What Drives Innovation? Lessons from COVID-19 R&D

Ruchir Agarwal and Patrick Gaule*

January 18, 2022

Abstract

This paper studies the global R&D effort to fight the deadliest diseases. We find: (1) the elasticity of R&D effort with respect to market size is about 1/2 in the cross-section of diseases; (2) given this elasticity, the R&D response to COVID-19 has been 4 to 26 times greater than that implied by its market size; (3) the aggregate short-term elasticity of science and innovation can be very large, as demonstrated by the aggregate flow of clinical trials increasing by 38% in 2020, with limited crowding out of trials for non-COVID diseases; and (4) public institutions and government-led incentives were a key driver of the COVID-19 R&D effort—with public research institutions accounting for 70 percent of all COVID-19 clinical trials globally. Overall, our work suggests that leveraging early-stage incentives, non-monetary incentives, and public institutions may be important for scaling up global innovation.

Keywords: COVID-19, Innovation, Market Size, Pharmaceutical Industry

*Agarwal: International Monetary Fund, RAgarwal@imf.org. Gaule: University of Bristol, patrick.gaule @bristol.ac.uk. Acknowledgements: We thank Pierre Azoulay, Dan Cao, Jeff Clemens, Annamaria Conti, Tom Cunningham, Ehsan Ebrahimy, Dominique Foray, Ina Ganguli, Stefano Giglio, Jonathan Gruber, Luigi Iovino, Miles Kimball, Nikolas Mittag, Morten Olsen, Ammon Salter, Rossella Salandra, Sarah Smith, Holger Spamann, Stan Veuger, Heidi Williams and Yanos Zylberberg. Further, as economists who study science, we are humbled by the ingenuity and tireless effort of scientists, health workers, regulators, industry participants and others who brought about an unprecedented scientific response to fight the COVID-19 pandemic. All errors are our own. The views expressed therein are those of the authors and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.

1 Introduction

The COVID-19 pandemic is one of the greatest challenges ever faced by modern medicine. When China publicly shared the genetic sequence of COVID-19 on January 12th, 2020, scientists quickly began working on vaccines, treatments, and diagnostics to fight SARS-CoV-2. However, the world was not hopeful for quick success.[1] On the vaccine front, the fastest any vaccine had previously been developed was four years, for mumps in the 1960s, and thus even predictions of success by the summer of 2021 seemed highly optimistic (Ball 2021). Similarly, on the treatment front, limited progress had been made on the other known coronaviruses that cause disease in humans despite decades of research.

The scientific community responded with a massive R&D effort to fight the COVID-19 pandemic. By December 2020, several vaccine candidates had excellent results in large trials, with two vaccines (developed by Pfizer/BionTech and Moderna) receiving emergency use authorization in several countries. Similarly, on the treatment front, researchers had identified various treatments—e.g. dexamethasone, IL-6 blockers, monoclonal antibodies—that could bring a moderate reduction in mortality rates. While major challenges remained in scaling up the production and distribution of vaccines and tackling the new virus strains, the pharmaceutical innovation response to COVID-19 has already turned out to be an unprecedented success in terms of product discovery and development within the first year of the pandemic.[2,3]

The COVID-19 experience demonstrates that drug development can potentially proceed rapidly—without compromising on safety—when there is a global emergency and sufficient resources available for research and development (R&D). This begs the question: what lessons can we draw from the innovation response to COVID-19 about the drivers of innovation, and how can it inform the

[1] Begum et al. (2020) and Saif (2020) listed multiple bottlenecks in COVID-19 vaccine development, including the absence of animal models that could appropriately mimic the human COVID-19 infection. Pillaiyar, Meenakshisundaram, & Manickam (2020) noted that, based on historical experience, it could take more than 10 years for a new potent anti-COVID-19 anti-viral agent to be approved. Abi Younes et al (2020) suggested that COVID-19 R&D effort might be hard to scale up in the short run due to scarcity of ideas or human capital.

[2] The closest historical equivalent to the COVID-19 R&D effort is perhaps the Manhattan project. Also, over the medium term, the COVID-19 innovation response may end up boosting research progress in multiple areas, including for messenger RNA (mRNA) medicines that effectively instruct a patient's own cells to produce proteins to prevent, treat, or cure disease.

[3] See Agarwal and Gopinath (2021) and Agarwal and Reed (2021) for issues related to equitable access to COVID-19 health tools.

global effort to scale up innovation to fight other deadly diseases? Shedding light on this question is the key focus of our paper.

The COVID-19 pandemic is a rare instance of a large discrete shift in global medical needs, and thus the market size for the pharmaceutical industry. This allows us to examine how the entire landscape of clinical trials altered in response to the COVID-19 shift. With that aim, we present four key results based on publicly available data on pharmaceutical clinical trials around the world and on our extension of the classic model of innovation and market size developed by Acemoglu and Linn (2004).

Our key findings are:

#1. We estimate that the elasticity of R&D effort with respect to market size is about $1/2$ in the cross-section of diseases. We establish this cross-sectional relationship between the market size of a disease (measured as the disease-level mortality risk at the national level weighted by national income levels) and R&D effort (measured as new clinical trials), by matching data on world-wide clinical trials to 75 broad disease categories that have a non-negligible death burden. We estimate an elasticity strictly less than 1 (and around $1/2$) across all disease categories and also within sub-categories (such as cancer or infectious diseases) with the R^2 of the univariate regressions ranging from about 0.3 to 0.5. As discussed in section 4.1, this result is consistent with previous empirical studies which have typically found that the elasticity of research effort with respect to market size is below one.

#2. We document that the R&D response to COVID-19 has been 4 to 26 times greater than that implied by the historical relationship between market size and R&D effort.

#3. Based on the COVID-19 episode we find that, even in the short term, the aggregate elasticity of science and innovation can be very large. We document that in the initial months of the pandemic up to 50% of newly started trials were directed towards COVID-19. And, despite the large increase in COVID-19 R&D, overall new clinical trials increased by 38% in 2020 with little crowding out of R&D effort for other diseases.

#4. We find that public research institutions were a key driver of the COVID-19 R&D effort—accounting for 70% of all COVID-19 clinical trials globally, and being 10 percentage points

more likely to conduct a COVID-19 trial relative to private firms. In addition, studying the speed of COVID-19 vaccine development, we find U.S. and Chinese candidates were on average 2 months faster than candidates from other countries. This crucial boost in speed was possibly due to greater provision of early-stage incentives by the policy response in these countries, including through programs such as Operation Warp Speed (OWS).

Our findings have three broad implications for the study of innovation. First, our findings suggest that *boosting market size by itself may not be an effective tool to scale up innovation to fight large diseases—due to diminishing effort.* That is, the presence of the concavity blunts the effect of the classic market size incentives favored in economics. Such a concavity could arise due to various mechanisms, and in our paper, we discuss three possible candidates: (a) decreasing returns to scale due to a scarcity of ideas or talent, (b) risk aversion among firms' management, and (c) disease-specific ex-post taxation (paradox of market size). From one perspective, the concavity may reflect a genuine technological constraint in the economy without creating an inefficiency (e.g. due to the scarcity of ideas). By contrast, from another perspective, some of these mechanisms could be associated with market failures (for instance disease-specific taxation or when firm risk aversion is greater than that of the social planner's), which could lead to inefficient outcomes and under-investment in R&D even beyond that suggested by classic economic theory.[4]

Second, our findings suggest that *public research institutions and non-monetary incentives can be important drivers of innovation*. Policy efforts to accelerate pharmaceutical innovation in certain areas have often focused on enhancing market size. For instance, to address insufficient incentives for diseases that affect a small number of patients, the U.S. Congress passed the Orphan Drug Act in 1993, whose main provision was a seven-year market exclusivity for companies that developed an orphan drug.[5] However, COVID-19 innovation proceeded rapidly without governments committing

[4]Classic economic theory suggests that market economies are likely to under-provide innovation because of the public good nature of knowledge (Arrow 1962, Bloom, van Reenen, & Williams, 2019). And, this problem may be exacerbated for particular types of innovation due to additional market or policy failures (Budish, Roin & Williams 2015, Kremer & Snyder 2015, Kremer 2000).

[5]Similarly, in 1997 the U.S, Congress enacted a new legislation providing six months exclusivity to incentivize manufacturers to conduct studies of drugs in children. Michael Kremer has consistently advocated for 'pull' mechanisms that would seek to create markets for vaccines for diseases primarily affecting developing countries (Kremer 2000, Kremer & Glennerster 2004, Kremer & Williams 2010). While these initiatives have often been found to be effective in promoting innovation (Lichtenberg & Waldfogel 2003, Olson & Yin 2008, Yin 2008, Kremer, Levin & Snyder 2020), our point is that enhancing market size is often seen as the main thing policymakers can do to increase innovation in a certain area.

to pay developers and manufacturers above average costs for effective vaccines or drug treatments.[6] We find a strong role of public research institutions in driving COVID-19 innovation, accounting for 70% of COVID-19 clinical trials. Moreover, when compared to private firms, we find that public research institutions were 10 percentage points more likely to pursue clinical trials to fight COVID-19. In addition, we observe several important COVID-19 pharmaceutical innovations being driven by public research institutions with no explicit monetary incentives.

Third, based on the experience of COVID-19 our work suggests that *the short-term aggregate elasticity of science and innovation is much larger than expected, raising the distinct possibility of scaling up global innovation in the future.* In this context, the experience of COVID-19 suggests that scaling up innovation may require a better understanding of non-monetary incentives or intrinsic motivations, and also a more active role by governments in providing early-stage R&D incentives.[7]

While we have discussed the size of R&D response and have also studied the subset of clinical trials for COVID-19 vaccines, we have paid less attention to the optimal composition and quality of research effort that may be affected when innovation is scaled up quickly (see Bryan et al. 2020). Also, one reason for the large COVID-19 R&D response could be the newness of COVID-19, which could make the expected marginal return from investing in the disease much higher than for other diseases (in the presence of decreasing returns to scale). While this could be an important contributing factor, our data does not allow us to further investigate this possibility as we have no new diseases of comparable scale in our dataset. Still, the rapid and large R&D response to COVID-19 suggests that scarcity of ideas and talent may not be a prohibitive barrier to scaling up innovation—at least for new diseases.

Overall, the COVID-19 R&D response raises the distinct possibility that global innovation in the

[6]The advance purchase contracts for various vaccine candidates in the U.S. Operation Warp Speed (OWS) were at below market price. Effectively, these advance contracts worked more as a reward for the government in case their initial support paid off (akin to the option to obtain the first batch of products in exchange for providing crowdfunding support on platforms such as Kickstarter).

[7]Toole (2012) finds that both market size and NIH funded basic research have economically and statistically significant effects on the entry of new drugs with the contribution of public basic research coming in the earliest stage of pharmaceutical drug discovery. Gross & Sampat (2021) describe the U.S. research effort during World War II and emphasize the role of the U.S. Office of Scientific Research and Development (OSRD) in funding and coordinating R&D efforts. In the context of pharmaceutical products for diseases afflicting developing countries, the public sector and philanthropic foundations have often provided early-stage incentives while seeking to retain control of the intellectual property to achieve their goals to develop new affordable products for neglected diseases. The underlying model—described in Moran (2005), Maurer (2006) and Munoz et al. (2015)—is often perceived to have been quite successful but has received relatively less attention among economists.

future can be scaled up significantly. Achieving that may require leveraging early-stage incentives, non-monetary incentives, and public institutions as a complement to market size incentives.

2 Theory

Our empirical framework is based on an extension of the model presented in Acemoglu and Linn (2004), which studies the impact of market size on innovation. In this section, we describe the key features of the model and its links to our empirical work, while presenting the full model in Appendix 2. Acemoglu and Linn (2004) build a model to link R&D effort (z_j) for disease j to the market size for the drug (Y_j). We make two key modifications to their model.

Disease-Specific Taxation: First, we allow for disease-specific taxation (τ_j) to take into account that the *de facto* tax rate faced by the firm on the profits from a drug may depend on the market size of the disease. Such a *de facto* tax structure may arise due to different reasons. For instance, governments often pay for drugs on behalf of the citizens—especially that is needed by a large fraction of the population (e.g. vaccines). In such situations, the price received by the drug-maker is determined by a negotiation process with a single buyer (the government), which needs to account for political economy, societal needs, and public image considerations. Moreover, similar considerations may influence the pricing of the drug even when the drug-maker is directly selling to the public—making it harder for the firm to charge a high price for a drug that serves a large segment of the population.

Non-Monetary Incentives for Innovation: Second, we consider the possibility that researchers may have non-monetary motivation (η_j), i.e. altruism motive or reputational incentives) to pursue R&D (discussed more below).

Other than these modifications, the setup of the model follows Acemoglu and Linn (2004) closely.[8]

In the model, the discounted value of profits for firms can be expressed by a standard dynamic programming recursion. $V_j(t \mid q_j)$, the value of a firm that owns the most advanced drug of quality

[8]To capture the effect of limited supply of top researchers, we also allow for economy-wide decreasing returns to scale to R&D effort (instead of disease-specific decreasing returns to scale as in the original model).

q_j in line j at time t, is

$$rV_j(t \mid q_j) - \dot{V}_j(t \mid q_j) = \pi_j(q_j(t)) - (1-\eta)\left[n_j(t)V_j(t \mid q_j)\right] \tag{1}$$

where $\pi_j(q_j(t))$ is the flow profits, r is the discount rate, and $z_j(t)$ is R&D effort at time t on this line by other firms. The value of owning the best technology in line j, $rV_j(t \mid q_j)$, is equal to the flow profits, $\pi_j(q_j(t))$, plus the potential appreciation of the value, $\dot{V}_j(t \mid q_j)$, and takes into account that at the flow rate $n_j(t)$ there will be an innovation, causing the current firm to lose its leading position and to make zero profits thereafter. The parameter $\eta \in [0,1]$ captures the non-monetary incentives and represents the degree to which the firm cares about bringing the innovation to market—independent of the possibility of its innovation being replaced by a new entrant at a future date. Another way to interpret the case of $\eta > 0$ is that if a firm gains reputation value from being first to bring an innovation to the market then it would still derive non-monetary benefits from the innovation even after it has been replaced by new entrants (thereby blunting the standard Arrow replacement effect as shown in Aghion and Howitt, 1992). Further, one possible source of non-monetary incentives or intrinsic motivation for scientists could be avoiding large social costs associated with certain diseases. For instance, COVID-19 was associated with lockdowns that imposed additional costs to society, and researchers and scientists could be motivated to exert R&D effort to help avoid such social costs.

To account for the disease-specific progressive taxation we suppose $\tau_j = 1 - k \cdot Y_j^{-\beta}$ for some non-negative constants k, β with $\beta < 1$. Further, we account for public investment in science and support for research (i.e., push incentives) that may increase R&D capacity and relax the capacity constraint by supposing that $C = \overline{C} \cdot (1+P)^a$, where P is a measure of public investment in science or push incentives, a is a constant between zero and one, and $\overline{C}$ is a positive constant. Then, under the model's assumptions, the equilibrium steady state effort for disease j can be represented as:

$$\ln z_j^S = \alpha_0 + \alpha_1 \cdot P + \eta + (1-\beta)\ln Y_j \tag{2}$$

That is, log R&D effort is a function of the log market size Y_j, non-monetary incentives η, public investment in science P, and a constant term. Here, the elasticity of R&D effort with respect to market size is given by $1-\beta$. This is the relationship we estimate in Table 2. Note that in the

Acemoglu and Linn (2004) model there is no role for taxes or non-monetary incentives, such that $\beta = 0$, $\eta = 0$, and $P = 0$. This implies that the elasticity of R&D effort with respect to market size is 1 in their baseline model.[9]

3 Data and Methods

Measuring R&D Effort. In much of our empirical work, the key outcome variable will be the disease-specific R&D effort. Following much of the literature on pharmaceutical innovation (see e.g. Yin 2008, Kyle & McGahan 2012, Blume-Kohout & Sood 2013, Dranove, Garthwaite & Hermosilla 2014, Budish, Roin & Williams 2015), we use the number of new clinical trials as a measure of R&D effort. Alternative measures include patents (Qian 2007), R&D spending (Ward & Dranove 1995, Giaccotto, Santerre & Vernon 2005), or the number of products in development but neither of these are readily available for COVID-19 R&D and for very recent periods more generally.[10] Alternatively, others (including Acemoglu & Linn 2004, Dubois et al. 2015) have used innovation outcomes such as new drug approvals by the FDA. However, these are also not well suited to study a very recent phenomenon.

Measuring Market Size. A key explanatory variable of interest for us is the potential market size for the disease. The literature has used different approaches to quantify potential market size. For instance, Dubois et al. (2015) measure expected market size as the lifetime revenue accruing to the average product in a given disease category. Lichtenberg (2005) and Civan & Maloney (2009) use Disability-Adjusted Life Years (DALYs) and disease mortality respectively. By contrast, Acemoglu & Linn (2004) derive an income-based measure of market size, which combines the number of consumers weighted by their incomes. In line with one strand of the literature, we construct an income-based measure of potential global market size. First, for the year 2017, we sum disease-level mortality data at the national level weighted by national GDP per capita and the value of statistical life (VSL) of USD 1 million for the mean global citizen.[11,12] The choice of this measure

[9] The relationship in (2) is consistent with Toole (2012), which found that both market size and public support for basic research had an effect on the entry of new drugs.

[10] For instance, patent applications are only observable when they are published 18 months after application. It is thus too early to observe a meaningful set of patent applications related to COVID-19

[11] The particular value of statistical life (VSL) we choose is inconsequential for our analysis since it just multiplies all diseases burden by a constant.

[12] Instead of using a deaths-based measure of market size, we could use a measure based on years of life lost (YLL)

is based on our formal framework in the appendix.

Clinical trials data. We use clinical trials data from clinicaltrials.gov, a registry of trials run by the U.S. National Library of Medicine. Initially meant as a resource for patients to identify experimental treatments, the registry has become a key source of information on clinical trials activity globally.[13] It currently includes more than 300,000 registered trials from 209 countries. Unlike competitive intelligence database sold by private vendors, it is freely available and can be downloaded in bulk. Data fields included in the clinicaltrials.gov records include the study title, condition, the intervention, outcomes, eligibility criteria, target enrollment, study phase, the lead sponsor, locations, registration date and state date and contact information of the lead investigators. Most relevant for us are the sponsor class (industry or not), the location of the clinical trials, the study phase, the starting date and the medical condition that the trials address. That field is a free form entry field and thus does not follow a standard vocabulary. However, trials for COVID-19 are specifically tagged by the National Library of Medicine,[14] and we manually developed a cross-walk to disease classifications for other diseases. For this study, we focus on clinical trials in phases 1, 2 and 3 and exclude phase 4 trials and trials not associated with a phase.[15]

Mortality data. We use disease-level mortality data from the Global Burden of Disease (GBD) Project. This project is a major data collection effort led by the Institute for Health Metrics and Evaluation and involving thousands of researchers (Global Burden of Disease Collaborative Network 2017). Data released in open access by the project includes estimates of the number of deaths by cause and country. We use the latest data available as of May 2020 which covers mortality in the 2017 calendar year. Causes of deaths in GBD are categorized in four levels from general to more specific. For instance, a death attributed to acute myeloid leukemia would be categorized as a death from non-communicable diseases (level 1), from a neoplasm (level 2), from leukemia (level 3) or acute myeloid leukemia (level 4). Our preferred classification is level 3 which includes

or disability adjusted life years (DALY). Given that the death burden of COVID-19 is skewed towards older individuals, such measures would result in a smaller market size estimate for COVID-19.

[13] Staring in 2005, the International Committee of Medical Journal Editors required clinical trials published in any of their affiliated journals to be pre-registered.

[14] Based upon the keywords 'COVID-19', 'COVID, 'SARS-CoV-2, 'severe acute respiratory syndrome coronavirus 2', '2019-nCoV', '2019 novel coronavirus', 'Wuhan coronavirus'

[15] Phase 1 clinical trials aim to assess safety and appropriate dosing while phase 2 and 3 assess efficacy and side effects. Phase 4 monitor the long-term effects of drugs that are already on the market. Trials not associated with a phase include monitoring of symptoms without intervention and non-pharmaceutical interventions.

170 causes of death. Since we are interested in diseases and their mortality, our analysis excludes injuries as well as diseases with negligible death burden (e.g., acne), resulting in a sample of 75 diseases, to which we add COVID-19 as a separate disease. From the raw data, we compute both the global burden of disease, and the share of deaths occurring in high-income countries, upper middle-income countries, lower middle-income countries, and low-income countries, based on the 2019 World Bank classification of countries. In addition to the 2017 global death burden, the GBD database also provides information on changes in the death burden from 2007 to 2017, which we incorporate in our dataset.

Vaccine development data. We use the World Health Organization (WHO) COVID-19 candidate vaccine landscape database, which compiles detailed information on COVID-19 vaccine candidates in development.[16] The landscape provides summary tables of COVID-19 vaccine candidates in both clinical and pre-clinical development; tracks the progress of each vaccine from pre-clinical, Phase 1, Phase 2, through to Phase 3 efficacy studies; and includes information on key attributes of each vaccine candidate. The database is updated periodically every few weeks and using the different vintages of the database (between February 2020 to December 2020) we can construct the timing of when the different candidates move from one clinical stage to another. In addition, based on a manual search for each vaccine candidate, we code the location of the developer, and whether each candidate is being developed by private sector entities, a public entity, or by a collaboration between public and private entities. Overall, we have 222 vaccine candidates in development by mid-December 2020, and our analysis of the speed and success of the COVID-19 vaccine R&D effort in Section 6 is based on this dataset. The dependent variables are (a) the month in which the vaccine candidate first appeared as a pre-clinical candidate in the dataset, (b) an indicator variable that takes the value of one if the vaccine candidate entered clinical trials (i.e. graduated from pre-clinical to Phase 1 or Phase 1/2), and (c) number of phase progressions experienced by the vaccine candidate in the year 2020, as a measure of vaccine R&D success. In addition, the vaccine developer location, the private/public status, and the month in which the candidate entered the pre-clinical phase are used as explanatory variables. Descriptive statistics for the vaccine data are in Appendix Table 6.

Disease cross-section data set. We form a cross-section of diseases, with research effort (aver-

[16]The data is available at https://www.who.int/publications/m/item/draft-landscape-of-covid-19-candidate-vaccines.

age yearly number of new trials from 2015 to 2019) and market size as previously defined. We include a number of additional variables, including Disability Adjusted Life Years (DALYs), whether the disease is an infectious disease, whether the disease is a cancer and the share of the disease burden in high-income countries and the change in disease burden from 2007 to 2017. Finally, as a measure of the difficulty of the disease in terms of R&D, we compute the ratio of phase III trials to phase I trials, roughly reflecting the share of drug candidates that progress to phase III from phase I. Descriptive statistics for the 75 diseases in our sample are displayed in Appendix Table A1.

COVID-19 entry data set. Based on the clinicaltrials.gov data, we form a cross-section of all organizations that conducted at least two clinical trials from 2015 to 2019. The original data includes information on whether the organization is in the industry (as opposed to hospitals and universities which also conduct clinical trials). We deduce the location of an organization based on the location of its clinical trials (complemented with manual checks). We measure the number of COVID-19 clinical trials started by the organization in 2020. We construct the total number of trials undertaken by the organization from 2015 to 2019 as well as an indicator variable for whether the firm conducted at least one trial related to lower respiratory or upper respiratory infections from 2015 to 2019. Descriptive statistics for the 1,773 organizations in our sample are displayed in Appendix Table 4.

4 Diminishing Effort in R&D

This section presents empirical evidence on the elasticity of R&D with respect to market size. We first present results from the previous literature, which we complement with our findings. Both previous studies and our results suggest an elasticity below one, i.e., diminishing effort in R&D. We end the section with a discussion of possible mechanisms that can lead to diminishing effort in R&D.

4.1 Previous Studies on Market Size and Research Effort

A number of prior studies have investigated how R&D and innovation respond to market size (summarized in Table 1). The majority of studies examine variation in market size across disease, while a subset use within-disease variation in market size—for instance due to demographic (Acemoglu

& Linn 2004) or policy changes (Finkelstein 2004). Giacotto, Santerre, & Vernon (2005) find that a 1% increase in the pharmaceutical price index leads to a 0.58% increase in R&D spending. Similarly, Civan & Maloney (2009) conclude that an 1% increase in expected US entry price leads to 0.5% increase in the number of drugs in the drug development pipeline. An older study by Ward & Dranove (1995) associates a 1 percent increase in demand in a therapeutic area with a 0.5-0.8 percent increase in R&D spending. In a cross section of cancers, Budish, Roin & Williams (2015) estimate an elasticity of research effort with respect to market size of 0.24 (a side result in their analysis of effective patent length on research effort). Also in a cross-section of cancers, Lichtenberg (2007) finds an elasticity of 0.53 for the number of chemotherapy regimens with respect to the number of cases.

A consistent result across these studies—using a variety of market size measures, R&D effort measures, and empirical strategies—is that the elasticity of R&D effort, as well as new drug products, is well below one. Dubois et al. (2015) note "The previous literature generally finds elasticities for new drug products [with respect to market size] to be in the vicinity of 0.5" while their preferred estimate is around 0.23. One notable exception is Blume-Kohout & Sood (2013) who find an elasticity of 2.8 of R&D effort to changes in market size due to Medicare expansion. However, as noted in Dubois et al. (2015), this may be due to the Medicare expansion also impacting cash flow in addition to changing market size for new products. Acemoglu & Linn (2004) estimate an elasticity of innovation with respect to market size rather than an elasticity of R&D with respect to market size and find a large elasticity (4 to 6), but as noted by Dubois et al. (2015) this is an outlier in the literature.

(insert Table 1 about here)

4.2 Results on the Cross-Section of Diseases

We complement the previous literature by investigating the relationship between market size and research effort using a cross-section of diseases for worldwide trials. The analysis is set up so that it can be used to predict how much R&D activity one might have expected to see for COVID-19 based on its market size. For the moment, however, we exclude COVID-19 from this analysis. We measure R&D effort as the average yearly number of new clinical trials by disease between

2015 and 2019. Market size is defined by the global burden of disease in terms of deaths with an adjustment for its geographic distribution, as discussed in the data section.

Regressing research effort on market size in the cross-section of 75 diseases,[17] we find that a 1% increase in market size is associated with a 0.43% increase in research effort (cf. Table 2 column 1), implying an R&D effort elasticity of 0.43.[18] To investigate the stability of this relationship within a more homogeneous set of diseases, we restrict the sample to cancers and then to infectious diseases (cf. Table 2 columns 2 and 3 respectively). In the cancer subsample, we get an elasticity of 0.61 and for infectious disease an elasticity of 0.47. The R^2 of the univariate regressions range from about 0.3 to 0.5.

One may worry about reverse causality such that market size could be the result of innovation, rather than its determinant. In this context, note that our measure of market size reflects income-weighted mortality risk instead of actual sales, which mitigates such concerns. Nevertheless, to the extent that mortality risk may be influenced by past innovation, we control for changes in mortality risk in column 4, along with a number of other plausible determinants of research effort.[19] While the change in disease burden is positively and significantly associated with higher research effort, the coefficient on market size is lower than in the main specification at 0.36. Overall, these results (also displayed graphically in Figure 1) paint a consistent picture with estimated elasticities of R&D effort to market size well below one, and distributed around $1/2$. Our results are consistent with the previous literature discussed in the previous subsection.

(insert Table 2 and Figure 1 about here)

4.3 Possible Mechanisms

Overall, the consistent finding in the literature and our paper is that the elasticity of R&D effort with respect to market size is about $1/2$. We next turn to discuss possible factors that may give rise to

[17] This regression directly corresponds to the equation linking market size to R&D effort in the model presented in Appendix 2.

[18] We also ran Poisson regressions with roughly similar estimates for market size.

[19] Disability adjusted life years (DALYs) is another common measure of disease burden. To reduce collinearity with market size we control DALYs over deaths rather than DALYSs directly. Products for diseases that have a higher share of their mortality in rich countries may be easier to monetize - above and beyond the direct mechanical effect on market size. As a measure of the difficulty of the disease in terms of R&D we compute the ratio of phase III trials to phase I trials, roughly reflecting the share of drug candidates that progress to phase III from phase I.

such a relationship. We present three broad possible mechanisms: (a) Scarcity of Talent and Ideas, (b) Risk Aversion, and (c) Disease-Specific Taxation and the Market Size Paradox.

First, there may be decreasing returns to scale, such that doubling the number of clinical trials can lead to an increase in average R&D cost per trial, or can reduce R&D quality per trial. Such decreasing returns to scale could arise if (good) ideas to pursue in R&D are scarce—a central theme in the work of Suzanne Scotchmer (see e.g. Scotchmer 2004).[20] Bloom et al. (2020) present evidence across a range of contexts that research effort is rising substantially, while research productivity is declining—consistent with ideas being scarce and getting harder to find.[21] Within the pharmaceutical sector, larger firms often seek to replenish their R&D pipeline through acquisitions (Higgins & Rodriguez 2006, Krieger, Li, & Thakor 2018). If ideas were not scarce, one might expect that pharmaceutical companies would rely primarily on in-house development in order to avoid information asymmetries associated with acquisitions. However, one piece of evidence that is difficult to reconcile with scarce ideas is that while the elasticity of research effort with respect to research effort is below one, the elasticity of innovation (as opposed to research effort) with respect to market size has been found to be well above one in a couple of notable studies (Acemoglu & Linn 2004, Finkelstein 2004). If an increase in market size pushes researchers to pursue marginal ideas of lower quality—as one would expect if ideas are scarce—then innovation should react relatively less to changes in market size.

An alternative source of decreasing returns to scale could come from a scarcity of skilled R&D personnel. If the number of talented researchers/scientists are in scarce supply, then firms would face an upward-sloping supply curve that makes it costlier to increase directed R&D towards specific diseases. For instance, in the case of COVID-19, many vaccine candidates that entered the race to find a successful vaccine early, also had prior experience with vaccine candidates for Ebola or SARS-COV-1. At the level of the individual scientist, the costs to switch to a different topic appear to be high (Myers 2020). Overall, either scarcity of talent and ideas could be one candidate explanation for the observed diminishing effort.

Second, the management of R&D firms may be risk-averse. For instance, recent work by Lo-

[20] Acemoglu & Linn (2004) consider such a case in an extension to their model. Also, see the model in Clemens and Olsen (2021).

[21] Along the same lines, there has also been much discussion around the pharmaceutical industry's 'productivity crisis' reflected in rising R&D cost per approved drug see e.g. Cockburn (2007).

vallo et al. (2020) documents that managers in large firms exhibit high degrees of risk aversion, and routinely quash risky ideas in favor of marginal improvement or safe investments. Moreover, as noted in Asplund (2002), even when owners wish to maximise expected profits, the delegation of control to a risk-averse manager may cause the firm to behave in a risk-averse manner. In our case, the presence of risk-averse firm management would imply that due to decreasing marginal utility over profits, the firm management may be less enticed by the incentive of capturing a very large market for drugs if it comes with the risk of having a more un-diversified R&D portfolio. Such concentration risk in the R&D portfolio can be a serious deterrent as not all candidate drugs graduate from the clinical stage to the licensed products stage, creating a tendency for management teams to avoid putting their eggs in a few baskets. This could give rise to diminishing effort at the firm level, which could in turn lead to diminishing effort in the aggregate. Another reason why risk aversion could lead to diminishing effort is the 'winner takes all' dynamics. If a few successful candidates (for instance those that manage to become the first few to obtain licensure) are likely to capture a dominant share of the profits for each disease, then that could deter risk-averse firms to enter the race to find cures ex-ante. Under certain market structures, this could lead to diminishing effort in the aggregate.

Third, we conjecture the empirical regularity of diminishing effort could arise due to the difficulty in monetizing innovation that benefits a large group of people. We call this the 'market size paradox.' The argument is developed more formally in Appendix 2, and we sketch the key intuition here. In the case of COVID-19, while the development of vaccines and treatments could generate billions of dollars for some pharmaceutical companies, there are already public concerns and accusations of exploiting the pandemic. This has led some companies, e.g. Johnson & Johnson and AstraZeneca, to pledge they will not profit from their vaccine, although they have suggested this would be limited to the time during the pandemic. This experience is illustrative of a type of ex-post drug-specific tax that pharmaceutical companies may face (as opposed to say information technology companies), whereby if the drug is likely to serve a large population then it becomes politically or socially untenable for the company to make large profits from the drug's success. Moreover, the bigger the market size the stronger the political pressure for governments is likely to be (due to the size of the affected electorate) to ensure the drug companies do not charge a high price for access to the vaccines or treatments. Thus, this force moderates the classic market size mechanism

that predicts bigger the market size the more incentives for innovation. By contrast, the paradox of market size suggests, the bigger the market size the higher the ex-post taxation of profits—thereby blunting the incentives for innovation. Therefore, one reason for the diminishing effort could be that the ex-post tax rate rises with the market size. That is, if disease-specific taxes are increasing with market size, we may have a concave relationship between market size and innovation: innovation effort normally increases with market size, but at a decreasing rate. An increase in market size may even reduce R&D due to this effect.

While bringing additional evidence to empirically disentangle the drivers of diminishing effort is beyond the scope of this paper, the discussion of potential mechanisms here could serve as a guide for future work.

5 The R&D Response to COVID-19

The main results presented in this section is that COVID-19 has not been subject to diminishing efforts observed for other large diseases, with the number of COVID-19 trials being 4-26 times greater than that implied by its market size; and the large increase in COVID-19 R&D did not lead to sizable crowding out of R&D effort for other diseases. These results are based on the findings of the next five sub-sections that examine (a) the size of COVID-19 R&D relative to total R&D, (b) how the overall R&D effort across all trials was impacted by COVID-19, (c) how much crowding out occurred due to COVID-19 R&D, (d) the actual COVID-19 R&D vs. the expected COVID-19 R&D effort implied by its market size under various epidemiological and economic assumptions, and (e) Implications for Aggregate R&D Elasticity.

5.1 What Was the Size of COVID-19 R&D Relative to Total R&D?

We begin by asking: how large is the COVID-19 R&D effort compared to other the R&D effort for other diseases? In our data, 1,433 trials devoted to COVID-19 were started in 2020.[22,23] To put this

[22] As discussed in the data section, the clinicaltrials.gov database specifically tags COVID-19 trials based on a combination of keywords. Since we restrict our attention to trials that are in phases 1, 2 or 3, we have a lower number of trials than the headline figure reported on clinicaltrials.gov.

[23] While we have discussed the size of R&D response and have also studied the subset of clinical trials for COVID-19 vaccines, we have paid less attention to the optimal composition of research effort between re-purposed drugs vs. novel drugs/vaccines. In this context, Bryan et al. (2020) build a model to illustrate that during crises like the COVID-19

number in perspective, the *total* number of trials started in 2019 was slightly above 3,500.

In Figure 2, we report the ratio of new COVID-19 trials to all newly started trials on a monthly basis. In March 2020, around 13% of newly started trials were on COVID-19. Just a month later, the same figure had jumped to close to 50% before gradually declining to around 25% by December 2020. Comparing broad geographic areas (Appendix Figure A1), the U.S. and Europe have followed similar trajectories although in Europe the share of newly started COVID-19 trials peaked as high as 75% in April 2020, while it never exceeded 50% in the U.S. Meanwhile, in China, the share of COVID-19 trials among newly started trials was around 40% in February but it quickly decreased to below 20%, in line with the early persistent decline in active COVID-19 cases in China.

(Insert Figure 2 around here)

The figures presented suggest that the COVID-19 R&D effort is large both in absolute and relative terms. The scale and speed of the COVID-19 R&D have also been noted by other observers (e.g. Bryan et al. 2020, Le et al. 2020).

5.2 How Did COVID-19 Impact the Overall R&D Effort for All Diseases?

While the size of the aggregate market for pharmaceutical innovation is constantly changing with demographic and economic growth, the COVID-19 pandemic is a rare instance of a discrete shift. Our model predicts that such a shift would increase total effort for all diseases as well as reduce effort directed towards diseases other than COVID-19. However, the relative magnitude of these phenomena is an empirical question. We examine the effect of the COVID-19 outbreak on total effort first, before turning our attention to the crowding out of effort directed towards other diseases in the next subsection.

(insert Figure 3 about here)

Figure 3 displays the number of new clinical started monthly (3-months moving average). From 2015 to 2019, the average number of new clinical trials started monthly worldwide was remarkably

pandemic, the direction of research endogenously affects market structure, such that there could be a situation in which the market for invention leads to too much work on "quick" projects like re-purposed drugs and too little work on long-run projects like vaccines.

stable around 300. In the second quarter of 2020, the number of new clinical trials started averaged slightly above 450, a 50% relative increase compared to the pre-pandemic average.[24] The total number of trials started in 2020 is up by 38% compared to the total number of trials started in 2019.

This data suggests a large response of total pharmaceutical effort to the shock in the demand for pharmaceutical innovation induced COVID-19. The increase in total R&D effort is all the more remarkable given that the pandemic could have affected the supply side of R&D, for instance by making it risky or difficult to bring patients to medical facilities to administer treatments or make measurements. Additionally, the pandemic may have disrupted the productivity of researchers through increased child care obligations and reduced access to certain facilities (Myers et al. 2020).

5.3 Did COVID-19 R&D Crowd Out R&D for Other Diseases?

As we have seen, a great deal of R&D effort has been directed towards COVID-19, and total pharmaceutical R&D has substantially increased. But how has the COVID-19 outbreak changed the intensity of R&D effort for other diseases?

Our model in the appendix predicts that economy-wide decreasing returns will lead to such crowding out. To investigate the possibility of crowding-out more systematically, we compare clinical trials at the disease and quarter level. We first construct a panel of diseases at the quarterly level from the first quarter of 2015 to the last quarter of 2020. We then estimate a Poisson model with new trials on the left-hand side.[25] Our variables of interest are indicator variables for the second, third and fourth quarter of 2020. We control for disease fixed effects; in one specification we also include a time trend and quarter of the year fixed effects to account for potential seasonality in new trials. Considering clinical trials worldwide for diseases other than COVID-19, we find that there was a reduction in the number of clinical trials of around 10% in the second and third quarters of 2020 (cf. Table 3 column 1). However, the coefficient for the fourth quarter of 2020 is no longer negative.

[24] Distinguishing among broad geographic regions (Appendix Figure A2), we observe similar patterns in the U.S. as in the world as a whole. In Europe, a sharp increase in new trials was followed by an equally sharp decrease while in China the number of new trials remained stable.

[25] For this exercise, we use Poisson Quasi-Maximum Likelihood as it can more easily accommodate zeroes than the log-linear model (Santos-Silva and Tenreyro 2006)

(insert Table 3 and Figure 4)

Finally we visualize the crowding-out by disease in Figure 4, with trials started from April to December 2019 on the horizontal axis and trials started from April to December 2020 on the vertical axis. The crowding out is overall modest in size, though noticeable for some of the larger diseases.

5.4 Actual vs. Expected COVID-19 R&D Based on Market Size

In this sub-section we compare the actual COVID-19 R&D effort to what one would expect based on its potential market size. To make this comparison, we need estimates of the perceived *direct* burden of disease for COVID-19 in the year 2020. In line with our measure of the market size for other diseases, we only focus on the direct burden of COVID-19 and do not include the large indirect costs arising due to lockdowns and other non-pharmaceutical interventions undertaken to limit COVID-19 mortality. This is because we are interested in the part of the surplus that R&D firms can capture in order to evaluate how they reacted to the outbreak of COVID-19.

To derive the expected number of COVID-19 trials based on potential market size, we combine an estimate of (a) the relationship between research effort and potential market size, with (b) an estimate of the potential market size for COVID-19. For the computation of (a), we use our regression of log research effort on log market size, which yields a point estimate of 0.43 (See Table 2 column 1). For sensitivity analysis, we also use a 'low' and a 'high' value based on the literature estimating the relationship between market size and R&D effort. For the low estimate, we take a value of 0.25, effectively the low end of the published estimates. Conversely, for the high estimate, we use a value of 0.6, roughly the higher end of published estimates.[26]

For (b), we conceptually think of the COVID-19 pandemic as a one-off R&D opportunity for firms, such that in the short run the global death burden would rise by x million due to COVID-19 deaths. In our exercise, this would create an incentive for firms to front-load COVID-19 R&D to fight the disease, and then after a couple of years return to working on other diseases. Then, the exercise simply boils down to finding an estimate of x and its worldwide distribution, to evaluate

[26] Giacotto, Santerre, & Vernon (2005) report an elasticity of R&D spending to the pharmaceutical price index of 0.58. Some studies have found higher elasticities - notably Acemoglu & Linn (2004) but for innovation with respect to market size, not R&D with respect to market size. See section 4.1 for a review of the literature and a discussion.

the expected number of trials based on this market size. Thus, for (b), we quantify the likely death burden of COVID-19 in the hypothetical case in which in the absence of non-pharmaceutical control measures 70% of the world population eventually would get infected. The 70% number accounts for the typical herd immunity estimates, whereby if the disease were allowed to spread, due to herd immunity the number of cases would fall after a certain threshold, thereby infecting only a fraction of the population.

We quantify the COVID-19 potential market size under three different scenarios (by varying the infection fatality rates (IFR)), which when combined with (a) gives us three different estimates of the expected number of COVID-19 trials based on its potential market size.[27]

The three different scenarios are used to account for uncertainties in the IFR. For the upper estimate scenario, we assume that the infection fatality ratio is 1%, on the higher side of published estimates.[28] This is somewhat of a worst-case scenario, as it implies that within a few years COVID-19 would account for about 60 million deaths worldwide, i.e. x=60.[29] For the mid-point estimate scenario, we take the IFR to be the bottom range of the estimate, i.e. 0.5%, which gives us a COVID-19 death burden of 30 million. Finally, for the lower estimate scenario, we consider an IFR of 0.1%, which is a tenth of our upper estimate and yields a total COVID-19 death burden of about 6 million. Note that this number is more than three times above the actual death burden from COVID-19 observed in 2020 (of about 1.9 million deaths)—in a period when no vaccines were commercially available and therapeutics began rolling out only in the second half of the year. Thus, the death burden in our lower estimate is still likely to be larger compared to the case in which one would extrapolate the death burden based on the experience of 2020. Further, note that here we are being somewhat conservative in assuming that all COVID-19 trials should be front-loaded in the first year under our three scenarios, instead of distributing them over multiple years.

To give a sense of the magnitude of the COVID-19 death burden under the three different sce-

[27] In the presence of the infectious disease externality one may imagine the procurement of drugs may happen through the government. This is consistent with our calculation since we conceptualize the demand for drugs arising from all individuals in the world.

[28] According to the World Health Organization, estimates of infection fatality rate converge to approximately 0.5 - 1% (`https://www.who.int/news-room/commentaries/detail/estimating-mortality-from-covid-19`, accessed 11 January 2021, Perez-Saez et al. 2020, Stringhini et al. 2020). Earlier studies (Wu et al. 2020, Verity et al. 2020) had slightly higher infection fatality rate estimates.

[29] We actually use the corresponding potential market size measure (weighting deaths by national GDP per capita). However, we present the more intuitive total deaths to illustrate the underlying assumptions.

narios (5.3, 26.6, and 53.2 million), note that the global burden from all diseases combined has been about 50 million deaths per year in recent years. The largest non-COVID-19 disease in our sample is coronary heart disease, which has an annual death burden of about 9 million. So, another way to interpret the COVID-19 death burdens under the three different scenarios is that it ranges from 0.6 to 6 times the death burden of coronary heart disease for our exercise.

Based on the methodology described here, we present our results on the actual vs. expected number of COVID-19 trials. Our baseline scenario is based upon our estimate of the elasticity of R&D to market size and the mid-point estimate of COVID-19 market size. The expected number of trials in this scenario is 150 while 1,433 trials were observed in 2020. Thus, there were almost ten times more trials than expected based on the assumptions described above. Figure 5 displays COVID-19 (under the baseline scenario) in relation to other diseases with respect to market size and R&D effort.[30]

(insert Figure 5 about here)

Appendix table A3 explores the sensitivity of this result to alternative assumptions about the elasticity of R&D with respect to market size, and alternative estimates of the COVID-19 market size. The values we get range from 26 times more COVID-19 trials than expected (low elasticity and low COVID-19 market size, upper left corner) to 3.7 more COVID-19 trials than expected (high elasticity and high COVID-19 market size, lower right corner). While this range is broad, it is clear that the number of even under the less favorable assumptions there was significantly more investment in COVID-19-related R&D than existing estimates of the elasticity of R&D with respect to market size.

One potential limitation of our empirical exercise is that we have characterized market size considering mortality alone without including indirect costs. In the case of COVID-19, the indirect costs arising due to lockdowns and other non-pharmaceutical interventions undertaken to limit COVID-19 mortality may have been particularly large.

[30]Figure A3 displays this relationship separately by phase.

5.5 Implications for Aggregate R&D Elasticity

In this sub-section, we ask what the COVID-19 R&D response implies for the aggregate R&D elasticity. In 2020, the overall R&D effort as measured by clinical trials increased by 38% worldwide. As discussed above, this is due to the significant increase in COVID-19 R&D and simultaneously due to only a modest crowding out of R&D for other diseases. Using the three different market sizes of COVID-19 (and the middle elasticity) from the previous sub-section, we can compute the implied aggregate elasticity under the three different scenarios.

Taking the three different estimates of the COVID-19 death burden, we get that the market size increased by 6.4, 32, and 64 percent under the three different scenarios, implying an aggregate short-term aggregate R&D elasticity of 6, 1.2, and 0.6 respectively.

This calculation suggests that if the elasticity of COVID-19 can be replicated for other diseases, we can expect global R&D to scale up significantly in a short time—with an aggregate short-term elasticity of at least about 0.6, and potentially much higher.

6 Who Had the Strongest R&D Response to COVID-19?

In this section, we examine which sponsors (either firms or public organizations) decided to engage in COVID-19 R&D. As noted by Gross & Sampat (2021), the elasticity of different organizations in their ability to pivot to crisis innovation problems on short notice is largely an open question and in this analysis, we take a step toward answering that question.[31] The main result in this section is that public research institutions were a key driver of the COVID-19 R&D effort—accounting for 70% of all COVID-19 clinical trials globally and being ten percentage points more likely to conduct a COVID-19 trial relative to private firms. In addition, we demonstrate that while experience with respiratory diseases is a strong predictor of COVID-19 entry, the vast majority of trials were

[31] Bryan, Lemus & Marshall (2020) analyze COVID-19 entrant characteristics, including firm age, experience with infectious diseases, vaccines and antivirals, and pipeline size, distinguishing between entry before and after March 11, and repurposed versus new COVID-19 drugs. Our results are not directly comparable given the different levels of analysis (clinical trials for us versus drug candidates for them) and focus. Further, Clemens and Rogers (2020) study crisis innovation in a historical context. They find that during the Civil War and World War I procurement shocks led to substantial increases in the quantity of prosthetic device patenting. Harris (2021) finds that the scientific know-how acquired in the decades-long effort to produce a workable HIV vaccine has served as the critical foundation for the development of vaccines against the SARS-CoV-2 virus.

undertaken by sponsors that did not have such experience.

We first present some descriptive statistics on COVID-19 trials (cf Table A2). Only around 30% of COVID-19 trials have an industry sponsor; close to 70% of COVID-19 trials were started by a university, hospital, or other public organization (for comparison purposes, only around 60% of trials started in 2019 had a non-industry sponsor). Around 12% of COVID-19 trials were undertaken by sponsors that had sponsored a trial on lower or upper respiratory infection prior to the pandemic. The U.S. accounts for around a third of COVID-19 trials started globally, with 23% of COVID-19 trials taking place in Europe and only 5% in China.

We analyze next how sponsor characteristics predict engaging in COVID-19 R&D in a large cross-section of potential sponsors.[32] We use as dependent variable either an indicator variable taking value one if the organization started at least one COVID-19 clinical trial or the number of COVID-19 clinical trials started by the organization. Our variables of interest are size, public/private status location and prior experience with respiratory diseases.[33]

(insert Table 5 about here)

Results are displayed in Table 5. Across the whole sample of organizations, around one in six has conducted one or more COVID-19 trials. Not surprisingly, organization size is a strong predictor of starting a COVID-19 trial and the number of COVID-19 trials. Despite controlling for size, prior expertise in respiratory diseases is associated with a 20 percentage points increase in the propensity to engage in COVID-19 R&D and a doubling in the number of COVID-19 trials.[34] Private firms were significantly less likely than public organizations to engage in COVID-19 R&D. While U.S. and European organizations were equally likely to engage in COVID-19 R&D, Chinese organizations were less likely to engage in COVID-19 and did fewer trials.[35] Overall, these results suggest that public organizations had a relatively stronger R&D response to the COVID-19 crisis

[32]The universe of sponsors that conducted at least two clinical trials from 2015 to 2019 in the clinicaltrials.gov database.

[33]We proxy size by the number of trials undertaken from 2015 to 2019. Prior experience with respiratory diseases is an indicator variable that takes value one if the firm has conducted at least one trial related to lower respiratory or upper respiratory infections from 2015 to 2019.

[34]The specification on the number of COVID-19 trials is estimated by a Poisson regression. Hence the point estimate of 0.7 translate into a 101% ((exp(0.7)-1) increase.

[35]However, distinguishing new clinical trials by quarter reveals that Chinese organizations were initially (in Q1) more engaged in COVID-19.

than private firms.

7 The Speed and Success of COVID-19 Vaccine Effort

The main result presented in this section is that the speed of COVID-19 vaccine development was on average 2 months faster for U.S. and Chinese candidates, possibly due to early-stage incentives provided by programs such as Operation Warp Speed.

Many observers have noted that COVID-19 vaccines have been developed at an unprecedented speed with progress across stages measured in months rather than years. However, even months of delays have considerable implications for the control of the pandemic, and there has been substantial heterogeneity across vaccine candidates in how fast they have progressed to and through clinical trials.

We are interested to compare the progress of COVID-19 vaccines according to the public/private status and the location of their lead sponsor. Our interest in the location of the lead sponsor stems from the fact that innovators faced different early-stage incentives depending on country-level actions taken by the respective governments to promote innovation. For instance, support from Operation Warp Speed was more readily available to American companies. However, vaccine developers faced a single global market, as a successful vaccine could be sold globally—especially in the initial stages of the pandemic due to limited supplies of the doses. Thus, in this context, we may see the late-stage market size incentives as essentially uniform across vaccine developers, while the early-stage incentives varied by country location.[36]

Our analysis is based on various vintages of the WHO vaccine landscapes supplemented by hand-curated data on COVID-19 vaccine developers and their location. We first analyze how many months it took for a COVID-19 vaccine candidate to enter the preclinical stage, conditional on entering the preclinical stage by the end of 2020. Then, controlling for when the vaccine entered the preclinical stage, we look at whether a COVID-19 vaccine candidate entered clinical trial before the end of 2020, and the number of phases a COVID-19 vaccine candidate went through by the end of 2020.[37] In all cases, our variables of interest are the public/private status of the vaccine

[36] Note that under OWS, the purchase price paid for vaccine doses were below market price.

[37] We code the number of phases as 1 for vaccines in phase 1 by the end of 2020, 1.5 for vaccines in phase 1/2, 2 for

lead sponsor(s), and the location of the lead sponsor (U.S., Europe, China, and rest of the world). We control for the technology type (mRNA, viral vector, protein subunit, attenuated, etc.) in all specifications.

(insert Table 7 about here)

Results are shown in Table 7. We have a total of 222 COVID-19 vaccine candidates which took on average six months to enter the preclinical stage. A quarter of these candidates had entered clinical trials by the end of 2020. We find that COVID-19 vaccines with a private lead sponsor, or involving a collaboration between public and private lead sponsor, moved to the preclinical stage 2 months earlier than those with a public lead sponsor. However, there is little difference between these sponsor types in the propensity to enter clinical trials or the number of phases completed by the end of December 2020.

Compared to European vaccine candidates, U.S. vaccine candidates moved faster to the preclinical stage and were more likely to enter clinical trials. Chinese vaccine candidates were also faster than European vaccine candidates in moving to the preclinical stage and entered clinical trials at a higher rate. Since market size—a late-stage incentive—is largely uniform across regions, these cross-country variations in the speed of vaccine R&D effort may reflect national differences in early-stage incentives such as Operation Warp Speed. We hope future work can develop a fuller understanding of the effectiveness of programs such as OWS.

8 Conclusion

We study the global R&D effort to fight the deadliest diseases to examine the drivers of innovation. We present four main results. First, the elasticity of R&D effort with respect to market size is about $1/2$ in the cross-section of diseases, i.e., diminishing effort in R&D. Second, the R&D response to COVID-19 has been 4 to 26 times greater than that implied by its market size. Third, the aggregate elasticity of science and innovation can be very large, as demonstrated by the aggregate flow of clinical trials rising by 38% in 2020, with limited crowding out of trials for non-COVID diseases.

vaccines in phase 2, 2.5 for vaccines in phase 2/3, 3 for vaccines in phase 3, and 4 for the two vaccines approved in the U.S. by the end of 2020.

And fourth, public institutions and government-led incentives were a key driver of the COVID-19 R&D effort—with public research institutions accounting for 70 percent of all COVID-19 clinical trials globally and being 10 percentage points more likely to conduct a COVID-19 trial relative to private firms.

While we emphasize the potential role of public institutions, the private sector is possibly more efficient at some stages of the development and commercialization process. For instance, we find that private companies were faster than public research institutions in moving to the pre-clinical stage for COVID-19 vaccines. Thus, harnessing the innovation potential of public research institutions may need to be complemented by engaging the private sector with appropriate incentives.

Various observers have long worried that scaling up innovation is difficult either due to scarcity of good ideas/talent and a corresponding decline in research productivity (Bloom et al. 2020), or due to an inherent lack of private incentives to pursue such effort (Kremer 2002). However, despite these barriers, the unprecedented global R&D response to fight COVID-19 shows that the aggregate innovation effort may be operating significantly below its potential. Indeed, innovators managed to increase the aggregate R&D effort (as measured by clinical trials) by 38% in 2020, and deliver several successful vaccine candidates about 10 times faster than previous efforts for other diseases. (This corresponds to a very large aggregate short-term R&D elasticity with respect to changes in market size of at least 0.6, and potentially much higher.) At the same time, this was done without significantly sacrificing R&D effort for other deadly diseases.

One may be tempted to attribute the large COVID-19 R&D response to the relative "easiness" of finding cures for the disease, such that its biology was relatively more amenable to a quick technological breakthrough. However, this would be in stark contrast to the consensus view of experts at the beginning of the pandemic. At the outbreak of COVID-19, the world still did not have any vaccines or effective treatments for the other known coronaviruses even after decades of work, and nearly no experts anticipated having a licensed COVID-19 vaccine before mid-2021. Alternatively, the large scale of the COVID-19 effort may also partly reflect the fact that it is a new disease. New diseases may elicit a differential R&D response, either because the stock of promising ideas to explore has not been depleted by previous effort, or because drugs for a new disease do not face competition from previously invented drugs. It is difficult to ascertain this

possibility empirically given that new diseases, especially of this severity, are rare.

We hope future research can examine three areas that remain beyond the scope of this paper. First, the extent to which our results can generalize beyond the pharmaceutical sector is an open question. The need to scale up innovation in domains such as clean energies and climate mitigation technologies is abundantly clear. Yet the type of the R&D problems and the nature of R&D effort to solve them is certainly different. One idiosyncratic factor that may have contributed to the scale of the COVID-19 R&D response is the existence of a large pool of human capital due to earlier policy interventions, such as the doubling of NIH R&D funding in the 2000s.

Second, there remains scope for future research to evaluate the quality of COVID-19 R&D. While we attempt to address this question in a limited way by studying the progression of clinical trials for COVID-19 vaccines through different stages, it is too early for us to conduct a comprehensive assessment of the quality of R&D. There are various reasons why average quality of R&D may have been impacted, including due to accelerated decision making by regulators and researchers, sheer need for speed more generally due to the large social costs associated with the pandemic, or researchers switching to study domains that are beyond their narrow areas of expertise. At the same time, COVID-19 brought together individuals from different domains to collaborate which could have opened the door for more creative ways to evaluate hard scientific problems.

Further, since only two years has elapsed since the onset of the pandemic, the assessment of the R&D response may change in the future. For instance, if more vaccines or therapeutics were approved in the coming months (such as Molnupiravir or Paxlovid) then that may change the ex-post assessment of the productivity of the R&D investments. Relatedly, it would be instructive to systematically compare the success rates of COVID-19 drug and vaccines candidates to those of drug and vaccines candidates started in non-crisis times. Such a comparison would form an objective though indirect measure of the quality of COVID-19 R&D and inform our understanding of the productivity of R&D spurred by demand-side shocks. Given that it typically takes multiple years for a new pharmaceutical product to move from initial discovery to product launch, this could be a subject for a future study when sufficient time has elapsed.

Third, we hope that our work encourages further research in understanding the role of non-monetary incentives and early-stage incentives as drivers of innovation. There is an intuitive under-

standing among many that reputational concerns, altruism, the sense of duty to the greater good, and space for creativity and curiosity facilitated by unconditional funding may play an important role in innovation. In this context, there is scope for future research to empirically evaluate the importance of these factors in different innovation domains.

In summary, the COVID-19 R&D response raises the distinct possibility that global innovation in the future can be scaled up significantly. However, our work suggests that leveraging early-stage incentives, non-monetary incentives, and public institutions may be important for scaling up global innovation.

References

Abi Younes, G., Ayoubi, C., Ballester, O., Cristelli, G., de Rassenfosse, G., Foray, D., Gaule, P, Pellegrino, G, van den Heuvel, M, Webster, E & Zhou, L. (2020). COVID-19: Insights from innovation economists. Science and Public Policy.

Acemoglu, D., & Linn, J. (2004). Market Size in Innovation: Theory and Evidence From the Pharmaceutical Industry. The Quarterly Journal of Economics. 119(3):1049-1090.

Agarwal, R., & Gopinath, G. (2021). A Proposal to End the COVID-19 Pandemic (No. 2021/004). International Monetary Fund.

Agarwal, R., & Reed, T. (2021). How to End the COVID-19 Pandemic by March 2022. World Bank Policy Research Working Paper.

Aghion, P., and Howitt, P. (1992). A Model of Growth through Creative Destruction, Econometrica, 323–351.

Arrow, K. (1962). Economic Welfare and the Allocation of Resources for Invention. In The Rate and Direction of Inventive Activity: Economic and Social Factors (pp. 609-626). Princeton University Press.

Asplund, M. (2002). Risk-averse firms in oligopoly. International Journal of Industrial Organization, 20(7): 995-1012.

Ball, P. (2021). The lightning-fast quest for COVID vaccines-and what it means for other diseases. Nature. 589(7840):16-18

Begum, J., Mir, N. A., Dev, K., Buyamayum, B., Wani, M. Y., & Raza, M. (2020). Challenges and prospects of COVID-19 vaccine development based on the progress made in SARS and MERS vaccine development. Transboundary and Emerging Diseases. 00:1–14

Bloom, N., Jones, C. I., Van Reenen, J., & Webb, M. (2020). Are ideas getting harder to find? American Economic Review. 110(4): 1104-44.

Bloom, N., Van Reenen, J., & Williams, H. (2019). A toolkit of policies to promote innovation. Journal of Economic Perspectives. 33(3): 163-84.

Blume-Kohout, M. E., & Sood, N. (2013). Market size and innovation: Effects of Medicare Part D on pharmaceutical research and development. Journal of Public Economics. 97: 327-336.

Bryan, K. A., Lemus, J. & Marshall, G. (2020). Innovation During a Crisis: Evidence from Covid-19. Mimeo, University of Toronto

Budish, E., Roin, B. N., & Williams, H. (2015). Do firms underinvest in long-term research? Evidence from cancer clinical trials. American Economic Review. 105(7): 2044-85.

Civan, A. & Maloney, M. (2009). The Effect of Price on Pharmaceutical R&D. The B.E. Journal of Economic Analysis & Policy. 9:1–24.

Clemens, J., & Olsen, M. (2021) Medicare and the Rise of American Medical Patenting: The Economics of User-Driven Innovation. Mimeo, the University of Copenhagen.

Clemens, J., & Rogers, P. (2020). Demand shocks, procurement policies, and the nature of medical innovation: Evidence from wartime prosthetic device patents. NBER Working Paper no 26679. National Bureau of Economic Research.

ClinicalTrials.gov [Internet]. Bethesda (MD): National Library of Medicine. Available from http://clinicatrials.gov

Dranove, D., Garthwaite, C., & Hermosilla, M. (2014). Pharmaceutical profits and the social value of innovation. NBER Working Paper No 20212. National Bureau of Economic Research.

Dubois, P., De Mouzon, O., Scott-Morton, F., & Seabright, P. (2015). Market size and pharmaceutical innovation. The RAND Journal of Economics. 46(4): 844-871.

Finkelstein, A. (2004). Static and dynamic effects of health policy: Evidence from the vaccine industry. The Quarterly Journal of Economics. 119(2): 527-564.

Giaccotto, C., Santerre, R.E., & Vernon, J.A. (2005) Drug Prices and Research and Development Investment Behavior in the Pharmaceutical Industry. Journal of Law and Economics, 48:195–214.

Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2017 (GBD 2017) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME), 2018. Available from http://ghdx.healthdata.org/gbd-results-tool.

Gross, D. P., & Sampat, B. N. (2021). The Economics of Crisis Innovation Policy: A Historical

Perspective. NBER Working Paper 28335. National Bureau of Economic Research.

Harris, J. E. (2021). The Repeated Setbacks of HIV Vaccine Development Laid the Groundwork for SARS-CoV-2 Vaccines. NBER Working Paper 28587. National Bureau of Economic Research.

Higgins, M. J., & Rodriguez, D. (2006). The outsourcing of R&D through acquisitions in the pharmaceutical industry. Journal of Financial Economics. 80(2): 351-383.

Kremer, M. (2000). Creating markets for new vaccines. Part I: rationale. Innovation Policy and the Economy. 1: 35-72.

Kremer, M. (2002). Pharmaceuticals and the developing world. Journal of Economic Perspectives. 16(4): 67-90.

Kremer, M., & Glennerster, R. (2004). Strong medicine: creating incentives for pharmaceutical research on neglected diseases. Princeton University Press.

Kremer, M., & Williams, H. (2010). Incentivizing innovation: Adding to the tool kit. Innovation Policy and the Economy. 10(1): 1-17.

Kremer, M., & Snyder, C. M. (2015). Preventives versus treatments. The Quarterly Journal of Economics. 130(3): 1167-1239.

Kremer, M., Levin, M., & Snyder, C.M. (2020). Advance Market Commitments: Insights from Theory and Experience. AEA Papers and Proceedings. 110:269-73.

Krieger, J., Li, X., & Thakor, R. T. (2018). Find and replace: R&D investment following the erosion of existing products. Available at SSRN 3240344

Kyle, M. K., & McGahan, A. M. (2012). Investments in pharmaceuticals before and after TRIPS. The Review of Economics and Statistics. 94(4): 1157-:1172.

Mahlich, J. C., & Roediger-Schluga, T. (2006). The determinants of pharmaceutical R&D expenditures: evidence from Japan. Review of Industrial Organization. 28(2): 145-164.

Maurer, S. M. (2006). Choosing the right incentive strategy for research and development in neglected diseases. Bulletin of the World Health Organization. 84(5): 376-381.

Moran, M. (2005). A breakthrough in R&D for neglected diseases: new ways to get the drugs

we need. PLoS Medicine. 2(9): e302

Munoz, V., Visentin, F., Foray, D., & Gaule, P. (2015). Can medical products be developed on a non-profit basis? Exploring product development partnerships for neglected diseases. Science and Public Policy. 42(3): 315-338

Le, T. T., Cramer, J. P., Chen, R., & Mayhew, S. (2020). Evolution of the COVID-19 vaccine development landscape. Nature Reviews Drug Discovery. 19(10): 667-8.

Lichtenberg, F. R. (2007). Importation and innovation. Economics of Innovation and New Technology, 16(6), 403-417.

Lichtenberg, F. R., & Waldfogel, J. (2003). Does Misery Love Company? Evidence from pharmaceutical markets before and after the Orphan Drug Act. NBER Working Paper 9750. National Bureau of Economic Research.

Lovallo, D., Koller, T., Uhlaner, R., & Kahneman, D. (2020). Your Company is Too Risk Averse. Harvard Business Review. 98(2): 104-111.

Myers, K. (2020). The Elasticity of Science. American Economic Journal: Applied Economics, 12(4), 103-34.

Myers, K., Tham, W. Y., Yin, Y., Cohodes, N., Thursby, J. G., Thursby, M., Schiffer, P.E., Walsh, J.T., Lakhani, K.R. & Wang, D. (2020). Quantifying the Immediate Effects of the COVID-19 Pandemic on Scientists. Available at SSRN 3608302.

Olson, M. K., & Yin, N. (2018). Examining firm responses to R&D policy: An analysis of pediatric exclusivity. American Journal of Health Economics. 4(3): 321-357.

Perez-Saez, F.J., Lauer, S.A., Kaiser, L., Regard, S., Delaporte, E., Guessous, I., Sringhini, S., Azman, A.S. (2020) Serology-informed estimates of SARS-CoV-2 infection fatality risk in Geneva, Switzerland. Lancet Infectious Diseases. 3099(20):30584-3

Pillaiyar, T., Meenakshisundaram, S., & Manickam, M. (2020). Recent discovery and development of inhibitors targeting coronaviruses. Drug Discovery Today 25(4): 668-688.

Qian, Y. (2007). Do National Patent Laws Stimulate Domestic Innovation in a Global Patenting Environment? A Cross-Country Analysis of Pharmaceutical Patent Protection, 1978–2002. The

Review of Economics and Statistics 89(3), 436–45

Santos-Silva, J., & Tenreyro, S. (2006). The Log of Gravity. The Review of Economics and Statistics, 88(4): 641-658.

Saif, L. J. (2020). Vaccines for COVID-19: Perspectives, prospects, and challenges based on candidate SARS, MERS, and animal coronavirus vaccines. European Medical Journal, 200324(10):33590.

Scotchmer, S. (2004). Innovation and Incentives. MIT press.

Stringhini, S., Wisniak, A., Piumatti, G., Azman, A.S., Lauer, S.A., Baysson, H., ..., & Guessous I. (2020) Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Geneva, Switzerland (SEROCoV-POP): a population-based study. The Lancet. 396:313-319

Toole, A. A. (2012). The impact of public basic research on industrial innovation: Evidence from the pharmaceutical industry. Research Policy, 41(1), 1-12.

Verity, R., Okell, L. C., Dorigatti, I., Winskill, P., Whittaker, C., Imai, N., Dighe, A., ..., & Ferguson, N. (2020). Estimates of the severity of coronavirus disease 2019: a model-based analysis. The Lancet Infectious Diseases. 20(6):669-677

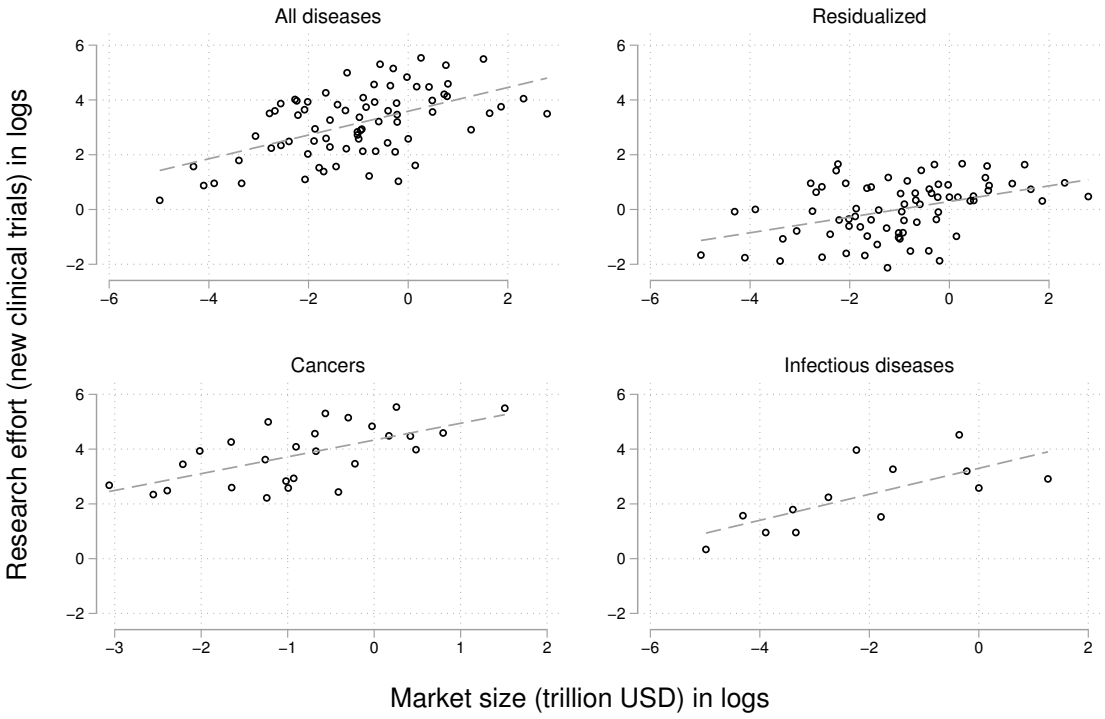
Ward, M. R., & Dranove, D. (1995). The Vertical Chain of Research and Development in the Pharmaceutical Industry. Economic Inquiry, 33(1), 70-87

World Health Organization (2020) DRAFT landscape of COVID-19 candidate vaccines.

Wu, J. T., Leung, K., Bushman, M., Kishore, N., Niehus, R., de Salazar, P. M., ... & Leung, G. M. (2020). Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. Nature Medicine, 26(4), 506-510.

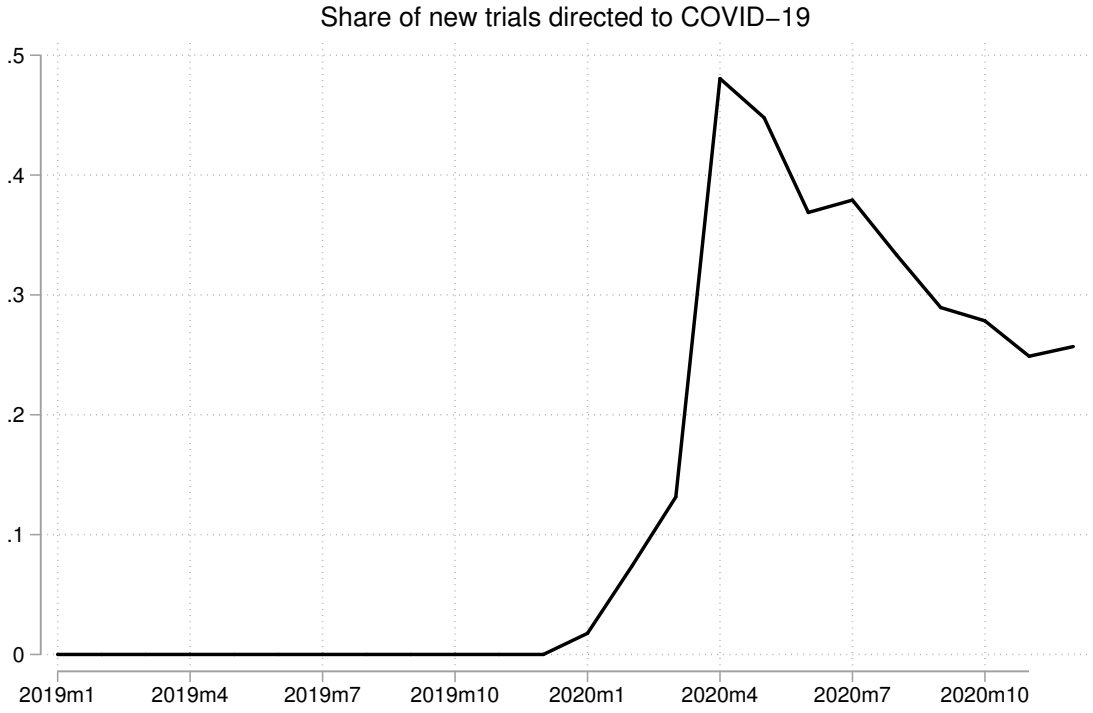
Tables and Figures

Figure 1: R&D Effort across Diseases and Diminishing Effort



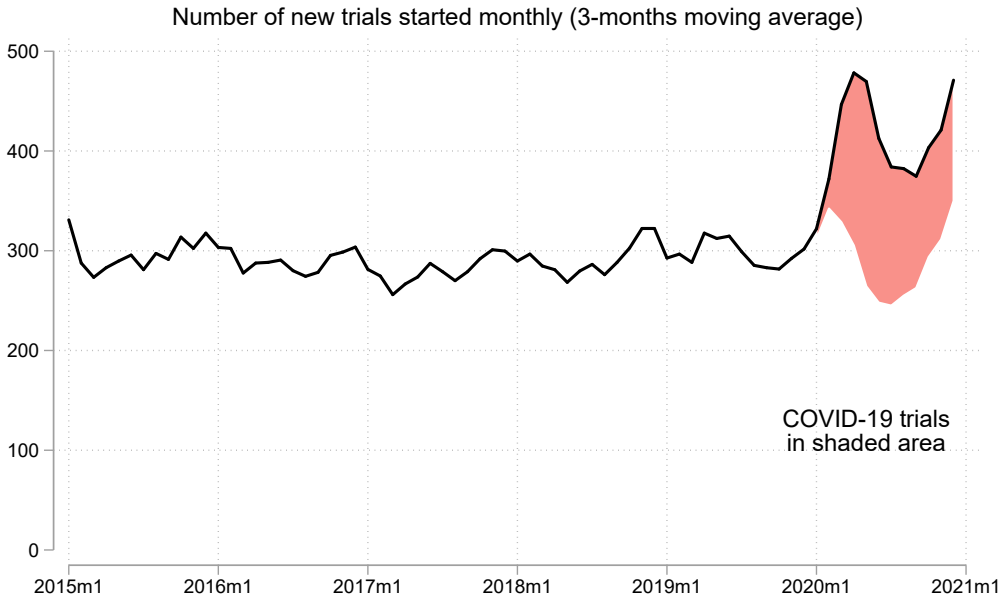
Notes: The figure displays the relationship between potential market size and research effort in the cross-section of 75 diseases in our sample (COVID-19 not included). Research effort is defined as the average yearly number of new trials per disease from 2015 to 2019, and expressed in logs. Potential market size is the 2017 disease-level mortality at the national level weighted by national GDP per capita and a value of statistical life of USD 1 million for the mean global citizen, expressed in logs. The upper left figure displays the overall relationship, while the lower left, and lower right figure restrict the sample to cancers and infectious diseases respectively. Finally, the upper right figure displays residualized research effort instead of actual research effort, where the residualized research effort comes from a regression of research effort on a broad set of potential determinants (except potential market size) detailed in Table 2 column 4. The regression equivalent of these figures are in Table 2.

Figure 2: Share of COVID-19 Trials among Newly Started Trials



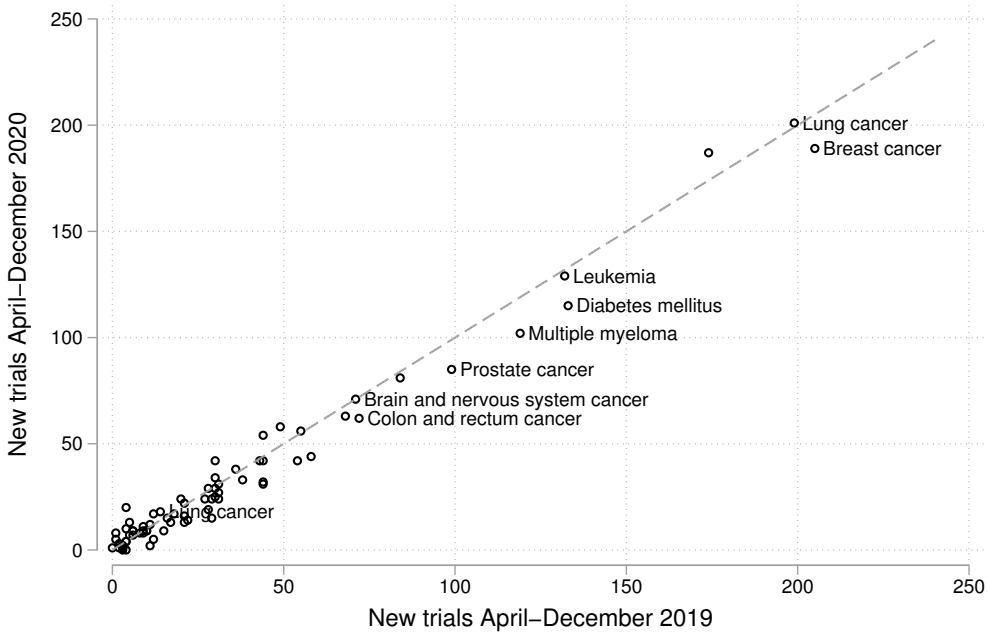
Notes: Based on clinicaltrials.gov data, the figure displays the share of new trials that were related to COVID-19 among all new trials on a monthly basis. For example, a number of .4 implies that 40% of the new trials in a given month were related to COVID-19. The COVID-19 trials are tagged as such by clinicaltrials.gov based upon a set of keywords related to COVID-19.

Figure 3: New Clinical Trials Started Monthly for All Diseases



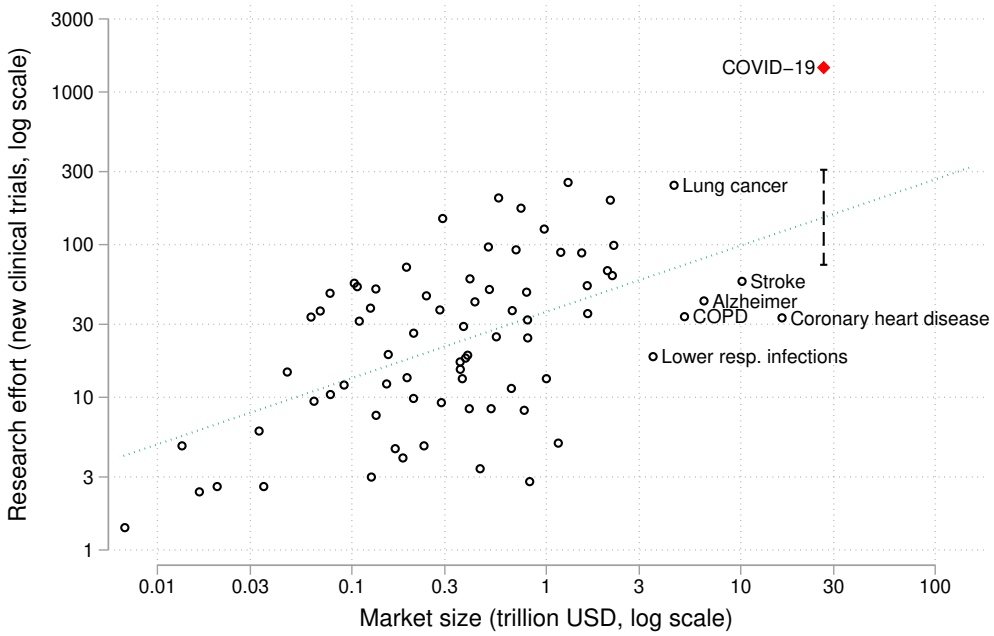
Notes: Based on clinicaltrials.gov data, the figure displays the total of new clinical trials worldwide (across all diseases) on a monthly basis. COVID-19 trials are showed in the shaded area. COVID-19 trials are tagged as such by clinicaltrials.gov based upon a set of keywords related to COVID-19.

Figure 4: Crowding Out of R&D Directed Towards Non-COVID-19 Diseases



Notes: In this figure, we compare the number of new trials by disease between 2019 (April-December) and 2020 (April-December). The sample includes all diseases except COVID-19. The choice of using the last three quarters for our comparison window is based on the spike in COVID-19 clinical trials observed after Q1 of 2020.

Figure 5: COVID-19 R&D Effort Compared with Other Diseases



Notes: The figure displays the relationship between potential market size and research effort in the cross-section of 75 diseases in our sample, plus COVID-19. Research effort is defined as the average yearly number of new trials per disease from 2015 to 2019; for COVID-19 we use the actual number of 2020 COVID-19 trials instead. Potential market size is the 2017 disease-level mortality at the national level weighted by national GDP per capita and a value of statistical life of USD 1 million for the mean global citizen, expressed in logs. For COVID-19, our estimate of potential market size is based on the hypothetical case in which in the absence of non-pharmaceutical control measures 70% of the world population eventually would get infected, and COVID-19 has an infection fatality rate of 0.5%. The dashed vertical line represents a 95% prediction interval for the number of COVID-19 trials. See main text for a detailed discussion of the COVID-19 potential market size and Table A3 for sensitivity to different elasticity estimates and COVID-19 market size assumptions.

Table 1: Elasticity of R&D Effort and Innovation to Market Size

Paper	Measure of Market Size	Measure of R&D Effort/New Products	Market	Elasticity
Panel A: Market Size and R&D				
Ward & Dranove (1995)	Demand in a therapeutic area	R&D spending	U.S. for all diseases	0.5-0.8
Giacotto, Santerre, & Vernon (2005)	Pharmaceutical price index	R&D spending	U.S. for all diseases	0.58
Mahlich & Roediger-Schluga (2006)	Industry profit margin	R&D spending	Japan for all diseases	0.17-0.53
Civan & Maloney (2009)	Expected U.S. entry price	Drugs in pipeline	U.S. for all diseases	0.5
Budish, Roin & Williams (2015)	Number of cancer patients	Clinical trials	U.S. for cancer	0.24
Panel B: Market Size and New Products				
Acemoglu & Linn (2004)	Income-weighted potential consumers	New molecular entities	U.S. for cancer	4-6
Lichtenberg (2007)	Number of cases	Chemotherapy regimens	U.S. for cancer	0.53
Dubois et al. (2015)	Global revenue of pharma. products	New molecular entities	Global for all diseases	0.23
Panel C: Estimates from this paper				
Table 2 Column 1	Income-weighted potential consumers	Clinical trials	Global for all diseases	0.43
Table 2 Column 3	Income-weighted potential consumers	Clinical trials	Global for infectious diseases	0.48
Table 2 Column 2	Income-weighted potential consumers	Clinical trials	Global for cancer	0.61

Notes: This table summarizes the elasticity of R&D effort to market size estimated in prior studies. For each study, the table lists the measure of market size, the measure R&D effort used in the study, and the geographic/pharmaceutical market studied. Panel A lists the estimates from prior studies regarding the elasticity of R&D to market size, panel B list the estimates from prior studies regarding the elasticity of new products to market size, and for comparison Panel C lists the estimates from this paper.

Table 2: R&D Effort across Diseases and Diminishing Effort

	(1)	(2)	(3)	(4)
	Dependent variable = New clinical trials (log)			
	All diseases	Cancers	Infectious diseases	All diseases
Potential market size (log)	0.434***	0.613***	0.475***	0.364***
	(0.074)	(0.095)	(0.127)	(0.068)
DALYs over deaths				0.002
				(0.001)
Change in disease burden				0.814**
(from 2007 to 2017)				(0.321)
Share of disease burden				-0.951
in high-income countries				(1.057)
Infectious				0.072
disease				(0.347)
Cancer				0.925***
				(0.269)
Disease difficulty				-0.180
				(0.121)
Observations	75	27	13	75
R2	0.287	0.396	0.514	0.482

Notes: The table investigates the relationship between potential market size and research effort in the cross-section of 75 diseases in our sample (COVID-19 not included). The data is set up as a cross section of 75 diseases, with a disease being the unit of observation. Regressions shown in column 1 and 4 are estimated using the full sample; regressions in columns 2 and 3 are estimated using the subsample of cancers and infectious diseases respectively. The dependent variable is the average yearly number of new trials per disease from 2015 to 2019, expressed in logs. The variable of interest, potential market size, is the 2017 disease-level mortality at the national level weighted by national GDP per capita and a value of statistical life of USD 1 million for the mean global citizen, expressed in logs. As a measure of 'disease difficulty' in terms of R&D we compute the ratio of phase III trials to phase I trials, roughly reflecting the share of drug candidates that progress to phase III from phase I. See main text for a description of the other control variables in column 4. Estimation is by OLS. Robust standard errors are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Crowding Out Results: Clinical Trials Directed Towards Non-COVID-19 Diseases

	(1)	(2)
	D.V.= New clinical trials	
Quarter = 2020Q2	-0.087**	-0.101***
	(0.037)	(0.034)
Quarter = 2020Q3	-0.131***	-0.118***
	(0.043)	(0.042)
Quarter = 2020Q4	0.069*	0.007
	(0.039)	(0.037)
Disease FE	Yes	Yes
Time trend	No	Yes
Quarter of the year FE	No	Yes
Diseases	75	75
Observations	1,800	1,800

Notes: In this table, we compare the number of new trials by disease in each quarter from 2015Q1 to 2020Q4. The sample includes all diseases except COVID-19. The dependent variable is the number of new clinical trials directed towards that disease in a quarter. Our variables of interest are indicator variables for the second, third and fourth quarter of 2020. Both specifications include disease fixed effects and are estimated by Poisson Quasi-Maximum Likelihood. The specification in column 2 also includes a linear time trend and dummies for each quarter of the year to account for potentially seasonality in new clinical trial starts. Robust standard errors are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Descriptive statistics on potential COVID-19 sponsors (n=1,773)

	Mean	S.D.	Min	Max
Nr of COVID-19 trials	0.41	1.28	0.00	27.00
Any COVID-19 trial	0.20	0.40	0.00	1.00
Sponsor size	7.92	18.54	2.00	281.00
Prior experience with respiratory diseases	0.06	0.23	0.00	1.00
Private sponsor	0.46	0.50	0.00	1.00
Sponsor location				
U.S.	0.35	0.48	0.00	1.00
Europe	0.25	0.43	0.00	1.00
China	0.16	0.36	0.00	1.00
ROW	0.25	0.43	0.00	1.00

Notes: This table presents descriptive statistics for the 1,773 organizations conducting at least two clinical trials listed in clinicaltrials.gov between 2015 and 2019. In addition to the number of COVID-19 trials and whether a sponsor had conducted any COVID-19 trial, the table presents information about the sponsor size measured as total trials between 2015 and 2019, whether the sponsor had prior experience with respiratory diseases, and whether the sponsor is a private pharmaceutical firm. Prior experience with respiratory diseases is an indicator variable that takes value one if the firm has conducted at least one trial related to lower respiratory or upper respiratory infections from 2015 to 2019. The second panel presents information about the geographic location of the sponsor based on the modal location of prior trials.

Table 5: COVID-19 R&D Effort and Sponsor Characteristics

	(1) Any COVID-19 trial	(2) # of COVID-19 trials
Trial sponsor location		
US	0.005	-0.123
	(0.022)	(0.196)
China	-0.105***	-0.886***
	(0.025)	(0.256)
ROW	0.082***	0.364*
	(0.026)	(0.190)
Europe omitted		
Other variables of interest		
Sponsor size	0.143***	0.700***
(log)	(0.011)	(0.048)
Prior experience with	0.272***	0.956***
respiratory diseases	(0.044)	(0.175)
Private sponsor	-0.101***	-0.710***
	(0.017)	(0.134)
Observations	1,773	1,773
Mean of dep. variable	0.196	0.408

Notes: In this table, we examine which type of sponsors had the strongest R&D response to COVID-19. The term 'sponsor' refers to any research entity (e.g. hospitals, firms, etc.) that are listed as a 'sponsor' on the NIH clinical trials registry. Here, the sample is a cross-section of all sponsors that conducted two or more clinical trials from 2015 to 2019. We use as dependent variable either an indicator variable taking value one if the firm started at least one COVID-19 clinical trial (column 1), or the number of COVID-19 clinical trials started by the firm (column 2). The key variables of interest are the location of the trial sponsor, size of the sponsor, whether the sponsor was a private or a for-profit entity, and whether the sponsor had prior experience with R&D directed towards respiratory diseases. We code the location of a sponsor based on the location of its clinical trials, and complemented the reported location with manual checks. The location category 'Europe' includes European Union Members as well as the United Kingdom and Switzerland. Sponsor size refers to the number of trials undertaken by the sponsor from 2015 to 2019 and enters the regression in log form. Prior experience with respiratory diseases is an indicator variable that takes value one if the firm has conducted at least one trial related to lower respiratory or upper respiratory infections from 2015 to 2019. Private sponsor is an indicator variable taking value one if the sponsor is tagged as 'Industry' in the clinicaltrials.gov data. Column 1 is estimated using OLS, and column 2 using a Poisson regression. Robust standard errors are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Descriptive statistics on COVID-19 vaccine candidates (n=222)

	Mean	S.D.	Min	Max
Months to preclinical	6.05	2.91	2.00	12.00
Entered clinical trials (0/1)	0.25	0.44	0.00	1.00
Nr of phase progressions	0.42	0.85	0.0	4.0
Vaccine developer location				
U.S.	0.20	0.40	0.00	1.00
China	0.09	0.29	0.00	1.00
Europe	0.25	0.43	0.00	1.00
ROW	0.45	0.50	0.00	1.00
Vaccine developer status				
Private organization	0.36	0.48	0.00	1.00
Public-private collaboration	0.16	0.37	0.00	1.00
Vaccine technology				
DNA	0.10	0.31	0.00	1.00
Protein subunit	0.35	0.48	0.00	1.00
RNA	0.13	0.34	0.00	1.00
Viral Vector (replicating)	0.09	0.29	0.00	1.00
Viral Vector (non-replicating)	0.13	0.33	0.00	1.00
Virus Like Particle	0.08	0.27	0.00	1.00
Inactivated Virus	0.08	0.27	0.00	1.00
Other	0.26	0.44	0.00	1.00

Notes: This table is based on various vintages of the WHO COVID-19 vaccine landscapes (https://www.who.int/publications/m/item/draft-landscape-of-covid-19-candidate-vaccines) supplemented by hand curated data on COVID-19 vaccine developers and their location. Vaccine progress and development is as of December 2020. The first panel presents the months it took for a COVID-19 vaccine candidate to enter the preclinical stage, conditional on entering the preclinical stage by the end of 2020, whether the candidate entered into a clinical trial, and number of phase progression (a phase progression can be from stage 1 to stage 2, or from stage 2 to stage 3, or from stage 3 to approval). The second panel presents the location of the vaccine developer by region of the world (with ROW referring to rest of the world). The third panel presents additional characteristics of the vaccine candidate based on whether it was developed by a private organization and whether the candidate was developed as part of public-private collaboration. The last panel presents information on the vaccine technology as reported by WHO.

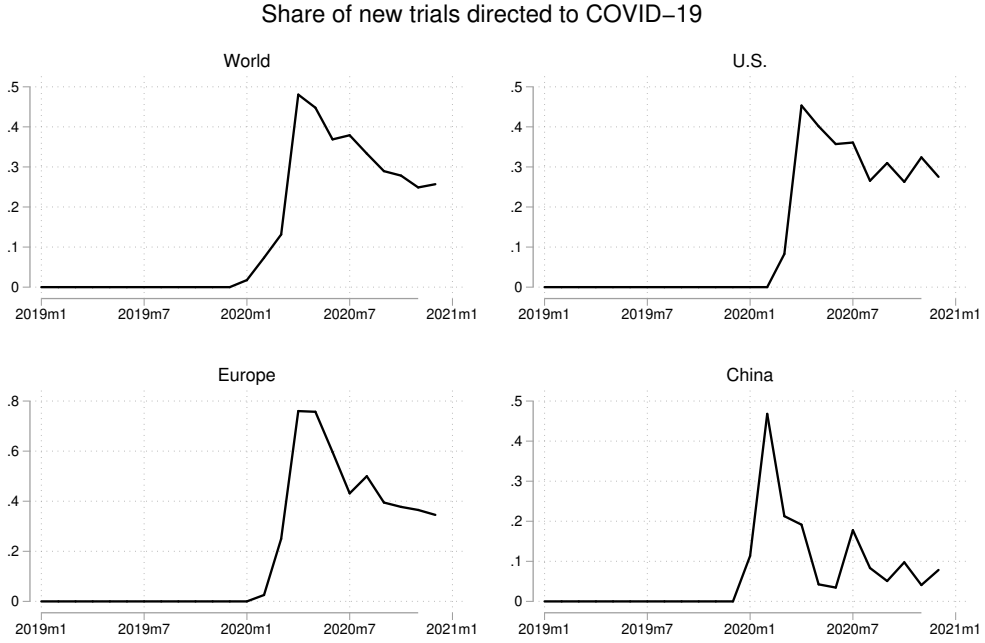
Table 7: Speed & Success of COVID-19 Vaccine R&D Effort

	(1) Months to preclinical	(2) Entered clinical trials (0/1)	(3) Nr of phase progressions
Vaccine developer location			
U.S.	-1.189**	0.163*	0.406
	(0.602)	(0.097)	(0.459)
China	-1.462*	0.341**	1.074**
	(0.757)	(0.140)	(0.530)
Rest of the World	-0.302	0.037	0.065
	(0.502)	(0.065)	(0.402)
Europe: omitted category			
Vaccine developer status			
Private Organization	-1.908***	0.057	0.294
	(0.427)	(0.069)	(0.252)
Public-Private Collaboration	-1.986***	-0.073	-0.091
	(0.516)	(0.075)	(0.394)
Control			
Preclinical Month		-0.015	-0.141**
		(0.012)	(0.058)
Observations	222	222	222
Mean of dep. var.	6.050	0.252	0.422

Notes: In this table, we examine the speed and success of COVID-19 vaccine development across the 222 COVID-19 vaccine candidates listed in the World Health Organization (WHO) COVID-19 candidate vaccine landscape database as of December 2020. The dependent variables are (a) the month in which the vaccine candidate first appeared as a pre-clinical candidate in the dataset (column 1), (b) an indicator variable that takes the value of one if the vaccine candidate entered clinical trials (column 2), and (c) number of phase progressions experienced by the vaccine candidate in the year 2020 (column 2). We code the number of phase progression as 1 for vaccines in phase 1 by the end of 2020, 1.5 for vaccines in phase 1/2, 2 for vaccines in phase 2, 2.5 for vaccines in phase 2/3, 3 for vaccines in phase 3, and 4 for the two vaccines approved in the U.S. by the end of 2020. The location of the lead developer was manually collected. The location category Europe includes European Union Members as well as the United Kingdom and Switzerland. Private sponsor is an indicator variable taking value one if the sponsor is tagged as 'Industry' in the clinicaltrials.gov data, whereas the Public-Private Collaboration takes the value of one when multiple sponsors are present with at least one industry sponsor and one non-industry sponsor. Columns 1 and 2 are estimated using OLS, and Column 3 is estimated using a Poisson regression. Robust standard errors are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

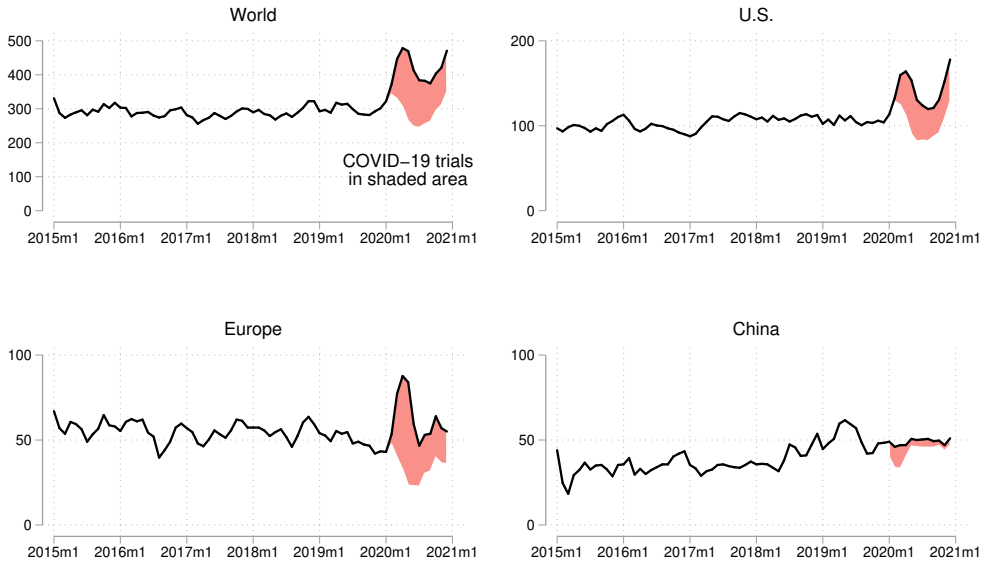
Appendix 1: Additional Figures and Tables

Figure A1: Share of COVID-19 Trials among Newly Started Trials, by Region



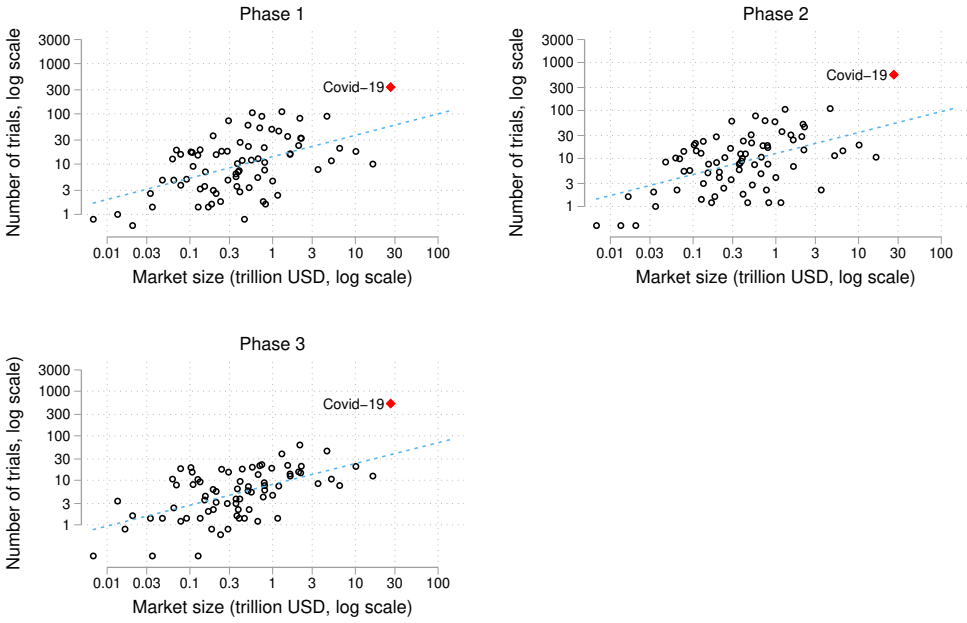
Notes: Based on clinicaltrials.gov data, the figure displays the share of new trials that were related to COVID-19 among all new trials on a monthly basis in the world, U.S., Europe, and China. Europe includes European Union Members as well as the United Kingdom and Switzerland. For example, a number of .4 implies that 40% of the new trials in a given month were related to COVID-19. The COVID-19 trials are tagged as such by clinicaltrials.gov based upon a set of keywords related to COVID-19.

Figure A2: New Clinical Trials Started Monthly for All Diseases, by Region



Notes: Based on clinicaltrials.gov data, the figure displays the total of new clinical trials (across all diseases) on a monthly basis in the world, U.S., Europe and China. Europe includes European Union Members as well as the United Kingdom and Switzerland. COVID-19 trials are showed in the shaded area. COVID-19 trials are tagged as such by clinicaltrials.gov based upon a set of keywords related to COVID-19.

Figure A3: COVID-19 R&D Effort Compared with Other Diseases, by Phase



Notes: This figure reproduces Figure 5 but distinguishes research effort by phases. Research effort is defined as the average yearly number of new trials per disease and phase from 2015 to 2019; for COVID-19 we use the actual number of 2020 COVID-19 trials instead. Potential market size is the 2017 disease-level mortality at the national level weighted by national GDP per capita and a value of statistical life of USD 1 million for the mean global citizen, expressed in logs. For COVID-19, our estimate of potential market size is based on the hypothetical case in which in the absence of non-pharmaceutical control measures 70% of the world population eventually would get infected, and COVID-19 has an infection fatality rate of 0.5%. This corresponds to 26.6 million COVID-19 deaths. See main text for a detailed discussion of the COVID-19 potential market size.

Table A1: Descriptive statistics on disease cross-section (n=75)

	Mean	S.D.	Min	Max
Total trials from 2015 to 2019	45.0	54.92	1.40	253.80
Deaths (millions)	0.6	1.33	0.01	8.90
Potential market size	1.1	2.38	0.01	16.25
DALYs (millions)	21.9	36.11	0.31	185.49
Change in mortality from 2007 to 2017	0.1	0.45	-0.91	1.48
Share of deaths in high-income countries	0.3	0.18	0.00	0.73
Infectious disease	0.2	0.38	0.00	1.00
Cancer	0.4	0.48	0.00	1.00
Ratio of Phase III to Phase 1 trials	0.4	0.64	0.00	4.00

Notes: This table presents the descriptive statistics for the 75 diseases in our sample. For each variable, the table presents the mean, standard deviation, and the range. Total trials corresponds to the number of clinical trials between 2015 and 2019. Deaths and Disability-Adjusted Life Years (DALYs) are measures of the disease burden measured in millions. Potential market size refers to our income-based measure of potential global market size of the disease measured in trillions of USD. Infectious disease and cancer represent the share of diseases that are classified as such. Change in mortality is measured in deaths per 100K. Share of deaths in high-income countries and Ratio of Phase III to Phase I trials are presented as a percentage.

Table A2: Descriptive statistics on COVID-19 trials (n=1,433)

	Mean	S.D.	Min	Max
Private sponsor	0.31	0.46	0.00	1.00
Sponsor has prior experience with respiratory diseases	0.12	0.32	0.00	1.00
Sponsor conducted more than 50 trials from 2015 to 2019	0.50	0.50	0.00	1.00
Trial location				
US	0.32	0.47	0.00	1.00
Europe	0.23	0.42	0.00	1.00
China	0.05	0.22	0.00	1.00
ROW	0.40	0.49	0.00	1.00

Notes: This table presents the descriptive statistics on COVID-19 trials in 2020. Private sponsor corresponds to whether the trial was sponsored by a private pharmaceutical firm. Characteristics of sponsors who conducted the COVID-19 trial are also represented in terms of whether the sponsor had prior experience with respiratory diseases and whether the sponsor had conducted more than 50 trials between the period 2015 and 2019. The second panel represents geographic information about the trial location.

Table A3: Robustness to market size and R&D elasticity assumptions

		COVID-19 Market Size		
		Low (0.001 IFR)	Mid-point (0.005 IFR)	High (0.01 IFR)
Elasticity	Low (0.25)	26.1	17.4	14.7
	Middle (0.43)	19.2 (*)	9.5 (**)	7.1 (***)
	High (0.6)	14.5	5.5	3.7

Notes: Each cell in the table presents the ratio of (1) Actual COVID-19 trials over (2) Predicted Number of COVID-19 trials based on market size. For instance, the number 3.7 in the lower right cell should be read as 3.7 times more trials compared to what would have been expected based on market size. To derive the expected number of COVID-19 trials based on potential market size, we combine an estimate of (a) the relationship between research effort and (b) potential market size with an estimate of the COVID-19 potential market size. The columns and rows correspond to different values of each. For the COVID-19 market size, the estimates are based upon different assumptions on the infection fatality rates (IFR) - 1 per 1000 in the low, 1 per 500 in the mid-point and 1 per 100 in the high estimate. The different values for the elasticity of R&D effort with respect to market size are 0.25, 0.43 and 0.6. The middle estimate comes from our regression of log research effort on log market size, which yields a point estimate of 0.43 (See Table 2). The low estimate of 0.25 and the high estimate of 0.6 are based upon the literature. In the case of the middle estimate for the elasticity -but not the others- we are able to produce a 95% prediction interval for the predicted research effort. (*) 95% prediction interval: 11.8-31.2 (**) 95% prediction interval: 4.7-19.5 (***) 95% prediction interval: 3.1-16.0.

Appendix 2: A Simple Model of Directed R&D

Model

Overview

This section builds on the model of Acemoglu and Linn (2004) to study the impact of market size on innovation, with a special focus on innovation directed towards COVID-19 innovation. The model makes three modifications to their model: (a) *Disease-Specific Taxation*: To take into account the feature that government's often pay for drugs on behalf of the citizens—especially that is needed by a large fraction of the population (e.g. vaccines)—we allow for disease-specific taxation, (b) *Non-Monetary Incentives for Innovation*: We consider the possibility that researchers may have non-monetary motivation (i.e. altruism motive or reputational incentives) to pursue R&D, and (c) *Economy-Wide Decreasing Returns*: To capture the effect of limited supply of top researchers, we allow for economy-wide decreasing returns to scale to R&D effort (instead of disease-specific decreasing returns to scale as in the original model). Other than these three modifications, the setup of the model is identical to Acemoglu and Linn (2004), which we follow closely.

Setup

Consider a small open economy consisting of a set I of infinitely lived individuals. Time is continuous $t \in [0,\infty)$. There are two types of goods in this economy. First, there is a basic good y, which can be used for consumption, for the production of other goods, or for research expenditure. Each individual i has an exogenously given endowment $y_i(t)$ at time t. Second, there are J drugs, $x_1,...,x_J$, each with a potentially time-varying "quality," $q_1(t),...,q_J(t)$. Each individual demands only one type of drug. Here the term drug is broadly used to refer to either vaccines, treatments, or therapeutics. We partition the set I of individuals into J disjoint groups, $G_1,...,G_J$ with $G_1 \cup G_2 \cup \ldots \cup G_J = I$, such that if $i \in G_j$, then individual i demands drug j. More specifically, if $i \in G_j$, then his preferences are given by:

$$\int_0^\infty \exp(-rt)\left[(c_i(t))^{1-\gamma}(q_j(t)x_{ij}(t))^{\gamma}\right]dt \tag{3}$$

where r is the discount rate of the consumers and the interest rate faced by the economy, $\gamma \in (0,1)$, $c_i(t)$ is the consumption of individual i of the basic good at time t, and $x_{ij}(t)$ is the consumption of drug j. The Cobb-Douglas functional form and the assumption that each individual only consumes one type of drug are for simplicity as in Acemoglu and Linn (2004). Normalizing the price of the basic good to 1 in all periods, and denoting the price of drug j at time t by $p_j(t)$, the demand of individual $i \in I$ for drug j is $x_{ij}(t) = \gamma y_i(t)/p_j(t)$ if $i \in G_j$, and $x_{ij}(t) = 0$ if $i \notin G_j$. Summing across individuals, total demand for drug j is:

$$X_j(t) = \frac{\gamma Y_j(t)}{p_j(t)} \tag{4}$$

where $Y_j(t) \equiv \sum_{i \in G_j} y_i(t)$ is the total income of the group of individuals consuming drug j.

At a given point in time, there is one firm with the best-practice technology for producing each type of drug, and it can produce one unit of this drug with quality $q_j(t)$ using one unit of the basic good. If there is an innovation for drug line j currently with quality $q_j(t)$, this leads to the discovery of a new drug of quality $\lambda q_j(t)$, with $\lambda > 1$. It is assumed that any innovation can be sold to consumers immediately (and is under patent protection indefinitely). There is free entry into R&D, and each firm has access to an R&D technology that generates a flow rate j of innovation for every dollar spent for research on drug j. So if R&D expenditure at time t is $z_j(t)$, the flow rate of innovation (and of entry of new drugs) for drug j is:

$$n_j(t) = \delta_j z_j(t) \phi(Z(t)) \tag{5}$$

Differences in $\delta_j > 0$ introduce the possibility that technological progress is scientifically more difficult in some lines than others.

In addition, $\phi'(Z) \leq 0$, and Z represents the total innovation across all drugs, i.e. $Z(t) = \sum_{j \in 1,\ldots,J} z_j(t)$. This assumes that there is within-period decreasing returns to R&D effort with respect to total R&D effort in the economy. The intuition behind this assumption is there is a limited supply of top researchers, so when economy-wide R&D effort gets scaled up, the average quality of R&D effort may decline. (By contrast Acemoglu and Linn (2004) consider the case of disease-specific decreasing returns, i.e. with respect to z_j and not overall Z). For simplicity, we assume that

the economy-wide decreasing function takes the form of:

$$\phi(Z(t)) = \frac{R\&D\,Capacity}{Total\,R\&D} = \frac{C(t)}{Z(t)} \tag{6}$$

with R&D capacity $C(t) > 0$ for all t. The R&D capacity is potentially a function of push incentives such as government funding for research. The R&D technology is directed in the sense that firms can devote their R&D to developing particular types of drugs. The demand curves in (5) have an elasticity equal to one, so an unconstrained monopolist would charge an arbitrarily high price. However, the firm with the best drug in line j is competing with the next best drug in that line. An arbitrarily high price would allow the next best firm to capture the entire market. Therefore, the firm with the best drug sets a limit price to exclude the next best firm—i.e., to ensure that consumers prefer its product rather than the next best drug supplied at the lowest possible price (i.e., equal to marginal cost, which is 1). As in Acemoglu and Linn (2004), the limit price for all j and t is given by:

$$p_j(t) = \lambda \tag{7}$$

The flow profits of the firm with the best product of quality $q_j(t)$ in line j at time t is given by:

$$\pi_j(q_j(t)) = (1-\tau_j)(\lambda-1)\gamma Y_j(t) \tag{8}$$

Here $0 \leq \tau_j < 1$ is a drug-specific tax rate imposed by the government, and $Y_j(t)$ corresponds to the market size (total sales) for drug j. Firms are forward-looking, and discount future profits at the rate r. The discounted value of profits for firms can be expressed by a standard dynamic programming recursion. $V_j(t \mid q_j)$, the value of a firm that owns the most advanced drug of quality q_j in line j at time t, is

$$rV_j(t \mid q_j) - \dot{V}_j(t \mid q_j) = \pi_j(q_j(t)) - (1-\eta)\left[\delta_j z_j(t)\phi(Z(t))V_j(t \mid q_j)\right] \tag{9}$$

where $\pi_j(q_j(t))$ is the flow profits, and $z_j(t)$ is R&D effort at time t on this line by other firms. The value of owning the best technology in line j, $rV_j(t \mid q_j)$, is equal to the flow profits, $\pi_j(q_j(t))$, plus the potential appreciation of the value, $\dot{V}_j(t \mid q_j)$, and takes into account that at the flow rate $n_j(t) = \delta_j z_j(t)\phi(Z(t))$ there will be an innovation, causing the current firm to lose its leading posi-

tion and to make zero profits thereafter. The parameter $\eta \in [0,1]$ captures the non-monetary incentives and represents the degree to which the firm cares about bringing the innovation to market—independent of the possibility of its innovation being replaced by a new entrant at a future date. Another way to interpret the case of $\eta > 0$ is that if a firm gains reputation value from being first to bring an innovation to the market then it would still derive non-monetary benefits from the innovation even after it has been replaced by new entrants. Note that the Acemoglu and Linn (2004) model has $\tau_j = 0$ and $\eta = 0$.

Free entry into R&D to develop better quality drugs implies zero profits; i.e., for all j and t:

$$if \quad z_j(t) > 0, \quad then \quad \delta_j \phi(Z(t)) V_j(t \mid q_j) = 1 \tag{10}$$

Meanwhile, if $z_j(t) = 0$, then $\delta_j V_j(t \mid q_j) \leq 1$ and there will be no equilibrium R&D for this drug.

Equilibrium

An equilibrium in this economy is given by sequences of prices $p_j(t) \mid_{j=1,\ldots,J}$ that satisfy (8), consumer demands for drugs $x_i(t) \mid_{i \in I}$ that satisfy (5) and R&D levels $z_j(t) \mid_{j=1,\ldots,J}$ that satisfy (11) with $V_j(\cdot)$ given by (10). Differentiating equation (11) with respect to time implies that $\dot{V}_j(t \mid q_j) = 0$ for all j and t, as long as $z_j(t) > 0$. Substituting this equation and (11) into (10), and assuming $Y_j(t) = Y_j$ for all t yields the levels of R&D effort in the unique steady state equilibrium as:

$$z_j^S = max \left\{ \frac{C\delta_j(1-\tau_j)(\lambda-1)\gamma Y_j - r\sum_{j \notin J} z_j^S}{(1-\eta)C\delta_j + r}; 0 \right\} \tag{11}$$

Let drugs J_{pos} represent the subset of drugs for which $z_j(t) > 0$. Then, the total research effort in the steady state is given by:

$$Z^S = \frac{C \cdot \sum_{j \in J_{pos}} ((1-\tau_j)(\lambda-1)\gamma Y_j)}{(1-\eta)C + r\sum_{j \in J_{pos}} \left(\frac{1}{\delta_j}\right)} \tag{12}$$

See appendix A.1 for derivations.

Implications for Directed Innovation

This model yields predictions about the direction of innovation in the pharmaceutical industry. These are presented below as six propositions to motivate empirical findings. Propositions 1 is consistent with the results of Acemoglu and Linn (2004), and do not depend on our extension of their model, while the other four propositions depend on our extensions of their model.

Proposition 1 characterizes the model's implication with respect to directed R&D effort and market size.

Proposition 1 (**Market Size Effect (with No Tax)**). *When tax does not depend on market size and firms are risk-neutral, a bigger market size leads to more directed R&D expenditure towards that drug by profit-maximizing firms, and this relationship is linear (i.e. there is no diminishing efforts). That is,* $\frac{dz_j^S}{dY_j} \geq 0$, *and* $\frac{d^2 z_j^S}{dY_j^2} = 0$.

This result states that R&D effort is likely to be directed towards developing drugs for diseases with larger market size—and this relationship is linear in absence of taxes or risk aversion. The concept of market size (Y_j) here corresponds to the total income of the group that demand drug j. Thus, the model predicts that R&D effort will more likely be directed towards diseases that affect a larger population and/or diseases that affect higher-income populations (e.g. those that live in advanced economies). Thus, according to Proposition 1 when the market size of COVID-19 drugs becomes bigger, more R&D effort will be directed towards COVID-19 as firms will find it more profitable to do so. What does this imply about the cross-effect on non-COVID-19 R&D effort, and on total R&D effort across all diseases? The model's implications for aggregate R&D effort and crowding out of other non-COVID R&D is presented in Proposition 2.

Proposition 2 (**Effect of Market Size on Aggregate R&D Effort and the Crowding Out Effect**). *An increase in R&D effort directed towards COVID crowds out R&D effort towards other diseases, but the size of crowding out is smaller when installed R&D capacity is bigger. That is, for all* $j \neq COVID$, $\frac{dz_j^S}{dz_{COVID}} \leq 0$, *and* $\mid \frac{d^2 z_j^S}{dz_{COVID} dC} \mid \leq 0$. *At the same time, aggregate R&D effort increases, i.e.* $\frac{dZ^S}{dz_{COVID}} \geq 0$.

As per Proposition 2, the model predicts that when the market size of COVID-19 drugs becomes larger it leads to a reduction in R&D effort for non-COVID-19 diseases (depending on the magnitude

of installed R&D capacity), while at the same time it leads to an increase in total R&D effort. This implies that under the model's assumptions, the crowding out effect of market size discussed in Proposition 1 is less than one-for-one.

Proposition 3 considers the role of non-monetary incentives in shaping R&D effort.

Proposition 3 (**Role of Non-Monetary Incentives**). *Firms with greater non-monetary incentives exert more R&D effort, i.e.* $\frac{dz_j^S}{d\eta} > 0$. *Moreover, if* $\frac{dz_j^S}{dY_j} \geq 0$, *then firms with greater non-monetary incentives exert more R&D effort for a drug with a larger market size, i.e.* $\frac{d^2 z_j^S}{dY_j d\eta} > 0$.

In our model, the non-monetary incentives act as an added motive to conduct R&D effort. To see this, first note that in the classic creative destruction models upon which the model of Acemoglu and Linn (2004) is based, the possibility of having your innovation being replaced by another competitor works as a deterrent to innovation. In our model, however, firms that have greater non-monetary incentives (i.e. higher η) are less likely to internalize the deterring effect of being replaced by another firm (as they are value being first to bring an innovation to market beyond the monetary incentives). Thus, high η firms are likely to exert more R&D effort. This same effect makes them more likely to exert R&D effort towards drugs with a larger marker size, as they get the benefits of market size while being relatively less impacted by being replaced compared to low η firms.

The first three propositions considered a case with no taxation. The next two propositions consider how disease-specific taxation impacts the relationship between market size and R&D effort.

Proposition 4 (**The Market Size Paradox and the Diminishing Effort**). *If after-tax profits from a drug are subject to progressive taxation, i.e.* $\frac{d\tau_j}{dY_j} > 0$ *and* $\frac{d^2(1-\tau_j)Y_j}{dY_j^2} < 0$, *then directed R&D expenditure for drug* j *is concave with respect to market size, i.e.* $\frac{d^2 z_j^S}{dY_j^2} < 0$.

Proposition 4 studies the case in which the ex-post taxation of a pharmaceutical firm's profits may depend on the market size. As discussed in the main text, we consider the case in which the pharmaceutical products may be subject to progressive taxation *de facto* due to political or social considerations. The notion of progressive taxation here relates to situation in which the tax rate rises with the market size and the after-tax profits are a concave function of the market size. A simple functional form for the tax rate that satisfies the progressive taxation assumption is: $\tau_j = 1 - k \cdot Y_j^{-\beta}$ for some non-negative constants k, β with $\beta < 1$. This proposition shows

why this could be one reason why we observe a concave relationship between the R&D effort directed towards development of drugs for a disease and its market size. Moreover, as discussed in proposition 5 below, when the convexity of taxation becomes relatively high, we can have a situation in which an increase in market size can lead to a decrease in R&D effort. We see the results in Proposition 4 and 5 as one possible mechanism (among others) by which one may generate the diminishing efforts in the model.

Proposition 5 (**The Market Size Paradox and the Reversal Effect**). *If the elasticity of taxation with respect to market size exceeds the inverse tax ratio at some point, i.e.* $\frac{d\ln\tau_j}{d\ln Y_j} > \frac{(1-\tau_j)}{\tau_j}$, *then we the market size paradox can lead to reversal effects, such that an increase in market size reduces directed R&D expenditure towards drug j, i.e.* $\frac{dz_j^S}{dY_j} < 0$.

The next section discusses how we map these propositions to our empirical propositions.

Mapping the Model to Data

The main predictions of the model relate to the R&D innovation in the pharmaceutical industry responds to (a) market size (Y_j), (b) non-monetary incentives, and (c) disease-specific taxation regimes (τ_j). To test the predictions with respect to market size we rely on variation in the income-weighted death burden of different diseases. Thus, we interpret j as a given disease (e.g. COVID-19, Malaria, etc.), and evaluate Y_j by summing up the national death burden weighted by the GDP per capita of each country. Acemoglu and Linn (2004) map the model to the data by taking a zeroth Taylor approximation of equation (12) around $r = 0$. We follow a similar approach to derive the estimation equation for Table 2, by taking a second-order multivariate Taylor approximation of equation (12) around $r = 0$ and $z_j^S = 0$ for $j \notin J$. To account for *de facto* progressive taxation we suppose $\tau_j = 1 - k \cdot Y_j^{-\beta}$ for some non-negative constants k, β with $\beta < 1$ (as in Proposition 4). Then, we can re-write equation (12) as:

$$\ln z_j^S = \ln(\lambda - 1) + \ln\gamma + \ln k - \ln(1-\eta) + (1-\beta)\ln Y_j - r/C\delta_j(1-\eta)$$

Further, one can account for public investment in science and support for research (i.e., push incentives) that may increase R&D capacity and relax the capacity constraint. Suppose that $C = \overline{C} \cdot (1+P)^a$, where P is a measure of public investment in science or push incentives, a is a constant

between zero and one, and $\overline{C}$ is a positive constant. Then, we can further re-write equation (12) as:

$$\ln z_j^S = \alpha_0 + \alpha_1 \cdot P + \eta + (1-\beta)\ln Y_j$$

where $\alpha_0 = \ln(\lambda - 1) + \ln\gamma + \ln k - r/\overline{C}\delta_j(1-\eta)$ and $\alpha_1 = ra/\overline{C}(1-\eta)$ and we have used a second order Taylor expansion around $P = 0$.

That is, log R&D effort is a function of the log market size Y_j, non-monetary incentives η, public investment in science P, and a constant term. Here, the elasticity of R&D effort with respect to market size is given by $1-\beta$. This is the relationship we estimate in Table 2.

Proofs and Derivations

A.1: Derivation of Steady State R&D Effort

Differentiating equation (11) with respect to time implies that $\dot{V}_j(t \mid q_j) = 0$ for all j and t, as long as $z_j(t) > 0$. Substituting this equation and (11) into (10), and assuming $Y_j(t) = Y_j$ for all t yields the levels of R&D effort in the unique steady state equilibrium as:

$$z_j^S = max\left\{\frac{(1-\tau_j)(\lambda-1)\gamma Y_j \delta\phi(Z^S) - r}{\delta\phi(Z^S)(1-\eta)}; 0\right\} \tag{13}$$

Based on (4) we have $\phi(Z^S) = \frac{C}{Z^S}$, and $Z^S = \sum_{j\in 1,\ldots,J} z_j^S$. Substituting these into (13) above we get:

$$z_j^S = max\left\{\frac{C\delta_j(1-\tau_j)(\lambda-1)\gamma Y_j - r\sum_{j\notin J} z_j^S}{(1-\eta)C\delta_j + r}; 0\right\} \tag{14}$$

Let drugs J_{pos} represent the subset of drugs for which $z_j(t) > 0$. Substituting $Z^S = \sum_{j\in 1,\ldots,J} z_j^S$ in the equation for $\phi(Z^S) = \frac{C}{Z^S}$, and summing over all j in (13) we can total research effort in steady state as:

$$Z^S = \sum_{j\in 1,\ldots,J_{pos}} z_j^S = \frac{C\cdot\sum_{j\in J_{pos}}((1-\tau_j)(\lambda-1)\gamma Y_j)}{(1-\eta)C + r\sum_{j\in J_{pos}}\left(\frac{1}{\delta_j}\right)} \tag{15}$$

A.2: Proof of Propositions

Proposition 1: The first proposition considers the case with $\frac{d\tau_j}{dY_j} = 0$. Under this case if $z_j^S > 0$, taking the derivative in (14) we get:

$$\frac{dz_j^S}{dY_j} = \frac{C\delta_j(1-\tau_j)(\lambda-1)\gamma}{(1-\eta)C\delta_j + r}$$

and the second derivative is zero: $\frac{d^2 z_j^S}{dY_j^2} = 0$.

The first derivative is positive since by assumption, $\lambda > 1, r \geq 0, C > 0, \delta_j > 0, 0 \leq \tau_j < 1, \eta \leq 1$. Thus, we have the result for Proposition 1: if $\frac{d\tau_j}{dY_j} = 0$ then $\frac{dz_j^S}{dY_j} \geq 0$ and $\frac{d^2 z_j^S}{dY_j^2} = 0$.

Proposition 2: For the second proposition consider the case with $z_j^S > 0$. Then taking the derivative in (14) we get for all $j \neq COVID$:

$$\frac{dz_j^S}{dz_{COVID}} = -\frac{r}{(1-\eta)C\delta_j + r} < 0$$

which is negative under our parameter assumptions. Thus, we have the result presented in Proposition 2: for all $j \neq COVID$, $\frac{dz_j^S}{dz_{COVID}} \leq 0$. Also, the magnitude of crowding is smaller when the installed R&D capacity (C) is larger such that: and $|\frac{d^2 z_j^S}{dz_{COVID} dC}| = -r(1-\eta)\delta_j = \leq 0$.

In addition, taking derivative of (15) with respect to z_{COVID} we get:

$$\frac{dZ^S}{dz_{COVID}} = \frac{dZ^S}{dY_{COVID}} \frac{dY_{COVID}}{dz_{COVID}} = \frac{(1-\tau_j)(\lambda-1)\gamma C}{(1-\eta)C + r\sum_{j \in J_{pos}} \left(\frac{1}{\delta_j}\right)} \left(\frac{dY_{COVID}}{dz_{COVID}}\right)$$

which under our assumptions and the result of Proposition 1 is positive. Thus, we have the result presented in Proposition 2: for all $\frac{dZ^S}{dz_{COVID}} \geq 0$.

Proposition 3: To derive the result for the third proposition we can take the derivative of R&D effort in (14) with respect to η and the cross-derivative with respect to η and market size Y_j to get:

$$\begin{aligned}
\frac{dz_j^S}{d\eta} &= \frac{C\delta_j \left[(1-\tau_j)(\lambda-1)\gamma Y_j C\delta_j - r\sum_{j \notin J} z_j^S\right]}{\left[(1-\eta)C\delta_j + r\right]^2} \\
\frac{d^2 z_j^S}{dY_j d\eta} &= \frac{C^2\delta_j^2(1-\tau_j)(\lambda-1)\gamma\left(1-\tau_j - Y_j\frac{d\tau_j}{dY_j}\right)}{\left[(1-\eta)C\delta_j + r\right]^2}
\end{aligned}$$

Thus, as represented in Proposition 3, under our parameter assumptions, $\frac{dz_j^S}{d\eta} > 0$. And, $\frac{d^2 z_j^S}{dY_j d\eta} > 0$ as long we do not have a paradox of market size effect, i.e. as long as $1 - \tau_j - Y_j \frac{d\tau_j}{dY_j} > 0$ or equivalently as long as $\frac{dz_j^S}{dY_j} > 0$.

Proposition 4: The fourth proposition considers the case τ_j a function of Y_j such that $\frac{d\tau_j}{dY_j} \geq 0$ and $(1-\tau_j)Y_j$ is a concave function of Y_j. The second condition is equivalent to having: $2\frac{d\tau_j}{dY_j} + Y_j \frac{d^2\tau_j}{dY_j^2} > 0$. Then note that when $z_j > 0$, taking the first and second derivatives of z_j^S with respect to Y_j in (14) we get:

$$\begin{aligned} \frac{dz_j^S}{dY_j} &= \frac{C\delta_j(1-\tau_j)(\lambda-1)\gamma\left(1-\tau_j - Y_j\frac{d\tau_j}{dY_j}\right)}{(1-\eta)C\delta_j + r} \\ \frac{d^2 z_j^S}{dY_j^2} &= - \frac{C\delta_j(1-\tau_j)(\lambda-1)\gamma\left(2\frac{d\tau_j}{dY_j} + Y_j\frac{d^2\tau_j}{dY_j^2}\right)}{(1-\eta)C\delta_j + r} \end{aligned}$$

Under our assumptions then we get $\frac{d^2 z_j^S}{dY_j^2} \leq 0$, which is the result in Proposition 4.

Proposition 5: The fifth proposition sets out the conditions under which $\frac{dz_j^S}{dY_j} < 0$. From the first derivative derived in the proof of proposition 4, we can see this is the case when:

$$1 - \tau_j - Y_j \frac{d\tau_j}{dY_j} < 0$$

This is equivalent to the condition $\frac{d\tau_j}{dY_j} > \frac{1-\tau_j}{Y_j}$. Substituting $\frac{d\ln\tau_j}{d\ln Y_j} = \frac{Y_j}{\tau_j}\frac{d\tau_j}{dY_j}$ we can re-write the condition as:

$$\frac{d\ln\tau_j}{d\ln Y_j} > \frac{(1-\tau_j)}{\tau_j}$$

which is the result in Proposition 5.