-\gamma_2) + (1-h\xi)\bar{\alpha}\beta}{\xi h(1-\gamma_2) + (1-h\xi)\bar{\alpha}}$. We need to check 3 cases.

- First, if at the optimal solution (62) is positive, i.e., $c_1 + \delta_1 - c_2 - \delta_2 \frac{\xi h(1-\gamma_2) + (1-h\xi)\bar{\alpha}\beta}{\xi h(1-\gamma_2) + (1-h\xi)\bar{\alpha}} > 0$, then $\frac{\partial}{\partial \gamma_1} \pi_1 > 0$, therefore Provider 1 must find it optimal to choose $\gamma_1 = \bar{\gamma}$ and $c_1 = c_1^{e1}$, $\delta_1 = \delta_1^{e2}$. Turning to (66) we will examine three subcases:

- First, if at the optimal solution $c_2 + \delta_2 > c_1^{e1} + \delta_1^{e2} \frac{h(1-\bar{\gamma}) + (1-h)\bar{\alpha}\beta}{h(1-\bar{\gamma}) + (1-h)\bar{\alpha}}$ then $\frac{\partial}{\partial \gamma_2} \pi_2 > 0$, therefore Provider 2 must find it optimal to choose $\gamma_2 = \bar{\gamma}$ and $c_2 = c_2^{e1}$, $\delta_2 = \delta_2^{e2}$. Since $\beta < 1$, at these choices the condition on (62) is indeed satisfied. However, for the condition on (66) to be satisfied the following condition must hold $c_2^{e1} + \delta_2^{e2} > c_1^{e1} + \delta_1^{e2} \frac{h(1-\bar{\gamma}) + (1-h)\bar{\alpha}\beta}{h(1-\bar{\gamma}) + (1-h)\bar{\alpha}}$.

- Second, if at the optimal solution $c_2 + \delta_2 < c_1^{e1} + \delta_1^{e2} \frac{h(1-\bar{\gamma}) + (1-h)\bar{\alpha}\beta}{h(1-\bar{\gamma}) + (1-h)\bar{\alpha}}$ then $\frac{\partial}{\partial \gamma_2} \pi_2 < 0$ therefore Provider 2 finds it optimal to choose $\gamma_2 = 0$ and $c_2 = c_2^*$, $\delta_2 = \delta_2^{e3}$. At these choices, the condition on (62) is indeed satisfied. However, for the condition on (66) to be satisfied the following condition must hold $c_2^* + \delta_2^{e3} < c_1^{e1} + \delta_1^{e2} \frac{h(1-\bar{\gamma}) + (1-h)\bar{\alpha}\beta}{h(1-\bar{\gamma}) + (1-h)\bar{\alpha}}$.

- Third, if at the optimal solution $c_2 + \delta_2 = c_1^{e1} + \delta_1^{e2} \frac{h(1-\bar{\gamma}) + (1-h)\bar{\alpha}\beta}{h(1-\bar{\gamma}) + (1-h)\bar{\alpha}}$ then $\frac{\partial}{\partial \gamma_2} \pi_2 = 0$. Therefore, Provider 2 will find it optimal to choose $\gamma_2 = \gamma_2^v$, where $0 < \gamma_2^v < \bar{\gamma}$ with c_2^v and δ_2^v such that (67) and (68) are equal to zero. γ_2^v is such that $c_2^v + \delta_2^v = c_1^{e1} + \delta_1^{e2} \frac{h(1-\bar{\gamma}) + (1-h)\bar{\alpha}\beta}{h(1-\bar{\gamma}) + (1-h)\bar{\alpha}}$. At these choices the condition on (62) is satisfied.

- Second, if at the optimal solution (62) is negative, i.e., $c_1 + \delta_1 - c_2 - \delta_2 \frac{\xi h(1-\gamma_2) + (1-h\xi)\bar{\alpha}\beta}{\xi h(1-\gamma_2) + (1-h\xi)\bar{\alpha}} < 0$ then $\frac{\partial}{\partial \gamma_1} \pi_1 \leq 0$, suggesting that Provider 1 would find it optimal to choose $\gamma_1 = 0$ and $c_1 = c_1^*$, $\delta_1 = \delta_1^{e3}$. Turning to (66) we will examine three subcases:

- First, if $c_2 + \delta_2 < c_1^* + \delta_1^{e3} \frac{h+(1-h)\bar{\alpha}\beta}{h+(1-h)\bar{\alpha}}$ then $\frac{\partial}{\partial \gamma_2} \pi_2 < 0$ therefore Provider 2 finds it optimal to choose $\gamma_2 = 0$ and $c_2 = c_2^*$, $\delta_2 = \delta_2^{e3}$. At these choices, the condition on (62) cannot be satisfied, leading to a contradiction. Therefore, this cannot be an equilibrium outcome.

- Second, if $c_2 + \delta_2 > c_1^* + \delta_1^{e3} \frac{h+(1-h)\bar{\alpha}\beta}{h+(1-h)\bar{\alpha}}$ then $\frac{\partial}{\partial \gamma_2} \pi_2 > 0$ therefore Provider 2 finds it optimal to choose $\gamma_2 = \bar{\gamma}$ and $c_2 = c_2^{e1}$, $\delta_2 = \delta_2^{e2}$. At these choices, the conditions on (62) and (66) are both satisfied if $c_1^* + \delta_1^{e3} < c_2^{e1} + \delta_2^{e2} \frac{\xi h(1-\bar{\gamma}) + (1-h\xi)\bar{\alpha}\beta}{\xi h(1-\bar{\gamma}) + (1-h\xi)\bar{\alpha}}$ and $c_2^{e1} + \delta_2^{e2} > c_1^* + \delta_1^{e3} \frac{h+(1-h)\bar{\alpha}\beta}{h+(1-h)\bar{\alpha}}$.

— Third, if $c_2 + \delta_2 = c_1^* + \delta_1^* \frac{h+(1-h)\bar{\alpha}\beta}{h+(1-h)\bar{\alpha}}$ then $\frac{\partial}{\partial \gamma_2} \pi_2 = 0$. Therefore, Provider 2 will find it optimal to choose $\gamma_2 = \gamma^{vv}$, where $0 < \gamma_2^{vv} < \bar{\gamma}$ with c_2^{vv} and δ_2^{vv} such that (67) and (68) are equal to zero. γ_2^{vv} is such that $c_1^* + \delta_1^{e3} \frac{h+(1-h)\bar{\alpha}\beta}{h+(1-h)\bar{\alpha}} = c_2^{vv} + \delta_2^{vv}$. At these choices condition (62) becomes $c_2^{vv} + \delta_2^{vv} \frac{h\xi(1-\gamma^{vv})+(1-h\xi)\bar{\alpha}\beta}{h\xi(1-\gamma^{vv})+(1-h\xi)\bar{\alpha}} > c_1^* + \delta_1^{e3}$. At these choices, adding up (62) and (66) gives $\delta_2^{vv}(1 - \frac{h\xi(1-\gamma^{vv})+(1-h\xi)\bar{\alpha}\beta}{h\xi(1-\gamma^{vv})+(1-h\xi)\bar{\alpha}}) + \delta_1^{e3}(1 - \frac{h+(1-h)\bar{\alpha}\beta}{h+(1-h)\bar{\alpha}}) < 0$, leading to a contradiction. Therefore, this cannot be an equilibrium outcome.

• Third, if at the optimal solution (62) is equal to zero, i.e., $c_1 + \delta_1 - c_2 - \delta_2 \frac{\xi h(1-\gamma_2)+(1-h\xi)\bar{\alpha}\beta}{\xi h(1-\gamma_2)+(1-h\xi)\bar{\alpha}} = 0$ then $\frac{\partial}{\partial \gamma_1} \pi_1 = 0$, suggesting that Provider 1 would find it optimal to choose $\gamma_1 = \gamma_1^m$ and $c_1 = c_1^m$, $\delta_1 = \delta_1^m$ such that conditions (63), (64) are equal to zero. Turning to (66) we will examine three subcases:

— First, if $c_2 + \delta_2 > c_1^m + \delta_1^m \frac{h(1-\gamma_1^m)+(1-h)\bar{\alpha}\beta}{h(1-\gamma_1^m)+(1-h)\bar{\alpha}}$ then $\frac{\partial}{\partial \gamma_2} \pi_2 > 0$ therefore Provider 2 finds it optimal to choose $\gamma_2 = \bar{\gamma}$ and $c_2 = c_2^{e1}$, $\delta_2 = \delta_2^{e2}$. At these choices, the condition on (66) implies that $c_2^{e1} + \delta_2^{e2} > c_1^m + \delta_1^m \frac{h(1-\gamma_1^m)+(1-h)\bar{\alpha}\beta}{h(1-\gamma_1^m)+(1-h)\bar{\alpha}}$. Given the condition on (62), this can be written as $\delta_1^m(1 - \frac{h(1-\gamma_1^m)+(1-h)\bar{\alpha}\beta}{h(1-\gamma_1^m)+(1-h)\bar{\alpha}}) + \delta_2^{e2}(1 - \frac{\xi h(1-\bar{\gamma})+(1-h\xi)\bar{\alpha}\beta}{\xi h(1-\bar{\gamma})+(1-h\xi)\bar{\alpha}}) > 0$, which is always satisfied. Turning to the condition on (62), there will exist a $\gamma_1^m < \bar{\gamma}$ to satisfy this condition only if $c_1^* + \delta_1^{e3} < c_2^{e1} + \delta_2^{e2} \frac{\xi h(1-\bar{\gamma})+(1-h\xi)\bar{\alpha}\beta}{\xi h(1-\bar{\gamma})+(1-h\xi)\bar{\alpha}}$.

— Second, if $c_2 + \delta_2 < c_1^m + \delta_1^m \frac{h(1-\gamma_1^m)+(1-h)\bar{\alpha}\beta}{h(1-\gamma_1^m)+(1-h)\bar{\alpha}}$ then $\frac{\partial}{\partial \gamma_2} \pi_2 < 0$ therefore Provider 2 finds it optimal to choose $\gamma_2 = 0$ and $c_2 = c_2^*$, $\delta_2 = \delta_2^*$. At these choices, the condition on (62) becomes $c_1^m + \delta_1^m = c_2^* + \delta_2^* \frac{\xi h+(1-h\xi)\bar{\alpha}\beta}{\xi h+(1-h\xi)\bar{\alpha}}$. This cannot be true as $c_1^m > c_2^*$, $\delta_1^m > \delta_2^{e3}$, leading to a contradiction. Therefore, this cannot be an equilibrium outcome.

— Third, if $c_2 + \delta_2 = c_1^m + \delta_1^m \frac{h(1-\gamma_1^m)+(1-h)\bar{\alpha}\beta}{h(1-\gamma_1^m)+(1-h)\bar{\alpha}}$ then $\frac{\partial}{\partial \gamma_2} \pi_2 = 0$. Therefore, Provider 2 will find it optimal to choose $\gamma_2 = \gamma_2^{mm}$, where $0 < \gamma_2^{mm} < \bar{\gamma}$ with c_2^{mm} and δ_2^{mm} such that (67) and (68) are equal to zero. γ_2^{mm} is such that $c_2^{mm} + \delta_2^{mm} = c_1^m + \delta_1^m \frac{h(1-\gamma_1^m)+(1-h)\bar{\alpha}\beta}{h(1-\gamma_1^m)+(1-h)\bar{\alpha}}$. At these choices, adding up (62) and (66) gives $\delta_2^{mm}(1 - \frac{h\xi(1-\gamma_2^{mm})+(1-h\xi)\bar{\alpha}\beta}{h\xi(1-\gamma_2^{mm})+(1-h\xi)\bar{\alpha}}) + \delta_1^m(1 - \frac{h(1-\gamma_1^m)+(1-h)\bar{\alpha}\beta}{h(1-\gamma_1^m)+(1-h)\bar{\alpha}}) = 0$, leading to a contradiction. Therefore, this cannot be an equilibrium outcome.

Collectively, these cases describe the two equilibria presented in the Proposition.

□