Online Appendix: Hospital Readmissions Reduction Program does not provide the right incentives: Issues and remedies

Contents

A. Assumptions

We next list the conditions that the total investment function R is assumed to satisfy throughout the proofs. We emphasize that these conditions are sufficient but not necessary for our results. Throughout we use the convention that the derivatives at the lower bound of the domain of a function is taken as the right-derivative and at the upper bound as the left-derivative, without further mention, for notational simplicity.

Assumption A1.
(i) Total investment is decreasing and sufficiently convex in both readmission rate and treatment cost, specifically, $R_r < 0$, $R_c < 0$, $R_{rr} > 0$, $R_{cc} > 0$, *and* $R_{rr}R_{cc} > (\lambda + R_{cr})^2$.
(ii) Reducing readmissions is more costly for lower values of treatment cost, i.e., $R_{rc} \geq 0$.
(iii) We assume the following boundary conditions hold.

$$
\begin{aligned}
&R_c(r_{max}, c_{min}) < -\lambda(1 + r_{max}),\\
&R_c(r_{min}, c_{max}) > -\lambda(1 + r_{min}),\\
&R_r(r_{max}, c_{min}) > -\lambda c_{min}, \text{ and}\\
&R_r(r_{min}, c_{max}) < -\lambda c_{max}(1 + 1/r_{min}).
\end{aligned}
$$

In line with the literature (e.g., Shleifer 1985, Savva et al. 2018, Andritsos and Tang 2018), Assumption A1(i) requires the investment cost $R(r,c)$ to be sufficiently convex to ensure that regulator's and hospitals' objective functions are concave in readmission rate r and treatment cost c. Assumption A1(ii) implies that when a hospital has improved the operational efficiency to reduce the treatment cost, it becomes even more costly to reduce readmissions by further operational improvement. We use Assumption A1(ii) to show that when the treatment cost is optimally determined by regulator or the hospital (i.e., $c = h(r)$), it decreases as the readmission rate increases, see Lemma A2 below. Finally Assumption A1(iii) guarantees that the regulator's as well as hospitals' optimal actions are not at the boundaries.

B. Proofs of the results in Section 4

B.1. Preliminary results

In this section we prove preliminary results that we utilize in proving the results in §4. We first show that given a readmission level there exists a unique associated marginal cost determined by (5).

Lemma A1. *For $r \in [r_{min}, r_{max}]$ there exists a unique $h(r) \in [c_{min}, c_{max}]$ that satisfies* (5)*. In addition $h : [r_{min}, r_{max}] \to [c_{min}, c_{max}]$ is differentiable and strictly decreasing.*

Proof of Lemma A1: Fix $r \in [r_{min}, r_{max}]$ and let

$$g(c) = \lambda(1+r) + R_c(r,c). \tag{A1}$$

Then

$$g'(c) = R_{cc}(r,c), \tag{A2}$$

where we use $g'(c) = \frac{dg(c)}{dc}$ throughout.

By Assumption A1(iii), $\lim_{c \downarrow c_{min}} g(c) < 0$ and $\lim_{c \uparrow c_{max}} g(c) > 0$ for any $r \in [r_{min}, r_{max}]$. Moreover, $g'(c) > 0$ since $R_{cc} > 0$ by Assumption A1(i). As a result, there exists a unique $h(r) \in (c_{min}, c_{max})$ that satisfies (5). By implicit function theorem and (5) ($R_{cc} > 0$ by Assumption A1(i))

$$h'(r) = -\frac{\lambda + R_{cr}}{R_{cc}} < 0, \tag{A3}$$

where the inequality follows from Assumption A1(i) and (ii). Thus, $h(r)$ is decreasing in r. $\square$

Next we show that the marginal treatment cost in any equilibrium under HRRP, HRRP-I and HRRP-II schemes is uniquely determined by the readmission rate. We utilize this result to focus only on readmission levels in determining equilibrium outcomes.

Lemma A2. *In any equilibrium under HRRP, HRRP-I, and HRRP-II, $c_i = h(r_i)$ for all $i = 1, \ldots, N$, where h is defined by* (5)*.*

Proof of Lemma A2: Given the choices of all other hospitals, the objective of hospital i can be written as

$$\Pi(r_i, c_i) = (\bar{c}_i - c_i)(1+r)\lambda - f(r_i, \bar{r}_i, \bar{c}_i) + \bar{R}_i - R(r_i, c_i), \tag{A4}$$

where $f(r_i, \bar{r}_i, \bar{c}_i) = \pi(r_i | \bar{r}_i, \bar{c}_i)$ under HRRP, see (7), $f(r_i, \bar{r}_i, \bar{c}_i) = \pi^{\text{I}}(r_i | \bar{r}_i, \bar{c}_i)$ under HRRP-I, see (9), and $f(r_i, \bar{r}_i, \bar{c}_i) = \pi^{\text{II}}(r_i | \bar{r}_i, \bar{c}_i)$ under HRRP-II, see (14). Clearly f does not depend on hospital's marginal cost c_i under these three payment systems.

Consider hospital i's objective under one of these payment schemes for a fixed readmission level $r \in [r_{min}, r_{max}]$, given the choices of all the other hospitals. Then the first derivative of the objective function with respect to c_i is equal to $-g(c_i)$, see (A1). Also by (A2) and Assumption A1(i), $c_i \equiv h(r)$ is the unique marginal treatment cost that optimizes hospital's objective. This implies that in any equilibrium under HRRP, HRRP-I or HRRP-II, $c_i = h(r_i)$ for all $i = 1, \ldots, N$. $\square$

For $r > \bar{r}$, define

$$\Phi^o(r|\bar{r}) = (h(\bar{r}) - h(r))(1+r)\lambda - R(r, h(r)) + R(\bar{r}, h(\bar{r})). \tag{A5}$$

In words, Φ^o is the objective function of a hospital when $P_{cap} = 0$ under HRRP, HRRP-I, and HRRP-II (note that these payment schemes are identical when $P_{cap} = 0$ by (7), (9), and (14)) assuming all the other hospitals pick $(\bar{r}, h(\bar{r}))$. We use the next result primarily to study the actions of hospitals when P_{cap} is small.

Lemma A3. *Fix $\bar{r} \in [r_{min}, r_{max})$.*
(i) For any $r \in [r_{min}, r_{max}]$

$$\frac{d\Phi^o(r|\bar{r})}{dr} > 0. \tag{A6}$$

(ii) For any $\bar{r} \le r_1 \le r_2 \le r_{max}$

$$\Phi^o(r_{max}|\bar{r}) \ge (r_{max} - r_2)(h(r_1) - h(r_2))\lambda. \tag{A7}$$

Proof of Lemma A3: Let $r \in [\bar{r}, r_{max}]$. By (5) and (A5),

$$\Phi^o_r(r|\bar{r}) = (h(\bar{r}) - h(r))\lambda - R_r(r, h(r)), \tag{A8}$$

where we set $\Phi^o_r(r|\bar{r}) = \frac{d\Phi(r|\bar{r})}{dr}$, and taking the derivative with respect to r again, we have

$$\Phi^o_{rr}(r|\bar{r}) = -(\lambda + R_{rc})h' - R_{rr} = \frac{(\lambda + R_{rc})^2}{R_{cc}} - R_{rr} < 0, \tag{A9}$$

where the second equality follows by plugging in h', see (A3), and the inequality follows from Assumption A1(i). From (A8) $\Phi^o_r(r|\bar{r}) > 0$ at $r = r_{max}$ because h is strictly decreasing by Lemma A1 and $R_r < 0$ by Assumption A1(i). This combined with (A9) gives (A6).

Now for any $\bar{r} \in [r_{min}, r_{max})$ and $\bar{r} \le r_1 \le r_2 \le r_{max}$

$$\begin{aligned}\Phi^o(r_{max}|\bar{r}) = \Phi^o(\bar{r}|\bar{r}) + \int_{\bar{r}}^{r_{max}} \Phi^o_r(u|\bar{r})du &\overset{\text{(r0)}}{=} \int_{\bar{r}}^{r_{max}} \Phi^o_r(u|\bar{r})du \\ &\overset{\text{(r1)}}{\ge} (r_{max} - r_2)(h(\bar{r}) - h(r_2))\lambda \\ &\overset{\text{(r2)}}{\ge} (r_{max} - r_2)(h(r_1) - h(r_2)))\lambda,\end{aligned} \tag{A10}$$

where (r0) above follows since $\Phi^o(r|r) = 0$ (see (A5)), (r1) follows from (A8), from $R_r < 0$ by Assumption A1(i), and the fact that h is strictly decreasing by Lemma A1, and the latter also gives (r2) because $\bar{r} \le r_1 \le r_2 \le r_{max}$. $\square$

Consider the following function.

$$\Psi\left(r|\bar{r}\right) = (h(\bar{r}) - h\left(r\right))(1+r)\lambda - h(\bar{r})\lambda(1+r)\left(\frac{r}{\bar{r}} - 1\right) - R(r, h\left(r\right)) + R(\bar{r}, h(\bar{r})). \tag{A11}$$

Note that Ψ is equivalent to the objective function of a hospital under HRRP-I by (1) and (9) if all the other hospitals pick actions $(\bar{r}, h(\bar{r}))$. Also, the objectives of a hospital under HRRP-II and HRRP are also equivalent to Ψ on a subset of their domain depending on the value of P_{cap}. We use the following result in establishing the equilibrium outcomes under all of these payment schemes.

Lemma A4. *$\Psi\left(r|\bar{r}\right)$ is concave in r and has a unique maximizer $r_\Psi\left(\bar{r}\right) \in (r_{min}, r_{max})$ satisfying*

$$d\Psi\left(r_\Psi\left(\bar{r}\right)|\bar{r}\right)/dr = 0.$$

Proof of Lemma A4: By (A11)

$$\frac{d\Psi\left(r|\bar{r}\right)}{dr} = (h(\bar{r}) - h(r))\lambda - \frac{1+2r-\bar{r}}{\bar{r}}h(\bar{r})\lambda - R_r(r, h(r)), \tag{A12}$$

$$\frac{d^2\Psi\left(r|\bar{r}\right)}{dr^2} = -(\lambda + R_{rc}(r, h(r)))h'(r) - \frac{2}{\bar{r}}h(\bar{r})\lambda - R_{rr}(r, h(r)). \tag{A13}$$

By Assumption A1(i) and (iii), $\lim_{r\downarrow r_{min}} d\Psi\left(r|\bar{r}\right)/dr > 0$ and $\lim_{r\uparrow r_{max}} d\Psi\left(r|\bar{r}\right)/dr < 0$. Moreover, $d^2\Psi\left(r|\bar{r}\right)/dr^2 < 0$ by Assumption A1(i). Thus, there exists a unique $r_\Psi\left(\bar{r}\right)$ such that $d\Psi\left(r_\Psi\left(\bar{r}\right)|\bar{r}\right)/dr = 0$. $\square$

B.2. Proof of Lemma 1

By (2), we have

$$\frac{\partial S(r,c)}{\partial c} = -(1+r)\lambda - R_c(r,c),$$

$$\frac{\partial^2 S(r,c)}{\partial c^2} = -R_{cc}(r,c).$$

By Lemma A1 there exists a unique $h(r) \in (c_{min}, c_{max})$ that satisfies (4) for a given $r \in [r_{min}, r_{max}]$. Since $S(r,c)$ is concave in c ($R_{cc} > 0$ by Assumption A1(i)),

$$S(r, h(r)) = \sup_{c\in[c_{min}, c_{max}]} S(r,c).$$

Next we show that there exits a unique $r^* \in (r_{min}, r_{max})$ that satisfies

$$S(r^*, h(r^*)) = \sup_{r\in[r_{min}, r_{max}]} S(r, h(r)).$$

Let $S(r) = S(r, h(r))$ for notational simplicity. Then

$$\frac{dS(r)}{dr} = -h'(r)(1+r)\lambda - h(r)\lambda - R_r(r, h(r)) - R_c h'(r) = -h(r)\lambda - R_r(r, h(r)). \tag{A14}$$

Differentiating again with respect to r, we have

$$\frac{d^2S(r)}{dr^2} = -(\lambda + R_{rc})h' - R_{rr} = \frac{(\lambda + R_{rc})^2}{R_{cc}} - R_{rr} < 0, \tag{A15}$$

where the second equality follows by plugging in h', see (A3), and the inequality follows from Assumption A1(i). Moreover, we have $\lim_{r\downarrow r_{min}} dS(r)/dr > 0$ and $\lim_{r\uparrow r_{max}} dS(r)/dr < 0$ by Assumption A1(iii). Thus there exists a unique $r^* \in (r_{min}, r_{max})$ that satisfies (3) with $c^* = h(r^*)$. □

B.3. Proof of Proposition 1

(a) We establish hospital i's best response given all the other hospital's actions, and for the rest of the proof we drop the hospital subscript "i" from our notation for simplicity.

First we argue that given $(\bar{r}, \bar{c})$ (see (6)) hospital i has a unique best response. By (1) and (9) hospital i's objective Π under HRRP-I is equal to Ψ defined in (A11). By Lemmas A1, A2, and A4 for any $(\bar{r}, \bar{c})$ hospital i has a unique best response (r, c), and $c = h(r)$.

Next we focus on symmetric equilibria. We first show that there exists at least one symmetric equilibrium. Let

$$\Gamma(r) = \frac{1+r}{r} h(r)\lambda + R_r(r, h(r)). \tag{A16}$$

By Lemma A4 and (A12) if

$$\Gamma(r) = 0, \tag{A17}$$

that is $r_\Psi(r) = r$, then $(r, h(r))$ is a symmetric equilibrium. By Assumption A1(iii), $\Gamma(r_{min}) < 0$ and $\Gamma(r_{max}) > 0$. Thus, by continuity of Γ on $[r_{min}, r_{max}]$, (A17) has at least one solution.

Next we prove that for any potential symmetric equilibrium $(\tilde{r}, \tilde{c})$, we have $\tilde{r} < r^*$. Consider the regulator's objective S defined by (2) and let $\Gamma_S(r) = -\frac{dS(r,h(r))}{dr}$ for notational simplicity. By (A14)

$$\Gamma_S(r) = h(r)\lambda + R_r(r, h(r)). \tag{A18}$$

Note that Γ_S is increasing by (A15) and $\Gamma_S(r^*) = 0$ by Lemma 1. Also $\Gamma(r) > \Gamma_S(r)$ for all $r > 0$ by (A16) and (A18). Thus

$$\Gamma(r) > \Gamma_S(r^*) = 0, \quad \text{for all } r \geq r^*, \tag{A19}$$

proving that $\tilde{r} < r^*$. By Lemma 1, $c^* = h(r^*)$, hence $\tilde{r} < r^*$ implies $\tilde{c} = h(\tilde{r}) > c^*$ by Lemma A1.

(b) Next we prove that no asymmetric equilibrium exists when $N = 2$. Suppose on the contrary that there exists an asymmetric equilibrium, where hospital i picks actions (r_i, c_i), $i = 1, 2$. We have $r_1 \neq r_2$ because otherwise we have $c_1 = h(r_1) = h(r_2) = c_2$ in any equilibrium by Lemma A1 and

(r_i, c_i), $i = 1, 2$, would not be an asymmetric equilibrium. Assume without loss of generality that $r_1 > r_2$. Then by (1) and (9) objective functions Π_1 and Π_2 of hospital 1 and 2 can be written as

$$\Pi_1 = (c_2 - c_1)(1 + r_1)\lambda - (r_1 - r_2)(1 + r_1)c_2\lambda/r_2 - R(r_1, c_1) + R(r_2, c_2),$$
$$\Pi_2 = (c_1 - c_2)(1 + r_2)\lambda - (r_2 - r_1)(1 + r_2)c_1\lambda/r_1 - R(r_2, c_2) + R(r_1, c_1)$$

by (A11). Thus

$$\Pi_1 + \Pi_2 = \frac{(r_1 - r_2)[c_1 r_2(1 + r_2 - r_1) - c_2 r_1(1 + r_1 - r_2)]\lambda}{r_1 r_2}.$$

Because $r_1 > r_2$, and $c_1 = h(r_1) < h(r_2) = c_2$, we have by (A3) $c_1 r_2(1 + r_2 - r_1) < c_2 r_1(1 + r_1 - r_2)$. Thus $\Pi_1 + \Pi_2 < 0$, which implies that at least one of the hospitals earns negative profits. However, both hospitals can earn zero profits if they take identical actions, thus (r_i, c_i), $i = 1, 2$ cannot be an equilibrium. □

B.4. Proof of Proposition 2

We break the proof into four steps. In step (a) we prove part (iii) of the proposition then we prove parts (i) and (ii) in steps (b)-(e). Specifically in part (b) we present the objective function of hospital 1 when all the other hospitals take identical actions. In part (c) we show that there exists no other symmetric equilibrium besides $\tilde{r}$ when $P_{cap} > 0$. In part (d) we present two preliminary results. Finally we complete the proof in part (e) by showing that $\tilde{r}$ is an equilibrium point only if P_{cap} is large enough.

Throughout the proof we only focus on potential symmetric equilibrium and the average readmission rate is identical to each hospital's readmission rate in any symmetric equilibrium, therefore, with a slight abuse of notation, we use $\bar{r}$ to denote a hospital's choice of readmission rate for notational simplicity. To prove that $(\tilde{r}, \tilde{c})$ is a symmetric equilibrium it is enough to show that for a given hospital the best action is $(\tilde{r}, \tilde{c})$ if all other hospitals choose $(\tilde{r}, \tilde{c})$. Without loss of generality, we focus on hospital 1 and drop the hospital index subscript from all mathematical expressions throughout the proof. By Lemmas A1, A2, and A4 for any $(\bar{r}, \bar{c})$ hospital i has a unique best response (r, c), and $c = h(r)$ under HRRP-II and since this relationship uniquely determines c we only focus on hospital's readmission choice r and drop c from the notation.

(a) First we prove part (iii) of the proposition. Fix $\bar{r} < r_{max}$. By (1) and (14) hospital 1's objective is given by $\Phi^o(\cdot|\bar{r})$ defined as in (A5). Then by Lemma A3(i), if $\bar{r} < r_{max}$, then

$$\Phi^o(r_{max}|\bar{r}) > 0.$$

In addition $\Phi^o(\bar{r}|\bar{r}) = 0$. Thus such $\bar{r} < r_{max}$ cannot be an equilibrium if $P_{cap} = 0$. Assume $\bar{r} = r_{max}$. By Lemma A3(i), hospital 1's best response is r_{max}. Therefore, $(r_{max}, h(r_{max}))$ is an equilibrium.

(b) Next we present preliminary results we use in proving parts (i) and (ii). Define

$$\Pi(r|\bar{r}) = \begin{cases} \Omega\left(r|\bar{r}\right), \text{ if } r < \bar{r}\left(1 - P_{cap}\right), \\ \Psi\left(r|\bar{r}\right), \text{ if } \bar{r}\left(1 - P_{cap}\right) \leq r < \bar{r}\left(1 + P_{cap}\right), \\ \Phi\left(r|\bar{r}\right), \text{ if } r \geq \bar{r}\left(1 + P_{cap}\right), \end{cases} \tag{A20}$$

where Ψ is defined as in (A11) and

$$\Omega\left(r|\bar{r}\right) = \left(h(\bar{r}) - h\left(r\right)\right)\lambda\left(1 + r\right) + h(\bar{r})\lambda\left(1 + r\right)P_{cap} - R\left(r, h\left(r\right)\right) + R(\bar{r}, h(\bar{r})), \tag{A21}$$

$$\Phi\left(r|\bar{r}\right) = \left(h(\bar{r}) - h\left(r\right)\right)\lambda\left(1 + r\right) - h(\bar{r})\lambda\left(1 + r\right)P_{cap} - R\left(r, h\left(r\right)\right) + R(\bar{r}, h(\bar{r})). \tag{A22}$$

By (1) and (14), hospital 1's objective is given by $\Pi(r|\bar{r})$ under HRRP-II, assuming all other hospitals choose $(\bar{r}, h(\bar{r}))$. Throughout the proof, we write $\Pi\left(r, P_{cap}|\bar{r}\right)$ (instead of just $\Pi\left(r|\bar{r}\right)$) when we need to make the dependence of this function on P_{cap} explicit and we follow the same convention with Ω and Φ.

(c) Next we show that there exists no other symmetric equilibrium besides $\tilde{r}$ when $P_{cap} > 0$. Recall that the symmetric equilibrium under HRRP-I is unique by Assumption 1. In addition by (9) the objective of a hospital under HRRP-I is equal to Ψ. Therefore by Lemma A4,

$$r_{\Psi}\left(\tilde{r}\right) = \tilde{r} \tag{A23}$$

and $\tilde{r}$ is the unique readmission level that satisfies this equality, since $\tilde{r}$ is the unique Nash equilibrium. By (A20), $\Pi(r|\bar{r}) = \Psi(r|\bar{r})$ for all $r \in (\bar{r}(1 - P_{cap}), \bar{r}(1 + P_{cap}))$. By Lemma A4

$$\frac{d\Psi\left(\bar{r}|\bar{r}\right)}{dr} \neq 0 \tag{A24}$$

if $\bar{r} \neq \tilde{r}$, because $\tilde{r}$ is the unique Nash equilibrium by Assumption 1. Hence $\bar{r} \neq \tilde{r}$ cannot be an equilibrium since hospital 1 can increase its profit by deviating from this point.

(d) To establish hospital 1's best response we need to compare the values of Ω, Ψ, and Φ, since they determine the hospital's profit under different actions. To do this we first prove two preliminary results. First result below (Lemma A5) shows that hospital 1 will never choose $r < \bar{r}\left(1 - P_{cap}\right)$ (assuming it is feasible), hence we can restrict our attention to $r \geq \bar{r}\left(1 - P_{cap}\right)$ when we establish hospital's best response. Then in Lemma A7 we show that for P_{cap} small enough hospital 1 can earn positive profits even when its actions are restricted to $r \geq \bar{r}\left(1 + P_{cap}\right)$. We later use this result to argue that for such P_{cap} there is no symmetric equilibrium because hospital 1's profit is equal to 0 in any symmetric equilibrium.

The following result shows that hospital 1 will never choose $r < \bar{r}\left(1 - P_{cap}\right)$ (assuming it is feasible).

Lemma A5. *For any fixed* $\bar{r} \in [r_{min}, r_{max}]$,

$$d\Omega\left(r|\bar{r}\right)/dr > 0,$$

i.e., $\Omega\left(r|\bar{r}\right)$ *is increasing in* $r \in [r_{min}, r_{max}]$.

Proof: Note that

$$\frac{d\Omega\left(r|\bar{r}\right)}{dr} = \left(h(\bar{r}) - h\left(r\right)\right)\lambda + h(\bar{r})\lambda P_{cap} - R_r\left(r, h\left(r\right)\right), \tag{A25}$$

$$\frac{d^2\Omega\left(r|\bar{r}\right)}{dr^2} = -R_{rr}(r, h\left(r\right)) + \frac{\left(\lambda + R_{rc}\left(r, h\left(r\right)\right)\right)^2}{R_{cc}\left(r, h\left(r\right)\right)} < 0, \tag{A26}$$

where inequality follows from Assumption A1(i). Hence Ω is concave. In addition, because $h(\bar{r}) > h\left(r_{max}\right)$ by Lemma A1 and by Assumption A1(i)

$$\frac{d\Omega\left(r_{max}|\bar{r}\right)}{dr} \geq 0.$$

This with concavity of Ω gives the desired result. $\square$

Next we consider the objective of hospital 1 for $r \geq \bar{r}\left(1 + P_{cap}\right)$. First for $\bar{r} \in [r_{min}, r_{max})$, let

$$P_m(\bar{r}) = \frac{r_{max}}{\bar{r}} - 1. \tag{A27}$$

Note that if $P_{cap} > P_m(\bar{r})$ then $\bar{r}(1 + P_{cap}) \geq r_{max}$, hence hospital 1's objective is never equal to Φ by (A20). So we only consider $P_{cap} \in [0, P_m(\bar{r})]$ in the next result. Fix $\bar{r} \in [r_{min}, r_{max})$ and let

$$\hat{\Phi}\left(P_{cap}\right) = \sup_{r \in [\bar{r}(1+P_{cap}), r_{max}]} \Phi\left(r, P_{cap}|\bar{r}\right), \quad \text{for } P_{cap} \in [0, P_m(\bar{r})]. \tag{A28}$$

In words, $\hat{\Phi}\left(P_{cap}\right)$ is hospital 1's maximum profit when its actions are restricted to $r \geq \bar{r}\left(1 + P_{cap}\right)$. We next establish two properties of this function, which we use below to prove Lemma A7.

Lemma A6. *For* $\bar{r} \in [r_{min}, r_{max})$, $\hat{\Phi}$ *is decreasing and right-continuous on* $[0, P_m(\bar{r}))$.

Proof Lemma A6: The fact that $\hat{\Phi}$ is decreasing follows from the definition of $\hat{\Phi}$ (see (A28)), from the fact that $\Phi\left(r, P_{cap}|\bar{r}\right)$ is decreasing in P_{cap} for fixed $r, \bar{r} \in [r_{min}, r_{max}]$ by (A22), and the fact that the set $[\bar{r}(1 + P_{cap}), r_{max}]$ becomes smaller with increasing P_{cap}.

Next we prove $\hat{\Phi}$ is right continuous. Let $P_n \downarrow P \in [0, P_m(\bar{r}))$ as $n \to \infty$. We next show that for any $\epsilon > 0$

$$\liminf_{n\to\infty} \hat{\Phi}\left(P_n\right) > \hat{\Phi}\left(P\right) - \epsilon. \tag{A29}$$

Because $\hat{\Phi}(P)$ is decreasing in P this implies it is right continuity since ϵ is arbitrary. Given $\epsilon > 0$, there exists $\delta > 0$ such that for

$$\hat{\Phi}^{\epsilon}\left(P\right) = \sup_{r \in [\bar{r}(1+P+\delta), r_{max}]} \Phi\left(r, P|\bar{r}\right)$$

we have

$$\hat{\Phi}^{\epsilon}(P) > \hat{\Phi}(P) - \epsilon \tag{A30}$$

since $\Phi(r, P|\bar{r})$ is continuous in r. Again by continuity, there exists $r^{\epsilon} \in [\bar{r}(1+P+\delta), r_{max}]$ such that

$$\hat{\Phi}^{\epsilon}(P) = \Phi(r^{\epsilon}, P|\bar{r}).$$

Hence for n large enough so that $P_n \leq P + \delta$

$$\liminf_{n\to\infty} \hat{\Phi}(P_n) \overset{\text{(r0)}}{\geq} \liminf_{n\to\infty} \Phi(r^{\epsilon}, P_n|\bar{r}) \overset{\text{(r1)}}{=} \Phi(r^{\epsilon}, P|\bar{r}) \overset{\text{(r2)}}{=} \hat{\Phi}^{\epsilon}(P), \tag{A31}$$

where $(r0)$ follows from the definition of $\hat{\Phi}$ and the fact that $P_n \leq P + \delta$, $(r1)$ follows from the continuity of Φ, and (r2) follows from the definition of $\hat{\Phi}^{\epsilon}$. Clearly, (A30) with (A31) implies (A29). $\square$

Next result shows that hospital 1 can earn positive profits even when its actions are restricted to $r \geq \bar{r}(1 + P_{cap})$.

Lemma A7. *Let $\bar{r} \in [r_{min}, r_{max})$. One of the following holds.*

i. For all $P_{cap} \in [0, P_m(\bar{r})]$

$$\hat{\Phi}(P_{cap}) > 0. \tag{A32}$$

ii. There exists $\bar{P}_{cap}(\bar{r}) \in [0, P_m(\bar{r})]$ such that if $P_{cap} \in [0, \bar{P}_{cap}(\bar{r}))$ then (A32) *holds and if $P_{cap} \in [\bar{P}_{cap}(\bar{r}), P_m(\bar{r})]$ then*

$$\hat{\Phi}(P_{cap}) \leq 0. \tag{A33}$$

Proof Lemma A7: We first argue that it is enough to show that there exists $P_{cap} \in (0, P_m(\bar{r})]$ such that

$$\hat{\Phi}(P_{cap}) > 0. \tag{A34}$$

This follows from the fact that if (A34) holds then, because $\hat{\Phi}(P)$ is decreasing by Lemma A6, $\hat{\Phi}(P) > 0$ for all $P \in [0, P_{cap}]$. Therefore, if $\hat{\Phi}(P_m(\bar{r})) > 0$ and (A34) holds, then (i) holds because $\hat{\Phi}(P)$ is decreasing by Lemma A6. If $\hat{\Phi}(P_m(\bar{r}))) \leq 0$ and (A34) hold, then for $\bar{P}_{cap}(\bar{r}) = \inf\{P : \hat{\Phi}(P) \leq 0\}$ part (ii) holds. (Note that $\hat{\Phi}\left(\bar{P}_{cap}(\bar{r})\right) \leq 0$ by right-continuity of $\hat{\Phi}(P)$ proved in Lemma A6.) Therefore it is enough to show that (A34) holds.

To prove (A34) we next show that given $\bar{r} \in [r_{min}, r_{max})$

$$\Phi(r_{max}, P_{cap}|\bar{r}) > 0 \tag{A35}$$

for $P_{cap} \in (0, P_m(\bar{r})]$ small enough. Clearly (A35) implies $\hat{\Phi}(P_{cap}) > 0$ by (A28).

Fix $\bar{r} \in [r_{min}, r_{max})$. By (A5) and (A22)

$$\Phi(r, P_{cap}|\bar{r}) = \Phi^o(r|\bar{r}) - P_{cap}(1+r)h(\bar{r})\lambda. \tag{A36}$$

Let $r_\Delta = \bar{r} + \frac{r_{max}-\bar{r}}{2}$, denote the midpoint between $\bar{r}$ and r_{max}. Because $\bar{r} \in [r_{min}, r_{max})$, we have $r_\Delta < r_{max}$. By Lemma A3(ii)

$$\Phi^o(r_{max}|\bar{r}) \geq (r_{max} - r_\Delta)(h(\bar{r}) - h(r_\Delta))\lambda \equiv \Delta > 0, \tag{A37}$$

where the inequality follows from $h(\bar{r}) - h(r_\Delta) > 0$ by Lemma A1 . Let

$$P_{cap} = \min\left\{\frac{\Delta}{2h(\bar{r})(1+r_{max})\lambda}, P_m(\bar{r})\right\}. \tag{A38}$$

By (A36)–(A38), we have

$$\Phi(r_{max}, P_{cap}|r) \geq \Delta/2 > 0,$$

giving (A35). $\square$

(e) Finally we finish the proof by showing that $\tilde{r}$ is an equilibrium only if P_{cap} is large enough. We show this by assuming all hospitals other than hospital 1 choose $\tilde{r}$ and then identifying conditions under which $\tilde{r}$ is hospital 1's best response.

Hospital 1's profit $\Pi(r|\tilde{r})$ (given all other hospitals choose $\tilde{r}$) is given by (A20), where $\bar{r} = \tilde{r}$. By Lemma A5, hospital 1's best response cannot be less than $\underline{r} = \max\left(\tilde{r}\left(1 - P_{cap}\right), r_{min}\right)$. Hence it is enough to consider $r \in [\underline{r}, r_{max}]$ to determine the best response of hospital 1.

We use Lemma A7 to complete the proof. First, for $\bar{r} = \tilde{r}$

$$\hat{\Phi}\left(P_m(\tilde{r})\right) \overset{(r0)}{=} \Phi\left(r_{max}, P_m(\tilde{r})|\tilde{r}\right) \overset{(r1)}{=} \Psi\left(r_{max}, P_m(\tilde{r})|\tilde{r}\right) \overset{(r2)}{\leq} \Psi\left(\tilde{r}, P_m(\tilde{r})|\tilde{r}\right) = 0, \tag{A39}$$

where (r0) follows from (A28), (r1) follows from (A11), and (r2) follows from Assumption 1, as otherwise $\tilde{r}$ would not be an equilibrium. Therefore, (A39) implies that Lemma A7(i) cannot hold for $\bar{r} = \tilde{r}$.

Set $\bar{P}_{cap} = \bar{P}_{cap}(\tilde{r})$, for $\bar{P}_{cap}(\tilde{r})$ defined as in Lemma A7(ii). By Lemma A7(ii) for $P_{cap} \geq \bar{P}_{cap}$, $\hat{\Phi}\left(P_{cap}\right) \leq \Psi(\tilde{r}|\tilde{r}) = 0$ and so by (A23) $\tilde{r}$ is the best response for hospital 1. Hence $\tilde{r}$ is a Nash equilibrium. If on the other hand $P_{cap} < \bar{P}_{cap}$, then $\hat{\Phi}\left(P_{cap}\right) > \Psi(\tilde{r}|\tilde{r}) = 0$ by Lemma A7(ii), hence $\tilde{r}$ cannot be a Nash equilibrium. $\square$

B.5. Proof of Proposition 3

We proceed as follows. In steps (a)-(c) we prove part (i), and show that part (iii) follows from part (i). In step (d) we prove part (ii) and finally in step (e) we prove part (iv). As in the proof of Proposition 2, we use $\bar{r}$ to denote a hospital's choice to keep the notation consistent throughout the proof, since in all symmetric equilibria each hospital's choice is equal to the average readmission rate. Also we focus on the objective of hospital 1 (without loss of generality), and use Π to denote it under HRRP defined as in (8), unless otherwise stated, and drop the hospital index from all mathematical expressions throughout the proof. To prove that $(\tilde{r}, \tilde{c})$ is a symmetric equilibrium it is enough to show that hospital 1's best action is $(\tilde{r}, \tilde{c})$ given that all other hospitals choose $(\tilde{r}, \tilde{c})$.

By Lemmas A1, A2, and A4 for any $(\bar{r}, \bar{c})$ hospital i has a unique best response (r, c) under HRRP, and $c = h(r)$ and since this relationship uniquely determines c we only focus on hospital's readmission choice r and drop c from the notation. Hence hospital 1's profit, if all the other hospitals pick $(\bar{r}, h(\bar{r}))$, is given by

$$\begin{aligned} \Pi(r|\bar{r}) = (h(\bar{r}) - h(r))(1+r)\lambda - \left(\min\left\{\frac{r-\bar{r}}{\bar{r}}, P_{cap}\right\}\right)^{+} h(\bar{r})(1+r)\lambda \\ - R(r, h(r)) + R(\bar{r}, h(\bar{r})) \end{aligned} \tag{A40}$$

by (8). Again, we write $\Pi(r, P_{cap}|\bar{r})$ instead of $\Pi(r|\bar{r})$ when we need to make the dependence on P_{cap} explicit.

(a) We start with a preliminary result, which shows that given all other hospitals choose the same readmission level $\bar{r}$, hospital 1's best response cannot be less than that readmission level under the HRRP scheme.

Corollary A1. *Fix $\bar{r} \in [r_{min}, r_{max}]$ and $P_{cap} > 0$. Under the HRRP scheme,*

$$\Pi(\bar{r}|\bar{r}) > \Pi(r|\bar{r}) \quad \textit{for any } r \in [r_{min}, \bar{r}), \tag{A41}$$

i.e., hospital 1's profit at $(\bar{r}, h(\bar{r}))$ is higher than that at any other point $(r, h(r))$ if $r \in [r_{min}, \bar{r})$.

This result follows from Lemma A3(i) because by (A40), $\Pi(r|\bar{r}) \equiv \Phi^o(r|\bar{r})$ for $r \in [r_{min}, \bar{r}]$ (see (A5) for the definition of Φ^o), since a hospital with lower-than-expected readmission rate does not receive rewards.

(b) We first show that any $(\bar{r}, h(\bar{r})) \in \mathcal{S}_p$ is a symmetric equilibrium. Assume that all hospitals except hospital 1 pick $(\bar{r}, h(\bar{r})) \in \mathcal{S}_p$. By Corollary A1 it is enough to show that hospital 1's profit at $(\bar{r}, h(\bar{r}))$ is higher than that at any other point $(r, h(r))$, for $r \in [\bar{r}, r_{max}]$, that is,

$$\Pi(\bar{r}|\bar{r}) \geq \Pi(r|\bar{r}) \quad \text{for any } r \in [\bar{r}, r_{max}]. \tag{A42}$$

By (A40) and because $(\bar{r}, h(\bar{r})) \in \mathcal{S}_p$, $\Pi(r|\bar{r}) = \Psi(r|\bar{r})$, (see (A11) for the definition of Ψ), hence the objective of hospital 1 is identical to that under HRRP-I if $(\bar{r}, h(\bar{r})) \in \mathcal{S}_p$. By Lemma A4, Ψ is concave hence it is enough to show that

$$\frac{d\Psi(\bar{r}|\bar{r})}{dr} < 0, \quad \text{for any } \bar{r} > \tilde{r}. \tag{A43}$$

To prove (A43) note that

$$\frac{d\Psi(\bar{r}|\bar{r})}{dr} = -\Gamma(\bar{r}), \tag{A44}$$

where Γ is defined as in (A16), by (A12).

By (A43) and (A44) it is enough to show that $\Gamma(r) > 0$ for all $r > \tilde{r}$. To prove this we use the following three results. (1) By Assumption 1, $\tilde{r}$ is the unique equilibrium under HRRP-I hence $\tilde{r}$ is the only readmission level that satisfies $\Gamma(\tilde{r}) = 0$. (2) By (A19), $\Gamma(r) > 0$ for all $r \geq r^*$. (3) Function Γ is continuous by the continuity of h and R_r. However if $\Gamma(r) < 0$ for some $r \in (\tilde{r}, r^*)$ then Γ cannot be continuous because $\tilde{r}$ is the unique point that satisfies $\Gamma(\tilde{r}) = 0$, a contradiction.

(c) Next we focus on the rest of the symmetric equilibria. First we explicitly define $\mathcal{S}_o \subset \mathcal{S}$ and then show that any $(\bar{r}, h(\bar{r})) \in \mathcal{S}_o$ is a symmetric equilibrium. Let

$$\mathcal{S}_o = \{(\bar{r}, \bar{c}) : \bar{c} = h(\bar{r}), \bar{r} \in [\tilde{r}, r_p], \sup_{r \in [(1+P_{cap})\bar{r}, r_{max}]} \Phi(r|\bar{r}) \leq 0\}, \tag{A45}$$

where function Φ is defined as in (A22). (In words, Φ is hospital 1's profit when its readmission level r is large enough so that the readmission penalty is capped.) If $\mathcal{S}_o = \emptyset$ then there is nothing to prove, hence assume for the rest of the proof that it is non-empty.

Assume that all hospitals except hospital 1 pick $(\bar{r}, h(\bar{r})) \in \mathcal{S}_o$. As in part (b) it is enough to show (A42) holds. First note that for $r \in [\bar{r}, (1 + P_{cap})\bar{r}]$, $\Pi(r|\bar{r}) = \Psi(r|\bar{r})$ by (A40). Then by (A43)

$$\Pi(\bar{r}|\bar{r}) \geq \Pi(r|\bar{r}) \quad \text{for any } r \in [\bar{r}, (1 + P_{cap})\bar{r}]. \tag{A46}$$

If on the other hand $r \in [(1+P_{cap})\bar{r}, r_{max}]$, then hospital 1's profit is given by $\Phi(r|\bar{r})$. By definition of $\mathcal{S}_o$, $\Phi(r|\bar{r}) \leq 0$ for any $r \in [(1 + P_{cap})\bar{r}, r_{max}]$. In addition $\Pi(\bar{r}|\bar{r}) = 0$ by (8). Hence any $(\bar{r}, h(\bar{r})) \in \mathcal{S}_o$ is a symmetric equilibrium.

(d) Next we prove that $(\bar{r}, \bar{c}) \notin \mathcal{S}_o \cup \mathcal{S}_p$ cannot be a symmetric equilibrium. Let $\bar{r} \in [r_{min}, r_{max}]$ and assume that all hospitals except hospital 1 choose $(\bar{r}, \bar{c}) \notin \mathcal{S}_o \cup \mathcal{S}_p$. If $\bar{c} \neq h(\bar{r})$, then $(\bar{r}, \bar{c})$ cannot be an equilibrium by Lemma A2, hence assume that $\bar{c} = h(\bar{r})$. Because $(\bar{r}, \bar{c}) \notin \mathcal{S}_o \cup \mathcal{S}_p$, either (1) $\bar{r} \in [r_{min}, \tilde{r})$, or (2) $r \in [\tilde{r}, r_p]$ and $(r, h(r)) \notin \mathcal{S}_o$, by definitions of $\mathcal{S}_o$ and $\mathcal{S}_p$.

Assume that $\bar{r} \in [r_{min}, \tilde{r})$. Then for $r \in (\bar{r}, (1 + P_{cap})\bar{r}]$, by (A40), $\Pi(r|\bar{r}) = \Psi(r|\bar{r})$. Hence it is enough to show that

$$\frac{d\Psi(\bar{r}|\bar{r})}{dr} > 0, \quad \text{for any } \bar{r} < \tilde{r}. \tag{A47}$$

Proof of (A47) is identical to that of (A43) using (A44) and the fact that $\Gamma(r_{min}) < 0$ (which follows from Assumption A1(iii)), hence we skip the details.

Assume now that $\bar{r} \in [\tilde{r}, r_p]$ and $\bar{r} \notin \mathcal{S}_o$, then as in part (c) of the proof, for $r \geq (1+P_{cap})\bar{r}$, hospital 1's profit is given by $\Phi(r|\bar{r})$. And by definition of $\mathcal{S}_o$, $\Phi(r_1|\bar{r}) > \Pi(\bar{r}, h(\bar{r})) = 0$, for some $r_1 \in [(1+P_{cap})\bar{r}, r_{max}]$. Hence such $\bar{r}$ cannot be an equilibrium.

Note that (a)–(c) prove part (i) of the proposition. Part (iii) follows from the fact that $r_p = \tilde{r}$ if $P_{cap} \geq P_{max}$.

(e) We next prove part (ii) of the proposition. First, if $P_{cap} < P_m(r^*)$, so that $r_p > r^*$, then $\mathcal{S}_p \subset \{(r, h(r)) : r \in (r^*, r_{max}]\}$ by (16). Hence for the rest of the proof we assume without loss of generality that $P_{cap} < P_m(r^*)$ so that $r_p > r^*$ and show that $\mathcal{S}_o \subset (r^*, r_{max}]$ for P_{cap} small enough. To prove $\mathcal{S}_o \subset \{(r, h(r)) : r \in (r^*, r_{max}]\}$, we show that there exists $\bar{\bar{P}}_{cap} \in (0, P_{max})$ such that

$$\Phi(r_{max}, P_{cap}|\bar{r}) > 0, \quad \text{for all } \bar{r} \in [\tilde{r}, r^*], \text{ and for } P_{cap} \leq \bar{\bar{P}}_{cap}. \tag{A48}$$

Note that (A48) implies that $\mathcal{S}_o \subset \{(r, h(r)) : r \in (r^*, r_{max}]\}$ for all $P_{cap} \leq \bar{\bar{P}}_{cap}$ by (A45). To prove (A48) we use an argument similar to that in the proof Lemma A7 .

By (A5) and (A22)

$$\Phi(r, P_{cap}|\bar{r}) = \Phi^o(r|\bar{r}) - P_{cap}(1+r)h(\bar{r})\lambda, \text{ for } r, \bar{r} \in [r_{min}, r_{max}] \text{ and } P_{cap} > 0. \tag{A49}$$

Let $r_\delta = r^* + \frac{r_{max} - r^*}{2}$, denote the midpoint between r^* and r_{max}. By Lemma 1, $r_\delta > r^*$ and $r_\delta < r_{max}$. Since h is strictly decreasing in r by Lemma A1, $h(r^*) - h(r_\delta) > 0$. Therefore by Lemma A3(ii)

$$\Phi^o(r_{max}|\bar{r}) \geq (r_{max} - r_\delta)\,(h(r^*) - h(r_\delta))\,\lambda \equiv \Delta > 0,$$

for any $\bar{r} \in [\tilde{r}, r^*]$. Let

$$\bar{\bar{P}}_{cap} = \min\left\{\frac{\Delta}{2h(\tilde{r})(1+r_{max})\lambda}, P_m(r^*)\right\}. \tag{A50}$$

By (A49)–(A50) and Lemma A1, we have

$$\Phi(r_{max}, P_{cap}|\bar{r}) \geq \Delta/2 > 0,$$

for any $\bar{r} \in [\tilde{r}, r^*]$ and $P_{cap} \leq \bar{\bar{P}}_{cap}$, giving (A48).

(e) Finally, we prove part (iv) of the proposition. Suppose hospitals take asymmetric actions (r_i, c_i), $i = 1, \ldots, N$. We prove that such a point cannot be an equilibrium. By Lemma A2, if $c_i \neq h(r_i)$ for some i then this point cannot be an equilibrium, so assume that $c_i = h(r_i)$ for all $i = 1, \ldots, N$. Without loss of generality, assume that $r_1 \leq r_2 \leq \ldots r_N$ and $r_1 < r_j$ for some $j \in \{2, \ldots, N\}$. By (8), hospital 1's objective function is

$$\Pi(r_1, h(r_1)) = (\bar{c}_1 - h(r_1))\,(1+r_1)\lambda - R(r_1, c_1) + \bar{R}_1,$$

where $\bar{c}_1$ and $\bar{R}_1$ are defined as in (6). Now because $h' < 0$ by Lemma A1 and $r_1 \le r_i$ for all $i \ne 1$, we have $\bar{c}_1 < h(r_1)$. Moreover, $dR(r,h(r))/dr < 0$ by our assumption, hence, $R(r_1,h(r_1)) > \bar{R}_1$ by (6) and so $\Pi(r_1,h(r_1)) < 0$. On the other hand $R(\bar{r}_1,\bar{c}_1) \le \bar{R}_1$, because R is convex by Assumption A1(i), and so (for hospital 1) $\Pi(\bar{r}_1,\bar{c}_1) \ge 0$ by (8). Hence these actions cannot constitute an equilibrium. □

C. Proofs of the results in Section 5

Proof of Proposition 4: The proof is based on the simple observation that under m-HRRP the difference between a hospital's objective and the regulator's objective is independent of that hospital's actions. More precisely, given the actions of all the other hospitals, by (1) and (17) hospital i's objective under m-HRRP is

$$\begin{aligned}\Pi(r_i,c_i) &= (\bar{c}_i - c_i)(1+r_i)\lambda + (\bar{r}_i - r_i)\,\bar{c}_i\lambda - R(r_i,c_i) + \bar{R}_i \\ &= \bar{c}_i(1+\bar{r}_i)\lambda - c_i(1+r_i)\lambda - R(r_i,c_i) + \bar{R}_i,\end{aligned} \tag{A51}$$

where $\bar{c}_i$ and $\bar{R}_i$ are defined as in (6). By (2) and (A51), we have

$$\Pi(r_i,c_i) - S(r_i,c_i) = \bar{c}_i(1+\bar{r}_i)\lambda + \bar{R}_i - V(\lambda).$$

Therefore the difference between the objective of the regulator and hospital i does not depend on r_i and c_i.

Since (r^*,c^*) is the unique maximizer of the social welfare S, it also maximizes hospital's profit Π under m-HRRP. That is, each hospital chooses r^* and c^* independent from other hospitals' decisions. Therefore, (r^*,c^*) constitutes the unique equilibrium under m-HRRP. □

Proof of Proposition 5: We prove the result by showing that in any symmetric equilibrium, FOCs of each hospital's objective under m-HRRPW payment scheme coincide with those of the regulator's objective.

By (24), social welfare is given by:

$$S(r_i,\mu_i,c_i) = \Lambda\int_{tW_i}^{\infty}(x - tW_i)\,d\Theta(x) - c_i(1+r_i)\lambda_i - R_i - c_e(\Lambda - \lambda_i),$$

where $W_i \equiv W(r_i,\mu_i)$ and $R_i \equiv R(r_i,\mu_i,c_i)$. Moreover, by (23) and (26), a hospital's profit under our proposed payment scheme is:

$$\Pi(r_i,\mu_i,c_i) = (\bar{c}_i(1+\bar{r}_i) - c_e)\bar{\lambda}_i - (c_i(1+r_i) - c_e)\lambda_i - t(W_i - \bar{W}_i)\lambda_i - R(r_i,\mu_i,c_i) + \bar{R}_i.$$

Then, for any decision variable $x_i \in \{r_i, \mu_i, c_i\}$, we have

$$\begin{aligned}
\frac{\partial S(r_i,\mu_i,c_i)}{\partial x_i} &= -t\Lambda\bar{\Theta}(tW_i)\frac{\partial W_i}{\partial x_i} - \frac{\partial\{[c_i(1+r_i)-c_e]\lambda_i + R_i\}}{\partial x_i} \\
&= -t\frac{\partial W_i}{\partial x_i}\lambda_i - \frac{\partial\{[c_i(1+r_i)-c_e]\lambda_i + R_i\}}{\partial x_i}, \\
\frac{\partial \Pi(r_i,\mu_i,c_i)}{\partial x_i} &= -t\frac{\partial W_i}{\partial x_i}\lambda_i - \frac{\partial\{[c_i(1+r_i)-c_e]\lambda_i + R_i\}}{\partial x_i} - t(W_i - \bar{W}_i)\frac{\partial \lambda_i}{\partial x_i}.
\end{aligned}$$

In any symmetric equilibrium, we have $W_i = \bar{W}_i$; therefore, $\frac{\partial \Pi}{\partial x_i} = \frac{\partial S_i}{\partial x_i}$ for all $x_i \in \{r_i, \mu_i, c_i\}$. The uniqueness of the symmetric equilibrium under m-HRRPW follows from the uniqueness of the social optimum (r^*, μ^*, c^*). Also in this symmetric equilibrium each hospital makes non-negative profit. □

D. Multiple monitored diseases

So far our analysis focused on a single disease model and we intrinsically assumed that if a regulator needs to monitor multiple diseases a similar payment scheme can be used for each disease separately. However the HRRP payment scheme CMS used actually monitors multiple diseases jointly and financial incentives are tied to the cumulative performance on readmission rates for these diseases. In this section we modify our model to study this payment scheme used by CMS and by leveraging our results for the single disease model in §4 demonstrate that with $J \geq 2$ monitored diseases there are again uncountably many symmetric equilibrium outcomes under HRRP.

We use the notation introduced in §3 except we now use j to index the monitored diseases. For example, λ_j denote the demand rate for patients with disease j and r_{ij} denotes the readmission rate at hospital i for disease j. For notational simplicity we set $\mathcal{J} = \{1, 2, \ldots, J\}$. As in our base model for the single disease model in §3, we assume that hospitals choose treatment costs and readmission rates for all diseases, and that there is ample capacity and all patients seek treatment.

We assume that investment cost associated with readmission and cost-reduction efforts is additively separable. More specifically if a hospital operates with readmission level $r_j \in [r_{min}, r_{max}]$ and marginal treatment cost $c_j \in [c_{min}, c_{max}]$ for disease $j \in \mathcal{J}$, then its total investment cost is given by

$$R(\mathbf{r}, \mathbf{c}) = \sum_{j \in \mathcal{J}} R^{(j)}(r_j, c_j),$$

where $\mathbf{r} = (r_1, r_2, \ldots, r_J)$, $\mathbf{c} = (c_1, c_2, \ldots, c_J)$, (we use bold letters to denote vectors) and cost functions $R^{(j)}$, $j \in \mathcal{J}$, are assumed to satisfy all conditions in Assumption A1. The additivity assumption enables us to utilize our results for the single disease case, specifically the equilibrium outcomes when we remove the penalty-only and penalty cap provisions, i.e., the payment system HRRP-I, the equilibrium outcomes are identical to the case when all the diseases are monitored separately.

Thence we are able to compare the impact of the cap in these two cases. Although it is possible to establish the equilibrium outcomes for more general investment functions, these outcomes depends on the specific assumptions on the impact of the interaction between multiple diseases on the total cost. We leave this for future research.

Regulator's objective: Similar to (2), total social welfare with multiple monitored diseases is given by

$$S(\mathbf{r},\mathbf{c}) = \sum_{j\in\mathcal{J}} \left[V_j(\lambda_j) - c_j(1+r_j)\lambda_j - R^{(j)}(r_j,c_j)\right]. \tag{A52}$$

Note that total social welfare is separable and results for social planner's problem from our single-disease model hold for each disease. Therefore socially-optimal actions $(\mathbf{r}^*, \mathbf{c}^*)$ satisfy

$$\begin{cases} (1+r_j^*)\lambda_j + R_c^{(j)}(r_j^*,c_j^*) = 0, \text{ for } j\in\mathcal{J}, \\ c_j^*\lambda_j + R_r^{(j)}(r_j^*,c_j^*) = 0, \text{ for } j\in\mathcal{J}, \end{cases} \tag{A53}$$

by Lemma 1. In addition, for any readmission level $r\in[r_{min}, r_{max}]$ the optimal treatment cost for disease j is given by $h^{(j)}(r)$, which satisfies

$$(1+r)\lambda_j + R_c^{(j)}(r, h^{(j)}(r)) = 0, \text{ for } j\in\mathcal{J}, \tag{A54}$$

by Lemma A1. For notational simplicity we set $h(\mathbf{r}) = \left(h^{(1)}(r_1), h^{(2)}(r_2), \ldots, h^{(J)}(r_J)\right)$.

Hospital's objective under HRRP: Next we present hospitals' objective with J monitored diseases under HRRP. Let $\mathbf{r}_i = (r_{ij}, j\in\mathcal{J})$, $\mathbf{c}_i = (c_{ij}, j\in\mathcal{J})$ denote the readmission rate and marginal cost of each hospital and disease, for $i=1,\ldots,N$ and $j\in\mathcal{J}$. Let $R_i^{(j)} \equiv R^{(j)}(r_{ij}, c_{ij})$ and

$$\bar{c}_{ij} = \frac{1}{N-1}\sum_{k\neq i} c_{kj}, \quad \bar{r}_{ij} = \frac{1}{N-1}\sum_{k\neq i} r_{kj}, \text{ and } \bar{R}_i^{(j)} = \frac{1}{N-1}\sum_{k\neq i} R_k^{(j)}, \text{ for } i=1,\ldots,N \text{ and } j\in\mathcal{J}, \tag{A55}$$

similar to (6). Also we set $\bar{\mathbf{r}}_i = (\bar{r}_{ij}, j\in\mathcal{J})$ and $\bar{\mathbf{c}}_i = (\bar{c}_{ij}, j\in\mathcal{J})$. Similar to (1) hospital i's profit with J monitored diseases can be written as

$$\hat{\Pi}(\mathbf{r_i},\mathbf{c_i}) = \sum_{j\in\mathcal{J}}\left[(\bar{c}_{ij} - c_{ij})\lambda_j(1+r_{ij})\right] - \hat{\pi}(\mathbf{r}_i|\bar{\mathbf{r}}_i, \bar{\mathbf{c}}_i) - \sum_{j\in\mathcal{J}} R^{(j)}(r_{ij},c_{ij}) + \sum_{j\in\mathcal{J}} \bar{R}_i^{(j)}, \tag{A56}$$

where

$$\hat{\pi}(\mathbf{r}_i|\bar{\mathbf{r}}_i, \bar{\mathbf{c}}_i) = \min\left(\sum_{j\in\mathcal{J}} \bar{c}_{ij}\lambda_j(1+r_{ij})\left(\frac{r_{ij}}{\bar{r}_{ij}} - 1\right)^+, P_{cap}\sum_{j\in\mathcal{J}} \bar{c}_{ij}\lambda_j(1+r_{ij})\right), \tag{A57}$$

and $P_{cap}\geq 0$ denotes the penalty cap as before.

Hospitals' objective is similar to the case with a single monitored disease except now it incorporates multiples diseases and in the incentive term $\hat{\pi}$ penalty cap is determined based on the cumulative readmission reduction effort of the hospital (in the single disease model penalty cap is determined separately for each disease, see term π in (8)). Term $\hat{\pi}$ is based on HRRP payment scheme implemented by CMS, see Zhang et al. (2016), that monitors multiple diseases. The total penalty is based on the "cost of excessive readmissions" for each disease, the term $\bar{c}_{ij}\lambda_j(1+r_{ij})\left(\frac{r_{ij}}{\bar{r}_{ij}}-1\right)^+$, but the total penalty is capped by the second term, $P_{cap}\sum_{j\in\mathcal{J}}\bar{c}_{ij}\lambda_j(1+r_{ij})$, based on the total payments for the monitored disease. Our model is slightly different from HRRP implemented in practice. Under HRRP the penalty cap is 3% of the total payments from Medicare, hence it does not depends solely on the payments for the monitored diseases. However as we discussed in Remark 1 this can be incorporated in our model by increasing P_{cap}.

HRRP-I with multiple monitored diseases: As in the case with a single monitored disease, the equilibrium outcomes under HRRP depends on the equilibrium outcomes when the penalty cap and no-reward provisions are removed from HRRP, and we refer to the resulting payment scheme again as HRRP-I. First by (A56), without the penalty cap ($P_{cap}=\infty$), hospital's profit is separable in each disease and maximizing a hospital's total profit is equivalent to maximizing its objective from each single disease. This implies that results from our single-disease model for HRRP-I scheme with reward and no penalty cap are valid for each disease. Specifically let $(\tilde{\mathbf{r}},\tilde{\mathbf{c}})$ be defined as follows

$$\begin{cases} (1+\tilde{r}_j)\lambda_j + R_c^{(j)}(\tilde{r}_j,\tilde{c}_j) = 0, \text{ for } j\in\mathcal{J}, \\ \frac{1+\tilde{r}_j}{\tilde{r}_j}\tilde{c}_j\lambda_j + R_r^{(j)}(\tilde{r}_j,\tilde{c}_j) = 0, \text{ for } j\in\mathcal{J}. \end{cases} \tag{A58}$$

By Proposition 1, under HRRP-I scheme, there exists symmetric (Nash) equilibrium and any symmetric equilibrium $(\tilde{\mathbf{r}},\tilde{\mathbf{c}})$ satisfy $\tilde{r}_j\le r_j^*$ and $\tilde{c}_j\ge c_j^*$ for $j\in\mathcal{J}$. We note that as in the single disease model, there might be multiple symmetric equilibria when multiple diseases are monitored. Therefore, in the remainder of this appendix, we assume that Assumption 1 holds, i.e., there exists a unique symmetric equilibrium $(\tilde{\mathbf{r}},\tilde{\mathbf{c}})$ under HRRP-I.

Equilibrium under HRRP with multiple monitored diseases: Next we characterize the equilibrium outcomes under HRRP scheme. Unlike HRRP-I, hospital's objective function is no longer separable because of the penalty cap and so our results for a single disease cannot be generalized in a straightforward manner. However, results in Section 4.3 can be extended to this case with redefining the equilibrium sets. We present the details next.

First for fixed $\bar{\mathbf{r}}=(\bar{r}_1,\bar{r}_2,\ldots,\bar{r}_J)\in[r_{min},r_{max}]^J$ we define

$$\mathbb{B}(\bar{\mathbf{r}})\equiv\left\{\mathbf{r}=(r_1,r_2,\ldots,r_J):\sum_{j\in\mathcal{J}}h^{(j)}(\bar{r}_j)\lambda_j(1+r_j)\left(\frac{r_j}{\bar{r}_j}-1\right)^+\ge P_{cap}\sum_{j\in\mathcal{J}}h^{(j)}(\bar{r}_j)\lambda_j(1+r_j)\right\}.$$

In words, set $\mathbb{B}(\bar{\mathbf{r}})$ includes all the readmission levels a hospital's penalty under HRRP is will be capped, assuming all the other hospital's pick actions $(\bar{\mathbf{r}},h(\bar{\mathbf{r}}))$. We also define

$$\hat{\mathcal{S}}_p\equiv\{(\bar{\mathbf{r}},\bar{\mathbf{c}}):\mathbb{B}(\bar{\mathbf{r}})=\emptyset,\bar{\mathbf{c}}=h\left(\bar{\mathbf{r}}\right),\bar{r}_j\in[\tilde{r}_j,r_{max}],\ \text{for } j\in\mathcal{J}\}.$$

We note that $\hat{\mathcal{S}}_p$ is non-empty–indeed it has uncountably many points–for $P_{cap} > 0$ because for $\bar{\mathbf{r}} = (\bar{r}_1, \bar{r}_2, \ldots, \bar{r}_J)$, if $\bar{r}_j \in [r_{max}/(P_{cap}+1), r_{max}]$, and $\bar{r}_j \geq \tilde{r}_j$, for $j \in \mathcal{J}$, then $(\bar{\mathbf{r}}, h(\bar{\mathbf{r}})) \in \hat{\mathcal{S}}_p$. Set $\hat{\mathcal{S}}_p$ corresponds to set $\mathcal{S}_p$ in the single monitored disease case, see (16).

We note that Lemma A2 still holds for each disease because $\hat{\pi}$ (see (A57)) does not depend on $\mathbf{c}$, hence $\bar{\mathbf{c}}_i = h(\bar{\mathbf{r}}_i)$ for each hospital i. Using this fact, a hospital's objective function can be written (with a slight abuse of notation) as

$$\begin{aligned}\hat{\Pi}(\mathbf{r} = (r_1, r_2, \ldots, r_J) | \bar{\mathbf{r}}) = \sum_{j \in \mathcal{J}} \left[\left(h^{(j)}(\bar{r}_j) - h^{(j)}(r_j) \right) \lambda_j (1 + r_j) \right] - \hat{\pi}(\mathbf{r} | \bar{\mathbf{r}}, h(\bar{\mathbf{r}})) \\ - \sum_{j \in \mathcal{J}} R^{(j)}(r_j, h^{(j)}(r_j)) + \sum_{j \in \mathcal{J}} \bar{R}_j(\bar{r}_j, h^{(j)}(\bar{r}_j)),\end{aligned} \tag{A59}$$

by (A56) if all the other hospitals pick $(\bar{\mathbf{r}}, h(\bar{\mathbf{r}}))$. And define

$$\hat{\mathcal{S}}_o \equiv \left\{ (\bar{\mathbf{r}}, \bar{\mathbf{c}}) : \mathbb{B}(\bar{\mathbf{r}}) \neq \emptyset, \bar{\mathbf{c}} = h(\bar{\mathbf{r}}), \bar{r}_j \in [\tilde{r}_j, r_{max}], \text{ for } j \in \mathcal{J}, \sup_{\mathbf{r} \in \mathbb{B}(\bar{\mathbf{r}})} \hat{\Pi}(\mathbf{r} | \bar{\mathbf{r}}) \leq 0 \right\}. \tag{A60}$$

This set corresponds to set $\mathcal{S}_o$ in the single monitored disease case, see (A45).

The following result extends our main result Proposition 3 to the case with multiple monitored diseases.

Proposition A1 (Equilibrium with multiple monitored diseases). *Under the HRRP scheme defined in* (A59) *with $J \geq 2$ monitored diseases*

(i) Any $(\mathbf{r}, \mathbf{c}) \in \hat{\mathcal{S}}_p \cup \hat{\mathcal{S}}_o$ is a symmetric equilibrium and there does not exist any other symmetric equilibrium.

(ii) If $P_{cap} = 0$, then $(\bar{\mathbf{r}} = (r_{max}, \ldots, r_{max}), h(\bar{\mathbf{r}}))$ is the unique symmetric equilibrium.

This results shows that there are uncountably many equilibria under HRRP with multiple monitored diseases, paralleling our result for the case with single monitored disease. In addition part (ii) shows that if $P_{cap} = 0$ hospitals have no incentive to invest in readmission reduction efforts. Also the set of equilibrium outcomes is especially large because for P_{cap} large enough

$$\hat{\mathcal{S}}_p \cup \hat{\mathcal{S}}_o \equiv \{ (\mathbf{r}, \mathbf{c}) \colon \mathbf{c} = h(\mathbf{r}), r_j \geq \tilde{r}_j \text{ for } j \in \mathcal{J} \},$$

as in Proposition 3(iii). However Proposition 3(iv) does not extend to the case with multiple diseases and we were unable to identify simple sufficient conditions.

Proof of Proposition A1: Throughout we focus on the symmetric equilibria. First, we focus on the symmetric equilibria and show that any $(\mathbf{r}, \mathbf{c}) \notin \hat{\mathcal{S}}_o$ cannot be a symmetric equilibrium and that any $(\mathbf{r}, \mathbf{c}) \in \hat{\mathcal{S}}_o$ is a symmetric equilibrium. Then we show that $(r_{max}, h(r_{max}))$ is the unique symmetric equilibrium if $P_{cap} = 0$. Without loss of generality we focus on the first hospital and drop the hospital index from our notation when it is clear from the context. Also recall that we assume that the symmetric equilibrium $(\tilde{\mathbf{r}}, \tilde{\mathbf{c}})$ under HRRP-I that satisfy (A58) is assumed to be unique, a fact we use throughout the proof.

To simplify the notation we define

$$\tilde{\mathcal{S}}_o \equiv \left\{ (\bar{\mathbf{r}}, \bar{\mathbf{c}}) : \bar{\mathbf{c}} = h\left(\bar{\mathbf{r}}\right), r_j \geq \tilde{r}_j \text{ for } j \in \mathcal{J}, \sup_{\mathbf{r} \in \mathbb{B}(\bar{\mathbf{r}})} \Pi(\mathbf{r}|\bar{\mathbf{r}}) \leq 0 \right\}, \tag{A61}$$

where we follow the convention $\sup_{\mathbf{r} \in \mathbb{B}(\bar{\mathbf{r}})} \Pi(\mathbf{r}|\bar{\mathbf{r}}) = -\infty$ if $\mathbb{B}(\bar{\mathbf{r}}) = \emptyset$. Set $\tilde{\mathcal{S}}_o$ is similar to set $\hat{\mathcal{S}}_o$ but does not impose the condition $\mathbb{B}(\bar{\mathbf{r}}) \neq \emptyset$. It is easy to check that $\tilde{\mathcal{S}}_o = \hat{\mathcal{S}}_p \cup \hat{\mathcal{S}}_o$.

(a) First, we will prove that any $(\bar{\mathbf{r}}, \bar{\mathbf{c}}) \notin \tilde{\mathcal{S}}_o$ cannot be a symmetric equilibrium. Suppose that all hospitals except hospital 1 choose $(\bar{\mathbf{r}}, \bar{\mathbf{c}}) \notin \tilde{\mathcal{S}}_o$. We show that hospital 1's best response cannot be $(\bar{\mathbf{r}}, \bar{\mathbf{c}})$, proving that $(\bar{\mathbf{r}}, \bar{\mathbf{c}})$ cannot be a symmetric equilibrium.

Lemma A2 still holds for each disease because $\hat{\pi}$ (see (A57)) does not depend on $\mathbf{c}$. Thus we can assume without loss of generality that $\mathbf{c} = h(\mathbf{r})$. Because $(\bar{\mathbf{r}} = (r_1, r_2, \ldots, r_J), \bar{\mathbf{c}}) \notin \tilde{\mathcal{S}}_o$ then either (1) $\bar{r}_j < \tilde{r}_j$ for at least one $j \in \mathcal{J}$ or (2) $\bar{r}_j \geq \tilde{r}_j$ for all $j \in \mathcal{J}$ and $\sup_{\mathbf{r} \in \mathbb{B}(\bar{\mathbf{r}})} \hat{\Pi}(\mathbf{r}|\bar{\mathbf{r}}) > 0$.

Assume first, without loss of generality, that $\bar{r}_1 < \tilde{r}_1$. Then there exists $\epsilon > 0$ such that for $\bar{r}^\epsilon = \{(r_1, \bar{r}_2, \ldots, \bar{r}_J) : r_1 \in [\bar{r}_1, \bar{r}_1 + \epsilon]\}$, we have $\bar{r}^\epsilon \cap \mathbb{B}(\bar{\mathbf{r}}) = \emptyset$. In words, there exists a neighborhood $\bar{r}^\epsilon$ of $\bar{\mathbf{r}}$ such that if hospital 1 chooses a readmission rate from this set $\bar{r}^\epsilon$ its financial penalty is below the cap. This implies by (A59) that for $\mathbf{r} \in \bar{r}^\epsilon$

$$\hat{\Pi}(\mathbf{r}|\bar{\mathbf{r}}) = \sum_{j \in \mathcal{J}} \Psi^{(j)}\left(r_j|\bar{r}_j\right),$$

where we obtain $\Psi^{(j)}$ after replacing R by $R^{(j)}$ in (A11). By Lemma A4, hospital 1 can improve its profit by increasing the readmission level for disease 1 since $\bar{r}_1 < \tilde{r}_1$.

If on the other hand $\bar{r}_j > \tilde{r}_j$ for all $j \in \mathcal{J}$ and $\sup_{\mathbf{r} \in \mathbb{B}(\bar{\mathbf{r}})} \hat{\Pi}(\mathbf{r}|\bar{\mathbf{r}}) > 0$, (implying $\mathbb{B}(\bar{\mathbf{r}}) \neq \emptyset$), because $\hat{\Pi}(\bar{\mathbf{r}}|\bar{\mathbf{r}}) = 0$, hospital 1's best response is in the set $\mathbb{B}(\bar{\mathbf{r}})$ (specifically it would choose the point that attains the supremum since $\hat{\Pi}$ is continuous), and clearly $\bar{\mathbf{r}} \notin \mathbb{B}(\bar{\mathbf{r}})$.

(b) Next we prove that any $(\mathbf{r}, \mathbf{c}) \in \tilde{\mathcal{S}}_o$ is a symmetric equilibrium. Assume that all hospitals except hospital 1 choose $(\bar{\mathbf{r}}, \bar{\mathbf{c}}) \in \tilde{\mathcal{S}}_o$. Then for $\mathbf{r} \notin \mathbb{B}(\bar{\mathbf{r}})$,

$$\hat{\Pi}(\mathbf{r} = (r_1, r_2, \ldots, r_J)|\bar{\mathbf{r}}) = \sum_{j \in \mathcal{J}} \Pi^{(j)}(r_j|\bar{r}_j),$$

where $\Pi^{(j)}$ is defined as in (A40) for $R = R^{(j)}$. As in the proof of Proposition 3, specifically by Corollary A1, (A43), and (A78), $\Pi^{(j)}(\bar{r}_j|\bar{r}_j) \geq \Pi^{(j)}(r_j|\bar{r}_j)$, for any $\mathbf{r} \notin \mathbb{B}(\bar{\mathbf{r}})$. If $\mathbf{r} \in \mathbb{B}(\bar{\mathbf{r}})$ on the other hand, by definition of the set $\hat{\mathcal{S}}_o$ in (A61)

$$\hat{\Pi}(\bar{\mathbf{r}}|\bar{\mathbf{r}}) = 0 \geq \hat{\Pi}(\mathbf{r}|\bar{\mathbf{r}}), \text{ for all } \mathbf{r} \in \mathbb{B}(\bar{\mathbf{r}}).$$

(c) Finally if $P_{cap}=0$ then

$$\hat{\Pi}(\mathbf{r}=(r_1,r_2,\ldots,r_J)|\bar{\mathbf{r}})=\sum_{j\in\mathcal{J}}\Phi^{o(j)}\left(r_j|\bar{r}_j\right),$$

for $\Phi^{o(j)}$ defined as in (A5) where R is replaced by $R^{(j)}$. Part (iii) then follows from Lemma A3(i). □

E. When targets are exogenous

To put our results in a better perspective, we establish hospitals' optimal actions when readmission target and reimbursement level are chosen exogenously at socially optimal levels and hospitals are reimbursed using a HRRP-like payment scheme with parameters set at these targets. (Similar schemes have been analyzed in Bastani et al. (2016), Adida et al. (2016), among others; see §2 for more details.)

Specifically we take the Modified HRRP payment scheme, set $\bar{c}_i=c^*$ and $\bar{r}_i=r^*$ for all $i=1,\ldots,N$, and find the optimal actions for a hospital when there is no bonus payment and penalty is capped. We focus on m-HRRP (see §5.1) to exclude the impact of the multiplier, which is already shown to distort the actions of hospitals. The objective of a hospital (we drop the hospital subscript for notational simplicity) under m-HRRP can be written as follows by (1) and (20).

$$\Pi(r,c)=(c^*-c)(1+r)\lambda-\left(\min\left\{\frac{r_i}{r^*}-1,P_{cap}\right\}\right)^+r^*c^*\lambda+R(r^*,c^*)-R(r,c), \tag{A62}$$

We first establish the optimal actions of a hospital when targets are exogenous and then compare these results with those when they are endogenous, see Remark 4. We have the following result. For the rest of this section we assume that Assumption A1 holds.

Proposition A2 (Optimal hospital actions under m-HRRP with exogenous targets). *Assume that the objective function of a hospital is given by* (A62). *Then there exists* $\bar{P}'_{cap}>0$ *such that*

i. If $P_{cap}\geq\bar{P}'_{cap}$*, then the optimal action for a hospital is to set* $r=r^*$ *and* $c=c^*$*, and*

ii. If $P_{cap}<\bar{P}'_{cap}$*, then the optimal action for a hospital is to set* $r=r_{max}$ *and* $c=h(r_{max})$*.*

Proof of Proposition A2. We break the proof into three steps. In steps (a)-(b) we characterize a hospital's locally optimal actions for $r\in[r_{min},r^*(1+P_{cap})]$ and $r\in[r^*(1+P_{cap}),r_{max}]$, respectively, and in step (c) we establish the hospital's globally optimal action. In addition, we have $c=h(r)$ in any equilibrium by (A62) and the proof of Lemma A2 and since this relationship uniquely determines c by Lemma A1 we only focus on hospital's readmission choice r and drop c from the notation.

(a) The hospital's objective $\Pi(r,h(r))$ defined as in (A62) reduces to $\Phi^o(r|r^*)$ defined as in (A5) for $r<r^*$ and reduces to the objective under m-HRRP defined as in (18) for $r\in[r^*,r^*(1+P_{cap})]$. By (A6) and Proposition 4, the hospital's optimal action for $r\in[r_{min},r^*(1+P_{cap})]$ is $r=r^*$.

(b) For $r\in[r^*(1+P_{cap}),r_{max}]$, the hospital's objective $\Pi(r,h(r))$ defined as in (A62) reduces to

$$\Pi(r,h(r))=(c^*-h(r))(1+r)\lambda-r^*c^*P_{cap}\lambda-R(r,h(r))+R(r^*,c^*), \tag{A63}$$

hence

$$\frac{d\Pi(r,h(r))}{dr}=(c^*-h(r))\lambda-R_r(r,h(r))>0 \text{ for } r\in[r^*(1+P_{cap}),r_{max}], \tag{A64}$$

where the inequality follows from: (i) $R_r<0$ by Assumption A1(i), and (ii) $c^*=h(r^*)\geq h(r)$ for any $r\geq r^*$ since h is decreasing by Lemma A1. By (A64), the hospital's optimal action for $r\in[r^*(1+P_{cap}),r_{max}]$ is $r=r_{max}$.

(c) If $P_{cap}\geq r_{max}/r^*-1$, $r\in[r_{min},r^*(1+P_{cap})]$ for all $r\in[r_{min},r_{max}]$ hence by part (a) the hospital's globally optimal action is $r=r^*$. If $P_{cap}=0$, it is optimal to choose $r=r^*$ for $r\in[r_{min},r^*]$ by part (a) and choose $r=r_{max}$ for $r\in[r^*,r_{max}]$ by part (b), hence the globally optimal action is $r=r_{max}$. If $P_{cap}\in(0,r_{max}/r^*-1)$, it is optimal to choose $r=r^*$ for $r\in[r_{min},r^*(1+P_{cap})]$ by part (a) and choose $r=r_{max}$ for $r\in[r^*(1+P_{cap}),r_{max}]$ by part (b), hence the firm chooses $r=r^*$ iff $\Pi(r^*,h(r^*))\geq\Pi(r_{max},h(r_{max}))$ or equivalently

$$P_{cap}\geqslant[(c^*-h(r_{max}))(1+r_{max})\lambda+R(r^*,c^*)-R(r_{max},h(r_{max}))]/(r^*c^*\lambda)\equiv\bar{P}'_{cap},$$

and chooses $r=r_{max}$ otherwise. Finally we prove $\bar{P}'_{cap}>0$. Since $\Pi(r_{max},h(r_{max}))$ is strictly decreasing in P_{cap} and is equal to 0 for $P_{cap}=\bar{P}'_{cap}$ (both by (A63)), it suffices to show that $\Pi(r_{max},h(r_{max}))>0$ for $P_{cap}=0$. This holds because $\Pi(r^*,h(r^*))=0$ by (A62) and $\frac{d\Pi(r,h(r))}{dr}>0$ for $r\in[r^*,r_{max}]$ by (A64). $\square$

We next show that no-bonus provision has no impact when the targets are set exogenously. The proof is similar to that of Proposition A2 and thus is omitted.

Corollary A2. *Assume that the objective function of a hospital is given by*

$$\Pi(r,c)=\begin{cases}(c^*-c)(1+r)\lambda-\min\left\{\frac{r}{r^*}-1,P_{cap}\right\}r^*c^*\lambda+R(r^*,c^*)-R(r,c), & \text{if } r\geq r^*,\\ (c^*-c)(1+r)\lambda+\min\left\{1-\frac{r}{r^*},P_{cap}\right\}r^*c^*\lambda+R(r^*,c^*)-R(r,c), & \text{if } r<r^*.\end{cases} \tag{A65}$$

Then the optimal actions of a hospital are identical to that given in Proposition A2.

In words, Proposition A2 and Corollary A2 imply that as long as P_{cap} is large enough, hospitals would take socially optimal actions if the regulator were able to set the cost and readmission targets to socially optimal levels (see footnote 1). Also, these results imply that no-bonus provision has no impact on hospital actions when targets are set exogenously. In addition, the regulator can adjust the value of P_{cap} if (some) hospitals end up having much larger readmission levels than the target, hence the negative impact of having low penalty caps can potentially be eliminated in practice.

These results are in stark contrast to those when these targets are set exogenously as done under HRRP (see Remark 4). First, we have shown that, in the endogenous case, there are multiple pure-strategy equilibria even when P_{cap} is (arbitrarily) large (see Proposition 3). Hence HRRP cannot restore socially optimal outcomes. Second, no-bonus provision has a significant impact on the ensuing equilibrium (see Propositions 2 and 3). Finally, because the equilibrium outcomes are not clear (e.g., under a mixed equilibrium), the regulator would not be able to determine the effective levels for P_{cap}.

F. Sufficient Conditions for Unique Symmetric Equilibrium under HRRP-I

In this section, we derive sufficient conditions that ensure the uniqueness of the symmetric equilibrium under HRRP-I so that Assumption 1 always holds true. Lemma A8 formalizes this.

Lemma A8. *Suppose that*

$$\frac{\partial}{\partial r}\left[\frac{rR_r(r,c)}{R(r,c)}\right] > 0, \frac{\partial}{\partial c}\left[\frac{cR_c(r,c)}{R(r,c)}\right] > 0, \ \frac{\partial}{\partial c}\left[\frac{R_r(r,c)}{R(r,c)}\right] \leq 0, \frac{\partial}{\partial r}\left[\frac{R_c(r,c)}{R(r,c)}\right] \leq 0. \tag{A66}$$

Then, the symmetric equilibrium under HRRP-I $(\tilde{r},\tilde{c})$ *is unique.*

The first two conditions in (A66) imply the investment cost function R being log-convex in r and c, respectively, see Lemma A9(i). Hence, they are similar to but more restrictive than log-convexity of R in each of its arguments. The last two conditions in (A66) are satisfied if, and only if, the natural logarithm of R decreases in each of its arguments less as the other argument increases (i.e., $\frac{\partial^2}{\partial r \partial c} \ln R(r,c) \leq 0$). In other words, the last two conditions require that, for lower values of treatment cost or readmission rate, the natural logarithm of the investment cost function (i.e., $\ln R(r,c)$) increases more when a hospital reduces its readmission rate or treatment cost, respectively. Also, it can be shown that all conditions in (A66) imply that $R(r,c)$ is jointly convex in r and c.

Lemma A9. *(i) If*

$$\frac{\partial}{\partial r}\left[\frac{rR_r(r,c)}{R(r,c)}\right] > 0 \text{ and } \frac{\partial}{\partial c}\left[\frac{cR_c(r,c)}{R(r,c)}\right] > 0,$$

then $R(r,c)$ *is log-convex in* r *and* c*, respectively.*

(ii) The investment cost function $R(r,c)$ *in* (10) *satisfies* (A66).

Lastly, Lemma A9(ii) shows that the investment cost function $R(r,c)$ in (10) satisfies conditions in (A66). Thus, the symmetric equilibrium in all numerical examples in §4.1.1 is unique.

F.1. Proofs of the results

Proof of Lemma A8: From the proof of Proposition 1, any symmetric equilibrium under HRRP-I $(\tilde{r},\tilde{c})$ must satisfy $\tilde{c}=h(\tilde{r})$ and $\Gamma(\tilde{r})=0$, where $\Gamma(r)$ is given by (A16). Note by (A16) that

$$\begin{aligned}
\Gamma(r) &= \frac{1+r}{r}h(r)\lambda + R_r(r,h(r)) \\
&= -\frac{h(r)}{r}R_c(r,h(r)) + R_r(r,h(r)) \\
&= \frac{R(r,h(r))}{r}\left[-\frac{h(r)R_c(r,h(r))}{R(r,h(r))} + \frac{rR_r(r,h(r))}{R(r,h(r))}\right],
\end{aligned}$$

where the second equality follows from (5). Then, by $R(r,h(r))>0$ for $r\in[r_{\min},r_{\max}]$, the symmetric equilibrium under HRRP-I $\tilde{r}$ must satisfy $\digamma_1(\tilde{r})=0$, where

$$\digamma_1(r) = -\frac{h(r)R_c(r,h(r))}{R(r,h(r))} + \frac{rR_r(r,h(r))}{R(r,h(r))}. \tag{A67}$$

Note that $\digamma_1(r_{\min}) = \Gamma(r_{min})\frac{r_{min}}{R(r_{min},h(r_{min}))} < 0$ and $\digamma_1(r_{\max}) = \Gamma(r_{max})\frac{r_{max}}{R(r_{max},h(r_{max}))} > 0$ by $\Gamma(r_{min})<0$ and $\Gamma(r_{max})>0$; see the proof of Proposition 1. Now, we show that $\digamma_1(r)$ is increasing in r by (A66) so that there exists a unique $\tilde{r}$ satisfying $\digamma_1(\tilde{r})=0$. To that end, we let

$$\begin{aligned}
u_1(r,c) &= c\frac{\partial}{\partial r}\left[\frac{R_c(r,c)}{R(r,c)}\right] = c\frac{R_{cr}(r,c)R(r,c)-R_c(r,c)R_r(r,c)}{R^2(r,c)} \le 0, \\
u_2(r,c) &= \frac{\partial}{\partial c}\left[\frac{cR_c(r,c)}{R(r,c)}\right] = \frac{(R_c(r,c)+cR_{cc}(r,c))R(r,c)-cR_c^2(r,c)}{R^2(r,c)} > 0, \\
u_3(r,c) &= r\frac{\partial}{\partial c}\left[\frac{R_r(r,c)}{R(r,c)}\right] = r\frac{R_{cr}(r,c)R(r,c)-R_c(r,c)R_r(r,c)}{R^2(r,c)} \le 0, \\
u_4(r,c) &= \frac{\partial}{\partial r}\left[\frac{rR_r(r,c)}{R(r,c)}\right] = \frac{(R_r(r,c)+rR_{rr}(r,c))R(r,c)-rR_r^2(r,c)}{R^2(r,c)} > 0,
\end{aligned}$$

where the inequalities follow from (A66). Taking the derivative of the first term on the right hand side (RHS) of (A67) with respect to r, we obtain:

$$\begin{aligned}
&\frac{d}{dr}\left[\frac{h(r)R_c(r,h(r))}{R(r,h(r))}\right] = h'(r)\frac{R_c(r,h(r))}{R(r,h(r))} \\
&+h(r)\frac{(R_{cr}(r,h(r))+h'(r)R_{cc}(r,h(r)))R(r,h(r))-R_c(r,h(r))(R_r(r,h(r))+h'(r)R_c(r,h(r)))}{R^2(r,h(r))} \\
&\qquad = u_1(r,h(r))+u_2(r,h(r))h'(r) < 0
\end{aligned}$$

where the inequality follows from $h'(r)<0$ (by Lemma A1) and (A66) (i.e., $u_1(r,h(r))\le 0, u_2(r,h(r))>0$). Similarly, taking the derivative of the second term on the RHS of (A67) with respect to r, we obtain:

$$\frac{d}{dr}\left[\frac{rR_r(r,h(r))}{R(r,h(r))}\right] = u_4(r,h(r))+u_3(r,h(r))h'(r) > 0$$

where the inequality follows from Lemma A1, and (A66) (i.e., $u_3(r,h(r)) \leq 0, u_4(r,h(r)) > 0$). Then, taking the derivative of $F_1(r)$, it follows that

$$\frac{dF_1(r)}{dr} = -\frac{d}{dr}\left[\frac{h(r)\,R_c(r,h(r))}{R(r,h(r))}\right] + \frac{d}{dr}\left[\frac{rR_r(r,h(r))}{R(r,h(r))}\right] > 0.$$

Thus, $F_1(r)$ is increasing in r. $\square$

Proof of Lemma A9: We prove each part of the lemma separately.

(i) Note that

$$\begin{aligned}\frac{\partial}{\partial r}\left[\frac{rR_r(r,c)}{R(r,c)}\right] &= \frac{\partial}{\partial r}\left[r\frac{\partial}{\partial r}(\ln R(r,c))\right]\\ &= \frac{\partial}{\partial r}(\ln R(r,c)) + r\frac{\partial^2}{\partial r^2}(\ln R(r,c))\end{aligned}$$

where the first equality follows from

$$\frac{\partial}{\partial r}(\ln R(r,c)) = \frac{R_r(r,c)}{R(r,c)}.$$

Thus, when $\partial\left[\frac{rR_r(r,c)}{R(r,c)}\right]/\partial r > 0$, this together with $R_r(r,c) < 0$ (by Assumption A1(i)) implies that $\partial^2(\ln R(r,c))/\partial r^2 > 0$ so that $R(r,c)$ is log-convex in r. Similarly, using $R_c(r,c) < 0$ in Assumption A1(i), one can show that if $\partial\left[\frac{cR_c(r,c)}{R(r,c)}\right]/\partial r > 0$, then $R(r,c)$ is log-convex in c.

(ii) By (10), we have

$$\begin{aligned}\frac{\partial}{\partial r}\left[\frac{rR_r(r,c)}{R(r,c)}\right] &= \frac{c\tau_1\tau_2}{(c\tau_1 + r\tau_2)^2} > 0,\\ \frac{\partial}{\partial c}\left[\frac{cR_c(r,c)}{R(r,c)}\right] &= \frac{r\tau_1\tau_2}{(c\tau_1 + r\tau_2)^2} > 0,\\ \frac{\partial}{\partial c}\left[\frac{R_r(r,c)}{R(r,c)}\right] &= \frac{\partial}{\partial r}\left[\frac{R_c(r,c)}{R(r,c)}\right] = -\frac{\tau_1\tau_2}{(c\tau_1 + r\tau_2)^2} < 0\end{aligned}$$

for $\tau_1, \tau_2 > 0$. Thus, $R(r,c)$ in (10) satisfies (A66). $\square$

G. Multiple Symmetric Equilibria under HRRP-I

In this section, we extend the analysis presented in §4.2-4.3 by establishing the equilibrium outcomes under HRRP-II and HRRP in cases where Assumption 1 does not hold, i.e., HRRP-I induces multiple symmetric equilibria. To this end, we adopt the following assumption in place of Assumption 1 throughout this section.

Assumption A2. *Under HRRP-I, there exist three symmetric equilibria.*

We focus on the case with the three symmetric equilibria because it is the next simplest case after the case with a single equilibrium; we show below that there cannot be two symmetric equilibria, see Lemma A10. The analysis in this section can be easily extended to cases with more symmetric equilibria and the results will not change qualitatively yet the required notation becomes much more complex.

Next we introduce the notation we use in this section. Let $(\tilde{r}_i, \tilde{c}_i)$, $i = 1, 2, 3$, denote the three equilibrium points and assume without loss of generality that

$$\tilde{r}_1 < \tilde{r}_2 < \tilde{r}_3. \tag{A68}$$

By Lemma A2, $\tilde{c}_i = h(\tilde{r}_i)$, for $i = 1, 2, 3$. To specify the equilibrium outcomes under HRRP-II we set

$$\mathcal{S}_m(P_{cap}) = \{(r, h(r)) : r \in \{\tilde{r}_1, \tilde{r}_2, \tilde{r}_3\}, \bar{P}_{cap}(r) \le P_{cap}\} \tag{A69}$$

and

$$\bar{\mathcal{P}}_{cap} = \{\bar{P}_{cap}(\tilde{r}_1), \bar{P}_{cap}(\tilde{r}_2), \bar{P}_{cap}(\tilde{r}_3)\},$$

where $\bar{P}_{cap}(r)$ is defined in Lemma A7 (we prove existence of $\bar{P}_{cap}(\tilde{r}_i)$, $i = 1, 2, 3$ in Lemma A13 below, hence $\bar{\mathcal{P}}_{cap}$ is non-empty). Let

$$\begin{aligned}
\bar{P}^1_{cap} &= \min \bar{\mathcal{P}}_{cap}, \\
\bar{P}^2_{cap} &= \min(\bar{\mathcal{P}}_{cap} / \bar{P}^1_{cap}), \\
\bar{P}^3_{cap} &= \max \bar{\mathcal{P}}_{cap}.
\end{aligned}$$

In words, $\bar{P}^i_{cap}$ is the i-th smallest element of $\bar{\mathcal{P}}_{cap}$ hence

$$\bar{P}^3_{cap} \ge \bar{P}^2_{cap} \ge \bar{P}^1_{cap}. \tag{A70}$$

The next proposition characterizes the equilibrium outcomes under HRRP-II defined as in (14).

Proposition A3 (Equilibrium under HRRP-II). *Suppose Assumption A2 holds. Under HRRP-II scheme:*

(i) Three symmetric equilibria (i.e., $(\tilde{r}_i, h(\tilde{r}_i))$ for $i = 1, 2, 3$) exist if $P_{cap} \ge \bar{P}^3_{cap}$.

(ii) Two symmetric equilibria (i.e., $(r, c) \in \mathcal{S}_m(P_{cap})$) exist if $\bar{P}^2_{cap} \le P_{cap} < \bar{P}^3_{cap}$.

(iii) A unique symmetric equilibrium (i.e., $(r, c) \in \mathcal{S}_m(P_{cap})$) exists if $\bar{P}^1_{cap} \le P_{cap} < \bar{P}^2_{cap}$.

(iv) No symmetric equilibrium exists if $0 < P_{cap} < \bar{P}^1_{cap}$.

(v) $(r_{max}, h(r_{max}))$ is the unique symmetric equilibrium if $P_{cap} = 0$.

When there are more than one symmetric equilibrium under HRRP-I, adding a cap on the readmission penalty/reward *reduces* the effect of readmission reduction financial incentive scheme instead of nullifying it (as in the case of a unique symmetric equilibrium; see Proposition 2). In fact, by parts (i)-(iv) the cap has no impact on the symmetric equilibria when it is large enough and eliminates them one by one as it decreases. To see the intuition, first recall from the discussion of Proposition 2 that a symmetric equilibrium is eliminated if the cap falls below a threshold such that a hospital earns higher profits by exerting no readmission reduction effort and paying the capped readmission penalty instead of choosing the equilibrium readmission level. In addition, the thresholds for different symmetric equilibria are different in general because they depend on equilibrium readmission levels which are different by (A68). As the cap decreases, these thresholds are reached separately in a series hence the corresponding symmetric equilibria are eliminated one by one. Finally, part (v) shows that if the cap decreases to zero, hospitals will exert no readmission reduction effort in equilibrium, as in the case of a unique symmetric equilibrium under HRRP-I.

Next we establish the equilibrium outcomes under the HRRP scheme defined in (7), where hospitals with lower-than-expected readmission rates do not receive bonus payments and the penalty is capped.

We need to introduce additional terminology to specify the equilibrium outcomes in this case. First let $r_{pi} = \max\{r_e, \tilde{r}_i\}$ for $i = 1, 3$, where r_e is defined in (15). To specify the equilibrium outcomes we set

$$\mathcal{S}'_p = \{(r, h(r)) : r \in [r_{p1}, \tilde{r}_2] \cup [r_{p3}, r_{max}]\} \tag{A71}$$

and note that, if $P_{cap} > 0$, $r_{pi} < r_{max}$ by Proposition 1 and so $\mathcal{S}'_p$ is non-empty. Finally let

$$\mathcal{S}' = \{(r, h(r)) : r \in ([\tilde{r}_1, \tilde{r}_2] \cap [\tilde{r}_1, r_{p1})) \cup ([\tilde{r}_3, r_{max}] \cap [\tilde{r}_3, r_{p3}))$$

and $P'_{max} = \frac{r_{max}}{\tilde{r}_1} - 1$. We have the following result.

Proposition A4 (Equilibrium under HRRP). *Suppose Assumption A2 holds. The following hold under the HRRP scheme:*

(i) *For any $P_{cap} \geq 0$, there exists $\mathcal{S}'_o \subset \mathcal{S}'$ (depending on P_{cap}), such that any $(r, c) \in \mathcal{S}'_o \cup \mathcal{S}'_p$ is a symmetric equilibrium and there is no other symmetric equilibrium.*

(ii) *There exists $\bar{\bar{P}}'_{cap} \in (0, P'_{max})$ such that for any $(r, c) \in S'_o \cup \mathcal{S}'_p$, $r > r^*$ and $c < c^*$ for $P_{cap} < \bar{\bar{P}}'_{cap}$.*

(iii) *If $P_{cap} \geq P'_{max}$, then any*

$$(r, c) \in \{(r, h(r)) : r \in [\tilde{r}_1, \tilde{r}_2] \cup [\tilde{r}_3, r_{max}]\} \; (= S' \cup \mathcal{S}'_p)$$

is a symmetric equilibrium.

(iv) *There is no asymmetric equilibrium if $dR(r, h(r))/dr < 0$ for all $r \in [r_{min}, r_{max}]$.*

Similar to the case of a unique symmetric equilibrium under HRRP-I, removing bonus payments generates uncountably many equilibrium outcomes (i.e., any $(r, c) \in \mathcal{S}'_o \cup \mathcal{S}'_p$) and the equilibrium set shrinks as the penalty cap becomes smaller (see Proposition 3 for a detailed discussion). In addition, the set of equilibrium outcomes is structurally more complicated in the case of multiple symmetric equilibria (under HRRP-I)–it splits into two disconnected sets that are subsets of $\{(r,h(r)): r\in[\tilde{r}_1,\tilde{r}_2]\}$ and $\{(r,h(r)): r\in[\tilde{r}_3,r_{max}]\}$ respectively. No point in $\{(r,h(r)): r\in(\tilde{r}_2,\tilde{r}_3)\}$ is an equilibrium because for $\bar{r}\in(\tilde{r}_2,\tilde{r}_3)$, a hospital earns higher profits by increasing r beyond $\bar{r}$ since the savings on readmission reduction cost outweigh the readmission penalty.

Propositions A3-A4 together show that if there are multiple equilibria under HRRP-I, HRRP-II and HRRP generate qualitatively similar equilibrium outcomes as when HRRP-I has a unique symmetric equilibrium. In addition, with multiple symmetric equilibria under HRRP-I, the impact of the penalty cap on the equilibrium outcomes under HRRP-II is more nuanced, and the set of equilibrium outcomes under HRRP is more complicated as it splits into disconnected sets.

G.1. Proofs of the results

G.1.1. Preliminary results In this section we prove preliminary results that we utilize in proving the results in Appendix G. We first show that the number of symmetric equilibria under HRRP-I must be odd.

Lemma A10. *Under HRRP-I the number of symmetric equilibria must be odd.*

Proof of Lemma A10. By Lemma A4 and (A12), $(\tilde{r}_i,\tilde{c}_i)$ is a symmetric equilibrium under HRRP-I iff $\tilde{r}_i$ satisfies

$$r_\Psi(\tilde{r}_i)=\tilde{r}_i \tag{A72}$$

and $\tilde{c}_i=h(\tilde{r}_i)$, i.e., the readmission rate in any symmetric equilibrium is a fixed point of r_Ψ. Below we show that the number of fixed points is odd.

Since r_Ψ is continuous by (A12), the number of fixed points of r_Ψ is equal to the times $y=r_\Psi(r)$ crosses the 45° line $y=r$ for $r\in[r_{min},r_{max}]$. Since there exists at least one symmetric equilibrium under HRRP-I by Proposition 1, $y=r_\Psi(r)$ crosses $y=r$ at least once. Moreover, by $\lim_{r\downarrow r_{min}} r_\Psi(r)>r_{min}$ from Lemma A4, the first time $y=r_\Psi(r)$ crosses $y=r$ is from above; by $\lim_{r\uparrow r_{max}} r_\Psi(r)<r_{max}$ from Lemma A4, the last time $y=r_\Psi(r)$ crosses $y=r$ is from above. Hence $y=r_\Psi(r)$ crosses $y=r$ with odd times. $\square$

Since HRRP-I generates an odd number of symmetric equilibria, the simplest case of multiple symmetric equilibria is to have three symmetric equilibria. We focus on this case by adopting Assumption A2 in Appendix G. The next lemma develops two properties of the symmetric equilibria $(\tilde{r}_i,\tilde{c}_i)$ for $i=1,2,3$.

Lemma A11. $\tilde{c}_i=h(\tilde{r}_i)$ *for* $i=1,2,3$ *and* $r_{min}<\tilde{r}_1<\tilde{r}_2<\tilde{r}_3<r^*<r_{max}$.

Proof of Lemma A11. $\tilde{c}_i = h(\tilde{r}_i)$ follows from Lemma A2. By Assumption A2 and monotonicity of $h(r)$ from Lemma A1, we have $\tilde{r}_i \neq \tilde{r}_j$ for $i \neq j$—otherwise $(\tilde{r}_i, h(\tilde{r}_i))$ and $(\tilde{r}_i, h(\tilde{r}_j))$ represent the same equilibrium. The proof is complete by

$$\tilde{r}_1 = r_\Psi(\tilde{r}_1) > r_{min},$$
$$\tilde{r}_3 = r_\Psi(\tilde{r}_3) < r^* < r_{max},$$

where the equalities follow from (A72) and the inequalities follow from Lemma A4 and Proposition 1. □

By Lemma A11, below we refer to the three symmetric equilibria as $(\tilde{r}_1, h(\tilde{r}_1))$, $(\tilde{r}_2, h(\tilde{r}_2))$, and $(\tilde{r}_3, h(\tilde{r}_3))$ satisfying $r_{min} < \tilde{r}_1 < \tilde{r}_2 < \tilde{r}_3 < r^* < r_{max}$ without loss of generality.

Next we present a preliminary result we will use in proving our main results in Propositions A3-A4.

Lemma A12. *$r_\Psi(r) > r$ for $r \in [r_{min}, \tilde{r}_1) \cup (\tilde{r}_2, \tilde{r}_3)$, $r_\Psi(r) < r$ for $r \in (\tilde{r}_1, \tilde{r}_2) \cup (\tilde{r}_3, r_{max}]$, and $r_\Psi(r) = r$ for $r \in \{\tilde{r}_1, \tilde{r}_2, \tilde{r}_3\}$.*

Proof of Lemma A12. The lemma follows from $r_{min} < \tilde{r}_1 < \tilde{r}_2 < \tilde{r}_3 < r_{max}$ by Lemma A11, the fact that $y = r_\Psi(r)$ crosses $y = r$ three times (by Assumption A2), and the first and last crossings are from above (by the proof of Lemma A10). □

G.1.2. Proof of Proposition A3 The proof of part (v) is the same as the proof of Proposition 2(iii), hence below we prove the cases in which $P_{cap} > 0$, i.e., parts (i)-(iv), in the following steps. In step (a) we show that there exists no other symmetric equilibrium besides $(\tilde{r}_i, h(\tilde{r}_i))$ for $i = 1, 2, 3$, and in step (b) we show that $(\tilde{r}_i, h(\tilde{r}_i))$ is an equilibrium point if and only if $P_{cap} \geq \bar{P}_{cap}(\tilde{r}_i)$ (defined as in Lemma A7(ii) and we prove its existence in Lemma A13). Finally in step (c) we use these results to prove parts (i)-(iv).

Throughout the proof we only focus on potential symmetric equilibria and the average readmission rate is identical to each hospital's readmission rate in any symmetric equilibrium, therefore, with a slight abuse of notation, we use $\bar{r}$ to denote a hospital's choice of readmission rate for notational simplicity. To prove that $(\tilde{r}_i, \tilde{c}_i)$ is a symmetric equilibrium it is enough to show that for a given hospital the best action is $(\tilde{r}_i, \tilde{c}_i)$ if all other hospitals choose $(\tilde{r}_i, \tilde{c}_i)$. Without loss of generality, we focus on hospital 1 and drop the hospital index subscript from all mathematical expressions throughout the proof. By Lemmas A1, A2, and A4 for any $(\bar{r}, \bar{c})$ hospital i has a unique best response (r, c), and $c = h(r)$ under HRRP-II and since this relationship uniquely determines c we only focus on hospital's readmission choice r and drop c from the notation.

(a) First we prove that any $r' \notin \{\tilde{r}_1, \tilde{r}_2, \tilde{r}_3\}$ cannot constitute a symmetric equilibrium because hospital 1 can profitably deviate from $r = r'$. For $\bar{r} = r'$, hospital 1's profit is $\Pi(r|r') = \Psi(r|r')$

for all $r \in (r'(1-P_{cap}), r'(1+P_{cap}))$ by (A20); by Lemma A4, $\Psi(r|r')$ is concave and the unique maximizer $r_\Psi \neq r'$ for $r' \neq \tilde{r}$ by Lemma A12. Hence

$$\frac{d\Psi(r'|r')}{dr} \neq 0$$

and hospital 1 can increase its profit by deviating from $r = r'$, therefore $r' \notin \{\tilde{r}_1, \tilde{r}_2, \tilde{r}_3\}$ cannot be a symmetric equilibrium.

(b) Next we characterize the necessary and sufficient conditions for $(\tilde{r}_i, h(\tilde{r}_i))$ to be a symmetric equilibrium by testing if hospital 1 can profitably deviate from $r = \tilde{r}_i$ for $\bar{r} = \tilde{r}_i$. By Lemma A5, hospital 1 will never choose $r < \bar{r}(1-P_{cap})$ (assuming it is feasible) hence we only consider $r \geq \bar{r}(1-P_{cap})$. For $r \in [\bar{r}(1-P_{cap}), \bar{r}(1+P_{cap}))$, hospital 1's profit is $\Pi(r|\bar{r}) = \Psi(r|\bar{r})$ by (A20) and $\Psi(r|\tilde{r}_i)$ is maximized at $r = r_\Psi(\tilde{r}_i) = \tilde{r}_i$ by Lemma A4 and (A72), hence it cannot profitably deviate from $r = \tilde{r}_i$ for $r \in [\bar{r}(1-P_{cap}), \bar{r}(1+P_{cap}))$. For $r \in [\bar{r}(1+P_{cap}), r_{max}]$, we only consider $P_{cap} \in [0, P_m(\bar{r})]$ because otherwise $[\bar{r}(1+P_{cap}), r_{max}] = \emptyset$ by (A27). For $P_{cap} \in [0, P_m(\bar{r})]$, hospital 1's maximum profit is $\hat{\Phi}(P_{cap}, \bar{r})$ when its actions are restricted to $r \geq \bar{r}(1+P_{cap})$, where $\hat{\Phi}$ is defined in (A28) and we make explicit its dependence on $\hat{r}$ for the rest of this proof. In addition, since $\Pi(\bar{r}|\bar{r}) = 0$ by (A4), hospital 1 can profitably deviate from $\tilde{r}_i$ iff

$$\hat{\Phi}(P_{cap}, \tilde{r}_i) > 0. \tag{A73}$$

Next result characterizes the condition for this to hold.

Lemma A13. *For $i = 1, 2, 3$, there exists $\bar{P}_{cap}(\tilde{r}_i) \in [0, P_m(\tilde{r}_i)]$ such that if $P_{cap} \in [0, \bar{P}_{cap}(\tilde{r}_i))$ then* (A73) *holds and if $P_{cap} \in [\bar{P}_{cap}(\tilde{r}_i), P_m(\tilde{r}_i)]$ then*

$$\hat{\Phi}(P_{cap}, \tilde{r}_i) \leq 0.$$

Proof of Lemma A13. We use Lemma A7 to prove this lemma. First, for $\bar{r} = \tilde{r}_i$

$$\hat{\Phi}\left(P_m(\tilde{r}_i)|\tilde{r}_i\right) \overset{\text{(r0)}}{=} \Phi\left(r_{max}, P_m(\tilde{r}_i)|\tilde{r}_i\right) \overset{\text{(r1)}}{=} \Psi\left(r_{max}, P_m(\tilde{r}_i)|\tilde{r}_i\right) \overset{\text{(r2)}}{\leq} \Psi\left(\tilde{r}_i, P_m(\tilde{r}_i)|\tilde{r}_i\right) = 0, \tag{A74}$$

where (r0) follows from (A28), (r1) follows from (A11), and (r2) follows from Assumption A2, as otherwise $\tilde{r}_i$ would not be an equilibrium. Therefore, (A74) implies that Lemma A7(i) cannot hold for $\bar{r} = \tilde{r}_i$. By Lemma A7(ii), $\hat{\Phi}\left(P_{cap}|\tilde{r}_i\right) \leq 0$ for $P_{cap} \geq \bar{P}_{cap}(\tilde{r}_i)$ and $\hat{\Phi}\left(P_{cap}|\tilde{r}_i\right) > 0$ for $P_{cap} < \bar{P}_{cap}(\tilde{r}_i)$. $\square$

By Lemma A13, $(\tilde{r}_i, h(\tilde{r}_i))$ is an equilibrium point if and only if $P_{cap} \geq \bar{P}_{cap}(\tilde{r}_i)$.

(c) Finally we prove parts (i)-(iv). In part (i), $P_{cap} \geq \bar{P}_{cap}(\tilde{r}_i)$ for $i = 1, 2, 3$ hence by part (b) $(\tilde{r}_i, h(\tilde{r}_i))$ for $i = 1, 2, 3$ are three symmetric equilibria. In part (ii), there is nothing to prove if $\bar{P}^2_{cap} = \bar{P}^3_{cap}$ hence assume $\bar{P}^2_{cap} < \bar{P}^3_{cap}$. By (A69) and $\bar{P}^2_{cap} \geq \bar{P}^1_{cap}$ from (A70), $\mathcal{S}_m(P_{cap})$ has two elements for $P_{cap} \in [\bar{P}^2_{cap}, \bar{P}^3_{cap})$ and they are symmetric equilibria by step (b). In part (iii), there is nothing to prove if $\bar{P}^1_{cap} = \bar{P}^2_{cap}$, hence assume $\bar{P}^1_{cap} < \bar{P}^2_{cap}$. By (A69) and $\bar{P}^3_{cap} \geq \bar{P}^2_{cap}$ from (A70), $\mathcal{S}_m(P_{cap})$ has one element for $P_{cap} \in [\bar{P}^1_{cap}, \bar{P}^2_{cap})$ and it is a symmetric equilibrium by step (b). In part (iv), $P_{cap} < \bar{P}_{cap}(\tilde{r}_i)$ for $i = 1, 2, 3$ hence no symmetric equilibrium exists by step (b). $\square$

G.1.3. Proof of Proposition A4 The proof of part (iv) is the same as the proof of Proposition 3(iv), hence below we prove parts (i)-(iii) in the following steps. In steps (a)-(c) we prove part (i), and show that part (iii) follows from part (i). In step (d) we prove part (ii). We use the same notation as in the proof of Proposition 3. To prove that (r, c) is a symmetric equilibrium it is enough to show that hospital 1's best action is (r, c) given that all other hospitals choose (r, c). By Lemmas A1, A2, and A4 for any $(\bar{r}, \bar{c})$ hospital i has a unique best response (r, c) under HRRP, and $c = h(r)$ and since this relationship uniquely determines c we only focus on hospital's readmission choice r and drop c from the notation. Hence hospital 1's profit, if all the other hospitals pick $(\bar{r}, h(\bar{r}))$, is $\Pi(r|\bar{r})$ given by (A40). Again, we write $\Pi(r, P_{cap}|\bar{r})$ instead of $\Pi(r|\bar{r})$ when we need to make the dependence on P_{cap} explicit.

(a) We first show that any $(\bar{r}, h(\bar{r})) \in \mathcal{S}'_p$ is a symmetric equilibrium. Assume that all hospitals except hospital 1 pick $(\bar{r}, h(\bar{r})) \in \mathcal{S}'_p$. By Corollary A1 it is enough to show that hospital 1's profit at $(\bar{r}, h(\bar{r}))$ is higher than that at any other point $(r, h(r))$, for $r \in [\bar{r}, r_{max}]$, that is,

$$\Pi(\bar{r}|\bar{r}) \geq \Pi(r|\bar{r}) \quad \text{for any } r \in [\bar{r}, r_{max}]. \tag{A75}$$

By (A40) and because $(\bar{r}, h(\bar{r})) \in \mathcal{S}'_p$, $\Pi(r|\bar{r}) = \Psi(r|\bar{r})$, (see (A11) for the definition of Ψ), hence the objective of hospital 1 is identical to that under HRRP-I if $(\bar{r}, h(\bar{r})) \in \mathcal{S}'_p$. By Lemma A4, Ψ is concave hence it is enough to show that

$$\frac{d\Psi(\bar{r}|\bar{r})}{dr} \leq 0, \qquad \text{for any } \bar{r} \in [\tilde{r}_1, \tilde{r}_2] \cup [\tilde{r}_3, r_{max}]. \tag{A76}$$

Since $\Psi(r|\bar{r})$ is concave and its unique maximizer r_Ψ satisfies $d\Psi(r_\Psi(\bar{r})|\bar{r})/dr = 0$ (both by Lemma A4), (A76) is equivalent to

$$r_\Psi(\bar{r}) \leq \bar{r}, \qquad \text{for any } \bar{r} \in [\tilde{r}_1, \tilde{r}_2] \cup [\tilde{r}_3, r_{max}],$$

which holds by Lemma A12.

(b) Next we focus on the rest of the symmetric equilibria. First we explicitly define $\mathcal{S}'_o \subset \mathcal{S}'$ and then show that any $(\bar{r}, h(\bar{r})) \in \mathcal{S}'_o$ is a symmetric equilibrium. Let

$$\mathcal{S}'_o = \{(\bar{r}, \bar{c}) : \bar{c} = h(\bar{r}), \bar{r} \in ([\tilde{r}_1, \tilde{r}_2] \cap [\tilde{r}_1, r_{p1})) \cup ([\tilde{r}_3, r_{max}] \cap [\tilde{r}_3, r_{p3})), \sup_{r \in [(1+P_{cap})\bar{r}, r_{max}]} \Phi(r|\bar{r}) \leq 0\}, \tag{A77}$$

where function Φ is defined as in (A22). If $\mathcal{S}'_o = \emptyset$ then there is nothing to prove, hence assume for the rest of the proof that it is non-empty.

Assume that all hospitals except hospital 1 pick $(\bar{r}, h(\bar{r})) \in \mathcal{S}'_o$. As in part (a) it is enough to show (A75) holds. First note that for $r \in [\bar{r}, (1+P_{cap})\bar{r}]$, $\Pi(r|\bar{r}) = \Psi(r|\bar{r})$ by (A40). Then by (A76)

$$\Pi(\bar{r}|\bar{r}) \geq \Pi(r|\bar{r}) \quad \text{for any } r \in [\bar{r}, (1+P_{cap})\bar{r}]. \tag{A78}$$

If on the other hand $r \in [(1+P_{cap})\bar{r}, r_{max}]$, then hospital 1's profit is given by $\Phi(r|\bar{r})$. By definition of $\mathcal{S}'_o$, $\Phi(r|\bar{r}) \leq 0$ for any $r \in [(1+P_{cap})\bar{r}, r_{max}]$. In addition $\Pi(\bar{r}|\bar{r}) = 0$ by (8). Hence any $(\bar{r}, h(\bar{r})) \in \mathcal{S}'_o$ is a symmetric equilibrium.

(c) Next we prove that $(\bar{r}, \bar{c}) \notin \mathcal{S}'_o \cup \mathcal{S}'_p$ cannot be a symmetric equilibrium. Let $\bar{r} \in [r_{min}, r_{max}]$ and assume that all hospitals except hospital 1 choose $(\bar{r}, \bar{c}) \notin \mathcal{S}'_o \cup \mathcal{S}'_p$. If $\bar{c} \neq h(\bar{r})$, then $(\bar{r}, \bar{c})$ cannot be an equilibrium by Lemma A2, hence assume that $\bar{c} = h(\bar{r})$. Because $(\bar{r}, \bar{c}) \notin \mathcal{S}'_o \cup \mathcal{S}'_p$, either (1) $\bar{r} \in [r_{min}, \tilde{r}_1) \cup (\tilde{r}_2, \tilde{r}_3)$, or (2) $(r, h(r)) \in \mathcal{S}'$ and $(r, h(r)) \notin \mathcal{S}'_o$, by definitions of $\mathcal{S}'_o$ and $\mathcal{S}'_p$.

Assume that $\bar{r} \in [r_{min}, \tilde{r}_1) \cup (\tilde{r}_2, \tilde{r}_3)$. Then for $r \in (\bar{r}, (1+P_{cap})\bar{r}]$, by (A40), $\Pi(r|\bar{r}) = \Psi(r|\bar{r})$. Hence it is enough to show that

$$\frac{d\Psi(\bar{r}|\bar{r})}{dr} > 0, \qquad \text{for any } \bar{r} \in [r_{min}, \tilde{r}_1) \cup (\tilde{r}_2, \tilde{r}_3). \tag{A79}$$

Since $\Psi(r|\bar{r})$ is concave and its unique maximizer r_Ψ satisfies $d\Psi(r_\Psi(\bar{r})|\bar{r})/dr = 0$ (both from Lemma A4), (A79) is equivalent to

$$r_\Psi(\bar{r}) > \bar{r} \qquad \text{for any } \bar{r} \in [r_{min}, \tilde{r}_1) \cup (\tilde{r}_2, \tilde{r}_3),$$

which holds by Lemma A12.

Assume now that $(r, h(r)) \in \mathcal{S}'$ and $(r, h(r)) \notin \mathcal{S}'_o$ $r \in \mathcal{S}'$ and $r \notin \mathcal{S}'_o$, then as in part (b) of the proof, for $r \geq (1+P_{cap})\bar{r}$, hospital 1's profit is given by $\Phi(r|\bar{r})$. And by definition of $\mathcal{S}'_o$, $\Phi(r_1|\bar{r}) > \Pi(\bar{r}, h(\bar{r})) = 0$, for some $r_1 \in [(1+P_{cap})\bar{r}, r_{max}]$. Hence such $\bar{r}$ cannot be an equilibrium.

Note that (a)–(c) prove part (i) of the proposition. Part (iii) holds because when $P_{cap} \geq P'_{max}$, $r_{pi} = \tilde{r}_i$ hence $S'_p = \{(r, h(r)) : r \in [\tilde{r}_1, \tilde{r}_2] \cup [\tilde{r}_3, r_{max}]\}$ and $S' = \emptyset$.

(d) We next prove part (ii) of the proposition. First, if $P_{cap} < P_m(r^*)$, so that $r_{p3} \geq r_{p1} > r^*$, then $\mathcal{S}'_p \subset \{(r, h(r)) : r \in (r^*, r_{max}]\}$ by (A71). Hence for the rest of the proof we assume without loss of generality that $P_{cap} < P_m(r^*)$ so that $r_{pi} > r^*$ and show that $\mathcal{S}'_o \subset \{(r, h(r)) : r \in (r^*, r_{max}]\}$ for P_{cap} small enough. By (A77), to prove $\mathcal{S}'_o \subset \{(r, h(r)) : r \in (r^*, r_{max}]\}$ it suffices to show the existence of $\bar{\bar{P}}'_{cap} \in (0, P'_{max})$ such that

$$\Phi(r_{max}, P_{cap}|\bar{r}) > 0, \quad \text{for all } \bar{r} \in [\tilde{r}_1, r^*], \text{ and for } P_{cap} \leq \bar{\bar{P}}'_{cap}. \tag{A80}$$

Proof of (A80) is identical to that of (A48) with $\tilde{r}$ replaced by $\tilde{r}_1$, hence we skip the details. $\square$

H. Spillover Between Hospitals

So far we assumed that hospitals operate in non-overlapping catchment areas, which implies that a patient requiring readmission (after an index hospitalization) would always return to the same hospital. However, in practice the patient could be readmitted to others hospitals, see Zhang et al. (2016). In this section we explore how our proposed payment schemes should be modified under this assumption. Specifically, we show that when hospital capacity is ample m-HRRP still elicits socially optimal actions from hospitals. However, when capacity is limited, hospitals operating in the same area should be benchmarked using hospitals that do not. We propose a version of m-HRRPW in the limited capacity setting which benchmarks each hospital i by hospitals from a different catchment area (i.e., patients whose index admission is at hospital i do not visit those hospitals for readmission purpose). We study spillovers under the simplifying (symmetry) assumption that that each patient whose index admission is at hospital i and who requires readmission, visits hospital i with probability ρ and each of the other hospitals with probability $(1-\rho)/(N-1)$.

H.1. Unlimited capacity

Under the symmetric spillover assumption hospital i's total arrival rate is given by

$$\lambda_i^e = \lambda + \rho r_i \lambda + \frac{1-\rho}{N-1}\lambda \sum_{j\neq i} r_j = \lambda + \rho r_i \lambda + (1-\rho)\bar{r}_i \lambda, \tag{A81}$$

where the second equality follows from (6). In words, the total arrival rate is the sum of the arrival rates of initial admissions, λ, readmissions whose index admission is at hospital i, $\rho r_i \lambda$, and readmissions whose index admission is at one of the other hospitals, $(1-\rho)\bar{r}_i\lambda$. Following (1), the objective function of hospital i is given by

$$\Pi(r_i, c_i) = T_i - c_i \lambda_i^e - R(r_i, c_i). \tag{A82}$$

As in (2), the social welfare contribution of patients treated in hospital i is given by

$$S_i = V(\lambda) - c_i \lambda_i^e - R(r_i, c_i) \tag{A83}$$

and the regulator's objective is to maximize total social welfare, as given by

$$S = \sum_{i=1}^{N} S_i. \tag{A84}$$

We note that, unlike the case without spillovers, i.e. (2), we need to consider the total welfare (sum of S_i) in this case because S_i now depends on other hospitals' actions through λ_i^e. For the rest of the analysis, we assume (analogously to Lemma 1) that there is a symmetric optimal solution for the regulator's optimization problem and we denote each hospital's optimal actions as r^* and c^* and that FOCs of regulator's objective function are necessary and sufficient for optimality.

Next we show that the proposed payment scheme m-HRRP still elicits socially optimal actions from hospitals. With spillovers, the transfer payment T_i to hospital i under m-HRRP is

$$T_i = \bar{c}_i \lambda_i^e + (\bar{r}_i - r_i)\bar{c}_i \lambda + \bar{R}_i, \tag{A85}$$

where $\bar{c}_i$, $\bar{r}_i$, and $\bar{R}_i$ are defined as in (6). This transfer payment is equal to that with no spillovers (see (17)) except that the total arrival rate changes from $(1 + r_i)\lambda$ into λ_i^e due to spillovers.

Proposition A5. *Under m-HRRP there exists a unique equilibrium and each hospital chooses the first-best readmission and cost levels (r^*, c^*) in this equilibrium.*

Proof of Proposition A5: The proof is based on the simple observation that under m-HRRP the difference between a hospital's objective and the regulator's objective is independent of that hospital's actions. More precisely, given the actions of all the other hospitals, by (A82)-(A85) hospital i's objective under m-HRRP is

$$\begin{aligned}\Pi(r_i, c_i) &= (\bar{c}_i - c_i)\,\lambda_i^e + (\bar{r}_i - r_i)\bar{c}_i\lambda - R(r_i, c_i) + \bar{R}_i, \\ &= \bar{c}_i\lambda + \bar{c}_i\frac{1-\rho}{N-1}\lambda\sum_{j\neq i} r_j + \bar{r}_i\bar{c}_i\lambda + \bar{c}_i\rho r_i\lambda - c_i\lambda_i^e - r_i\bar{c}_i\lambda - R(r_i, c_i) + \bar{R}_i\end{aligned} \tag{A86}$$

where $\bar{c}_i$ and $\bar{R}_i$ are defined as in (6). By (A81)-(A84),

$$S = NV(\lambda) - c_i\lambda_i^e - \sum_{j\neq i}\lambda c_j\left(1 + \rho r_j + \frac{1-\rho}{N-1}\sum_{k\neq i,j} r_k\right) - (1-\rho)r_i\bar{c}_i\lambda - \sum_{i=1}^{N} R(r_i, c_i). \tag{A87}$$

By (A86)-(A87),

$$\begin{aligned}\Pi(r_i, c_i) - S =& \bar{c}_i\lambda + \frac{1-\rho}{N-1}\lambda\sum_{j\neq i} r_j + \bar{r}_i\bar{c}_i\lambda + \sum_{j\neq i}\lambda c_j\left(1 + \rho r_j + \frac{1-\rho}{N-1}\sum_{k\neq i,j} r_k\right) \\ &+ \sum_{j\neq i} R(r_j, c_i) + \bar{R}_i - NV(\lambda).\end{aligned}$$

Therefore the difference between the objective of the regulator and hospital i does not depend on r_i and c_i.

Since (r^*, c^*) is the unique maximizer of the social welfare S, it also maximizes hospital's profit Π under m-HRRP. That is, each hospital chooses r^* and c^* independent from other hospitals' decisions. Therefore, (r^*, c^*) constitutes the unique equilibrium under m-HRRP. $\square$

H.2. Limited capacity

Next we focus on the setting of limited capacity. In this case a direct extension of m-HRRPW (see (26)) does not elicit socially optimal outcomes because of the interlink between hospitals' arrival rates. In particular, if a hospital increases its readmission rates, this would increase the arrival rate of all the hospitals operating in the same catchment area. This in turn would reduce other hospitals' waiting time performance, increasing the waiting time target for the hospital itself. Hence if we use m-HRRPW without altering the waiting time benchmarks, it would introduce perverse incentives. Instead, we benchmark each hospital using the performance of hospitals operating in a different catchment area, an idea first proposed in Savva et al. (2018).

For the rest of this section we use the notation we introduced in §5.2.1 with the following natural extensions: the readmission rate, capacity, and cost per patient at each hospital are given by $\boldsymbol{r} = (r_1, r_2, \ldots, r_N)$, $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_N)$, $\boldsymbol{c} = (c_1, c_2, \ldots, c_N)$, respectively, and the vector $\boldsymbol{\Sigma} = (\boldsymbol{r}, \boldsymbol{\mu}, \boldsymbol{c})$ captures all this information. We denote the (effective) patient arrival rate at hospital i in equilibrium by $\lambda_i(\boldsymbol{\Sigma})$, and let $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_N)$. For notational simplicity we set $\boldsymbol{x}_{-i} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots x_N)$ for $x \in \{\lambda, r, \mu, c\}$.

We start by characterizing patients' equilibrium joining rate $\boldsymbol{\lambda}$ for fixed $\boldsymbol{\Sigma}$. Due to spillovers, the total arrival rate at hospital i is given by

$$\lambda_i^e = \lambda_i + \rho r_i \lambda_i + \frac{1-\rho}{N-1} \sum_{j \neq i} r_j \lambda_j. \tag{A88}$$

As in the setting of unlimited capacity, the total arrival rate is the sum of the arrival rates of initial admissions, λ_i, readmissions whose index admission is at hospital i, $\rho r_i \lambda_i$, and readmissions whose index admission is at one of the other hospitals, $\frac{1-\rho}{N-1} \sum_{j \neq i} r_j \lambda_j$.

At fixed $\boldsymbol{\lambda}_{-i}$, hospital i's equilibrium arrival rate λ_i is the unique solution of

$$\lambda_i = \Lambda \bar{\Theta}(tW(\lambda_i^e, r_i, \mu_i)). \tag{A89}$$

Since λ_i^e depends on $\boldsymbol{\lambda}$ and $\boldsymbol{r}$ by (A88), (A89) determines hospital i's equilibrium arrival rate λ_i as a function of $\boldsymbol{\lambda}_{-i}$, $\boldsymbol{r}$, and μ_i, which we denote by

$$\lambda_i = \mathcal{L}(\boldsymbol{\lambda}_{-i}, \boldsymbol{r}, \mu_i). \tag{A90}$$

In words, due to the spillover effect the arrival rate at each hospital depends on the arrival rates at all other hospitals, hence patients' joining decisions form a simultaneous-move game, which we characterize as below.

Let

$$\boldsymbol{\mathcal{L}}(\boldsymbol{x}, \boldsymbol{r}, \boldsymbol{\mu}) = [\mathcal{L}(\boldsymbol{\lambda}_{-1}, \boldsymbol{r}, \mu_1), \ldots, \mathcal{L}(\boldsymbol{\lambda}_{-N}, \boldsymbol{r}, \mu_N)].$$

For given $\boldsymbol{r}$ and $\boldsymbol{\mu}$, the equilibrium arrival rates $\boldsymbol{\lambda}$ is a fixed point of $\boldsymbol{\mathcal{L}}$, i.e.,

$$\boldsymbol{\lambda} = \boldsymbol{\mathcal{L}}(\boldsymbol{\lambda}, \boldsymbol{r}, \boldsymbol{\mu}). \tag{A91}$$

By continuity of Θ and W and using Brouwer's Fixed Point Theorem (Ok 2007), at least one solution to (A91) exists. In addition, we assume that the solution is unique.

Hospitals' and regular's objectives. Hospital i chooses readmission rate r_i, marginal cost c_i, and capacity μ_i to maximize total profit Π as given by

$$\Pi(r_i, \mu_i, c_i) = T_i - c_i\lambda_i^e - R(r_i, \mu_i, c_i), \tag{A92}$$

and the regular's objective is to maximize the total social welfare consisting of total patient utility for those who access care, total cost of providing treatment, and the cost of some patients unable to access care due to excess delays. Hence the social welfare contribution from hospital i is given by:

$$S_i = \Lambda \int_{tW(\lambda_i^e, r_i, \mu_i)}^{\infty} (x - tW(\lambda_i^e, r_i, \mu_i))d\Theta(x) - c_i\lambda_i^e - R(r_i, \mu_i, c_i) - c_e(\Lambda - \lambda_i) \tag{A93}$$

and total social welfare is given by:

$$\boldsymbol{S}(\boldsymbol{r}, \boldsymbol{c}, \boldsymbol{\mu}) = \sum_{i=1}^{N} S_i. \tag{A94}$$

We assume that there is a unique solution $(\boldsymbol{r}^*, \boldsymbol{c}^*, \boldsymbol{\mu}^*)$ to the regulator's problem in which different hospitals have the same actions, i.e., $r_i = r^*$, $c_i = c^*$, and $\mu_i = \mu^*$ for all $i = 1, \ldots, N$. We also assume that FOCs are necessary and sufficient for optimality.

Reimbursement scheme. The proposed reimbursement scheme in the no-spillover setting (i.e., m-HRRPW, see (26)) cannot restore social optimum, because the waiting-time benchmark for each hospital depends also on its own actions due to patient spillovers to other hospitals. To circumvent this issue while still using a payment scheme that is relatively easy to implement, we propose to benchmark a hospital using the performance of hospitals operating in a different catchment area, as proposed in Savva et al. (2018) in a setting where hospitals compete for customers who are not (endogenously) readmitted.

To demonstrate, consider a case with $2N$ identical hospitals, where hospitals 1 through N serve a different catchment area than hospitals $(N+1)$ through $2N$ in that patients whose index admissions are from hospitals in one catchment area do not seek treatment for readmission purpose from any hospital in a different catchment area, possibility due to proximity consideration. For each hospital i, we define sets $\mathbb{S}_i$ and $\mathbb{D}_i$ as the indices of hospitals who operate in the same catchment area as hospital i, and of those in the other catchment area, respectively, for example, $\mathbb{S}_1 = \{1, \ldots, N\}$

and $\mathbb{D}_1 = \{N+1, \ldots, 2N\}$. Consider the reimbursement scheme with hospital i's transfer payment equal to

$$T_i = c_e\lambda_i - t\left(\bar{W}_i^S - \bar{W}_i^D\right) N\lambda_i + \bar{R}_i + \left(\sum_{j\in\mathbb{D}_i} c_j\lambda_j^e - \sum_{j\in\mathbb{S}_i, j\neq i} c_j\lambda_j^e\right) - c_e\left(\sum_{j\in\mathbb{D}_i}\lambda_j - \sum_{j\in\mathbb{S}_i, j\neq i}\lambda_j\right), \quad \text{(A95)}$$

where $\bar{W}_i^S$, $\bar{W}_i^D$, and $\bar{R}_i$ are given by

$$\bar{W}_i^S = \frac{1}{N}\sum_{j\in\mathbb{S}_i} W_j, \quad \bar{W}_i^D = \frac{1}{N}\sum_{j\in\mathbb{D}_i} W_j, \quad \bar{R}_i = \frac{1}{N}\sum_{j\in\mathbb{D}_i} R_j, \quad \text{(A96)}$$

and we use W_i and R_i to denote the expected waiting time and total lump-sum cost for hospital i. There are two main differences between this payment scheme and the scheme proposed for the no-spillover setting, see (26): (i) the waiting-time benchmark $\bar{W}_i^D$ and the lump-sum payment $\bar{R}_i$ are based on a set of hospitals which patients of hospital i do not visit for readmission purpose, and (ii) the relative performance-based financial incentive for waiting time is multiplied by the number of hospitals in the same group, i.e., N. The first difference reinstates the power of relative benchmarking which spillovers had eroded. The second difference amplifies the financial incentives for a hospital to reduce expected waiting time, which due to spillovers, affects the expected waiting times of all other hospitals from the same catchment area.

Proposition A6. *Under the reimbursement scheme given in* (A95)*, the unique symmetric equilibrium is for each hospital i to pick $r_i = r^*$, $\mu_i = \mu^*$, and $c_i = c^*$.*

Proof of Proposition A6. We prove the result by showing that in any symmetric equilibrium, FOCs of each hospital's objective under m-HRRPW payment scheme are equal to the corresponding FOCs of the regulator's objective.

By (A93)-(A94), total social welfare is given by:

$$\boldsymbol{S} = \sum_{i=1}^{2N} S_i = \sum_{i=1}^{2N}\left\{\Lambda\int_{tW_i}^{\infty}(x - tW_i)\, d\Theta(x) - c_i\lambda_i^e - R_i - c_e(\Lambda - \lambda_i)\right\},$$

where $W_i \equiv W(\lambda_i^e, r_i, \mu_i)$, $R_i \equiv R(r_i, \mu_i, c_i)$, and λ_i^e is given by (A88) and (A90). For hospital i's any decision variable $x_i \in \{r_i, \mu_i, c_i\}$,

$$\begin{aligned}
\frac{\partial \boldsymbol{S}}{\partial x_i} &\overset{\text{(r0)}}{=} \frac{\partial S_i}{\partial x_i} + \sum_{j\in\mathbb{S}_i, j\neq i}\frac{\partial S_j}{x_i} \\
&\overset{\text{(r1)}}{=} -t\Lambda\bar{\Theta}(tW_i)\frac{\partial W_i}{\partial x_i} - \frac{\partial\{c_i\lambda_i^e - c_e\lambda_i + R_i\}}{\partial x_i} + \sum_{j\in\mathbb{S}_i, j\neq i}\left\{-t\Lambda\bar{\Theta}(tW_j)\frac{\partial W_j}{\partial x_i} - \frac{\partial\{c_j\lambda_j^e - c_e\lambda_j\}}{\partial x_i}\right\} \\
&\overset{\text{(r2)}}{=} -t\lambda_i\frac{\partial W_i}{\partial x_i} - \frac{\partial\{c_i\lambda_i^e - c_e\lambda_i + R_i\}}{\partial x_i} + \sum_{j\in\mathbb{S}_i, j\neq i}\left\{-t\lambda_j\frac{\partial W_j}{\partial x_i} - \frac{\partial\{c_j\lambda_j^e - c_e\lambda_j\}}{\partial x_i}\right\}
\end{aligned}$$

$$\overset{(r3)}{=} -t\lambda_i \frac{\partial W_i}{\partial x_i} - \sum_{j\in\mathbb{S}_i, j\neq i} t\lambda_j \frac{\partial W_j}{\partial x_i} - \sum_{j\in\mathbb{S}_i} \frac{\partial (c_j \lambda_j^e)}{\partial x_i} + c_e \sum_{j\in\mathbb{S}_i} \frac{\partial \lambda_j}{\partial x_i} - \frac{\partial R_i}{\partial x_i}, \tag{A97}$$

where (r0) follows from (A94) and the fact that any hospital j with $j \in \mathbb{D}_i$ operates in a different area from hospital i, (r1) follows from (A93), (r2) follows from (A89), and (r3) follows by reorganizing the terms.

By (A92) and (A95), hospital i's objective is

$$\Pi(r_i, \mu_i, c_i) = -t\left(\bar{W}_i^S - \bar{W}_i^D\right) N\lambda_i + \left(\sum_{j\in\mathbb{D}_i} c_j\lambda_j^e - \sum_{j\in\mathbb{S}_i} c_j\lambda_j^e\right) - c_e\left(\sum_{j\in\mathbb{D}_i}\lambda_j - \sum_{j\in\mathbb{S}_i}\lambda_j\right) - R_i + \bar{R}_i, \tag{A98}$$

hence

$$\begin{aligned}
\frac{\partial \Pi}{\partial x_i} &= -t\frac{\partial \bar{W}_i^S}{\partial x_i} N\lambda_i - t(\bar{W}_i^S - \bar{W}_i^D) N \frac{\partial \lambda_i}{\partial x_i} - \sum_{j\in\mathbb{S}_i} \frac{\partial (c_j\lambda_j^e)}{\partial x_i} + c_e \sum_{j\in\mathbb{S}_i} \frac{\partial \lambda_j}{\partial x_i} - \frac{\partial R_i}{\partial x_i} \\
&= -\sum_{j\in\mathbb{S}_i} t\lambda_i \frac{\partial W_j}{\partial x_i} - \sum_{j\in\mathbb{S}_i} \frac{\partial (c_j\lambda_j^e)}{\partial x_i} + c_e \sum_{j\in\mathbb{S}_i} \frac{\partial \lambda_j}{\partial x_i} - \frac{\partial R_i}{\partial x_i} - t(\bar{W}_i^S - \bar{W}_i^D) N \frac{\partial \lambda_i}{\partial x_i},
\end{aligned} \tag{A99}$$

where the second equality follows from (A96).

In any symmetric equilibrium, $x_i = x_j$ for any $x_i \in \{r_i, \mu_i, c_i\}$ and $j \neq i$, hence

$$\lambda_i = \lambda_j, \quad \lambda_i^e = \lambda_j^e, \quad R_i = R_j, \text{ and } \; W_i = W_j, \text{ for any } j \neq i \tag{A100}$$

by (A88)-(A89), and

$$\bar{W}_i^S = \bar{W}_i^D \text{ and } \; R_i = \bar{R}_i \text{ for } i = 1, \ldots, 2N \tag{A101}$$

by (A96) and (A100). By (A97)-(A101),

$$\frac{\partial \boldsymbol{S}}{\partial x_i} = \frac{\partial \Pi}{\partial x_i}$$

for all $x_i \in \{r_i, \mu_i, c_i\}$ and $i = 1, \ldots, 2N$. Because FOCs are assumed to be necessary and sufficient to obtain hospitals' optimal actions, (r^*, μ^*, c^*) is a symmetric equilibrium. Because (A97) determines a unique solution, so does (A99); hence (r^*, μ^*, c^*) is the unique symmetric equilibrium. Finally, by (A98), (A100), and (A101), each hospital earns null profit in any symmetric equilibrium. □

I. Inpatient Prospective Payment System

In this appendix, we explain the connection between our model and the Inpatient Prospective Payment System (IPPS) and HRRP CMS uses. First, in §I.1 we delineate how CMS determines the reimbursement amounts per discharge under IPPS and explain how $\bar{c}_i$ and $\bar{R}_i$ in our model are obtained from CMS's reimbursement payment per discharge under our assumptions. Then, in §I.2, we explain the connection between the HRRP penalty CMS uses and the one we use in our model (also referred to as HRRP in the paper).

I.1. Reimbursement payment per discharge under IPPS

The reimbursement amounts under IPPS for each diagnosis-related group (DRG) are based on (i) operating and capital base rates, and (ii) DRG weights. Operating and capital base rates, respectively, represent the average operating and capital cost for a typical Medicare inpatient stay and do not involve case-mix, area wages, and teaching costs (Guterman and Dobson 1986). Base rates were initially set based on average cost information and are updated annually.[1] Under IPPS, CMS adjusts the base rates by DRG weights to determine the per-admission payment for each DRG. Through a process called recalibration, DRG weights are determined annually based on average standardized charges for each DRG. These averages are calculated by summing the standardized charges for all cases in a DRG for CMS patients and then dividing that amount by the number of cases classified in that specific DRG.[2] In addition, CMS further adjusts the base rates for each hospital for various other factors, e.g., geographic location, disproportionately indigent patient populations, use of high-cost technologies in treatment, etc., see CMS (2019).

In line with the literature (e.g., Shleifer 1985, Tangerås 2009), we make simplifying assumptions in modelling IPPS to make the analysis tractable. First, to abstract away from all the adjustments regarding patient population, geographic location, and other factors, we assume that all hospitals are identical. Second, we only focus on a single DRG. (See Remark 1 in the paper for our reasons to make these assumptions.) Under these assumptions, the reimbursement amount per discharge for the DRG we consider (i.e., DRG weight$\times$ operating base rate) is equal to $\bar{c}_i$, which is the average DRG-adjusted operating cost across all hospitals. In addition, we use $\bar{R}_i$ to capture the capital payments.

There are three main differences between per-discharge reimbursement payment in our model and the one in practice. First, CMS uses all discharges to find the average cost whereas we use hospital averages, following Shleifer (1985). This simplifies the analysis and notation, and does not affect our results as the symmetric equilibria in our model and under CMS's approach are the same. Second, when calculating base rates, unlike our model, CMS uses a mix of average cost information and updates based on annual expected cost increase. However, CMS's approach is similar, in spirit, to using average cost information as in our model (MedPAC 2017). Third, unlike our model, CMS does not make lump-sum payments and reimburses capital expenses per discharge basis. However, our lump-sum payment can be made per discharge by dividing it by number of discharges, i.e., $\bar{R}_i/(\lambda(1+r_i))$, and it can be interpreted as the capital payments made by CMS to hospitals as explained above.

[1] See Federal Register, Sections D&M, 42 CFR Part 412, 2004, `https://ecfr.io/Title-42/cfr412_main`.

[2] See Federal Register , Section C, Vol. 64, No. 146, 1999, page 41498, `http://govinfo.gov/content/pkg/FR-1999-07-30/pdf/99-19334.pdf`

I.2. HRRP penalty

Our model is based on the version of HRRP implemented between 2012 and 2018 financial year (FY) (old-HRP, hereafter), which was when this project was conducted. A new version of HRRP (new-HRRP, hereafter) came into effect in 2019 FY. We did not modify our model since the new-HRRP only changes the way risk-adjustment is done under old-HRRP and does not affect our analysis under homogenous hospitals assumption (see below for more details). The new-HRRP primarily uses the same incentive scheme as old-HRRP (which is the main focus of our work). To explain this in detail, we first clarify the connection between our model and old-HRRP below. Then, we explain the changes new-HRRP brought and their implications for our analysis.

A hospital's profit function in (8) is the same under old- and new-HRRP, except the penalty term π. Let $\pi^{old-HRRP}$ and $\pi^{new-HRRP}$ denote the penalty under old- and new-HRRP, respectively.

I.2.1. Old HRRP. The penalty of a hospital under old-HRRP is given by (see CMS (2017) for a general overview and MedPAC (2013) for a detailed explanation):

$$\pi^{old-HRRP} = DRGP \times \min\left\{ \frac{\sum_{j\in C} DRGP_j \times (ERR_j - 1.0)^+}{DRGP}, P_{cap} \right\}, \tag{A102}$$

where $(a)^+ = max\{a, 0\}$, C is the set of conditions monitored by CMS, [3] $DRGP$ is the total DRG-based operating payments to the hospital for all (monitored and unmonitored) conditions, $DRGP_j$ is the total DRG-based operating payments to the hospital for condition j, P_{cap} is the penalty cap (currently equal to 3%), ERR_j is the excess readmission ratio of the hospital for condition j.[4]

Next, we argue that, under our assumptions, $\pi^{old-HRRP}$ in (A102) reduces to HRRP penalty π in (7). We again use subscript i for hospital i. Recall from §I.1 that per-admission operating DRG payment to hospital i is equal to the average operating cost of all other hospitals $\bar{c}_i$. Also, our base model assumes identical hospitals and focuses on a single disease/condition so that λ is the demand rate for all hospitals and there is only one element in C. (Recall that we extend our model to multiple diseases in Appendix G and to heterogeneous hospitals in §5.3.) As a result, $DRGP = DRGP_1 = \bar{c}_i\lambda(1 + r_i)$ when hospital i's readmission level is r_i.

CMS uses a logistic Hierarchical Generalized Linear Model (HGLM) to determine the relative hospital performance ERR_j (see Section 2.6 in Horwitz et al. (2012)), whereas we use $r_i/\bar{r}_i$ in our model.[5] In our model, we ignore the potential errors that the estimation procedure introduces and

[3] As of 2019 FY, CMS monitors only six conditions under HRRP, namely, Acute Myocardial Infarction (AMI), Heath Failure, Pneumonia, Chronic Obstructive Pulmonary Disease (COPD), Total Hip/Knee Arthroplasty (THA/TKA), and Coronary Artery Bypass Graft (CABG) Surgery.

[4] Excess readmission ratio (ERR) is the adjusted actual number of readmissions predicted at the hospital over the expected number (at an average hospital) of readmissions.

[5] Under certain assumptions, the estimates of HGLM are consistent, i.e., the parameter estimates of the model converges to the true values in probability as the sample size goes to infinity, see Nie (2006).

assume that estimates from the model are accurate estimates of hospitals' true readmission rates. Also, we assume that hospitals treat homogenous patients so that $r_i/\bar{r}_i$ is the true estimate of the hospital i's relative performance –see, for example, Zhang et al. (2016), Andritsos and Tang (2018), Adida and Bravo (2019), for similar assumptions.

Under these assumptions, the penalty for hospital i under old-HRRP in (A102) reduces to:

$$\begin{aligned}\pi_i^{old-HRRP} &= \min\left\{\left(\frac{r_i}{\bar{r}_i}-1\right)^+, P_{cap}\right\}\bar{c}_i\lambda(1+r_i)\\ &= \left(\min\left\{\frac{r_i-\bar{r}_i}{\bar{r}_i}, P_{cap}\right\}\right)^+\bar{c}_i\lambda(1+r_i), \end{aligned} \tag{A103}$$

which is identical to $\pi(r_i|\bar{r}_i,\bar{c}_i)$ in (7).

I.2.2. New HRRP. In order to improve the risk-adjustment procedure used in old-HRRP, new-HRRP stratifies hospitals into five peer groups based on the proportion of dual eligible[6] patients a hospital treats. Then, a hospital's performance is assessed relative to the performance of the hospitals within the same peer group. Under the simplifying assumptions stated above, specifically under the identical hospitals and homogenous patients assumptions, all hospitals in our model would fall into the same stratified group. Another change new-HRRP introduced is that a relative performance target is set using the median hospital performance but not the average performance as is the case in old-HRRP (see below for more details). In a symmetric equilibrium, the median and mean readmission rates within a peer group are identical. Thus, if all hospitals are assumed to be in the same peer group, any symmetric equilibrium under old-HRRP in our model is also a symmetric equilibrium under the new-HRRP so that all our results in the paper continue to hold under the new-HRRP.

Furthermore, our results continue to hold, even when one assumes that hospitals are divided into different peer groups, under the assumptions that hospitals in the same group are identical and are treating a homogenous group of patients. We will demonstrate this next.

First, new-HRRP changes the relative performance comparison from $ERR_j - 1.0$ used in old HRRP to $ERR_j - ERR_{j,pgmed}$, where $ERR_{j,pgmed}$ is the median ERR of all hospitals within the peer group for condition j. Also, new-HRRP applies a neutrality modifier (NM), which is a multiplicative factor that, when applied to hospital payment adjustment factors, equates to total Medicare savings under the old and new HRRP (CMS 2017). To calculate NM, CMS estimates the total Medicare savings across all hospitals under the old HRRP as well as under the new HRRP (in the absence of a modifier). The penalty of a hospital under new-HRRP is given by:

$$\pi^{new-HRRP} = DRGP \times \min\left\{\frac{\sum_{j\in C} \mathrm{NM} \times DRGP_j(ERR_j - ERR_{j,pgmed})^+}{DRGP}, P_{cap}\right\}. \tag{A104}$$

[6] A hospital's dual proportion is the proportion of Medicare fee-for-service (FFS) and managed care stays where the patient was dually eligible for Medicare and full-benefit Medicaid.

Now, assume that hospitals are divided into five peer groups, where hospitals in the same peer group treat homogenous patients, and focus on one peer group (i.e., group 1). Under above assumptions, $\pi^{new-HRRP}$ in (A104) reduces to:

$$\pi_i^{new-HRRP} = \left(\min\left\{\mathrm{NM} \times \left(\frac{r_i - \bar{r}_i}{\bar{\bar{r}}_i}\right), P_{cap}\right\}\right)^+ \bar{c}_i \lambda (1 + r_i), \tag{A105}$$

where $\bar{r}_i$ is the median readmission rate of hospitals in peer group 1 and $\bar{\bar{r}}_i$ is the average readmission rate of all hospitals other than hospital i. There are three differences between $\pi^{new-HRRP}$ and $\pi_i^{old-HRRP}$ (see (A103)); (i) $\pi^{new-HRRP}$ has a new term NM; (ii) old-HRRP uses $\frac{r_i - \bar{r}_i}{\bar{r}_i}$ to determine the relative performance of hospital i whereas new-HRRP uses $\frac{r_i - \bar{r}_i}{\bar{\bar{r}}_i}$; and (iii) in new-HRRP $\bar{r}_i$ is the median readmission rate instead of average used in old-HRRP.

We next argue that, under certain assumptions, our main results based on old-HRRP continue to hold when the penalty calculation follows $\pi_i^{new-HRRP}$. First, because median and average are identical in a symmetric equilibrium, our analysis focusing on symmetric equilibria does not need to be modified for using medians. Also, we assume that NM is exogenous to hospital i's actions, in a way similar to that we assume $\bar{r}_i$ and $\bar{c}_i$ are. In addition, we assume that $\bar{c}_i$ and $\bar{R}_i$ for hospital i are calculated using data only from hospitals in the same peer group –CMS does indeed increase the reimbursement amounts based on a hospital's share of low-income patients, see CMS (2019).

Under these assumptions, it is easy to demonstrate the differences between old- and new-HRRP by focusing on the payment scheme obtained from each one after we remove the cap and the no-bonus provisions. In particular, if we remove the cap and the no-bonus provisions from (A105), the new penalty term becomes

$$\pi_i^{new-HRRP} = (r_i - \bar{r}_i)\, \bar{c}_i \lambda \frac{(1 + r_i) \times \mathrm{NM}}{\bar{\bar{r}}_i}, \tag{A106}$$

corresponding to HRRP-I. Hence, the multiplier is equal to $\frac{(1+r_i)\times \mathrm{NM}}{\bar{\bar{r}}_i}$ under new-HRRP –recall that it is equal to $\frac{(1+r_i)}{\bar{r}_i}$ under old-HRRP, see $\pi^I(r_i|\bar{r}_i, \bar{c}_i)$ in (9).

Using this, one can easily show that Proposition 1 still holds, if $\frac{\mathrm{NM}(1+r_i)}{\bar{\bar{r}}_i} > 1$. This is a reasonable condition under the current estimates; NM for 2019 FY is estimated to be 0.9481 (CMS 2018) and readmission rates are typically around 20%, see, for example Chen and Savva (2018). Based on this result, using the same arguments in the paper and assuming Assumption 1 is valid for some (potentially different) $\tilde{r}(< r^*)$, it can be shown that Propositions 2 and 3 continue to hold.

References

Adida, E. and F. Bravo (2019). Contracts for healthcare referral services: Coordination via outcome-based penalty contracts. *Management Science 65*(3), 1322–1341.

Adida, E., H. Mamani, and S. Nassiri (2016). Bundled payment vs. fee-for-service: Impact of payment scheme on performance. *Management Science 63*(5), 1606–1624.

Andritsos, D. A. and C. S. Tang (2018). Incentive programs for reducing readmissions when patient care is co-produced. *Production and Operations Management 27*(6), 999–1020.

Bastani, H., M. Bayati, M. Braverman, R. Gummadi, and R. Johari (2016). Analysis of Medicare pay-for-performance contracts. `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2839143`, last accessed on December 03, 2017.

Chen, C. and N. Savva (2018). Unintended consequences of hospital regulation: The case of the Hospital Readmissions Reduction Program. Technical report, London Business School.

CMS (2017). New stratified methodology hospital-level impact file user guide. `https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Downloads/HRRP_StratMethod_ImpctFile_UG.PDF`, last accessed on August 12, 2018.

CMS (2018). Details for title: FY 2019 IPPS proposed rule: Hospital readmissions reduction program supplemental data file. `https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Archived-Supplemental-Data-Files/FY2019-IPPS-Proposed-Rule-HRRP-Supplemental-Data-File.html?DLPage=1&DLEntries=10&DLSort=1&DLSortDir=descending`, last accessed on August 19, 2019.

CMS (2019). Acute care hospital inpatient prospective payment system. `https://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNProducts/downloads/AcutePaymtSysfctsht.pdf`, last accessed on August 8, 2019.

Guterman, S. and A. Dobson (1986). Impact of the medicare prospective payment system for hospitals. *Health Care Financing Review 7*(3), 97–114.

Horwitz, L., C. Partovian, Z. Lin, and et al. (2012). Hospital-wide all-cause unplanned readmission measure: Final technical report. `https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/downloads/MMSHospital-WideAll-ConditionReadmissionRate.pdf`, last accessed on August 12, 2019.

MedPAC (2013). Report to congress: Medicare and the health care delivery system. online appendix 4b: Computation of how penalty is not reduced with a decline in readmission rates. `http://www.medpac.gov/docs/default-source/reports/jun13_ch04_appendix.pdf?sfvrsn=0`, last accessed on August 19, 2019.

MedPAC (2017). Hospital acute inpatient services payment system. `http://medpac.gov/docs/default-source/payment-basics/medpac_payment_basics_17_hospital_final65a311adfa9c665e80adff00009edf9c.pdf`, last accessed on August 19, 2019.

Nie, L. (2006, Apr). Strong consistency of the maximum likelihood estimator in generalized linear and nonlinear mixed-effects models. *Metrika 63*(2), 123–143.

Ok, E. A. (2007). *Real analysis with economic applications*, Volume 10. Princeton University Press.

Savva, N., T. Tezcan, and O. Yildiz (2018). Can yardstick competition reduce waiting times? To appear in *Management Science*.

Shleifer, A. (1985). A theory of yardstick competition. *The RAND Journal of Economics 16*(3), 319–327.

Tangerås, T. P. (2009). Yardstick competition and quality. *Journal of Economics & Management Strategy 18*(2), 589–613.

Zhang, D. J., I. Gurvich, J. A. Van Mieghem, E. Park, R. S. Young, and M. V. Williams (2016). Hospital Readmissions Reduction Program an economic and operational analysis. *Management Science 62*(11), 3351–3371.