Online Appendix
"Common Ownership and Innovation Efficiency"

Xuelin Li

Tong Liu

Lucian A. Taylor

July 15, 2022

This Internet Appendix explains the sample construction process in detail and presents additional empirical results, most for checking the robustness of our results presented in the paper. The motivation and summary of these robustness results are in the paper.

Contents

Section A5 Results with 3-digit ICD Codes

- ◇ Table A5.1 Version of Table 2 Using 3-digit ICD Codes
- ◇ Table A5.2 Version of Table 3 Using 3-digit ICD Codes
- ◇ Table A5.3 Version of Table 4 Using 3-digit ICD Codes

Section A6 Robustness of Table 3

- ◇ Table A6.1 Project Outcomes After Excluding Bad Phase I Outcomes
- ◇ Table A6.2 Version of Table 3 Following 2SLS without Probit
- ◇ Table A6.3 Version of Table 3 Including ICD × Quarter Fixed Effects
- ◇ Table A6.4 Version of Table 3 Controlling for Technological Similarity
- ◇ Table A6.5 Version of Table 3 Controlling for Same MSA
- ◇ Table A6.6 Version of Table 3 Dropping Boston
- ◇ Table A6.7 Version of Table 3 Dropping San Francisco
- ◇ Table A6.8 Version of Table 3 Alternative Instrument using the Corporate Opportunity Waivers
- ◇ Table A6.9 Cross-Sectional Version of Table 3
- ◇ Table A6.10 Version of Table 3 Using *Suspend* as the Dependent Variable

Section A7 Robustness of Table 4

- ◇ Table A7.1 Version of Table 4 Including Startup×Quarter Fixed Effects
- ◇ Table A7.2 Lead vs. Non-lead VC
- ◇ Table A7.3 Version of Table 4 Controlling for Technological Similarity
- ◇ Table A7.4 Version of Table 3 and Table 4 Controlling for $CommonOwn_{it}$
- ◇ Table A7.5 Address Concerns About Poaching Nearby Investors
- ◇ Table A7.6 Version of Table 4 Alternative Instrument using the Corporate Opportunity Waivers

Section A8 Cox Proportional Hazard Model

- ◇ Table A8.1 Cox Proportional Hazards Regressions

Section A9 Robustness of Table 9

A9.1 Oster (2019) test for omitted variable bias

- ◇ Table A9.1 Innovation Efficiency: Adding Controls Gradually
- ◇ Table A9.2 Innovation Efficiency: Alternative Specifications
- ◇ Table A9.3 Impacts Measured by Forward Citations of Drug Patents

Section A10 Coefficients of Control Variables in Table 7 and Table 10

- ◇ Table A10.1 Coefficients of Control Variables in Table 7
- ◇ Table A10.2 Coefficients of Control Variables in Table 10

A1 Case Study: Common Ownership at a Prominent VC Firm

New England Associates (NEA) is a Maryland-based VC firm that, in 2018, was ranked second in Forbes' list of most-active healthcare VCs.[1] In July 2012, NEA was an investor in two Boston-based startups, Intarcia and Rhythm Pharmaceuticals. Both startups at that time had drug projects that were targeting obesity and that were in Phase I clinical trials. Rhythm's obesity-related project, setmelanotide, corresponds to the pioneer project in our paper. That project progressed from Phase I to Phase II in December 2012. This event corresponds to the "shock" in our main tests. Rhythm eventually went public in 2017, shortly after beginning Phase III trials, and setmelanotide obtained FDA approval in 2020. Intarcia's obesity-related project, named "weight regulating human endocrine peptide" in the Cortellis database, corresponds to the lagging project in our paper. That project never progressed to Phase II, and updates on the project stopped in 2016, indicating the project was abandoned. NEA's last investment in Intarcia occurred in November 2012, so NEA stopped funding Intarcia after the "shock" from Rhythm. Since then, Intarcia has mainly focused on its diabetes pipeline related to the drug exenatide. In 2015 Intarcia redirected some of its earlier obesity-related research into two new projects, "exenatide + optimized peptide 1" and "exenatide + optimized peptide 2," both of which target "type 2 diabetes and/or obesity."

[1] https://www.forbes.com/sites/michelatindera/2018/12/27/these-10-vc-firms-made-the-most-investments-in-healthcare-startups-this-year/?sh=5f4321485498

A2 Model Appendix

A2.1 Technical Parameter Restrictions

We impose the following parameter restrictions:

$$p_2\left(\left(1-\lambda\right)W+\lambda D\right)-c_2>0, \tag{Assumption 1}$$

$$p_1\left(\frac{p_2\left(\left(1-\lambda\right)W+\lambda D\right)-c_2}{r+2p_2}\right)-c_1>0, \tag{Assumption 2}$$

and

$$\lambda<\frac{p_1}{p_1+p_2}\frac{p_2}{r+p_2},\text{ and }D>L\geq c_2/p_2. \tag{Assumption 3}$$

These assumptions ensure startups have an incentive to join the game, so they represent participation constraints. Assumption 1 implies that a Phase II startup is willing to continue when its competitor is also in Phase II. Similarly, Assumption 2 ensures that a Phase I startup is willing to continue when its competitor is already in Phase II. We formally prove these statements in Lemma A2.1 below. Lastly, Assumption 3 ensures the probability and profitability of catching up is sufficiently small, otherwise players can free-ride the other competitor's success.

A2.2 Proofs of Section 1

To prove these results, we first derive the startups' continuation values. When both startups are in Phase II, each startup's Bellman equation follows from

$$rV_2=-c_2+p_2\left(\left(1-\lambda\right)W+\lambda D-V_2\right)+p_2\left(\lambda\left(D-L\right)-V_2\right),$$

where V_2 is the value of each startup in this scenario. To understand the above equation, the VC's flow value rV_2 in startup A, for example, has three components. First, it spends the R&D costs. Second, with probability p_2, A finishes Phase II first. But it only receives W with probability $1-\lambda$, otherwise it enters into the duopoly competition. So the expected jump in payoff is $\left(1-\lambda\right)W+\lambda D-V_2$. Lastly, with probability p_2, startup B finishes Phase II first, and thus startup A receives net duopoly profit $D-L$ with probability λ. Rearranging this equation generates

$$V_2=\frac{p_2\left(\left(1-\lambda\right)W+\lambda 2D\right)-p_2\lambda L-c_2}{r+2p_2}.$$

When one pioneering startup is in Phase II and the other is lagging in Phase I, the VC's continuation value in the pioneering startup V_P follows

$$rV_P=-c_2+p_2\left(\left(1-\lambda\right)W+\lambda D-V_P\right)+p_1\left(V_2-V_P\right),$$

or equivalently,

$$V_P=\frac{p_2\left(\left(1-\lambda\right)W+\lambda D\right)+p_1V_2-c_2}{r+p_2+p_1}.$$

The owner's continuation value in the lagging startup V_L follows

$$rV_L=-c_1+p_1\left(V_2-V_L\right)+p_2\left(\lambda\left(D-L\right)-V_L\right),$$

or equivalently,

$$V_L = \frac{p_2\lambda\left(D-L\right)+p_1V_2-c_1}{r+p_2+p_1}.$$

The following Lemma shows that both startups are willing to participate in all phases:

Lemma A2.1. *Given Assumption 1 and Assumption 2, a startup owned by a non-common VC will continue the patent race in both phases. In particular, a non-common VC will continue the lagging startup with probability one.*

Proof. It is easy to verify that Assumption 1 ensures $V_2 > 0$. If both startups are in Phase I, a Phase I non-commonly owned startup's continuation value is

$$rV_1 = -c_1 + p_1\left(V_P - V_1\right) + p_1\left(V_L - V_1\right),$$

which is equivalent to

$$V_1 = \frac{p_1\left(V_P+V_L\right)-c_1}{r+2p_1}.$$

The participation constraint is simply $V_1 \geq 0$, or equivalently

$$\left(r+2p_2\right)\left(r+2p_1+p_2\right)\frac{c_1}{p_1} \leq \left(r+2p_1+2p_2\right)\left(p_2\left(\left(1-\lambda\right)W+\lambda 2D\right)-p_2\lambda L-c_2\right).$$

This holds given Assumption 2 since

$$\frac{r+2p_1+2p_2}{r+2p_1+p_2}\left(p_2\left(\left(1-\lambda\right)W+\lambda 2D\right)-p_2\lambda L-c_2\right) > p_2\left(\left(1-\lambda\right)W+\lambda D\right)-c_2.$$

Next, a non-common VC will continue the lagging startup if $V_L \geq 0$, which requires

$$V_2 = \frac{p_2\left(\left(1-\lambda\right)W+\lambda 2D\right)-p_2\lambda L-c_2}{r+2p_2} \geq \frac{c_1}{p_1} - \frac{p_2\lambda\left(D-L\right)}{p_1}.$$

This equation implies that

$$c_2 \leq \bar{c} = p_2\left(\left(1-\lambda\right)W+\lambda 2D\right)-p_2\lambda L - \frac{\left(r+2p_2\right)c_1}{p_1} + \frac{\left(r+2p_2\right)p_2\lambda\left(D-L\right)}{p_1}.$$

Notice that Assumption 2 implies

$$c_2 \leq C < p_2\left(\left(1-\lambda\right)W+\lambda D\right) - \frac{\left(r+2p_2\right)c_1}{p_1} = \bar{c} - \frac{\left(r+p_1+2p_2\right)p_2\lambda\left(D-L\right)}{p_1}.$$

This ensures that $C < \bar{c}$ and a non-common VC will continue the lagging startup. Lastly, it is easy to verify that Assumption 2 ensures $V_P > 0$. □

Next, consider the common VC that invests in both the pioneering and lagging startup. Once it stops investment in the lagging startup, the continuation value in the pioneering startup, denoted V_M, becomes

$$rV_M = -c_2 + p_2\left(W - V_M\right), \tag{A1}$$

or equivalently,

$$V_M = \frac{p_2W-c_2}{r+p_2}.$$

The common VC will continue the lagging startup only if

$$V_M \leq V_P + V_L. \tag{A2}$$

Proof of the Proposition in Section 1

Proof. Define $\Pi = (1-\lambda) W + \lambda 2D - \lambda L$ to simplify notation. Equation (A2) is equivalent to

$$\begin{aligned} \frac{p_2 W - c_2}{r + p_2} &\leq \frac{p_2 \Pi + 2p_1 \frac{p_2 \Pi - c_2}{r + 2p_2} - c_1 - c_2}{r + p_1 + p_2} \\ \Leftrightarrow (r + p_1 + p_2)\, p_2 W - (r + p_1 + p_2)\, c_2 &\leq (r + p_2)\, p_2 \Pi + p_1 \frac{2r + 2p_2}{r + 2p_2} p_2 \Pi \\ &\quad - p_1 \frac{2r + 2p_2}{r + 2p_2} c_2 - (r + p_2)\, c_1 - (r + p_2)\, c_2. \end{aligned} \tag{A3}$$

Notice that

$$p_1 \frac{2r + 2p_2}{r + 2p_2} + r + p_2 = \frac{p_1 r}{r + 2p_2} + r + p_1 + p_2.$$

Thus Equation (A3) is equivalent to

$$\frac{p_1 r}{r + 2p_2} c_2 \leq \frac{p_1 r}{r + 2p_2} p_2 \Pi + (r + p_1 + p_2)\, p_2\, (\Pi - W) - (r + p_2)\, c_1,$$

which implies $c_2 \leq \underline{c}$. □

Proof of Corollary 1 in Section 1

Proof. Recall that Lemma A2.1 implies a non-common VC will always continue the lagging startup, i.e. $C < \bar{c}$ and $\Pr\{\text{Continue}|\text{No Common Ownership}\} = 1$. Define $\Delta c = \bar{c} - \underline{c}$. It is easy to show that

$$\Delta c = \frac{p_2}{r} \frac{(r + 2p_2)\, c_1}{p_1} + \left(1 + \frac{p_1}{r} + \frac{p_2}{r}\right) \frac{(r + 2p_2)\, p_2 \lambda\, (\delta + L)}{p_1} + \frac{(r + 2p_2)\, p_2 \lambda\, (D - L)}{p_1} > 0.$$

Therefore, a common-VC is (weakly) more likely to hold back the lagging project. Focusing on the inner solution with $\underline{c} < C$, the lagging project is held back with probability $\beta = F(\underline{c}) - 1 < 0$ given common ownership. □

Proof of Corollary 2 in Section 1

Proof. $\beta_A = F_A(\underline{c}) - 1 \leq F_B(\underline{c}) - 1 = \beta_B$. □

Proof of Corollary 3 in Section 1

Proof. This is because

$$\frac{\partial \beta}{\partial \delta} = f(\underline{c}) \frac{\partial \underline{c}}{\partial \delta} = -f(\underline{c}) \left(1 + \frac{p_1}{r} + \frac{p_2}{r}\right) \frac{(r + 2p_2)\, p_2 \lambda}{p_1} < 0,$$

where $f(\cdot)$ is the probability density function and the inequality is strictly negative since $f(c) > 0$. □

Discussion of Assumptions

Assumption 2 implies that $C < \bar{c}$ and $\Pr\{\text{Continue}|\text{No Common Ownership}\} = 1$. Suppose this assumption does not hold, i.e. $\bar{c} \leq C$. Then Corollary 1 still generally holds. This is because $\Delta c > 0$ and

$$\beta = F(\underline{c}) - F(\bar{c}) = -\int_{\underline{c}}^{\bar{c}} f(c)dc < 0.$$

A2.3 Micro Foundation of δ

To micro-found δ, the competition-loss parameter in our model, we extend the model to include duopoly competition with product differentiation. This extension follows the representative consumer model in Mas-Colell, Whinston, and Green (1995), pg. 399. To allow for outside competitors other than the two drugs A and B, we assume there exist $N \geq 0$ additional drugs in the market. A representative consumer has the following linear-quadratic utility function over all $N + 2$ drugs:

$$U(q) = a\sum_i q_i - \frac{1}{2}\left(b\sum_i q_i^2 + \sum_i \sum_{j\neq i} s q_i q_j\right).$$

In the above function, q_i is the quantity of drug i consumption. Parameter s, assumed to be weakly smaller than b, captures drug similarity. This utility function implies that if the consumer purchases a fixed total quantity of these drugs, she will be weakly better off with consuming more diversified drugs. In the case that $b = s$, these all drugs are perfectly substitutable, i.e., there exists no product differentiation, and competition is greatest.

The first-order condition of $U(q)$ generates the following inverse demand function:

$$p_i(q_i, q_{-i}) = a - bq_i - s\sum_{j\neq i} q_{-i}.$$

For simplicity, we assume the startups have a linear production cost function kq_i. The following standard assumption will guarantee an inner solution: $a > k > 0$.

We solve for the symmetric Nash equilibrium of a Cournot duopoly model, where each startup i maximizes its profit given the output level of the competitor q_{-i}:

$$\max_{q_i \geq 0} p_i(q_i, q_{-i})\, q_i - kq_i.$$

Each startup i' best response function is

$$b_i(q_{-i}) = \frac{a - k - s\sum_{j\neq i} q_{-i}}{2b}.$$

The symmetric equilibrium can be found by setting $b_i(q^\star) = q^\star$. We can then back out the equilibrium price $p^\star$ through the inverse demand function, which generates

$$q^\star = \frac{a-k}{2b + (N+1)s},\ p^\star = \frac{b(a+k) + (N+1)sk}{2b + (N+1)s}. \tag{A4}$$

These equilibrium values yield the flow duopoly profit d in our model

$$d = \frac{b}{(2b + (N+1)s)^2}(a-k)^2.$$

The monopoly output and price are simply given by replacing $N + 1$ with N in the above equations,

which forces the lagging startup's output to equal zero, yielding

$$q^\star = \frac{a-k}{2b+Ns},\ p^\star = \frac{b\,(a+k)+Nsk}{2b+Ns}. \tag{A5}$$

So the flow monopoly profit w is given by

$$w = \frac{b}{(2b+Ns)^2}\,(a-k)^2\,.$$

And finally the duopoly loss is

$$\delta = W - 2D = \frac{(a-k)^2\,b}{r}\left(\frac{1}{(2b+Ns)^2} - \frac{2}{(2b+(N+1)\,s)^2}\right).$$

In the above equation, we require the following equation holds to make $\delta \geq 0$:

$$\frac{s}{2b+Ns} \geq \sqrt{2}-1.$$

To understand the above equation, if N is very large, then the marginal market power received by killing a commonly-owned competitor is close to zero. In this case, the common-VC is better off with having more approved drugs and will not benefit from holding back the lagging company.

Next, we generate corollaries that establish the following statements in the paper: *δ increases in the degree of similarity between the startups' products and decreases in the number of preexisting competitors in the market.*

Corollary A2.1. *$\partial\delta/\partial s > 0$ and the common VC is more likely to hold back the lagging startup if two drugs become more similar, i.e. $\partial\beta/\partial s < 0$.*

Proof. To show $\partial\delta/\partial s > 0$,

$$\partial\delta/\partial s = \frac{(a-k)^2\,b}{r}\left(\frac{4\,(N+1)}{((N+1)\,s+2b)^3} - \frac{2N}{(Ns+2b)^3}\right),$$

which is obviously positive if $N=0$. If $N>0$, it is positive if

$$\frac{4\,(N+1)}{2N} > \left(1+\frac{s}{Ns+2b}\right)^3.$$

The RHS strictly increases with s. Setting $s=b$ generates

$$\frac{4\,(N+1)}{2N} > \left(1+\frac{1}{N+2}\right)^3,$$

which holds for all $N \geq 1$. Recall that $\partial\beta/\partial\delta < 0$, thus the second statement holds.

□

Corollary A2.2. *$\partial\delta/\partial N < 0$ and the common VC is less likely to hold back the lagging startup if there are more outside competitors when δ is non-negative, i.e. $\partial\beta/\partial N > 0$.*

Proof. To show $\partial\delta/\partial N < 0$,

$$\partial\delta/\partial N = \frac{(a-k)^2\,2sb}{r}\left(\frac{2}{(s\cdot(N+1)+2b)^3} - \frac{1}{(sN+2b)^3}\right),$$

which is negative if

$$2 < \left(1 + \frac{s}{Ns + 2b}\right)^3,$$

or equivalently

$$\frac{s}{2b + Ns} > \sqrt[3]{2} - 1.$$

This holds since $\delta \geq 0$ implies that

$$\frac{s}{2b + Ns} \geq \sqrt{2} - 1 > \sqrt[3]{2} - 1.$$

□

A2.4 Proofs of Section 4

This section proves the theoretical claims in Section 4 of the paper. In particular,

- Corollary A2.3 proves the statement, "in our model, producer surplus is unambiguously higher under common ownership than separate ownership."
- Corollary A2.4 proves the statement, "our model predicts higher innovation efficiency under common ownership than separate ownership."
- Corollary A2.5 proves the statement, "the model implies that consumer surplus is higher if two rather than one drug completes the patent race."

Corollary A2.3. *Producer surplus is always weakly higher given common ownership.*

Proof. Producer surplus in our model is equivalent to the startups' combined total value when they are both in Phase I initially. Recall that in the non-common VC case, each startup's initial value is

$$V_1 = \frac{p_1 (V_P + V_L) - c_1}{r + 2p_1}.$$

A common VC will optimally stop the lagging startup only when $V_P + V_L \leq V_M$. So put differently, the producer surplus is

$$V_1^{Common} = \frac{p_1 V_M - c_1}{r + 2p_1} \geq V_1.$$

This result also implies that entrepreneurs are willing to accept a common VC, because that VC will only hold back the lagging startup when *ex ante* this behavior increases producer surplus. □

Corollary A2.4. *Define innovation efficiency as*

$$IE = \frac{\mathbb{E}\,(\textit{Number of drugs in the market})}{\mathbb{E}\,(\textit{R\&D Cost in the market})}.$$

If $\underline{c} < c_2 < C$, then innovation efficiency is higher if both startups are commonly invested.

Proof. If startups A and B are commonly owned, then $\mathbb{E}\,(\text{Number of drugs in the market}) = 1$. In the case where these two are owned by different VCs, $\mathbb{E}\,(\text{Number of drugs in the market}) = 1 + \lambda$ since the lagging startup can possibly catch up.[2]

[2] In our model, common ownership clearly reduces total investment, but it also reduces the expected number of drugs that reach the market. Recall that without common ownership, both startups continue investing, and the expected number of drugs to reach market is $1 + \lambda$. We impose a restriction on λ such that cost improvements dominate the effect of fewer drugs reaching market. In a more general model without that parameter restriction, common ownership does not necessarily improve innovation efficiency.

We then calculate the expected costs in both cases. First consider the common VC case. We consider the case in our model where, whenever one startup enters into Phase II, the other startup is stopped. So C_2^{Common}, the present value of expected costs in this case, follows from

$$rC_2^{Common} = c_2 + p_2\left(-C_2^{Common}\right),$$

where the second term captures when the startup succeeds and all future costs become zero. This implies

$$C_2^{Common} = \frac{c_2}{r+p_2}.$$

Intuitively, this is the second term of V_M. $\mathbb{E}$(R&D Cost in the market) is simply the present value of expected future total costs when both startups are in Phase I, denoted C_1^{Common}, which follows

$$rC_1^{Common} = 2c_1 + 2p_1\left(C_2^{Common} - C_1^{Common}\right),$$

where the second term captures when one startup succeeds and future total costs become C_2^{Common}, given that the lagging startup will be killed. So

$$C_1^{Common} = \frac{2c_1}{r+2p_1} + \frac{2p_1}{r+2p_1}\frac{c_2}{r+p_2}.$$

Now consider the non-common VC case. We have to first consider the case when both startups are in Phase II, in which case the present value of expected future costs, denoted C_2^{NC}, follows

$$rC_2^{NC} = 2c_2 + 2p_2\left(\lambda L - C_2^{NC}\right).$$

Notice that the term λL captures the potential lump-sum cost due to catching-up possibility. This equation is equivalent to

$$C_2^{NC} = \frac{2c_2 + 2p_2\lambda L}{r+2p_2}.$$

Next, consider the case when the two startups are in different phases. The present values of the startups' combined expected costs, denoted C_{12}^{NC}, follows

$$rC_{12}^{NC} = c_1 + c_2 + p_1\left(C_2^{NC} - C_{12}^{NC}\right) + p_2\left(\lambda L - C_{12}^{NC}\right),$$

or equivalently

$$C_{12}^{NC} = \frac{c_1 + c_2 + p_2\lambda L}{r+p_1+p_2} + \frac{c_2 + p_2\lambda L}{r+2p_2}\frac{2p_1}{r+p_1+p_2}.$$

Finally, $\mathbb{E}$(R&D Cost in the market) in the non-common VC case is C_1^{NC}, the present value of expected costs when both startups are in Phase I, which follows from

$$rC_1^{NC} = 2c_1 + 2p_1\left(C_{12}^{NC} - C_1^{NC}\right),$$

which implies

$$C_1^{NC} = \frac{2c_1}{r+2p_1} + \frac{2p_1}{r+2p_1}\left(\frac{c_1 + c_2 + p_2\lambda L}{r+p_1+p_2} + \frac{c_2 + p_2\lambda L}{r+2p_2}\frac{2p_1}{r+p_1+p_2}\right).$$

For the corollary statement to hold, it requires

$$C_1^{NC} > (1+\lambda)\, C_1^{Common}.$$

To see why this inequality holds, it is sufficient to show

$$\begin{aligned} \frac{2c_1}{r+2p_1}\left(\lambda - \frac{p_1}{r+p_1+p_2}\right) \;\leq\; & \frac{2p_1}{r+2p_1}(1+\lambda)\left(\frac{c_2}{r+p_1+p_2} + \frac{2p_1c_2}{r+2p_2}\frac{1}{r+p_1+p_2}\right) \\ & - (1+\lambda)\frac{2p_1}{r+2p_1}\frac{c_2}{r+p_2}, \end{aligned} \tag{A6}$$

where we utilize the fact that $p_2 L \geq c_2$ in Assumption 3. The RHS of Equation (A6) is equivalent to

$$\frac{2p_1c_2}{r+2p_1}(1+\lambda)\frac{r}{r+2p_2}\frac{p_1}{r+p_1+p_2}\frac{1}{r+p_2} > 0.$$

For the LHS,

$$\lambda - \frac{p_1}{r+p_1+p_2} < 0$$

if $\lambda < p_1/(r+p_1+p_2)$, which holds by Assumption 3 since

$$\lambda < \frac{p_1}{p_1+p_2}\frac{p_2}{r+p_2} < \frac{p_1}{r+p_1+p_2}.$$

□

Corollary A2.5. *Consumer surplus is always higher when the two startups enter into duopoly competition.*

Proof. We use the extended model presented in Section A2.3. We first compute CS^D, the flow consumer surplus when the two startups enter into duopoly competition.

$$CS^D = U(q^\star) - (N+2)\,p^\star q^\star = \left(\frac{a-k}{2b+(N+1)\,s}\right)^2 \frac{N+2}{2}\left((N+1)\,s+b\right).$$

In the monopoly case,

$$CS^M = \left(\frac{a-k}{2b+Ns}\right)^2 \frac{N+1}{2}\left(Ns+b\right).$$

Note that

$$CS^D/CS^M = \left(\frac{2b+Ns}{2b+(N+1)\,s}\right)^2 \frac{N+2}{N+1}\frac{(N+1)\,s+b}{Ns+b},$$

which is strictly greater than one given $s \leq b$. □

A3 Data Appendix

A3.1 Clean Cortellis Raw Data

First, we downloaded the file "drugrecords_devstatus" from the WRDS website, on November 18, 2018. The file is located at "wrds/Clarivate Analytics/Drug Development Status". We downloaded the entire data available on that date. Table A3.1 lays out the data structure and lists variables useful in this project.

This data is unbalanced panel data. Each observation documents the development status of a company's specific drug targeting one certain indication on a status date. Therefore, we shall call a *companyid-drugid-indicationid* combination as a "project". While most variables in Table A3.1 are interpretable by the variable label, we want to show all the development statuses and their corresponding order (the number following "—"). They are "Outlicensed — 1", "No Development Reported — 3", "Discontinued — 4", "Withdrawn — 5", "Suspended — 6", "Discovery — 7", "Clinical — 8", "Phase 1 Clinical — 9", "Phase 2 Clinical — 10", "Phase 3 Clinical — 11", "Pre-registration — 12", "Registered — 13", "Launched — 14".

The development history of a successful drug usually starts from clinical trials, then walks through Phase I, II, and III, lastly registers and launches. The developing company can suspend or outlicense the project at any time, resulting in development failure. Therefore, the above order is inconsistent with a regular R&D timeline. We redefine the variable *developmentstatussortorder* as follows: "Discovery — 7", "Clinical — 8", "Phase 1 Clinical — 9", "Phase 2 Clinical — 10", "Phase 3 Clinical — 11", "Pre-registration — 12", "Registered — 13", "No Development Reported — 14", "Launched — 15", "Outlicensed — 16", "Suspended — 17", "Discontinued — 18", "Withdrawn — 19".

We now briefly explain how the file "drugrecords_devstatus" is constructed by Clarivate Analytics ("Clarivate" hereafter). Employees identify a set of pharmaceutical and biotech companies worldwide. They then search these companies' drug project updates from FDA Orange Book, clinicaltrials.gov, SEC Edgar, and news coverage, etc. Whenever they locate an event, they will add an observation in "drugrecords_devstatus" and document the development status on the even date (the variable *statusdate*). There are two specific issues worth mentioning. First, even if a project is being actively developed in some quarter, it will not show up in the file unless there are some updates reported. Second, though news related to changes in development statuses is timely documented, there are also other updates such as patent information and FDA regulatory decisions. Therefore many observations are not "updating" statuses from previous events. As a result, the raw "drugrecords_devstatus" data is very unbalanced.

We clean the raw data and transform it into balanced panel data, which we shall name the "USQtr" data. For each project, the "USQtr" data covers its development status in every quarter from the initiation quarter until the first quarter (included) in which it is either approved or suspended. We take the following steps. Table A3.2 illustrates the number of observations, companies, and drugs after each step, and lists all variables in the "USQtr" data.

Step 1. Country Location *country* indicates where the approval process takes place, and where the drug will be commercialized. Since we focus on the U.S. market and follow the institutional details of the FDA, we keep the projects whose variable *country* is either "US" or "North America". Notice this does not necessarily imply that the headquarter of the company is in the U.S.

Step 2. Fill the Gap For a project's all unique *developmentstatussortorder* value, we keep its earliest *statusdate*. In other words, we locate the first date that this project enters into different

statuses. We then extract the *year* and *quarter* information out of *statusdate*. We fill the gaps in between two different quarters by adding quarterly observations whose development statuses follow the earlier quarter's status. For the last status documented, we extend it until 2018Q4 by assuming that the status does not change. For example, if the project's last observation is in "Phase 3 Clinical — 11" in 2017Q4, then we add four more observations in 2018Q1–Q4 while keeping its status in Phase III.

Step 3. Identify the End We identify two types of projects with "obvious endings" and create two dummy variables *approved* and *suspended*. *approved* is 1 if the status is "Launched — 15" and 0 otherwise. *suspended* is 1 if the status is one of the following, "Outlicensed — 16", "Suspended — 17", "Discontinued — 18", "Withdrawn — 19", and 0 otherwise. For each project with an obvious ending, we truncate the observations in the first quarter when it is approved or suspended.

Step 4. Zombie Projects While companies are willing to disclose good news in time, they may provide no follow-up updates on the projects that cannot progress. Such projects then become "zombie projects". Clarivate identifies the first period when the update stops, turning the project into the status "No Development Reported — 14" ("ndr" hereafter). Though there is no doubt that zombie projects are in the end discarded secretly, it is unclear exactly when the suspension happens. We assume that on average, the suspension happens 4 years after the first ndr quarter. For a zombie project, we truncate the observations at the $16th$ ndr quarter and change *suspended* to 1 in that quarter.

Table A3.1: Cortellis "drugrecords_devstatus" Data Structure

Variables	Variable Label	Dataset Duration	# of Obs	# of Companies (by companyid)	# of Drugs (by drugid)
drugid	Drug ID	05/08/1942 - 10/13/2018	393,582	15,233	70,689
drugname	Drug Name				
companyname	Company Name				
companyid	Company ID				
country	Country				
developmentstatus	Development Status				
developmentstatussortorder	Development Status Order				
indication	Indication				
indicationid	Indication ID				
statusdate	Status Date				

Table A3.2: "USQtr" Data Construction and Variables

Panel A: Panel Construction Step	# of Obs	# of Companies	# of Drugs
1. Country Location	152,989	8,193	38,322
2. Fill the Gap	3,732,112		
3. Identify the End	2,739,506		
4. Zombie Projects	2,089,626		

Panel B: Variables in the "USQtr" data Variable	Variable Label
drugid	Drug ID
drugname	Drug Name
companyname	Company Name
companyid	Company ID
country	Country
developmentstatus	Development Status
developmentstatussortorder	Development Status Order
indication	Indication
indicationid	Indication ID
year	Year
quarter	Quarter
suspended	Suspension Indicator
approved	Approval Indicator

A3.2 VC Holding Data

This section introduces the VC investment data as well as how we match a VC-invested company to a *companyid* in Cortellis.

We first downloaded a full list of VC deals (hereafter "VCall" file) from the SDC Platinum VentureXpert, requiring the investment time between 01/01/1980 and 12/31/2017, portfolio company's location in the U.S., and type of the deal to be venture investment. This generates a sample of 370,991 deal records. Table A3.3 reports the key variables we used in our paper.

We assign an id, denoted by *companyid_vc*, to each unique combination of *company_name*, *company_county*, *company_add* and *company_zip* in the "VCall" file. We then downloaded the "company_basic" file from "wrds/Clarivate Analytics", at the same time with "drugrecords_devstatus". This file contains basic information on healthcare companies, including company name, headquarter address, company website, etc. This file also uses the same *companyid* with the with "drugrecords_devstatus" file.

We use the customized version of Stata command "stn_compname" to standardize company names in both the "VCall" file and the "company_basic" file. We then exploit the Stata command "reclink2" to fuzzy-match standardized names. Both commands are developed by Wasi (2015). We manually check each matched pair by confirming information on company state, city, county, street address, and ZIP codes. This generates 6,177 matched pairs of $companyid\_vc - companyid$ links.

We exploit Jay Ritter's IPO database and SDC M&A dataset to determine each portfolio company's IPO or acquisition date. As before, we first fuzzy-match the standardized company names and then manually check each candidate pair on other complementary information. Among the 6,177 pairs, we recover IPO dates for 933 companies and M&A exit dates for 1,945 companies.

For each *companyid_vc*, we take the following steps to construct VC holding records, which we shall name the "Biovchold" data. Consistent with the "USQtr" data, we also keep quarterly observations. Table A3.4 summarizes the number of observations, portfolio companies, and VC firms after each step, and lists all variables in the "Biovchold" data.

Step 1. Matched Pair For all portfolio companies in the "VCall" file, we only keep those whose *companyid_vc* is successfully linked to a *companyid* in Cortellis. For each remaining company, we sort the investment records by the variable *round_num*.

Step 2. Determine Exit Date If we have recovered its IPO or M&A date for a portfolio company, then we use this date as the exit date for this company. If we have not, we follow the literature and assume that a portfolio company is written off from a VC's balance sheet if it does not receive any fundings 5 years after the most recent round. In this case, we assume a VC holds the company in its portfolio until the written-off date. This projected date is truncated by 12/31/2018.

Step 3. Expand into Quarterly We expand the dataset into quarterly observations. For each VC-company pair, the observation starts in the first quarter in which this VC participates in a syndicate. The observation ends at the quarter of the exit date defined in Step 2.

Step 4. Round Information In each quarterly observation of a VC-company pair, we document the following information of the most recent round: round date (*round_date*), number of investors in the round (*round_num_inv*) and round amount (*round_amount*).

Table A3.3: SDC Platinum VentureXpert Deal Data Structure

Variable	Variable Label
deal_num	Deal Number
disbid	Disbursement ID
round_date	Round Date
round_num	Round Number
round_num_inv	Number of Investors in a Round
round_amount	Disclosed Round Amount ($ Thousand)
round_est_amount	Estimated Round Amount ($ Thousand)
company_name	Portfolio Company Name
company_ipo	Whether Company Had IPO
company_ipodate	Company IPO Date if Applicable
company_cusip	Company 6-digit CUSIP Number
company_state	Company State Code
company_city	Company City
company_county	Company County
company_msa	Company MSA Code
company_add	Company Street Address
company_foundyear	Company Founded Year
company_sic	Company Primary SIC Code
company_stage	Company Stage Level 1 at Each Round Date
company_val	Disclosed Post-Round Company Valuation ($ Thousand)
firm_name	VC Firm Name
firm_state	Firm State Code
firm_city	Firm County
firm_msa	Firm MSA Code
firm_nation	Firm Nation Code
firm_foundyear	Firm Founded Year
firm_geopref	Firm Geography Preference
firm_indpref	Firm Industry Preference
firm_rolepref	Firm Preferred Investment Role (Code)
firm_stagepref	Firm Investment Stage Preference
firm_type	Firm Type
firm_zip	Firm ZIP Code
fund_state	Fund State Code Operated by VC Firm
fund_city	Fund City
fund_county	Fund County
fund_msa	Fund MSA Code
fund_nation	Fund Nation Code
fund_name	Fund Name
fund_size	Fund Size ($ Million)
fund_status	Fund Raising Status
fund_stage	Fund Stage
fund_year	Fund Year
fund_zip	Fund ZIP Code

Table A3.4: "Biovchold" Data Construction and Variables

Panel A: Panel Construction

Step	# of Obs	# of Companies	# of VC firms
1. Matched Pair	81,783	6,177	4,209
2. Determine Exit Date	81,783		
3. Expand into Quarterly	782,012		
4. Round Information	782,012		

Panel B: Variables in the "Biovchold" data

Variable	Variable Label
firmid	VC Firm Identifier
firm_name	Name of the VC Firm
companyid	Company ID from Clarivate
first_round_date	First Investment Date of the VC
year	Year
qtr	Quarter
companyname	Company Name
companyid_vc	Author-Created Company ID in the VC Data
IPO_date	IPO Exit Date, if any
M&A_date	M&A Exit Date, if any
round_date	Round Date (Most Recent)
round_num	Round Number (Most Recent)
round_num_inv	Number of Investors in a Round (Most Recent)
round_amount	Disclosed Round Amount ($ Thousand, Most Recent)

A3.3 The Phase I Sample

In this section, we explain how to generate the sample of Phase I projects used in our analysis. The "USQtr" data contains over 2 million observations, including foreign companies outside the U.S., non-startup companies, and development history beyond Phase I. We take the following steps to construct the "PhaseI" data. Table A3.5 illustrates the number of observations, companies, and drugs after each step, and lists all variables in the "PhaseI" data.

Step 1. U.S. Startups We delete companies not covered in the VentureXpert database. The remaining companies are startups located in the U.S.

Step 2. ICD-9 We cannot use the indications from Cortellis to define competing projects, because indication names are not standardized and are sometimes too specific or too vague. Instead, we manually match Cortellis indications to the second chapter level of International Classification of Diseases (9^{th} Revision, "ICD"), and we refer to each chapter as an ICD category. We generate the variable *icd_id* to save the ICD category id for each project. Then we drop projects that (*i*) cannot be accurately assigned to a category, or (*ii*) belong to a 3-digit ICD code that contains the keyword "unspecified." See Table A3.6 for the complete list of ICD codes included.

Step 3. Phase I We first drop projects that were initiated before 2005Q1, due to concerns of missing data. Then we delete the observations earlier than the first quarter that the project's status is in "Phase 1 Clinical — 9". The number of observations for each project varies in the following three cases:

- If the project ever progresses to Phase II, then we keep all observations until the first quarter that the project's status is in "Phase 2 Clinical — 10" (included). We generate a dummy variable *progress*, which is 1 if the project jumps to Phase II at a certain quarter, and 0 otherwise.
- If the project is suspended before entering into Phase II, then we keep all observations until the quarter that the project is suspended, i.e., the variable *suspended* is 1. Notice this includes suspensions of zombie projects.
- If the project never progresses to Phase II and is not suspended, we keep the observations until 2018Q4. These projects tend to be recently initiated, so we drop those initiated after the first quarter of 2016. We do so because insufficient time has passed for these projects to reach any outcome.

Step 4. Project ID We generate a variable *projectid* by assigning a unique id to each *companyid-drugid-indicationid* combination. Lastly, we delete some variables that will never be used again.

Step 3′. Phase II We first drop projects that were initiated before 2005Q1. Then we delete the observations earlier than the first quarter that the project's status is in "Phase 2 Clinical — 10". The number of observations for each project varies in the following three cases:

- If the project ever progresses to Phase III, then we keep observations all until the first quarter that the project's status is in "Phase 3 Clinical — 11" (included). We generate a dummy variable *progress*, which is 1 if the project jumps to Phase III at a certain quarter, and 0 otherwise.

- If the project is suspended before entering into Phase III, then we keep observations all until the quarter that the project is suspended, i.e., the variable *suspended* is 1.
- If the project never progresses to Phase III and is not suspended, we keep the observations until 2018Q4 and drop those initiated after the first quarter of 2016.

Table A3.5: "PhaseI" Data Construction and Variables

Panel A: Panel Construction			
Step	# of Obs	# of Companies	# of Drugs
1. U.S. Startups	642,140	1,671	12,306
2. ICD9	604,871	1,659	11,825
3. Phase I	31,749	619	1,499

Panel B: Variables in the "PhaseI" data	
Variable	Variable Label
projectid	Project ID
drugid	Drug ID
drugname	Drug Name
companyname	Company Name
companyid	Company ID
icd_id	ICD Category ID
indication	Indication
indicationid	Indication ID
year	Year
quarter	Quarter
suspended	Suspension Indicator
progress	Progressing Indicator

Table A3.6: List of Sample ICD Groups and Codes

This table lists the details of the 78 ICD categories included in our sample. In the first column, an ICD category is the second chapter level of the International Classification of Diseases, 9th revision. For each category, we list all 3-digit ICD codes that belong to this category in the parenthesis. In the second column, we list the 3-digit ICD codes that belong to this category and exist in our sample. Some 3-digit ICD codes are missing either because they are not developed by our sample firms, or because their names contain "unspecified," and therefore dropped.

ICD Category Name	**List of 3-digit ICD Codes Included**
Intestinal infectious diseases (001 - 009)	(1) Cholera
Tuberculosis (010 - 018)	(11) Pulmonary tuberculosis
Zoonotic bacterial diseases (020 - 027)	(22) Anthrax
Human immunodeficiency virus (HIV) infection (042 - 044)	(42) Human immunodeficiency virus (hiv) disease
Poliomyelitis and other non-arthropod-borne viral diseases of central nervous system (045 - 049)	(46) Slow virus infection and prion diseases of central nervous system
Viral diseases accompanied by exanthem (050 - 059)	(54) Herpes simplex
Arthropod-borne viral diseases (060 - 066)	(60) Yellow fever, (61) Dengue, (66) Other arthropod-borne viral diseases
Other diseases due to viruses and chlamydiae (070 - 079)	(70) Viral hepatitis, (78) Other diseases due to viruses and chlamydiae
Mycoses (110 - 118)	(110) Dermatophytosis, (112) Candidiasis, (117) Other mycoses
Malignant neoplasm of digestive organs and peritoneum (150 - 159)	(151) Malignant neoplasm of stomach, (153) Malignant neoplasm of colon, (154) Malignant neoplasm of rectum rectosigmoid junction and anus, (155) Malignant neoplasm of liver and intrahepatic bile ducts, (157) Malignant neoplasm of pancreas, (158) Malignant neoplasm of retroperitoneum and peritoneum
Malignant neoplasm of respiratory and intrathoracic organs (160 - 165)	(162) Malignant neoplasm of trachea bronchus and lung, (163) Malignant neoplasm of pleura
Malignant neoplasm of bone, connective tissue, skin, and breast (170 - 175)	(170) Malignant neoplasm of bone and articular cartilage, (171) Malignant neoplasm of connective and other soft tissue, (172) Malignant melanoma of skin, (173) Other malignant neoplasm of skin, (174) Malignant neoplasm of female breast
Kaposi's sarcoma (176 - 176)	(176) Kaposi's Sarcoma
Malignant neoplasm of genitourinary organs (179 - 189)	(183) Malignant neoplasm of ovary and other uterine adnexa, (185) Malignant neoplasm of prostate, (188) Malignant neoplasm of bladder
Malignant neoplasm of other and unspecified sites (190 - 199)	(190) Malignant neoplasm of eye, (191) Malignant neoplasm of brain, (193) Malignant neoplasm of thyroid gland, (194) Malignant neoplasm of other endocrine glands and related structures
Malignant neoplasm of lymphatic and hematopoietic tissue (200 - 208)	(200) Lymphosarcoma and reticulosarcoma, (201) Hodgkin's disease, (202) Other malignant neoplasms of lymphoid and histiocytic tissue, (203) Multiple myeloma and immunoproliferative neoplasms, (204) Lymphoid leukemia, (205) Myeloid leukemia
Neuroendocrine tumors (209 - 209)	(209) Neuroendocrine tumors
Diseases of other endocrine glands (249 - 259)	(250) Diabetes mellitus, (251) Other disorders of pancreatic internal secretion, (253) Disorders of the pituitary gland and its hypothalamic control, (255) Disorders of adrenal glands, (257) Testicular dysfunction, (259) Other endocrine disorders
Nutritional deficiencies (260 - 269)	(264) Vitamin a deficiency
Other metabolic and immunity disorders (270 - 279)	(270) Disorders of amino-acid transport and metabolism, (271) Disorders of carbohydrate transport and metabolism, (272) Disorders of lipoid metabolism, (273) Disorders of plasma protein metabolism, (274) Gout, (275) Disorders of mineral metabolism, (276) Disorders of fluid electrolyte and acid-base balance, (278) Overweight, obesity and other hyperalimentation, (279) Disorders involving the immune mechanism
Anemia (280 - 285)	(282) Hereditary hemolytic anemias, (283) Acquired hemolytic anemias, (284) Aplastic anemia

Continued.

ICD Group Name	List of 3-digit ICD Codes Included
Coagulation/hemorrhagic (286 - 287)	(287) Purpura and other hemorrhagic conditions
Other diseases of blood and blood-forming organs (288 - 289)	(288) Diseases of white blood cells, (289) Other diseases of blood and blood-forming organs
Organic psychotic conditions (290-294)	(290) Dementias, (294) Persistent mental disorders
Other psychoses (295-299)	(295) Schizophrenic disorders, (296) Episodic mood disorders
Neurotic disorders, personality disorders, and other nonpsychotic mental disorders (300 - 316)	(300) Anxiety, dissociative and somatoform disorders, (303) Alcohol dependence syndrome, (304) Drug dependence, (305) Nondependent abuse of drugs, (307) Special symptoms or syndromes not elsewhere classified, (311) Depressive disorder, (314) Hyperkinetic syndrome of childhood
Inflammatory diseases of the central nervous system (320-326)	(321) Meningitis due to other organisms
Hereditary and degenerative diseases of the central nervous system (330-337)	(331) Other cerebral degenerations [including Alzheimer's disease], (332) Parkinson's disease, (333) Other extrapyramidal disease and abnormal movement disorders [including Huntington's chorea], (334) Spinocerebellar disease, (335) Anterior horn cell disease
Other disorders of the central nervous system (340-349)	(340) Multiple sclerosis, (344) Other paralytic syndromes, (345) Epilepsy, (346) Migraine, (348) Other conditions of brain
Disorder of the peripheral nervous system (350-359)	(358) Myoneural disorders, (359) Muscular dystrophies and other myopathies
Disorders of the eye and adnexa (360-379)	(362) Other retinal disorders, (364) Disorders of iris and ciliary body, (365) Glaucoma, (370) Keratitis, (372) Disorders of conjunctiva, (373) Inflammation of eyelids, (379) Other disorders of eye
Diseases of the ear and mastoid process (380-389)	(382) Suppurative otitis media, (386) Vertiginous syndromes and other disorders of vestibular system, (388) Other disorders of ear
Hypertensive disease (401 - 405)	(401) Essential hypertension
Ischemic heart disease (410 - 414)	(410) Acute myocardial infarction, (414) Other forms of chronic ischemic heart disease
Diseases of pulmonary circulation (415 - 417)	(416) Chronic pulmonary heart disease
Other forms of heart disease (420 - 429)	(425) Cardiomyopathy, (427) Cardiac dysrhythmias, (428) Heart failure
Cerebrovascular disease (430 - 438)	(434) Occlusion of cerebral arteries, (435) Transient cerebral ischemia
Diseases of arteries, arterioles, and capillaries (440 - 449)	(443) Other peripheral vascular disease
Diseases of veins and lymphatics, and other diseases of circulatory system (451 - 459)	(453) Portal vein thrombosis
Acute respiratory infections (460 - 466)	(466) Acute bronchitis and bronchiolitis
Other diseases of the upper respiratory tract (470 - 478)	(471) Nasal polyps, (477) Allergic rhinitis
Pneumonia and influenza (480 - 488)	(481) Pneumococcal pneumonia, (482) Other bacterial pneumonia, (487) Influenza
Chronic obstructive pulmonary disease and allied conditions (490 - 496)	(491) Chronic bronchitis, (493) Asthma
Other diseases of respiratory system (510 - 519)	(516) Other alveolar and parietoalveolar pneumonopathy
Diseases of oral cavity, salivary glands, and jaws (520 - 529)	(523) Gingival and periodontal diseases, (528) Diseases of the oral soft tissues excluding lesions specific for gingiva and tongue
Diseases of esophagus, stomach, and duodenum (530 - 539)	(530) Diseases of esophagus, (536) Disorders of function of stomach
Noninfectious enteritis and colitis (555 - 558)	(555) Regional enteritis, (556) Ulcerative enterocolitis
Other diseases of intestines and peritoneum (560 - 569)	(564) Functional digestive disorders
Other diseases of digestive system (570 - 579)	(571) Chronic liver disease and cirrhosis, (572) Liver abscess and sequelae of chronic liver disease, (576) Other disorders of biliary tract, (577) Diseases of pancreas, (579) Intestinal malabsorption
Nephritis, nephrotic syndrome, and nephrosis (580 - 589)	(580) Acute glomerulonephritis, (581) Nephrotic syndrome, (583) Nephritis and nephropathy not specified as acute or chronic, (585) Chronic kidney disease (ckd)

Continued.

ICD Group Name	List of 3-digit ICD Codes Included
Other diseases of urinary system (590 - 599)	(595) Cystitis, (596) Other disorders of bladder, (597) Urethritis not sexually transmitted and urethral syndrome, (599) Other disorders of urethra and urinary tract
Diseases of male genital organs (600 - 608)	(600) Hyperplasia of prostate
Inflammatory Disease Of Female Pelvic Organs (614 - 616)	(614) Inflammatory disease of ovary fallopian tube pelvic cellular tissue and peritoneum
Other disorders of female genital tract (617 - 629)	(622) Noninflammatory disorders of cervix, (625) Pain and other symptoms associated with female genital organs, (627) Menopausal and postmenopausal disorders
Complications mainly related to pregnancy (640 - 649)	(642) Hypertension complicating pregnancy childbirth and the puerperium
Other inflammatory conditions of skin and subcutaneous tissue (690 - 698)	(691) Atopic dermatitis and related conditions, (692) Contact dermatitis and other eczema, (695) Erythematous conditions, (696) Psoriasis and similar disorders, (698) Pruritus and related conditions
Other diseases of skin and subcutaneous tissue (700 - 709)	(701) Other hypertrophic and atrophic conditions of skin, (704) Diseases of hair and hair follicles, (706) Diseases of sebaceous glands, (709) Other disorders of skin and subcutaneous tissue
Arthropathies and related disorders (710 - 719)	(710) Diffuse diseases of connective tissue, (714) Rheumatoid arthritis and other inflammatory polyarthropathies, (715) Osteoarthrosis and allied disorders
Dorsopathies (720 - 724)	(720) Ankylosing spondylitis and other inflammatory spondylopathies
Rheumatism, excluding the back (725 - 729)	(728) Disorders of muscle ligament and fascia, (729) Other disorders of soft tissues
Osteopathies, chondropathies, and acquired musculoskeletal deformities (730 - 739)	(733) Other disorders of bone and cartilage, (738) Other acquired musculoskeletal deformity
Nervous system (740-742)	(741) Spina bifida
Urinary system (753-753)	(753) Congenital anomalies of urinary system
Integument (757-757)	(757) Congenital anomalies of the integument
Symptoms (780 - 789)	(780) General symptoms, (783) Symptoms concerning nutrition metabolism and development, (785) Symptoms involving cardiovascular system, (786) Symptoms involving respiratory system and other chest symptoms, (787) Symptoms involving digestive system, (788) Symptoms involving urinary system
Sprains and strains of joints and adjacent muscles (840 - 848)	(848) Other and ill-defined sprains and strains
Internal injury of thorax, abdomen, and pelvis (860 - 869)	(864) Injury to liver
Contusion with intact skin surface (920 - 924)	(921) Contusion of eye and adnexa
Poisoning by drugs, medicinal and biological substances (960 - 979)	(964) Poisoning by agents primarily affecting blood constituents
Toxic effects of substances chiefly nonmedicinal as to source (980 - 989)	(989) Toxic effect of other substances chiefly nonmedicinal as to source
Other and unspecified effects of external causes (990 - 995)	(995) Certain adverse effects not elsewhere classified
Complications of surgical and medical care, not elsewhere classified (996 - 999)	(996) Complications peculiar to certain specified procedures
V codes - Supplementary classification of factors influencing health status and contact with health services	(V04) Need for prophylactic vaccination and inoculation against certain viral diseases, (V09) Infection with drug-resistant microorganisms, (V25) Encounter for contraceptive management, (V42) Organ or tissue replaced by transplant

A3.4 Merge the Phase I Sample with the VC Holding Data

We merge the "PhaseI" data with the "Biovchold" Data by matching *companyid*, *year*, and *quarter*. We drop observations with VC $firm\_name$ equal to "Undisclosed Firm". The matched sample, which we shall call the "VCPhaseI" data, has 57,316 observations, 481 unique startup companies, 771 drugs, and 764 VC firms. Each observation in the matched sample is a $firmid$-*companyid-projectid* combination in quarter t.

The "VCPhaseI" data can be illustrated with a tree structure as Figure A3.1. In a given quarter t, we document all VC firms investing in the biotech industry. For each VC, we document all companies in its portfolio. For each company, we document all active Phase I projects. Each project then becomes an observation, documenting its developing *companyid*, the investing $firmid$, and project information including *icd_id*, *suspended*, *progress*, etc. Notice that one project can appear in multiple observations in a single quarter, as "Project3" does in Figure A3.1. This is because multiple VCs can invest in its developing company at the same time.

The Project Regression

In our project-level regressions, each quarterly observation contains information from a unique project. Therefore, we have to collapse *projectid* duplicates due to multiple investing VCs in the "VCPhaseI" Data. We briefly walk through how to construct $Lagging \times SharedVC$ and $Lagging$ in this case. First, we construct a variable icd_shock_{it}, which is 1 for project i if there exists at least one Phase I project with a *different projectid*, developed by a *companyid* different from project i's but sharing the same *icd_id* with project i, progressing to Phase II at t, and 0 otherwise. Put it simply, icd_shock_{it} indicates that there is at least one project i's competing drug progressing to Phase II at t. Next, we generate a variable $sv\_icd\_shock_{it}$, indicating that the progressing competing drug shares at least one VC investor with project i. Specifically, denote V_{it} as the set of unique $firmid$ that are associated with project i in quarter t. Conditional on $icd\_shock_{it} = 1$, $sv\_icd\_shock_{it}$ is 1 for project i if there exists at least one progressing competing project j such that $V_{jt} \cap V_{it} \neq \emptyset$, and 0 otherwise. After we construct icd_shock_{it}, $sv\_icd\_shock_{it}$ and other control variables, we keep only one observation for each project i in quarter t. Lastly, $Lagging_{it}$ is 1 if there exists a quarter $s < t$ such that $icd\_shock_{is} = 1$, and 0 otherwise. $Lagging \times SharedVC_{it}$ is 1 if there exists a quarter $s < t$ such that $sv\_icd\_shock_{is} = 1$, and 0 otherwise.

The VC-Funding Regressions

In our VC-funding regressions, each quarterly observation is a unique *firmid-companyid* combination, indicating the investment relationship between them two. However, there are duplicates of observations for each *firmid-companyid* in the "VCPhaseI" Data because one company can develop multiple projects, for example $VC1 - Com1$ in Figure A3.1. Again, we briefly walk through how to construct $Lagging \times SharedVC$ and $Lagging$ in this case. First, we construct a variable icd_shock_{jt}, which is 1 for company j if there exists at least one company with a *different companyid*, which has at least one Phase I project sharing the same *icd_id* with company j's products and progressing to Phase II at t, and 0 otherwise. Put it simply, icd_shock_{jt} indicates there is at least one of company j's competitors having a project progressing to Phase II at t. Next, we generate a variable $sv\_icd\_shock_{jkt}$, indicating that the progressing competitor shares VC k with company j. Specifically, conditional on $icd\_shock_{jt} = 1$, $sv\_icd\_shock_{jkt}$ is 1 for the "company j-VC k" combination if the progressing competing company

is also invested by VC k, and and 0 otherwise. After we construct icd_shock_{jt}, $sv\_icd\_shock_{jkt}$ and other control variables, we keep only one observation for each VC k's portfolio company j in quarter t. Lastly, $Lagging_{jkt}$ is 1 if there exists a quarter $s < t$ such that $icd\_shock_{js} = 1$, and 0 otherwise. $Lagging \times SharedVC_{jkt}$ is 1 if there exists a quarter $s < t$ such that $sv\_icd\_shock_{jks} = 1$, and 0 otherwise.

We do not go through detailed constructions of all other variables involved in the above regressions, as we believe the variable label is self-explaining. Instead, we list all variables in Table A3.7.

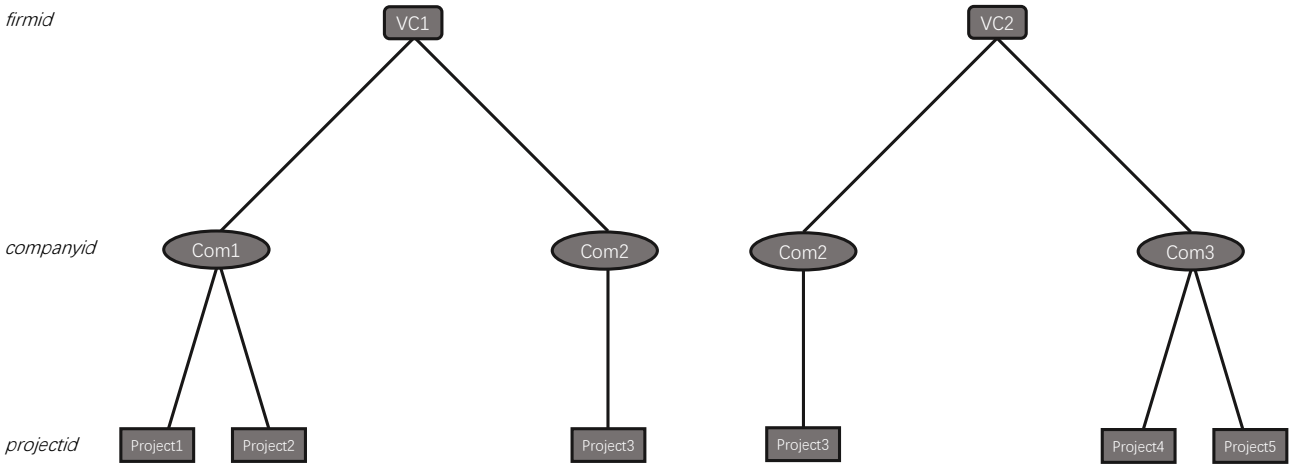

Figure A3.1: "VCPhaseI" Data Structure

Table A3.7: Variables in the Project and VC Funding Regression

Panel A: Variables in the Project Regression Variable	Variable Label
projectid	Project ID
companyid	Company ID
icd_id	ICD Category ID
year	Year
quarter	Quarter
suspended	Suspension Indicator
progress	Progressing Indicator
icd_shock	1 if a competing product progresses in a given quarter
sv_icd_shock	1 if a VC-sharing competing product progresses in a given quarter
Lagging	1 if a competing product has progressed before a given quarter
Lagging×SharedVC	1 if a VC-sharing competing product has progressed before a given quarter
Age	log(1+quarters since initation)
NProjects	Number of Phase I projects being developed in project's i's company
NVCs	Number of VC firms investing in project i's developing company
Panel B: Variables in the Funding Regression	
firmid	VC Firm ID
companyid	Company ID
year	Year
quarter	Quarter
ExtendFunds	1 if this VC invests in this company in a given quarter
icd_shock	1 if a competor's project progresses in a given quarter
sv_icd_shock	1 if the progressing competitor is also invested by this VC firm
Lagging	1 if a competor's project has progressed before a given quarter
Lagging×SharedVC	1 if for a competitor in this VC firm's portoflio, its competing project has progressed before a given quarter
SelfProgress	1 if this company has progressed a Phase I drug before a given quarter
NCats	Number of ICD categories invested by this company
Duration	log(1+Years since previous funding from this VC)
NProjects	Number of Phase I projects being developed by this company
NVCs	Number of VC firms investing in this company
PortfolioSize	Number of biotech startups in this VC's portfolio
PrevRoundSize	log(funding amount in the previous deal from any (syndicates of) VC)

A3.5 Other Data in Cortellis

For our instrumental variables, we extract the company's headquarter address from the variable *hqaddress* in the file "company_basic", which contains the zip code information. Using the "Zip Code Tabulation Area (ZCTA) Relationship Files" on census.gov, we assign each company to a metropolitan statistical area. We then cross-verify this information with the *company_msa* variable from the VentureXpert data.

For technology similarity, we extract patent information from the variable $patentfamilynumber$ and $patentfamilytitle$ in the file "drugrecords_patentfamilies". The former contains the United States Patent Application number, and the latter contains the patent title. Using these two variables, we create a scraping algorithm on Google Patent to extract patent citation information.

A4 Additional Results Supporting Subsection 2.2

Table A4.1: Compare Treated and Control Groups Conditional on $Lagging = 1$

This table summarizes the characteristics of lagging projects in the quarter before the pioneering projects progress. We split the lagging projects in the first four columns by whether they share VCs with the pioneering projects. In the last four columns, we split the lagging projects by their distances to the pioneering project. The *Close* (*far*) group are those with distances below (above) the median. *ProjectAge* is the number of quarters since project initiation. *CompanyAge* is the number of quarters since the company founding quarter. *ComProgressNum* is the total number of projects ever progressing from Phase I to Phase II by a given startup. *ComLateProgressNum* is the total number of projects ever progressing from Phase II to Phase III by a given startup. *SharedCite* is 1 if the lagging project shares patent citation with the pioneering project and 0 otherwise. *Boston* (*SanFrancisco*) is 1 if the project is in Boston (San Francisco) and 0 otherwise. *Distance* is the distance between the lagging and pioneering projects (in miles). *Proximity* is our instrumental variable. $t-statistic$ of the difference is in the parenthesis.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Mean				**Mean**			
	$ShareVC = 0$	$ShareVC = 1$	*Difference*	*P-Value*	*Close*	*Far*	*Difference*	*P-Value*
ProjectAge	8.716	9.957	-1.241	0.129	9.158	8.614	0.543	0.339
			(-1.5)				(0.95)	
CompanyAge	11.41	12.446	-1.036	0.301	11.118	12.003	-0.885	0.204
			(-1.05)				(-1.25)	
ComProgressNum	0.517	0.413	0.104	0.466	0.458	0.550	-0.093	0.351
			-0.75				(-0.95)	
ComLateProgressNum	0.048	0.087	-0.04	0.142	0.049	0.058	-0.009	0.605
			(-1.45)				(-0.5)	
SharedCite	0.072	0.087	-0.015	0.624	0.067	0.083	-0.016	0.439
			(-0.5)				(-0.75)	
Boston	0.178	0.174	0.005	0.916	0.185	0.172	0.013	0.649
			(-0.1)				(0.45)	
SanFrancisco	0.152	0.196	-0.044	0.287	0.137	0.180	-0.044	0.122
			(-1.05)				(-1.55)	
Distance	959.744	759.5	200.244***	0.009	300.178	1569.023	-1268.846***	0.000
			(-2.6)				(-63.95)	
Proximity	0.622	0.736	-0.114***	0.003	0.904	0.370	0.533***	0.000
			(-3.05)				(34.35)	

Table A4.2: Compare Investor Characteristics Conditional on $Lagging = 1$

This table compares characteristics of VC investors for the lagging projects in the quarter before the pioneering projects progress. We sort these projects into three groups, $Close$, $Medium$, and Far, based on their distance to the leading projects. In the event that a project has multiple VC investors when it became lagged, we calculate the average characteristics across the investors (Panel A) or only keep the lead VC (Panel B). *PortSize* is the number of startup companies in VC i's portfolio at quarter t. *VCAge* is the number of quarters since VC firm founded. *ActivePortSize* is the number of unique startups invested by VC j in the past 5 years. *AUM (million)* is asset under management of VC i. *EarlyStage* is a dummy equal to one if VC i prefers investing in early stage startups. *HighDiversify* is a dummy equal to one if the average active portfolio size of VC i is above the sample median (19.6) as defined in Table 6. *Ncats* is the number of ICD areas that VC i ever invested. *SuccessRate* is the ratio of successful exits (IPOs or M&As) among VC i's investments. $t - statistic$ of the difference is in the parenthesis.

Panel A: Average VC Characteristics

	(1)	(2)	(3)	(5)	(6)
		Mean			
	$Close$	$Medium$	Far	$Close - Far$	$P\text{-}Value$
PortSize	3.846	3.947	4.224	-0.378	0.300
				(-1.038)	
VCAge	78.723	74.374	75.174	3.549	0.506
				(0.666)	
ActivePortSize	34.780	28.732	32.469	2.311	0.467
				(0.729)	
AUM (million)	542.476	461.534	467.623	74.853	0.331
				(0.976)	
EarlyStage	0.304	0.246	0.254	0.049	0.201
				(1.284)	
HighDiversify	0.543	0.474	0.524	0.018	0.690
				(0.399)	
Ncats	4.819	4.918	5.054	-0.235	0.583
				(-0.550)	
SuccessRate	0.220	0.237	0.236	-0.016	0.333
				(-0.970)	

Panel B: Lead VC Characteristics

	(1)	(2)	(3)	(5)	(6)
		Mean			
	Close	*Medium*	*Far*	*Close − Far*	*P-Value*
PortSize	4.971	4.651	5.631	-0.659	0.316
				(-1.007)	
VCAge	78.856	75.848	74.573	4.283	0.535
				(0.622)	
ActivePortSize	48.740	40.524	48.291	0.449	0.927
				(0.092)	
AUM (million)	693.883	433.768	673.793	20.090	0.885
				(0.145)	
EarlyStage	0.305	0.311	0.252	0.052	0.403
				(0.839)	
HighDiversify	0.714	0.547	0.689	0.025	0.696
				(0.392)	
Ncats	5.695	5.377	6.262	-0.567	0.405
				(-0.834)	
SuccessRate	0.218	0.233	0.232	-0.015	0.391
				(0.859)	

Table A4.3: Examining the Exclusion Condition

This table reports the project-level OLS regression results on a subsample with $CommonOwn = 0$ for all projects, where $CommonOwn$ is a dummy equal to one if project i shares a VC in common with any other projects in the same ICD in quarter t. The dependent variable is $Progress$ dummy, and the key independent variable is $Lagging \times Proximity$. All other details are the same as in Table 3.

	$Progress$
$Lagging \times Proximity$	0.001
	(0.30)
$Lagging$	-0.002
	(-0.37)
$\ln(Age)$	0.023***
	(3.90)
$NProjects$	0.001
	(0.75)
$NVCs$	0.008**
	(2.55)
$NProjectsperICD$	-0.064***
	(-3.45)
Startup FE	Yes
Yr-Qtr FE	Yes
ICD FE	Yes
N	6,242
Adj. R^2	0.116

Table A4.4: Compare Compliers and Noncompliers

This table compares characteristics of complier-VCs and noncomplier-VCs in a subsample of 120 common VC investors which invest in both the leading and lagging projects. We sort these VCs into three groups, *Close*, *Medium*, and *Far*, based on their distance to the leading projects. Compliers are defined as the *Close* subgroup while noncompliers are defined as the *Far* subgroup. *PortSize* is the number of startup companies in VC i's portfolio at quarter t. *VCAge* is the number of quarters since VC firm founded. *ActivePortSize* is the number of unique startups invested by VC j in the past 5 years. *AUM (million)* is asset under management of VC i. *EarlyStage* is a dummy equal to one if VC i prefers investing in early stage startups. *HighDiversify* is a dummy equal to one if the average active portfolio size of VC i is above the sample median (19.6) as defined in Table 6. *Ncats* is the number of ICD areas that VC i ever invested. *SuccessRate* is the ratio of successful exits (IPOs or M&As) among VC i's investments. *LeadVC* is a dummy equal to one if VC i prefers being leading VC in investments. $t-statistic$ of the difference is in the parenthesis.

	(1)	(2)	(3)	(5)	(6)
		Mean			
	Close	*Medium*	*Far*	*Close − Far*	*P-Value*
PortSize	7.525	8.800	9.925	-2.400**	0.043
				(2.06)	
VCAge	72.800	88.700	86.750	-13.950	0.122
				(1.56)	
ActivePortSize	50.650	55.200	63.400	-12.750*	0.079
				(1.78)	
AUM (million)	780.564	895.864	1136.562	-355.997	0.250
				(1.18)	
EarlyStage	0.250	0.125	0.125	0.125	0.156
				(1.433)	
HighDiversify	0.750	0.825	0.900	-0.150*	0.080
				(1.78)	
Ncats	8.825	9.925	11.075	-2.250	0.109
				(1.62)	
SuccessRate	0.288	0.300	0.295	-0.007	0.672
				(0.43)	
LeadVC	0.425	0.225	0.400	0.025	0.823
				(0.22)	

Table A4.5: Collection of Results for the Probit Stage prior to 2SLS in Table 3, A6.1, and 4

This table collects the probit stage results prior to 2SLS in the main text. Columns 1 to 3 reports the probit stage results for Table 3, A6.1, and 4, respectively. All other details are the same as in those tables.

	Probit prior to 2SLS		
	(1)	(2)	(3)
	$Lagging \times SharedVC$	$Lagging \times SharedVC$	$Lagging \times SharedVC$
$Lagging \times Proximity$	0.447***	0.308***	0.269***
	(7.17)	(3.45)	(8.83)
$\ln(Age)$	0.208**	0.224**	
	(2.06)	(2.10)	
$NProjects$	0.001	0.014	
	(0.04)	(0.56)	
$NVCs$	0.084***	0.091***	
	(7.26)	(5.00)	
$NProjectsperICD$	-0.061	-0.005	
	(-0.62)	(-0.05)	
$SelfProgress$			0.238
			(1.43)
$NCats$			-0.004
			(-0.13)
$NProjects$			0.003
			(0.09)
$NVCs$			0.014
			(1.10)
$PortfolioSize$			0.096***
			(2.91)
$Duration$			0.018
			(0.31)
$PrevRoundSize$			0.006
			(0.12)
Startup FE	No	No	No
Startup× Qtr. FE	No	No	No
VC Firm FE	No	No	No
Yr-Qtr FE	No	No	No
ICD FE	No	No	No
N	12,481	5,601	32,552

A5 Results with 3-digit ICD Codes

Table A5.1: Version of Table 2 Using 3-digit ICD Codes

This table replicates Table 2, except we use 3-digit ICD codes, a finer classification of diseases, to define drug product markets. All other details are the same as in Table 2.

Percent of startups with a close competitor	80.8%
Percent of startups with a close competitor held by same VC	25.3%
Average number of VCs per drug category	9.83
Average number of drug categories per VC	2.82
Percent of competing startup pairs with a common VC	8.5%

Table A5.2: Version of Table 3 Using 3-digit ICD Codes

This table replicates Table 3, except we use 3-digit ICD codes, a finer classification of diseases, to define drug product markets. All other details are the same as in Table 3.

	(1)	(2)	(3)	(4)	(5)
	OLS *Progress*	**Probit** *Lagging× SharedVC*	**1st Stage** *Lagging× SharedVC*	**2SLS** *Progress*	**OLS** *Progress*
$Lagging \times SharedVC$	-0.022***			-0.087***	-0.021**
	(-2.71)			(-2.87)	(-2.54)
$P\,(Lagging \times Proximity)$			0.756***		
			(7.06)		
$Lagging \times Proximity$		0.541***			
		(5.85)			
$Lagging$	-0.004		0.119***	0.006	
	(-0.96)		(3.72)	(0.94)	
$\ln(Age)$	0.010***	0.334***	-0.017***	0.010**	0.005
	(2.66)	(2.65)	(-3.06)	(2.57)	(1.34)
$NProjects$	0.002**	-0.024	0.000	0.002**	0.002**
	(2.05)	(-1.41)	(0.04)	(2.26)	(2.19)
$NVCs$	0.001	0.066***	-0.009	0.001	-0.002
	(0.30)	(4.33)	(-0.85)	(0.20)	(-0.43)
$NProjectsperICD$	-0.070***	0.018	0.015	-0.070***	
	(-4.35)	(0.20)	(0.64)	(-4.30)	
1*st stage F-stat*			49.78		
			(0.000)		
Startup FE	Yes	No	Yes	Yes	Yes
Yr-Qtr FE	Yes	No	Yes	Yes	No
ICD FE	Yes	No	Yes	Yes	No
ICD × Qtr. FE	No	No	No	No	Yes
N	12,469	12,481	12,469	12,469	9,621
Adj. R^2	0.085		0.570		0.146

Table A5.3: Version of Table 4 Using 3-digit ICD Codes

This table replicates Table 4, except we use 3-digit ICD codes, a finer classification of diseases, to define drug product markets. All other details are the same as in Table 4.

	(1)	(2)	(3)	(4)	(5)
	OLS	**Probit**	1*st* **Stage**	**2SLS**	**OLS**
	ExtendFunds	*Lagging*× *SharedVC*	*Lagging*× *SharedVC*	*ExtendFunds*	*ExtendFunds*
Lagging × *SharedVC*	-0.053***			-0.314***	-0.055***
	(-2.93)			(-3.70)	(-3.59)
P (*Lagging* × *Proximity*)			0.764***		
			(4.55)		
Lagging × *Proximity*		0.329***			
		(5.75)			
Lagging	0.057***		0.026***	0.063***	0.201***
	(6.44)		(3.87)	(6.62)	(7.99)
SelfProgress	0.025**	0.334*	-0.002	0.026**	0.161***
	(2.12)	(1.74)	(-0.27)	(2.17)	(5.68)
NCats	-0.003**	-0.033*	0.002	-0.003**	-0.003***
	(-2.52)	(-1.65)	(1.33)	(-2.25)	(-3.14)
NProjects	0.001	0.104***	-0.005	0.003	
	(0.30)	(2.94)	(-1.12)	(0.63)	
NVCs	0.004	0.018	-0.002	0.004	
	(1.51)	(1.30)	(-1.06)	(1.28)	
PortfolioSize	0.004*	0.134***	-0.006	0.005*	0.003*
	(1.87)	(4.66)	(-1.57)	(1.90)	(1.91)
Duration	-0.206***	0.036	-0.008	-0.208***	-0.266***
	(-26.67)	(0.50)	(-1.47)	(-26.32)	(-27.37)
PrevRoundSize	-0.002	-0.102*	0.004	-0.003	
	(-0.64)	(-1.71)	(0.90)	(-4.91)	
1*st stage F-stat*			20.70		
			(0.000)		
VC Firm FE	Yes	No	Yes	Yes	Yes
Startup FE	Yes	No	Yes	Yes	No
Yr-Qtr FE	Yes	No	Yes	Yes	No
Startup× Qtr. FE	No	No	No	No	Yes
N	32,537	32,552	32,537	32,537	34,414
Adj. R^2	0.334		0.274		0.584

A6 Robustness of Table 3

Table A6.1: Project Outcomes After Excluding Bad Phase I Outcomes

This table repeats the estimation in Table 3, except now we restrict the sample to drug projects that have no adverse readouts from Phase I clinical trials. Specifically, we match each drug project in the sample with detailed clinical trials information from the Clinical Trials Database of Cortellis. We exclude projects with adverse events (e.g., death of trial participants) in the process of clinical trials and those that failed to reach their primary endpoints. In total, 6,880 observations involving 525 drug projects are dropped. All other details are the same as Table 3.

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	***1st Stage***	***2SLS***	**Biprobit**	**Biprobit**	***OLS***
	Progress	*Lagging× SharedVC*	*Progress*	*Lagging× SharedVC*	*Progress*	*Progress*
$Lagging \times SharedVC$	-0.039**		-0.207**		-0.135*	-0.042**
	(-2.36)		(-2.61)		(-1.76)	(-2.19)
$P\,(Lagging \times Proximity)$		0.838***				
		(3.34)				
$Lagging \times Proximity$				0.301***		
				(3.14)		
$Lagging$	0.004	0.114***	0.025		0.002	
	(0.57)	(3.22)	(1.66)		(0.03)	
$\ln(Age)$	0.029***	-0.013	0.030***	0.241**	0.003	0.025**
	(3.99)	(-1.13)	(4.41)	(2.12)	(0.04)	(2.50)
$NProjects$	0.003	-0.004	0.002	0.012	-0.015	0.002
	(1.47)	(-1.59)	(1.17)	(0.51)	(-1.08)	(0.97)
$NVCs$	0.008	-0.014**	0.008	0.092***	0.035**	0.005
	(1.21)	(-2.47)	(1.11)	(5.06)	(2.00)	(0.64)
$NProjectsperICD$	-0.082***	0.002	-0.079**	-0.002	-0.039	
	(-2.82)	(0.03)	(-2.58)	(-0.03)	(-0.95)	
1st stage F-stat		11.12				
		(0.001)				
Startup FE	Yes	Yes	Yes	No	No	Yes
Yr-Qtr FE	Yes	Yes	Yes	No	No	No
ICD FE	Yes	Yes	Yes	No	No	No
ICD × Qtr. FE	No	No	No	No	No	Yes
N	5,592	5,592	5,592	5,601	5,601	4,578
Adj. R^2	0.088	0.563				0.098

Table A6.2: Version of Table 3 Following 2SLS without Probit

This table replicates columns 2–3 of Table 3, except we directly estimate a 2SLS model using $Lagging \times Proximity$ as the instrumental variable. All other details are the same as in Table 3.

	(1) **1st Stage** *Lagging× SharedVC*	(2) **2SLS** *Progress*
$Lagging \times SharedVC$		-0.073***
		(-3.08)
$Lagging \times Proximity$	0.072***	
	(6.55)	
$Lagging$	0.070**	0.007
	(2.37)	(1.48)
$\ln(Age)$	-0.009	0.010***
	(-1.21)	(3.02)
$NProjects$	0.005*	0.002**
	(1.69)	(2.52)
$NVCs$	-0.003	0.000
	(-0.43)	(0.05)
$NProjectsperICD$	-0.018	-0.034**
	(-0.42)	(-2.49)
1*st stage F-stat*	42.97	
	(0.000)	
Startup FE	Yes	Yes
Yr-Qtr FE	Yes	Yes
ICD FE	Yes	Yes
N	12,469	12,469
Adj. R^2	0.540	

Table A6.3: IV Version of Table 3 Including ICD × Quarter Fixed Effects

This table shows the IV analysis of Table 3 with ICD × Quarter fixed effects. Column 1 below matches column 7 in Table 3 in the paper. Note that Biprobit analysis does not allow fixed effects, so we do not need to replicate columns 4 and 5 in Table 3. All other details are the same as in Table 3.

	(1)	(2)	(3)	(4)
	OLS *Progress*	***Probit*** *Lagging× SharedVC*	1*st* ***Stage*** *Lagging× SharedVC*	***2SLS*** *Progress*
$Lagging \times SharedVC$	-0.018**			-0.067***
	(-2.37)			(-3.76)
$P\,(Lagging \times Proximity)$			0.709***	
			(4.412)	
$Lagging \times Proximity$		0.431***		
		(7.87)		
$\ln(Age)$	0.006*	0.214**	0.018	0.012***
	(1.74)	(2.13)	(1.46)	(3.56)
$NProjects$	0.002**	-0.000	0.001	0.002**
	(2.06)	(-0.01)	(0.23)	(2.54)
$NVCs$	-0.001	0.083***	-0.009	0.000
	(-0.43)	(7.24)	(-1.15)	(0.10)
1*st stage F-stat*			56.042	
			(0.000)	
Startup FE	Yes	No	Yes	Yes
ICD × Qtr. FE	Yes	No	Yes	Yes
N	11,507	12,481	11,507	11,507
Adj R^2	0.078		0.558	

Table A6.4: IV Version of Table 3 Controlling for Technological Similarity

This table replicates the IV analysis in Table 3, except we include $Lagging \times SharedCite$ in order to control for the similarity of technologies between the lagging project and the progressing project. All other details are the same as in Table 3.

	(1)	(2)	(3)	(4)	(5)
	Probit *Lagging× SharedVC*	***1st Stage*** *Lagging× SharedVC*	***2SLS*** *Progress*	**Biprobit** *Lagging× SharedVC*	**Biprobit** *Progress*
$Lagging \times SharedVC$			-0.160**		-0.075**
			(-2.24)		(-2.44)
$P\,(Lagging \times Proximity)$		0.402**			
		(2.19)			
$Lagging \times Proximity$	0.837***			0.822***	
	(5.88)			(5.62)	
$Lagging \times SharedCite$	0.831***	0.011	0.024	0.832***	0.480**
	(2.81)	(0.15)	(1.38)	(2.82)	(2.39)
$Lagging$		0.093***	0.016*		0.059
		(3.31)	(1.83)		(0.75)
$\ln(Age)$	0.391***	-0.004	0.010***	0.408***	-0.008
	(3.89)	(-0.46)	(2.99)	(3.94)	(-0.14)
$NProjects$	-0.003	0.005	0.002**	-0.004	-0.020
	(-0.17)	(1.43)	(2.37)	(-0.20)	(-1.13)
$NVCs$	0.090***	-0.007	0.000	0.090***	0.035***
	(8.36)	(-1.14)	(0.18)	(8.36)	(3.30)
$NProjectsperICD$	-0.035	-0.019	-0.037**	-0.033	-0.093**
	(-0.31)	(-0.41)	(-2.35)	(-0.29)	(-2.06)
1st-Stage F-stat		4.81			
		(0.031)			
Startup FE	No	Yes	Yes	No	No
Yr-Qtr. FE	No	Yes	Yes	No	No
ICD FE	No	Yes	Yes	No	No
N	12,481	12,469	12,469	12,481	12,481
Adj. R^2		0.526			

Note the F-statistic in column 2 is lower than in most our analyses. However, its associated p-value, 0.031, rejects the null of weak instruments at the 5% confidence level.

Table A6.5: Version of Table 3 Controlling for Same MSA

This table replicates main results in Table 3, except we include $Lagging \times SameMSA$ in order to control for the case when the pioneering and lagging startups are in the same MSA, i.e. $SameMSA = 1$. All other details are the same as in Table 3.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS *Progress*	***Probit*** *Lagging× SharedVC*	***1st Stage*** *Lagging× SharedVC*	***2SLS*** *Progress*	**Biprobit** *Lagging× SharedVC*	**Biprobit** *Progress*	***OLS*** *Progress*
$Lagging \times SharedVC$	-0.019***			-0.076***		-0.047**	-0.016**
	(-2.952)			(-3.971)		(-2.145)	(-2.263)
$P\,(Lagging \times Proximity)$			0.892***				
			(9.875)				
$Lagging \times Proximity$		0.325***			0.322***		
		(5.286)			(5.153)		
$Lagging \times SameMSA$	-0.005	1.365***	-0.060***	-0.005	1.363***	0.221	-0.224***
	(-0.735)	(6.702)	(-5.085)	(-0.763)	(6.639)	(1.638)	(-6.243)
$Lagging$	0.004		0.099***	0.012		-0.072	0.178***
	(0.719)		(4.068)	(1.550)		(-0.508)	(5.812)
$\ln(Age)$	0.009***	0.153	-0.012*	0.010***	0.162	-0.057	0.009**
	(2.878)	(1.205)	(-1.722)	(3.024)	(1.243)	(-1.474)	(2.373)
$NProjects$	0.002**	-0.006	0.005	0.002**	-0.006	-0.023	0.001
	(2.201)	(-0.353)	(1.545)	(2.487)	(-0.367)	(-1.397)	(0.949)
$NVCs$	0.000	0.085***	-0.015**	0.000	0.085***	0.024**	-0.001
	(0.106)	(6.763)	(-2.214)	(0.053)	(6.734)	(2.490)	(-0.257)
$NProjectsperICD$	-0.033**	-0.149	0.004	-0.034**	-0.147	-0.106***	
	(-2.510)	(-1.433)	(0.105)	(-2.455)	(-1.395)	(-2.788)	
1*st stage F-stat*			97.53				
			(0.000)				
Startup FE	Yes	No	Yes	Yes	No	No	Yes
Yr-Qtr FE	Yes	No	Yes	Yes	No	No	No
ICD FE	Yes	No	Yes	Yes	No	No	No
ICD × Qtr. FE	No	No	No	No	No	No	Yes
N	12,469	12,481	12,469	12,469	12,481	12,481	11,507
Adj. R^2	0.073		0.559		0.153		

Table A6.6: Version of Table 3 Dropping Boston

This table repeats the estimation of Eq. (2), except now we drop projects located in the Boston metropolitan area (MSA code equals 1120). In total, 1,866 observation involving 82 biotechs are dropped. All other details are the same as in Table 3

	(1)	(2)	(3)	(4)	(5)
	OLS *Progress*	***Probit*** *Lagging× SharedVC*	***1st Stage*** *Lagging× SharedVC*	***2SLS*** *Progress*	***OLS*** *Progress*
Lagging × SharedVC	-0.015**			-0.089**	-0.015*
	(-2.19)			(-2.06)	(-1.84)
P (Lagging × Proximity)			0.507***		
			(2.71)		
Lagging × Proximity		1.035***			
		(6.68)			
Lagging	0.001		0.078***	0.010	
	(0.40)		(2.96)	(1.54)	
ln(*Age*)	0.009**	0.440***	-0.001	0.010***	0.005
	(2.56)	(4.11)	(-0.14)	(2.76)	(1.35)
NProjects	0.002**	-0.001	0.004	0.002**	0.002**
	(2.33)	(-0.04)	(1.24)	(2.46)	(2.09)
NVCs	0.000	0.086***	-0.013*	-0.000	0.000
	(0.13)	(6.52)	(-1.68)	(-0.02)	(0.03)
NProjectsperICD	-0.029**	-0.055	0.031	-0.028*	
	(-2.22)	(-0.44)	(0.60)	(-1.96)	
1*st stage F-stat*			7.36		
			(0.008)		
Startup FE	Yes	No	Yes	Yes	Yes
Yr-Qtr FE	Yes	No	Yes	Yes	No
ICD FE	Yes	No	Yes	Yes	No
ICD × Qtr. FE	No	No	No	No	Yes
N	10,606	10,615	10,606	10,606	9,628
Adj. R^2	0.074		0.548		0.086

Note the F-statistic in column 3 is lower than in most our analyses. However, its associated p-value, 0.008, rejects the null of weak instruments at the 1% confidence level.

Table A6.7: Version of Table 3 Dropping San Francisco

This table repeats the estimation of Eq. (2), except now we drop projects located in the San Francisco metropolitan area (MSA code equals 7360). In total, 2,108 observation involving 74 biotechs are dropped. All other details are the same as in Table 3

	(1)	(2)	(3)	(4)	(5)
	OLS *Progress*	***Probit*** *Lagging× SharedVC*	***1st Stage*** *Lagging× SharedVC*	***2SLS*** *Progress*	***OLS*** *Progress*
$Lagging \times SharedVC$	-0.021***			-0.097**	-0.020*
	(-2.68)			(-2.51)	(-1.86)
$P\,(Lagging \times Proximity)$			0.601***		
			(2.65)		
$Lagging \times Proximity$		0.829***			
		(5.91)			
$Lagging$	0.001		0.078***	0.009	
	(0.16)		(2.81)	(1.44)	
$\ln(Age)$	0.009**	0.416***	-0.014*	0.009**	0.006
	(2.55)	(3.97)	(-1.75)	(2.59)	(1.56)
$NProjects$	0.001**	-0.017	0.003	0.002**	0.002**
	(2.18)	(-0.84)	(1.60)	(2.29)	(2.64)
$NVCs$	0.001	0.105***	-0.001	0.001	-0.001
	(0.21)	(9.07)	(-0.08)	(0.56)	(-0.53)
$NProjectsperICD$	-0.037**	-0.021	-0.033	-0.040**	
	(-2.33)	(-0.17)	(-0.57)	(-2.33)	
1*st stage F-stat*			7.04		
			(0.010)		
Startup FE	Yes	No	Yes	Yes	Yes
Yr-Qtr FE	Yes	No	Yes	Yes	No
ICD FE	Yes	No	Yes	Yes	No
ICD × Qtr. FE	No	No	No	No	Yes
N	10,363	10,373	10,363	10,363	9,447
Adj. R^2	0.069		0.546		0.075

Note the F-statistic in column 3 is lower than in most our analyses. However, its associated p-value, 0.010, rejects the null of weak instruments at the 1% confidence level.

Table A6.8: Version of Table 3 Alternative Instrument using the Corporate Opportunity Waivers

This table replicates our project-level results using the the staggered adoption of laws across eight states that enable corporations to adopt corporate opportunity waivers (COWs). $Progress_{it}$, $Lagging_{it}$, $SharedVC_{it}$ and control variables are defined in the same way as Table 3. COW_{it} indicates whether startup i's state allows corporations to adopt COWs by t. Column 1 reports the probit model with dependent variable $Lagging \times SharedVC$ and independent variable $Lagging \times CCW$. The predicted probability is $P\left(Lagging \times CCW\right)$. Column 2 reports results from the first stage of the 2SLS regression. The table reports the Kleibergen-Paap Wald rk F statistic, with its associated p-value in parenthesis. Standard errors are computed by two-way clustering at the ICD category and startup company levels in the OLS and 2SLS regressions. t-statistics are in parentheses. FEs are noted in the bottom row. ***,**, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)
	Probit $Lagging\times SharedVC$	**1st Stage** $Lagging\times SharedVC$	**2SLS** $Progress$
$Shocked \times SharedVC$			-0.122*
			(-1.771)
$P(Lagging \times COW)$		0.559**	
		(2.105)	
$Shocked \times COW$	5.289***		
	(13.405)		
$Lagging$		0.113***	0.013
		(3.876)	(1.653)
COW	-4.880***	0.020	-0.004
	(-24.415)	(0.462)	(-0.496)
$\ln(Age)$	0.528***	-0.017	0.010***
	(5.637)	(-1.316)	(3.167)
$NProjects$	0.000	0.004	0.002**
	(0.005)	(1.435)	(2.394)
$NVCs$	0.084***	-0.010	-0.000
	(7.327)	(-1.489)	(-0.017)
$NProjectsperICD$	0.063	-0.019	-0.035**
	(0.497)	(-0.410)	(-2.389)
$1st\ stage\ F\text{-}stat$		4.43	
		(0.038)	
Startup FE	No	Yes	Yes
Yr-Qtr FE	No	Yes	Yes
ICD FE	No	Yes	Yes
N	12,481	12,469	12,469
Adj. R^2		0.522	

Table A6.9: Cross-Sectional Version of Table 3

This table reports results from a cross-sectional version of Equation (2), with one observation per project. The dependent variable $Progress_i$ measures whether project i ultimately progresses to Phase II. $Lagging_i$ is an indicator for whether another project in the same ICD category as project i progresses to Phase II between project i's initiation quarter and end of Phase I. $SharedVC_i$ is an indicator for whether project i shares a VC with the pioneering project. We generate control variables by taking their average value within a project. All other details are the same as in Table 3.

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS *Progress*	**Probit** *Lagging*× *SharedVC*	**1st Stage** *Lagging*× *SharedVC*	**2SLS** *Progress*	**Biprobit** *Lagging*× *SharedVC*	**Biprobit** *Progress*
Lagging × *SharedVC*	-0.169*			-0.910***		-0.445***
	(-1.83)			(-4.25)		(-4.74)
P (*Lagging* × *Proximity*)			0.825***			
			(6.00)			
Lagging × *Proximity*		0.448***			0.446***	
		(11.36)			(11.25)	
Lagging	-0.197***		0.163***	-0.019		-0.260***
	(-4.39)		(2.95)	(-0.26)		(-2.79)
MeanNProjects	-0.001	-0.015	-0.007	-0.010	-0.022	-0.071**
	(-0.21)	(-0.67)	(-0.77)	(-0.81)	(-0.98)	(-2.25)
MeanNVCs	-0.054**	0.091***	0.001	-0.040	0.092***	0.046***
	(-2.36)	(7.30)	(0.07)	(-1.60)	(7.42)	(4.17)
MeanNProjectsperICD	-0.337***	-0.074	0.032	-0.296**	-0.057	-0.121*
	(-2.98)	(-0.87)	(0.42)	(-2.24)	(-0.69)	(-1.95)
1*st stage F-stat*			36.04			
			(0.000)			
Startup FE	Yes	No	Yes	Yes	No	No
ICD FE	Yes	No	Yes	Yes	No	No
Initial Qtr FE	Yes	No	Yes	Yes	No	No
N	794	1,045	794	794	1,045	1045
Adj R^2	0.325		0.429			

Interpretation of magnitudes: The slope coefficients are considerably larger in magnitude than in the panel regressions, which makes sense given that this test aggregates all of a project's quarters. Column 1 implies the probability of a lagging project ever reaching Phase II is 0.169 lower if it shares a VC with the pioneer. The corresponding magnitudes for columns 4 and 6 are 0.910 and 0.445, respectively. For comparison, the dependent variable's mean and standard deviation are 0.260 and 0.439, respectively.

Table A6.10: Version of Table 3 Using *Suspend* as the Dependent Variable

This table contains estimates of Equation (2), replacing $Progress_{it}$ with $Suspend_{it}$, an indicator for whether project i is suspended in quarter t. We define $Suspend_{it}$ in two steps. First, we check whether Cortellis records a project as being explicitly discontinued, withdrawn, or out-licensed in a given quarter. Second, there are many projects that are never officially discarded and yet continue to be listed in the drug portfolio without further trial updates. We assume these "zombie projects" are suspended three years after the first quarter when Cortellis designates them as "no development reported." For zombie projects without such designation, we assume they are suspended five years after project initiation. Columns 1 and 3 include *ICD* fixed effects and *Quarter* fixed effects, corresponding to Table 3's columns 1 and 3 respectively. Columns 2 and 4 include $ICD \times Qtr$ fixed effects, corresponding to Table 3's columns 6 and Table A6.3's column 4 respectively.

	(1)	(2)	(3)	(4)
	OLS	***OLS***	***2SLS***	***2SLS***
	Suspend	*Suspend*	*Suspend*	*Suspend*
$Lagging \times SharedVC$	0.025***	0.031***	0.178***	0.257***
	(2.714)	(3.122)	(4.034)	(3.614)
$Lagging$	0.006		-0.013*	
	(1.045)		(-1.692)	
$\ln(Age)$	0.015***	0.019***	0.015***	0.008*
	(5.965)	(6.320)	(5.795)	(1.817)
$NProjects$	0.001	0.000	-0.000	0.000
	(0.762)	(0.089)	(-0.058)	(0.178)
$NVCs$	-0.002	-0.004	-0.002	-0.004
	(-1.226)	(-1.263)	(-0.681)	(-1.026)
$NProjectsperICD$	-0.018		-0.015	
	(-1.225)		(-0.943)	
Startup FE	Yes	Yes	Yes	Yes
Yr-Qtr. FE	Yes	Yes	Yes	Yes
ICD FE	Yes	No	Yes	No
ICD $\times$ *Qtr. FE*	No	Yes	No	Yes
N	12,469	11,507	12,469	11,507
Adj. R^2	0.036	0.041		

A7 Robustness of Table 4

Table A7.1: Version of Table 4 Including Startup×Quarter Fixed Effects

This table replicates Table 4, except we include *Startup* × *Quarter* fixed effects in order to control for the startup's funding demand. All other details are the same as in Table 4.

	(1) **OLS** *ExtendFunds*	(2) **Probit**	(3) 1*st* **Stage**	(4) **2SLS** *ExtendFunds*
Lagging × *SharedVC*	-0.028**			-0.241***
	(-2.321)			(-3.130)
P (*Lagging* × *Proximity*)			1.568***	
			(7.792)	
Lagging × *Proximity*		0.206***		
		(5.715)		
SelfProgress	0.211***	0.343**	0.021	0.233***
	(7.702)	(2.249)	(0.554)	(7.872)
NCats	-0.005***	-0.006	0.007**	-0.004***
	(-3.776)	(-0.261)	(2.008)	(-3.022)
PortfolioSize	0.005**	0.101***	-0.022***	0.005**
	(2.447)	(3.580)	(-4.756)	(2.542)
Duration	-0.262***	0.064	-0.018*	-0.266***
	(-26.955)	(1.061)	(-1.923)	(-27.006)
1*st stage F-stat*			60.72	
			(0.000)	
VC Firm FE	Yes	No	Yes	Yes
Startup× *Qtr. FE*	Yes	No	Yes	Yes
N	34,414	36,042	34,414	34,414
Adj. R^2	0.578		0.292	

Table A7.2: Lead vs. Non-lead VC

This table presents results from estimating the VC-funding regression in subsamples based on the VC's financial commitment to the startup. The dependent variable is an indicator for whether VC j extends funding to startup i in quarter t. Columns 1 and 2 compare results across subsamples of non-lead and lead VCs, where a lead VC is defined as the VC whose total amount invested to date is the highest across all the startup's VCs. Our data report the amount invested by each VC syndicate but not by each syndicate member. In cases where these missing data create ambiguity about the lead-VC measure, we assume all syndicate members invest equal amounts. This table reports OLS estimates. Remaining details and variable definitions are the same as in Table 4.

	(1)	(2)
	Non-Lead	**Lead**
$Lagging \times SharedVC$	-0.005	-0.083***
	(-0.40)	(-3.35)
$Lagging$	0.052***	0.099***
	(8.23)	(8.49)
$SelfProgress$	0.012*	0.004
	(1.76)	(0.22)
$NCats$	-0.002	-0.006
	(-1.28)	(-1.55)
$NProjects$	0.000	0.006
	(0.22)	(0.88)
$NVCs$	0.011***	0.022***
	(6.58)	(2.97)
$PortfolioSize$	-0.001	0.010***
	(-0.28)	(3.20)
$Duration$	-0.099***	-0.175***
	(-17.86)	(-13.83)
$PrevRoundSize$	-0.009***	-0.006
	(-2.74)	(-0.59)
VC Firm FE	Yes	Yes
Startup FE	Yes	Yes
Qtr FE	Yes	Yes
N	24,532	7,988
Adj. R^2	0.139	0.229

Table A7.3: Version of Table 4 Controlling for Technological Similarity

This table replicates columns 2–6 in Table 4, except we include $Lagging \times SharedCite$ in order to control for the similarity of technologies between the lagging startup and the startup owning the progressing project. All other details are the same as in Table 4.

	(1) **Probit** *Lagging× SharedVC*	(2) 1*st* **Stage** *Lagging× SharedVC*	(3) **2SLS** *ExtendFunds*	(4) **Biprobit** *Lagging× SharedVC*	(5) **Biprobit** *ExtendFunds*
$Lagging \times SharedVC$			-0.430***		-0.152***
			(-3.42)		(-8.49)
$P\,(Lagging \times Proximity)$		0.749***			
		(6.36)			
$Lagging \times Proximity$	0.448***			0.457***	
	(5.05)			(5.46)	
$Lagging \times SharedCite$	0.795***	-0.026	0.020	0.793***	0.101
	(4.04)	(-1.09)	(1.18)	(4.01)	(1.02)
$Lagging$		0.005	0.081***		0.494***
		(0.67)	(7.27)		(7.93)
$SelfProgress$	0.116	-0.009	0.012	0.144	0.230***
	(0.65)	(-0.86)	(0.99)	(0.81)	(3.81)
$NCats$	0.001	0.005	-0.001	0.000	0.001
	(0.05)	(1.49)	(-0.68)	(0.01)	(0.07)
$NProjects$	0.062	-0.005	0.003	0.061	0.022
	(1.28)	(-0.84)	(0.53)	(1.27)	(0.84)
$NVCs$	0.008	-0.001	0.012***	0.009	0.003
	(0.74)	(-0.30)	(4.46)	(0.84)	(0.56)
$PortfolioSize$	0.087**	-0.009**	0.002	0.090***	0.041***
	(2.57)	(-2.13)	(0.89)	(2.75)	(2.91)
$Duration$	0.145**	-0.012**	-0.126***	0.152**	-0.745***
	(2.30)	(-2.19)	(-18.70)	(2.55)	(-19.08)
$PrevRoundSize$	-0.033	-0.008	-0.012***	-0.042	-0.047***
	(-0.66)	(-1.48)	(-2.63)	(-0.85)	(-2.77)
1*st stage F-stat*		40.47			
		(0.000)			
VC Firm FE	No	Yes	Yes	No	No
Startup FE	No	Yes	Yes	No	No
Yr-Qtr. FE	No	Yes	Yes	No	No
N	31,313	31,298	31,298	31,313	31,313
Adj. R^2		0.323			

Table A7.4: Version of Table 3 and Table 4 Controlling for $CommonOwn_{it}$

This table replicates column 1 in Tables 3 and 4. As comparison, columns 1 and 3 repeat the OLS analysis of column 1 in Tables 3 and 4. Columns 2 and 4 replicates the OLS regression, except we add a control $CommonOwn_{it}$, which is a dummy equal to one if project i shares a VC in common with any other projects in the same ICD in quarter t. All other details are the same as in Tables 3 and 4.

	(1) *Progress*	(2) *Progress*	(3) *ExtendFunds*	(4) *ExtendFunds*
Lagging × *SharedVC*	-0.019***	-0.019***	-0.030**	-0.030**
	(-2.95)	(-2.85)	(-2.26)	(-2.25)
Lagging	0.000	0.000	0.098***	0.098***
	(0.16)	(0.14)	(9.91)	(9.90)
CommonOwn		0.003		-0.007
		(0.76)		(-1.11)
ln(*Age*)	0.009***	0.009***		
	(2.89)	(2.94)		
NProjects	0.002**	0.002**		
	(2.20)	(2.22)		
NVCs	0.000	0.000		
	(0.10)	(0.06)		
NProjectsperICD	-0.034**	-0.034**		
	(-2.53)	(-2.54)		
SelfProgress			0.022*	0.022*
			(1.75)	(1.76)
NCats			-0.005***	-0.006***
			(-3.03)	(-3.13)
NProjects			0.001	0.002
			(0.31)	(0.41)
NVCs			0.002	0.002
			(0.93)	(0.93)
PortfolioSize			0.005**	0.006**
			(2.22)	(2.40)
Duration			-0.213***	-0.214***
			(-27.00)	(-27.02)
PrevRoundSize			-0.001	-0.001
			(-0.32)	(-0.30)
Startup FE	Yes	Yes	Yes	Yes
Yr-Qtr FE	Yes	Yes	Yes	Yes
ICD FE	Yes	Yes	No	No
VC Firm FE	No	No	Yes	Yes
N	12,469	12,469	32,537	32,537
Adj. R^2	0.073	0.073	0.344	0.344

Table A7.5: Address Concerns About Poaching Nearby Investors

This table replicates columns 1–3 of Table 5 in a subsample where $SharedVC = 0$. We estimate all models by OLS. Instead of $Lagging \times SharedVC_{jt}$, the key independent variable in all regressions is $Lagging \times Proximity_{jt}$. Standard errors are clustered at the startup company level. FEs are noted in the bottom row. ***,**, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)
	$ExtendFundsAgg$	$AmountRaised$	$NewInvestors$
$Lagging \times Proximity$	0.027*	-0.010	-0.001
	(1.80)	(-0.40)	(-0.07)
$Lagging$	0.112***	0.023*	-0.011
	(6.02)	(1.83)	(-0.92)
$SelfProgress$	0.007	0.020	-0.036**
	(0.27)	(1.06)	(-2.18)
$NProjects$	0.008	-0.006	-0.001
	(0.80)	(-0.85)	(-0.15)
$NVCs$	0.026***	0.044***	0.026***
	(3.20)	(4.56)	(4.57)
$Duration$	-0.229***	-0.034***	-0.042***
	(-21.24)	(-3.65)	(-6.73)
$PrevRoundSize$	-0.005		0.013***
	(-0.73)		(3.11)
$\log(CumFunds_{i,t-1})$		-0.264***	
		(-3.85)	
Startup FE	Yes	Yes	Yes
Yr-Qtr FE	Yes	Yes	Yes
N	6,125	6,070	6,125
Adj. R^2	0.374	0.296	0.092

Table A7.6: Version of Table 4 Alternative Instrument using the Corporate Opportunity Waivers

This table replicates our VC funding results using the the staggered adoption of laws across eight states that enable corporations to adopt corporate opportunity waivers (COWs). $ExtendFunds_{ijt}$, $Lagging_{it}$, $SharedVC_{ijt}$ and control variables are defined in the same way as Table 4. COW_{it} indicates whether startup i's state allows corporations to adopt COWs by t. Column 1 reports the probit model with dependent variable $Lagging \times SharedVC$ and independent variable $Lagging \times CCW$. The predicted probability is $P\left(Lagging \times CCW\right)$. Column 2 reports results from the first stage of the 2SLS regression. The table reports the Kleibergen-Paap Wald rk F statistic, with its associated p-value in parenthesis. Standard errors are computed by two-way clustering at the ICD category and startup company levels in the OLS and 2SLS regressions. t-statistics are in parentheses. FEs are noted in the bottom row. ***,**, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)
	Probit $Lagging\times SharedVC$	1*st* **Stage** $Lagging\times SharedVC$	**2SLS** $ExtendFunds$
$Lagging \times SharedVC$			-0.781***
			(-3.766)
$P\left(Lagging \times COW\right)$		0.749***	
		(4.674)	
$Lagging \times COW$	4.426***		
	(16.540)		
$Lagging$		0.015**	0.080***
		(2.118)	(6.492)
COW	-4.193***	0.017	0.027
	(-22.798)	(0.555)	(1.168)
$SelfProgress$	0.168	-0.008	0.021
	(0.915)	(-0.732)	(1.532)
$NCats$	0.016	0.004	0.000
	(0.592)	(1.390)	(0.128)
$NProjects$	0.105*	-0.008	0.004
	(1.949)	(-1.053)	(0.539)
$NVCs$	0.007	-0.001	0.012***
	(0.483)	(-0.340)	(3.831)
$PortfolioSize$	0.070**	-0.009**	0.002
	(2.236)	(-2.062)	(0.566)
$Duration$	0.275***	-0.018***	-0.127***
	(3.472)	(-3.469)	(-16.561)
$PrevRoundSize$	-0.053	-0.007	-0.016***
	(-0.951)	(-1.326)	(-2.642)
1*st stage F-stat*		22.44	
		(0.000)	
VC Firm FE	No	Yes	Yes
Startup FE	No	Yes	Yes
Yr-Qtr FE	No	Yes	Yes
N	32,552	32,537	32,537
Adj. R^2		0.310	

A8 Cox Proportional Hazard Model

Table A8.1: Cox Proportional Hazards Regressions

This table reports regressions results of Cox Proportional Hazard model. Column 1 repeats our analysis of drug project outcomes akin to column 1 of Table 3. Column 2 repeats our analysis of company funding outcomes akin to column 1 of Table 4. The robust standard errors are calculated and corresponding t-values are reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)
	Progress	*ExtendFunds*
$Lagging \times SharedVC$	-0.716*	-0.641***
	(-1.92)	(-3.37)
$Lagging$	0.943***	3.960***
	(5.04)	(48.78)
$\ln(Age)$	0.136**	
	(2.34)	
$NProjects$	-0.095	
	(-1.62)	
$NVCs$	0.055***	
	(3.34)	
$NProjectsperICD$	-0.388***	
	(-7.00)	
$SelfProgress$		0.851***
		(7.18)
$NCats$		-0.038*
		(-2.13)
$NProjects$		-0.088**
		(-2.38)
$NVCs$		0.053***
		(7.36)
$PortfolioSize$		0.080***
		(3.15)
$Duration$		-40.116***
		(-13.92)
$PrevRoundSize$		-0.123***
		(-4.88)
Startup FE	No	No
Startup× Qtr. FE	No	No
VC Firm FE	No	No
Yr-Qtr FE	No	No
ICD FE	No	No
N	12,469	21,319
Log-Likelihood	-1,387.847	-4,977.660

Interpretation: Column 1 shows that the hazard rate of progressing from Phase I to Phase II, conditional on seeing a close competitor reach Phase II, is significantly lower if the pioneering and lagging projects share a common VC. This is the same message that emerges from Table 3 in the paper. Column 2 shows that the hazard rate of being funded by a VC, conditional on seeing a close competitor reach Phase II, is significantly lower if the pioneering and lagging startups share a VC. This is the same message that emerges from Table 4 in the paper.

A9 Robustness of Table 9

A9.1 Oster (2019) test for omitted variable bias

Oster (2019) provides a test for omitted variable bias that uses the information on the change in coefficient and the change in R^2 when moving from uncontrolled to controlled regression. The test's main assumption is that the relation between treatment and unobservables can be fully recovered from the relation between treatment and controls. Oster's method generates an identified set for Table 9's coefficient on *Common Ownership Rate*. We perform this test using information in columns 1 and 2 of Table 9 in the paper. Using Oster's notation, her recommended identified set is $[\beta^*(\Pi, \tilde{R}, \delta), \tilde{\beta}]$, where $\delta = 1$, Π ranges from 1.25 to 2, and

$$\beta^*(\Pi, \tilde{R}, \delta) = \tilde{\beta} - \delta[\beta^o - \tilde{\beta}]\frac{\Pi\tilde{R} - \tilde{R}}{\tilde{R} - R^o}.$$

Using Π=1.25, the recommended set is [0.012, 0.025], which safely excludes zero and therefore rejects that the effect of common ownership on innovation efficiency is driven by omitted variables. Using Oster's more conservative value, $\Pi = 2$, the recommended set is [-0.025, 0.025], which contains zero and therefore does not reject that claim.

Table A9.1: Innovation Efficiency: Adding Controls Gradually

This table shows alternative specifications of Table 9 by adding the control variables gradually in each column. The other details are the same as Table 9.

	(1)	(2)	(3)	(4)	(5)
Common Ownership Rate	0.031***	0.026***	0.026***	0.025**	0.025**
	(3.62)	(2.70)	(2.70)	(2.39)	(2.35)
Duration to Phase III		-0.017	-0.015	-0.015	-0.015
		(-1.24)	(-1.01)	(-0.98)	(-0.98)
Prob. Reach Phase III			0.001	0.001	0.001
			(0.35)	(0.38)	(0.37)
Num. VCs per Startup				0.003	0.003
				(0.40)	(0.38)
VC Holding Duration					-0.005
					(-0.14)
N	94	94	94	94	94
R^2	0.125	0.139	0.140	0.142	0.142
Adj. R^2	0.115	0.120	0.112	0.103	0.093

Table A9.2: Innovation Efficiency: Alternative Specifications

This table shows alternative specifications of Table 9. In columns (1) and (2), we use *Common Ownership Dummy* as the explanatory which indicates whether common ownership exists in this ICD. In columns (2) and (3), we use Phase III efficiency as the outcome variable, defined as the number of drugs reaching Phase III divided by total VC investment in this ICD. In columns (4) and (5), we use *CompanyCORate* as the explanatory variable, defined as the probability that a startup has at least one commonly-invested competitor in this ICD. In column (6), we drop small ICDs by requiring this market has strictly more than 5 competing startups. In column (7), we use Survival Rate as the outcome variable, defined as the number of approved drugs over the total number of initiated drugs.

	(1) *Approval Efficiency*	(2) *Phase III Efficiency*	(3) *Phase III Efficiency*	(4) *Approval Efficiency*	(5) *Approval Efficiency*	(6) *Approval Efficiency*	(7) *Survival Rate*
Common Ownership Dummy	0.001***	0.003***					
	(3.40)	(2.70)					
Common Ownership Rate			0.080*			0.025**	0.167**
			(1.79)			(2.06)	(2.61)
CompanyCORate				0.003***	0.003***		
				(5.06)	(3.89)		
Duration to Phase III					-0.002	-0.012	-0.060
					(-0.15)	(-0.63)	(-0.66)
Prob. Reach Phase III					0.002	0.002	0.007
					(0.84)	(0.67)	(0.71)
Num. VCs per Startup					0.002	0.005	0.007
					(0.22)	(0.49)	(0.13)
VC Holding Duration					-0.013	0.006	-0.050
					(-0.39)	(0.09)	(-0.26)
N	94	92	92	95	95	82	94
Adj. R^2	0.102	0.065	0.024	0.207	0.183	0.054	0.092

Table A9.3: Impacts Measured by Forward Citations of Drug Patents

We estimate OLS regressions where each observation corresponds to a drug patent. We use two dependent variables to proxy for the patent's impact: Columns 1–2 use the dependent variable ln(*Cite*), the natural logarithm of patent citation counts. Columns 3–4 use the dependent variable ln(*Normalized Cite*), which is the natural logarithm of citation counts normalized by the average citation counts of patents issued in the same year following Hall et al (2005). Citation counts are measured as of June 2021. We only include patent-to-patent citations, so we exclude family-to-family citations. *Common Ownership Dummy* is an indicator for whether the patent's associated drug ever shared a VC in common with a close competitor, as defined in the paper. In columns 2 and 4, we add following variables as controls: *Num. ICD* is the number of ICD categories a drug covers. *Num. Project* is the number of projects being developed by the developer during the drug's lifespan. *Num. Competing Project* is the total number of competing projects being developed in the drug's life. Standard errors are clustered at the drug level. FEs are noted in the bottom row. t-statistics are in parentheses. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels, respectively.

	(1)	(2)	(3)	(4)
	ln(Cite)		ln(Normalized Cite)	
Common Ownership Dummy	0.049	0.002	0.016	-0.009
	(0.61)	(0.02)	(0.35)	(-0.20)
Num. ICD		0.099**		0.071***
		(2.21)		(2.52)
Num. Project		-0.005		-0.002
		(-0.74)		(-0.58)
Num. Competing Project		0.004		0.003**
		(1.28)		(2.07)
Drug Initiated Time FE	No	Yes	No	Yes
N	3,941	3,939	3,941	3,939
Adj. R^2	-0.000	0.024	-0.000	0.052

A10 Coefficients of Control Variables in Table 7 and Table 10

Table A10.1: Coefficients of Control Variables in Table 7

This table shows the coefficients of control variables in Table 7.

	(1) *Progress*	(2) *ExtendFunds*	(3) *Progress*	(4) *ExtendFunds*
$Lagging \times SharedVC$	-0.017**	-0.016	-0.017**	-0.009
	(-2.53)	(-1.31)	(-2.50)	(-0.89)
$Lagging \times SharedVC \times LessCompetitiveICD$	-0.046**	-0.052**		
	(-2.10)	(-2.02)		
$Lagging \times LessCompetitiveICD$	0.013	-0.010		
	(0.83)	(-0.60)		
$LessCompetitiveICD$	-0.007	0.016		
	(-0.75)	(0.93)		
$Lagging \times SharedVC \times SharedCite$			-0.020*	-0.103***
			(-1.84)	(-3.60)
$Lagging \times SharedCite$			0.012	0.001
			(1.12)	(0.08)
$Lagging$	-0.001	0.071***	-0.000	0.069***
	(-0.20)	(6.40)	(-0.08)	(6.77)
$NProjects$	0.002**	0.002	0.002**	0.001
	(2.18)	(0.36)	(2.24)	(0.27)
$NVCs$	0.000	0.012***	0.000	0.012***
	(0.11)	(4.38)	(0.11)	(4.34)
$\ln(Age)$	0.009***		0.009***	
	(2.91)		(2.82)	
$NProjectsperICD$	-0.036**		-0.034**	
	(-2.58)		(-2.58)	
$SelfProgress$		0.019*		0.019*
		(1.77)		(1.67)
$NCats$		-0.004**		-0.004**
		(-2.19)		(-2.20)
$PortfolioSize$		0.003		0.003
		(1.33)		(1.25)
$Duration$		-0.122***		-0.122***
		(-19.28)		(-19.52)
$PrevRoundSize$		-0.009**		-0.009**
		(-2.43)		(-2.53)
VC Firm FE	No	Yes	No	Yes
Startup FE	Yes	No	Yes	No
Yr-Qtr. FE	Yes	No	Yes	No
ICD FE	Yes	No	Yes	No
N	12,469	32,537	12,469	32,537
Adj. R^2	0.073	0.179	0.073	0.180

Table A10.2: Coefficients of Control Variables in Table 10

This table shows the coefficients of control variables in Table 10.

	(1) *Progress*	(2) *ExtendFunds*
$Lagging \times SharedVC$	-0.015**	-0.021**
	(-2.15)	(-2.04)
$Lagging \times SharedVC \times Noncompete$	-0.002	-0.008
	(-0.28)	(-0.48)
$Lagging$	0.001	0.265***
	(0.11)	(7.26)
$Lagging \times Noncompete$	0.000	0.013
	(0.06)	(0.30)
$Noncompete$	0.008	
	(1.01)	
$\ln(Age)$	0.010***	
	(2.98)	
$NProjects$	0.002**	
	(2.23)	
$NVCs$	0.000	
	(0.14)	
$SelfProgress$		0.091***
		(3.61)
$NCats$		-0.003**
		(-2.13)
$PortfolioSize$		0.002
		(1.36)
$Duration$		-0.172***
		(-18.96)
Startup FE	Yes	No
Startup× Qtr. FE	No	Yes
VC Firm FE	No	Yes
Yr-Qtr FE	Yes	No
ICD FE	Yes	No
N	12,469	34,414
Adj. R^2	0.071	0.465

References

Mas-Colell, Andreu and Whinston, Michael Dennis and Green, Jerry R. "Microeconomic Theory", *Oxford University Press New York*, 1995.

Oster, Emily, "Unobservable selection and coefficient stability: Theory and evidence", *Journal of Business & Economic Statistics* 37(2), 187–204, 2019.

Wasi, Nada and Flaaen, Aaron, "Record linkage using Stata: Preprocessing, linking, and reviewing utilities ", *The Stata Journal* 15(3), 672–697, 1995.