Supplementary Appendix for "Designing Pull Funding for a COVID-19 Vaccine"

This appendix provides technical details supporting the analysis in the manuscript.

A1. Formal Model

This section provides a formal description of the model. Besides complementing the more conceptual description in the text, the notation introduced here will be useful in the derivation of the optimal mechanism in the next section.

Let n denote the number of potential suppliers, indexed by $i \in \{1, \dots, n\}$. Vaccine firms can undertake a project combining at-risk investment (investment before success in earlier clinical trials is learned) in both phase-3 trials and capacity installation. Let c_{3i} and c_{4i} denote the respective fixed costs of those stages. Conditional on success, the firm expends additional manufacturing costs to produce $q_i \leq \overline{q}$ units, where $\overline{q}$ denotes firm capacity. Letting c_{5i} denote the variable manufacturing cost (sometimes referred to as "cost of goods sold," abbreviated COGS) per dose, total variable manufacturing costs are $c_{5i}q_i$. We simplify the firm's problem by assuming that the firm produces up to full capacity $\overline{q}$ if it produces at all. This assumption is without loss of generality in the scenarios examined in the text since virtually all candidates are selected and net benefits increase with capacity. Let $I_i \in \{0,1\}$ denote the firm's investment decision. Then its output equals $q_i = \overline{q}I_i$.

Let A_i be an indicator for the success of firm i's project, independent draws from a Bernoulli distribution, with $s = \Pr(A_i = 1)$.

Let c_i be firm i's combined cost for all three stages (phase-3 trials, capacity, and COGS) in ex ante expected value terms. At-risk investments are added into c_i on a dollar for dollar basis. COGS are only expected conditional on success, so with probability s. Thus, COGS add into c_i weighted by s. Hence, $c_i = c_{3i} + c_{4i} + sc_{5i}q_i$. Assume that c_i factors in a sufficient profit margin so that if c_i is covered by the award in the mechanism, the firm is willing to undertake investment.

Assume the firm learns c_i prior to any investment. Costs are private information for firms, independent and identically distributed (iid) draws from a distribution with support $[0, \overline{c}]$, probability density function (pdf) $f(c_i)$, and cumulative distribution function (cdf) $F(c_i)$, where the upper bound on the support $\overline{c}$ can be infinite. Throughout the appendix, we use boldface to denote vectors of variables for each firm. Thus, for example, $\mathbf{c} = (c_1, \dots, c_n)$ and $\mathbf{A} = (A_1, \dots, A_n)$.

A2. Derivation of Optimal Mechanism

Building on Myerson's (1981) seminal work on optimal auction design, Chaturvedi and Martínez-de-Albéniz (2011) (hereafter CM) characterize the optimal procurement mechanism of multiple units from multiple suppliers who can suffer failure risk. In this section, we show how the optimal funding mechanism in our model can be derived as a corollary of their general characterization.

A2.1. Generalizing the Model

CM's model is more general than ours. As indicated by an added firm subscript, they allow for heterogeneity in firms' success probabilities, s_i, and investment-cost pdfs, $f_i(c_i)$, and cdfs, $F_i(c_i)$. Rather than studying a fixed investment project, they allow firms to produce a variable quantity, q_i. To accommodate this change, rather than regarding c_i as a fixed investment cost, they regard it as the constant marginal cost for each unit of q_i.

CM define a mechanism as a triple $(\mathbf{y}, \mathbf{x}, \mathbf{q})$, where $\mathbf{y} = (y_1, \ldots, y_n)$ is a vector of ex ante payments $y_i(\mathbf{c}, \mathbf{s})$ to firms (unconditional on success), $\mathbf{x} = (x_1, \ldots, x_n)$ is a vector of ex post payments $x_i(\mathbf{c}, \mathbf{s})$ to firms (conditional on success), and $\mathbf{q} = (q_1, \ldots, q_n)$ is a vector of quantities $q_i(\mathbf{c}, \mathbf{s})$ procured from firms. Let $\mathbf{z} = (z_1, \ldots, z_n)$ denote the vector of expected payments to firms

$$z_i(\mathbf{c}, \mathbf{s}) = y_i(\mathbf{c}, \mathbf{s}) + s_i x_i(\mathbf{c}, \mathbf{s}) \tag{1}$$

CM define payoffs as follows. The expected gross benefit of the funder (denoted player 0) is $U_0(\mathbf{q}, \mathbf{s}) = \mathbb{E}_{\mathbf{A}}\left(V\left(\sum_{i=1}^{n} A_i q_i\right)\right)$, where $\mathbb{E}$ is the expectations operator, here taken over the vector of Bernoulli variables $\mathbf{A}$, and V is some increasing, concave function over realized output. To obtain the funder's net payoff will require payments to be subtracted. Firm i's expected net payoff is $U_i(\mathbf{c}, \mathbf{s}) = z_i(\mathbf{c}, \mathbf{s}) - c_i q_i(\mathbf{c}, \mathbf{s})$.

A2.2. Theoretical Results

CM's main relevant result imposes the following regularity condition on the cost distribution.

Definition (Regularity). *A distribution with pdf $f(c)$ and cdf $F(c)$ is* regular *if and only if $F(c)/f(c)$ is nondecreasing.*

CM derive the following theorem as a corollary of Myerson (1981).

Theorem 1 [Chaturvedi and Martínez-de-Albéniz (2011)]. *Assume $f_i(c_i)$ is regular for all i. Then $(\mathbf{y}, \mathbf{x}, \mathbf{q})$ represents an optimal mechanism in dominant strategy equilibrium if and only if for all $\mathbf{c}$ it satisfies*

$$\mathbf{q}(\mathbf{c}, \mathbf{s}) = \underset{\mathbf{q}}{\operatorname{argmax}} \left[U_0(\mathbf{q}, \mathbf{s}) - \sum_{i=1}^{n} M_i(c_i) q_i(\mathbf{c}, \mathbf{s}) \right], \tag{2}$$

where

$$M_i(c_i) = c_i + \frac{F_i(c_i)}{f_i(c_i)} \tag{3}$$

$$z_i(\mathbf{c}, \mathbf{s}) = c_i q_i(\mathbf{c}, \mathbf{s}) + \int_{c_i}^{\overline{c}_i} q_i(\mathbf{c}_{-i}, t_i, \mathbf{s}) dt_i. \blacksquare \tag{4}$$

We will derive the optimal funding mechanism in our model as a corollary of Theorem 1. The first step is to translate certain variables, recognizing that c_i is the total cost of a 0-1 decision in our model, whereas CM take it as a marginal cost per unit. To make the translation, we will substitute I_i for q_i in Theorem 1. Equation (2) becomes

$$\mathbf{I}(\mathbf{c},\mathbf{s}) = \underset{\mathbf{I}}{\operatorname{argmax}} \left[U_0(\mathbf{I},\mathbf{s}) - \sum_{i=1}^{n} M_i(c_i) I_i(\mathbf{c},\mathbf{s}) \right], \tag{5}$$

where $\mathbf{I} = (I_1, \ldots, I_n)$ is the vector of 0-1 investment decisions; and (4) becomes

$$z_i(\mathbf{c},\mathbf{s}) = c_i I_i(\mathbf{c},\mathbf{s}) + \int_{c_i}^{\overline{c}_i} I_i(\mathbf{c}_{-i}, t_i, \mathbf{s}) dt_i. \tag{6}$$

We next derive a convenient expression for $U_0(\mathbf{I},\mathbf{s})$. Let $w(\mathbf{c},\mathbf{s}) = \sum_{i=1}^{n} I_i(\mathbf{c},\mathbf{s})$, interpreted as the number of winning suppliers, i.e., selected by the mechanism to be eligible to invest and win an award. Let $u(w)$ be the funder's gross surplus (surplus not subtracting off expenditures) if it selects w winners. This is only a function of the number of winners because selected firms produce up to exogenous capacity $\overline{q}$ and are homogeneous except for cost, which only determines payment, not the gross benefit that the funder obtains from them. The binomial distribution of successes from the pool of winners yields the following formula:

$$u(w) = \sum_{i=1}^{w} \binom{w}{i} s^i (1-s)^{w-i} GS(i\overline{q}), \tag{7}$$

where

$$\binom{w}{i} = \frac{w!}{(w-i)!\, i!} \tag{8}$$

and $GS(Q)$ is the funder's surplus, gross of expenditures, when Q units are available each month in the period during which pandemic harm is experienced. In our simulations, $u(w)$ turns out to be concave. We conjecture this is true in general based on the logconcavity of binomial coefficients and concavity of $GS(Q)$. One can show that $GS(Q)$ is concave since it is the sum of gross consumer surpluses—well known to be concave—derived from the vaccine demand curve each month during the pandemic period, summed across those months weighted by pandemic intensity in those months.

Defining marginal benefit $\Delta(w) = u(w) - u(w-1)$, we have $u(w) = \sum_{i=1}^{w} \Delta(i)$. Then

$$U_0(\mathbf{I},\mathbf{s}) = u(w(\mathbf{c},\mathbf{s})) = \sum_{i=1}^{w(\mathbf{c},\mathbf{s})} \Delta(i). \tag{9}$$

The next step is to express the term that is subtracted in equation (5) in more concrete terms. Since costs are iid, the subscript on M_i in equation (5) can be dropped, writing

$$M(c_i) = c_i + \frac{F(c_i)}{f(c_i)}. \tag{10}$$

This is an increasing function since it is the sum of an increasing function and a nondecreasing function, the second function nondecreasing because of the regularity of the distribution. Since $U_0(\mathbf{I}, \mathbf{s})$ is determined by $w(\mathbf{c}, \mathbf{s})$, it is apparent that the term subtracted in objective (5) can be minimized leaving $U_0(\mathbf{I}, \mathbf{s})$ unchanged by selecting the $w(\mathbf{c}, \mathbf{s})$ lowest-cost suppliers for the winners. To provide notation for this selection, arrange supplier costs in ascending order: $c_{[1]} \leq c_{[2]} \leq \cdots \leq c_{[n]}$, where bracketed subscripts are used to indicate order statistics. Then using the preceding insights and substituting from equation (9), (5) can be rewritten

$$\mathbf{I}(\mathbf{c}, \mathbf{s}) = \underset{\mathbf{I}}{\operatorname{argmax}} \left\{ \sum_{i=1}^{w(\mathbf{c},\mathbf{s})} \left[\Delta(i) - M(c_{[i]}) \right] \right\}, \tag{11}$$

implying

$$w(\mathbf{c}, \mathbf{s}) = \underset{w}{\operatorname{argmax}} \left\{ \sum_{i=1}^{w} \left[\Delta(i) - M(c_{[i]}) \right] \right\}. \tag{12}$$

The solution to equation (12) is straightforward: since $\Delta(i)$ is decreasing in i and $M(c_{[i]})$ is nondecreasing in i, the bracketed expression is decreasing in i. Bracketed terms should be added until they become negative. Hence,

$$w(\mathbf{c}, \mathbf{s}) = \max\left\{ i \in \{0, \dots, n\} \middle| \Delta(i) \geq M(c_{[i]}) \right\}. \tag{13}$$

Equivalently, upon inverting,

$$w(\mathbf{c}, \mathbf{s}) = \max\left\{ i \in \{0, \dots, n\} \middle| M^{-1}(\Delta(i)) \geq c_{[i]} \right\}. \tag{14}$$

Equation (14) embodies the procedure for selecting the number of winning firms described in the text and shown in the example in Exhibit 3. We now have the notation to express the procedure more formally. The funder demand schedule, drawn as the light blue curve, is given by $\Delta(i)$. Applying the inverse function M^{-1} gives the reserve-price schedule. Equation (14) states that the optimal number of winning firms is given by the intersection of the reserve-price schedule $M^{-1}(\Delta(i))$, drawn as the dark blue curve, and the cost schedule $c_{[i]}$, drawn as the red curve.

The last step is to derive a concrete expression for the expected award payment to winning firms in equation (6) fixing $w^* = w(\mathbf{c}, \mathbf{s})$. Now $I_i(\mathbf{c}_{-i}, t_i, \mathbf{s}) = 1$ if both $t_i < c_{[w^*+1]}$ and $t_i < M^{-1}(\Delta(w^*))$, while $I_i(\mathbf{c}_{-i}, t_i, \mathbf{s}) = 0$ if either $t_i > c_{[w^*+1]}$ or $t_i > M^{-1}(\Delta(w^*))$. Thus,

$$\int_{c_i}^{\overline{c}_i} I_i(\mathbf{c}_{-i}, t_i, \mathbf{s}) dt_i = \min\left\{ c_{[w^*+1]}, M^{-1}(\Delta(w^*)) \right\} - c_i. \tag{15}$$

Substituting equation (15) into (6) yields award

$$z_i(\mathbf{c}, \mathbf{s}) = \min\{c_{[w^*+1]}, M^{-1}(\Delta(w^*))\} \quad (16)$$

for winning firms and 0 for losing firms. Equation (16) is the payment in a $w^* + 1$ price Vickrey auction with reserve price $M^{-1}(\Delta(w^*))$. Graphically, this award is the highest point in the intersection between the reserve-price schedule (the dark blue curve in Figure 1) and the cost schedule (the red curve).

As in CM, the division of the expected award into ex ante (y_i) versus ex post (x_i) payments is not pinned down in our model. We will take them to be ex post payments to address the concern coming from outside the model that suppliers would default on a supply commitment if paid up front. To obtain the correct expected payment in equation (1), the ex post payment must be scaled up by the reciprocal of the success probability: $x_i(\mathbf{c}, \mathbf{s}) = z_i(\mathbf{c}, \mathbf{s})/s$.

A3. Global Demand

A3.1. Expressing Mortality Losses in Monetary Terms

As stated in the text, our mortality-loss estimates are based on projections by the Imperial College COVID-19 Response Team (Walker *et al*. 2020) of deaths by country. Several steps are required to express these losses in monetary terms. The medical study by Hanlon et al. (2020) suggests that after accounting for age and comorbidities, each COVID death results in an average of 12 years of lost life (YLL). Since this estimate already allocates shorter lifespans to people with comorbidities, we assume one YLL translates into one disability adjusted life year (DALY) without need for further downward adjustment to reflect a proportion of years lived with a disability. To convert DALYs into monetary values, we multiply DALYs lost in a country by three times that country's 2019 GDP per capita, reflecting World Health Organization (WHO) standards for a cost-effective health intervention in a country stated in Marseille et al (2015). According to this standard, a health intervention is cost effective if the cost per DALY saved is less than three times that country's per-capita GDP. We divide the result by the country's population to obtain per-capita figures and further divide by 24 to convert into monthly figures.

A3.2. Descriptive Statistics

Exhibit A1 provides descriptive statistics for the country data used to estimate global vaccine demand. The dataset includes 191 observations, corresponding to the countries for which we have available data. Descriptive statistics for countries' population, GDP, and sources of COVID harm are reported.

A3.3. Static Demand

The last row of Exhibit A1 provides descriptive statistics for the sum of health and economic harms due to COVID across the 191 countries in our dataset. This variable can be translated into a global demand curve for a vaccine that would allow this harm to be avoided. We do this by arranging countries in descending order of total harm and plotting each country's harm against the running total countries' populations plotted so far. The demand curve thus constructed is drawn as the black curve in Exhibit A2.

A3.4. Dynamic Demand

The monthly demand curve is not static but can expand or contract as the pandemic varies in intensity over the timeframe considered. Exhibit A3 illustrates the intensity path assumed in the baseline scenario. The underlying assumption generating this path is that the arrival of herd immunity, an effective treatment, or some other event generates a 5% chance that the pandemic ends each month.

The initial (month-12) cross section in Exhibit A3 is identical to the black curve in Exhibit A2. In subsequent months, the dynamic demand curve in Exhibit A3 shrinks over time because the demand for avoiding expected future harm (harm times the probability that it is experienced) declines.

A3.5. Heterogeneous Population Within Country

The black curve in Exhibit A2 is based on the assumption that the population within a country is homogenous. The added blue curve shows how demand changes when we introduce heterogeneity within each country in the form of a vulnerable subpopulation, constituting a third of its population but accounting for two thirds of its total harm.

Let P_i and H_i be the population and harm for country i in the model with within-country homogeneity. For the model with within-country heterogeneity, let P_i^v and H_i^v be the corresponding variables for the vulnerable subpopulation and P_i^l and H_i^l those for the less-vulnerable subpopulation. One can show the following system of equations yields the assumed one-third/two-thirds concentration of harm: $P_i^v = P_i/3$, $P_i^l = 2P_i/3$, $H_i^v = 2H_i$, and $H_i^l = H_i/2$.

Note the blue demand curve is approximately a rotation of the black one, concentrating more of the value toward the vertical axis, allowing more harm to be relieved with fewer vaccine doses.

A4. Supplier Cost Distribution

Our estimates of the distribution of suppliers' investment costs use proprietary data from an on-going survey conducted by the Coalition for Epidemic Preparedness Innovations (CEPI). The initial wave of the survey was analyzed by Gouglas et al. (2018). In comparison, our analysis covers an accreted sample for phase-3 trial costs and capacity costs and COGS for a focused survey of firms with COVID-19 vaccine candidates.

A4.2. CEPI Survey Details

Cost data included in this study draw from two sources: First, a survey-based cost data collection process (Gouglas *et al*. 2018) for a) Phase IIb/III efficacy testing, Chemistry, Manufacturing and Control (CMC) and regulatory activities associated with development of vaccines against Epidemic Infectious Diseases (EIDs); and b) development of stockpiles of up to 1 million doses of EID vaccines for emergency use in outbreak conditions. Second, CEPI's own data on activities associated with large scale manufacturing capacity expansion to support phase IIb/III studies, scale up and scale out production in response to the COVID-19 pandemic.

Specifically, from September 2017 to January 2018 a cost data collection survey was launched by which 414 organizations conducting EID vaccine R&D were approached, of whom

64 submitted responses, covering 313 vaccine candidates for EIDs in total. Organizations were surveyed if they appeared as EID vaccine development partners in published literature, websites and press releases per the systematic search process described in Gouglas et al 2018.

The definition of full R&D costs for phase IIb/III included whether reported costs covered all or most critical clinical, CMC and other non-clinical activities associated with this R&D phase, including: a) Phase III trial planning and execution costs and justification for these i.e. number of clinical trial sites/locations, number of clinical trial subjects, cost per subject including direct costs and/or assumptions associated with clinical trial conduct; b) scale up manufacturing costs for production of investigational material to support large scale efficacy testing, including pilot plant manufacturing setup costs where these were needed; c) indirect and other program-management costs and/or assumptions associated with clinical trial conduct, etc.

The definition of costs for a stockpile of up to one million doses included whether reported costs covered the production of engineering, consistency and stockpile batches, drug substance manufacturing but not drug product manufacturing and fill finish.

Based on these criteria, a set of 99 vaccine candidate cost entries was constructed, where full R&D costs for phase IIb/III had been reported; and 90 cost entries were constructed for developing up to one million dose EID vaccine stockpiles for emergency use. For these entries, Request for Information- based follow ups with vaccine developers took place by CEPI between 2018 and 2020 to clarify the nature and key drivers of this cost information. No follow ups were pursued for other data out of the 313 EID vaccine candidate sample, where the responses were either incomplete on phase IIb/III or stockpile development, or where development efforts concerned pathogens outside the WHO Blueprint list. For more details on survey design, see Gouglas *et al*. 2018.

In addition to the above, this study draws from current CEPI estimates of manufacturing scale up, scale out and production for pandemic use based on its current portfolio of COVID-19 vaccine project investments.

The definition of large-scale manufacturing setup includes whether costs capture scale-up process and consistency lots for antigen, or scale-out process including technology transfer and consistency lots for antigen. In case of scale-out costs this includes reservation fees, start-up and leasing costs as well as scale-up process for antigen. Costs for building new manufacturing facilities from scratch are excluded.

Large-scale manufacturing rollout accounts for monthly, fill-finished vaccine dose rollout (drug product available for distribution) using current yield, dose estimates and production timing, where yield, dose and timing are varied over a given range across a multiplicity of manufacturing sites of different production scale; and where glass vials are assumed to be in sufficient supply for fill-finish.

In terms of manufacturing rollout costs, COGS per dose is assumed to be a sufficient, overall estimate of full manufacturing rollout cost per dose for the purposes of this study, including Drug Substance and Drug Product (added cost of vial and filling). Its estimation takes into account assumptions on: per dose equivalents of costs to build stockpiles of up to 1 million doses of EID vaccines, after accounting for the production of four engineering, consistency and stockpile batches; no economies of scale i.e. COGS per dose is fixed irrespective of volume of production; additional $2 per dose for vial and filling, assuming vial capacity exists. Delivery costs are excluded.

In terms of variability and spread of phase IIb/III and manufacturing cost estimates in the reported data sample, indirect costs, sectoral affiliation (industry versus non-industry) and

geographical location of product development and manufacturing are all explanatory factors. Moreover, manufacturing costs vary because of how much in-house manufacturing capacity already exists, though this variability does not suggest corresponding differences in capacity to manufacture different volumes of vaccine.

The target pathogens, number of sites and/or subjects of the different phase IIb/III trials for which cost data is included in this study sample are diverse, and consequently the range of phase IIb/III costs is quite wide. Given the current uncertainty around how COVID-19 phase III studies will be structured, and consequently costed, assuming a broader range of phase IIb/III costs associated with EID vaccines is, in the authors' view, a reasonable assumption to make at this point in time. This assumption is further strengthened by the fact that CEPI vaccine project budgets over the past three years—e.g. for Lassa, MERS, and Nipah (see for instance Table 1 in Gouglas *et al*. 2019)—have fallen within a wide range of costs resembling closely the cost estimates reported in Gouglas *et al*. 2018.

A4.2. Descriptive Statistics

Exhibit A4 provides descriptive statistics for survey responses for costs of separate vaccine stages.

A4.3. Maximum Likelihood Estimates of Lognormal Parameters

Assuming investment cost c_{ti} in stage $t \in \{1,2\}$ for firm i has a lognormal distribution, its probability density function (pdf) is given by

$$f_t(c_{ti}) = \frac{1}{\sqrt{2\pi} v_t c_{ti}} \exp\left(-\frac{(\ln c_{ti} - m_t)^2}{2v_t^2}\right) = \phi\left(\frac{\ln c_{ti} - m_t}{v_t}\right)\frac{1}{v_t c_{ti}}, \tag{17}$$

where m_t denotes the lognormal distribution's scale parameter, v_t its shape parameter, and ϕ the standard normal pdf. One can express the lognormal distribution's mean μ_t and variance σ_t^2 as a function of these parameters using standard formulas:

$$\mu_t = \exp\left(m_t + \frac{v_t^2}{2}\right) \tag{18}$$

$$\sigma_t^2 = [\exp(v_t^2) - 1]\exp(2m_t + v_t^2). \tag{19}$$

The method of maximum likelihood (ML) searches for parameters maximize the sum over observations of the natural logarithm of the likelihood of each observation. For the lognormal distribution, the likelihood of an observation is given simply by the pdf in equation (17). Exhibit A5 provides results using the LOGNFIT routine in Matlab for a bias-corrected ML estimation of lognormal parameters. Besides bias correction, these routines have the advantage of providing useful ancillary statistics such as confidence intervals.

A4.4. Fenton-Wilkinson Approximation of Combined Distribution

We seek a method of combining the cost distributions for the two stages into a single distribution representing their sum. There is no general guarantee that the sum of lognormals is lognormal. Following the argument in the text that the lognormal is an attractive functional form for an investment-cost distribution, we will impose the lognormal form on the combined cost distribution. A decision remains, however, because there is no single accepted way of approximating a sum of lognormals with a lognormal. In their comparison of various approaches, Beaulieu, Abu-Dayya, and McLane (1995) find that one of the earliest and simplest, the Fenton-Wilkinson (FW) approximation [after Fenton (1960)] has the best performance under many circumstances. We adopt this approximation here.

Intuitively, the FW approximation matches the first two moments of the lognormal distribution. For independent random variables, the sum of the means equals the mean of the sum (first moment) and the sum of the variances equals the variance of the sum (second moment). These equalities can be manipulated to obtain formulas for the resulting scale and shape parameters, m and v, of the lognormal distribution for the combined investment.

Formally, let μ and σ^2 denote the mean and variance of the combined cost distribution and $\hat{\mu}$ and $\hat{\sigma}^2$ denote their respective estimates. The FW approximation sets

$$\hat{\mu} = \hat{\mu}_1 + \hat{\mu}_2 \tag{20}$$

$$\hat{\sigma}^2 = \hat{\sigma}_1^2 + \hat{\sigma}_2^2. \tag{21}$$

Substituting from equations (18) and (19) into (20) and (21) and solving yields

$$\hat{v} = \sqrt{\ln\left(1 + \frac{\hat{\xi}_1}{\hat{\xi}_0^2}\right)} \tag{22}$$

$$\hat{m} = \ln \hat{\xi}_0 - \frac{\hat{v}^2}{2}, \tag{23}$$

where

$$\hat{\xi}_0 = \exp\left(\hat{m}_1 + \frac{\hat{v}_1^2}{2}\right) + \exp\left(\hat{m}_2 + \frac{\hat{v}_2^2}{2}\right) \tag{24}$$

$$\hat{\xi}_1 = [\exp(\hat{v}_1^2) - 1]\exp(2\hat{m}_1 + \hat{v}_1^2) + [\exp(\hat{v}_2^2) - 1]\exp(2\hat{m}_2 + \hat{v}_2^2). \tag{25}$$

These formulas can be applied iteratively to add more than two lognormals. In the general case of adding k lognormals, we have

$$\hat{\xi}_0 = \sum_{t=1}^{k} \exp\left(\hat{m}_t + \frac{\hat{v}_t^2}{2}\right) \tag{26}$$

$$\hat{\xi}_1 = \sum_{t=1}^{k} [\exp(\hat{v}_t^2) - 1]\exp(2\hat{m}_t + \hat{v}_t^2). \tag{27}$$

Substituting the values from Exhibit A5 into equations (26) and (27) yields $\hat{m} = 0.788$ and $\hat{v} = 0.833$. By construction [equations (20) and (21)], the moments for the combined distribution are $\hat{\mu} = 3.111$ and $\hat{\sigma} = 3.111$. All these estimates are measured in units of billion dollars.

A4.6. Regularity of Lognormal

Our optimal mechanism is derived as a corollary of CM's Theorem 1, which requires the cost distribution to be regular. Thus, we need to prove the lognormal distribution is regular to use it in simulations of the optimal mechanism derived as a corollary of Theorem 1. By Table 3 of Bagnoli and Bergstrom (2005), a lognormal cdf $F(c)$ is logconcave, implying

$$\frac{d^2}{dc^2}\ln F(c) = \frac{f'(c)F(c) - f(c)^2}{F(c)^2} < 0, \tag{28}$$

in turn implying $f(c)^2 - f'(c)F(c) > 0$. But the preceding inequality implies

$$\frac{d}{dc}\left[\frac{F(c)}{f(c)}\right] = \frac{f(c)^2 - f'(c)F(c)}{f(c)^2} > 0, \tag{29}$$

proving that the cost distribution is regular.

A5. Inversion for Reserve Price

Letting Δ be funder demand for the marginal winning firm's investment, the associated reserve price is given by $M^{-1}(\Delta)$, the inverse of the function $M(c)$ defined in equation (10). Lacking a closed-form solution for this inverse, we resort to numerical methods. Letting $c = M^{-1}(\Delta)$, we have $M(c) = \Delta$, implying

$$\phi(c, \Delta) = 0, \tag{30}$$

defining $\phi(c, \Delta) = M(c) - \Delta$. Since $M(c)$ is increasing, $\phi(c, \Delta)$ is increasing in c as well. We use the procedure FSOLVE in Matlab to solve for c in equation (30).

A6. Detail on Baseline Scenario

For reference, Exhibit A6 lists the assumptions in the baseline scenario. Exhibit A7 provides histograms showing the entire distribution across the million simulations for each outcome variable in the baseline scenario. This complements the first row of results in Exhibit 3, which just report means across the simulations in the baseline scenario. As Exhibit A7 shows, outcomes can vary widely across simulations. For example, funder surplus ranges from 0 (in simulations in which no selected candidate is successful) to over $4 trillion.

References

Bagnoli M, Bergstrom T. (2005) "Log-concave probability and its applications," *Economic Theory* 26(2): 445–69.

Beaulieu NC, Abu-Dayya AA, McLane PJ. (1995) "Estimating the distribution of a sum of independent lognormal random variables," *IEEE Transactions on Communications* 43(12): 2869–73.

Chaturvedi A, Martínez-de-Albéniz V. (2011) "Optimal procurement design in the presence of supply risk," *Manufacturing & Service Operations Management* 13(2): 227–43.

Fenton LF. (1960) "The sum of lognormal probability distributions in scatter transmission," *IEEE Transactions on Communication Systems* 8(1): 57–67.

Gouglas D, Le TT, Henderson K, Kaloudis A, Danielsen T, Hammersland NC, *et al*. (2018) "Estimating the cost of vaccine development against epidemic infectious diseases: a cost minimization study." *Lancet Global Health* 6(12): e1386–96.

Gouglas D, Christodoulou M, Plotkin SA, Hatchett R. (2019) "CEPI: driving toward epidemic preparedness and response," *Epidemiologic Reviews* 41(1): 28–33.

Hanlon P, Chadwick F, Shah A, Wood R, Minton J, McCartney G, *et al*. (2020) "COVID- 19—exploring the implications of long-term condition type and extent of multimorbitity on years of life lost: a modelling study," Wellcome Open Research 5:75 [Preprint]. April 23, 2020 [cited May 22, 2020]. Available from: https://wellcomeopenresearch.org/articles/5-75.

International Monetary Fund. (2020) *World Economic Outlook.* Washington (DC); April.

Marseille E, Larson B, Kazi DS, Kahn JG, Rosen S. (2015) "Thresholds for the cost-effectiveness of interventions: alternative approaches," *Bulletin of the World Health Organization* 93(2): 118–24.

Myerson RD. (1981) "Optimal auction design," *Mathematics of Operations Research* 6(1): 58–73.

Walker PGT, Whittaker C, Watson O, Baguelin M, Ainslie KEC, Bhatia S, *et al*. "The global impact of COVID-19 and strategies for mitigation and suppression." London: Imperial College COVID-19 Response Team working paper. March 26, 2020. Article accessed May 26, 2020 on https://dsprdpub.cc.ic.ac.uk:8443/handle/10044/1/77735.

World Bank. (2020a) "GDP per capita (current US$)," *World Development Indicators* [data file]. Available from https://data.worldbank.org/indicator/NY.GDP.PCAP.CD.

World Bank. (2020b) “Population, total,” *World Development Indicators* [data file]. Available from https://data.worldbank.org/indicator/SP.POP.TOTL.

World Health Organization. (2016) *An R&D Blueprint for Action to Prevent Epidemics.* Geneva.

EXHIBIT A1

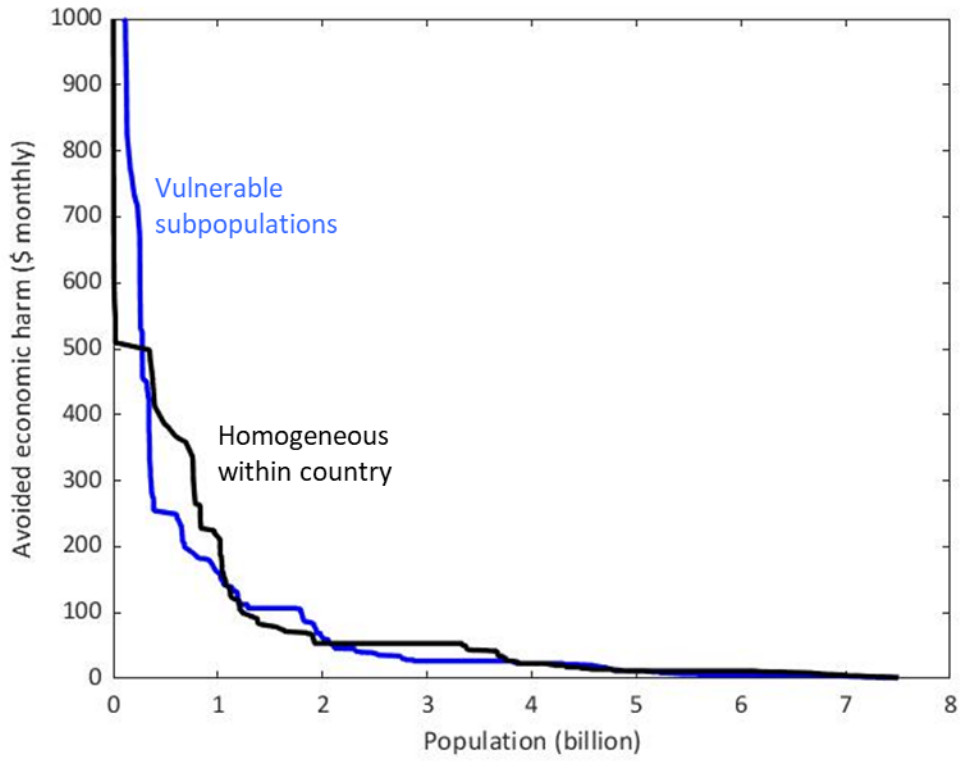
Descriptive statistics for country-level data on used in demand estimation

Vaccine stage	Obs.	Units	Mean	Median	Std. dev.
Country population (2018)	191	Million	39.3	8.5	145.2
Country per-capita GDP (2019)	191	Thousand $	15.1	6.2	20.7
Monthly per-capita harm in country					
Economic-output loss	191	$ per capita	77.5	28.2	111.7
Mortality loss	191	$ per capita	30.2	9.5	45.8
Total	191	$ per capita	107.7	41.4	156.3

SOURCES Country population in 2018 from World Bank (2020a). GDP per capita computed from layering IMF (2020) growth statistics for 2019 on World Bank (2020b) data on 2018 GDP per capita. Economic-output loss from authors' calculation based on IMF (2020) projections layered 2019 per-capita GDP estimates. Mortality loss from authors' calculations based on Walker et al. (2020) mortality projections.

EXHIBIT A2

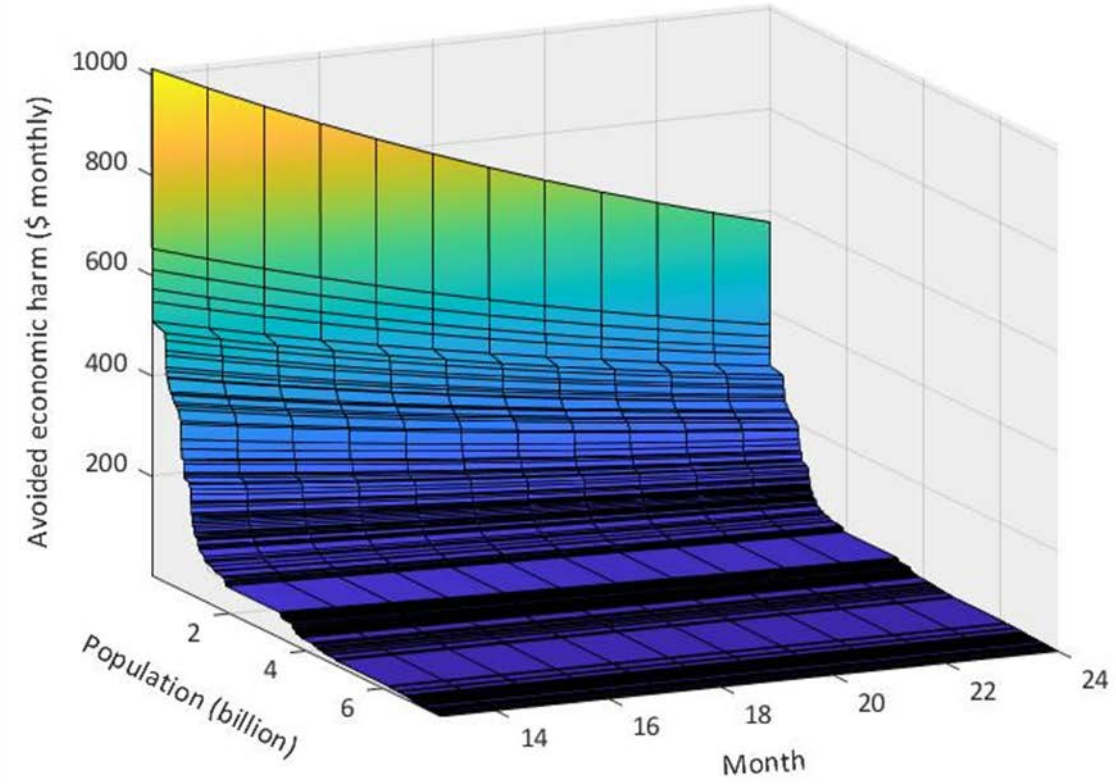
Estimated global demand for COVID vaccine



SOURCE Authors' calculations based on harm data. **NOTES** Black demand curve assumes population is homogeneous within each country. Blue demand curve concentrates 2/3 of harm in the vulnerable subpopulation in each country, constituting 1/3 of the population. Vertical axis censored at 1,000 for legibility.

EXHIBIT A3

Evolution of global demand during COVID pandemic



SOURCE Authors' calculations based on IMF forecasts of economic harm and Imperial College COVID-19 Response Team forecasts of mortality losses. **NOTES** The 12 months on the front-facing axis represent period from first availability of vaccines to end of forecast horizon. Initial (month 12) cross section identical to black curve in Exhibit A2, reflecting total monthly harm, the sum of economic-output losses and mortality losses. Demand cross sections shrink over time reflecting assumed 5% probability that pandemic ends each month due to arrival of herd immunity, a treatment, or some other event that curtails the pandemic, reducing expected harm.

EXHIBIT A4

Descriptive statistics for CEPI survey data used to estimate cost distributions

Vaccine stage	Obs.	Units	Mean	Median	Std. dev.
Phase-3 trials	99	Million $	83.1	30.0	125.3
Capacity installation (per site)	22	Million $	78.3	60.0	63.5
Variable manufaturing costs (per dose)	90	$	13.3	7.0	13.5

SOURCE CEPI survey of firms' investment costs. **NOTES** Capacity installation cost quoted for one site. Two sites assumed to be required for baseline capacity of 750 million doses annually.Variable manufacturing costs, also referred to as COGS, scaled up by the 750 million annual doses to get a total to add into firm's combined cost.

EXHIBIT A5

Maximum-likelihood estimates of lognormal cost distributions

	Vaccine stage, *t*		
	Phase-3 trials (*t* = 3)	**Capacity (*t* = 4)**	**Variable manufacturing costs (COGS) (*t* = 5)**
Units	One set	Two sites	750 million doses
Parameter estimates			
Scale, m_t			
Estimate	-3.119	-2.142	1.871
95% confidence interval	[-3.328, -2.909]	[-2.507, -1.777]	[1.686, 2.055]
Shape, v_t			
Estimate	1.049	0.823	0.880
95% confidence interval	[0.921, 1.220]	[0.633, 1.176]	[0.768, 1.302]
Observations, *n*	99	22	90
Derived moments			
Mean, μ_t	0.083	0.157	9.951
Standard deviation, σ_t	0.125	0.127	10.094

SOURCE Author calculations using CEPI survey of firms' costs. **NOTES** Bias-corrected maximum likelihood estimates produced using the LOGNFIT command in Matlab programming language. Estimates and derived moments in units of billion dollars.

EXHIBIT A6

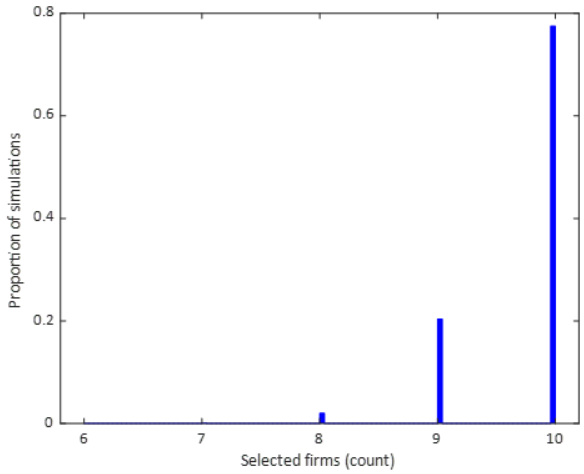
List of assumptions in baseline scenario

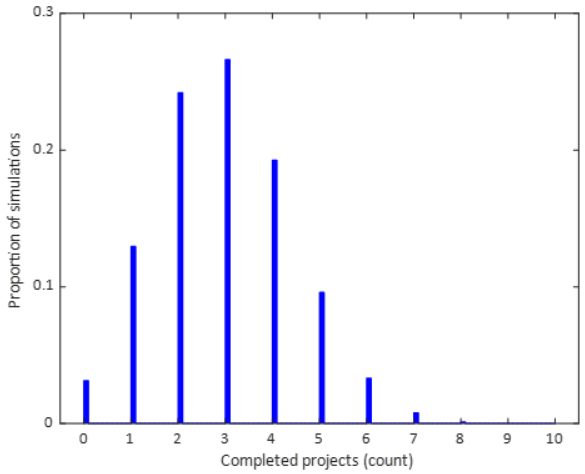
Category	Value
Parameters	
Number of candidate firms	10
Annual per-firm capacity	750 million doses
Probability of success for a firm	30%
Doses required for immunity	1
Vaccine efficacy	100%
Probability pandemic ends each month	5%
Percentage expenditure at risk	
Phase-3 trials	100%
Manufacturing capacity	100%
Cost of goods sold (COGS)	0%
Population	
Across countries	Heterogeneous
Within country	Homogeneous
Mechanism	
Countries participating	All
Funding	100% pull
Allocation	Efficient (highest harm first)

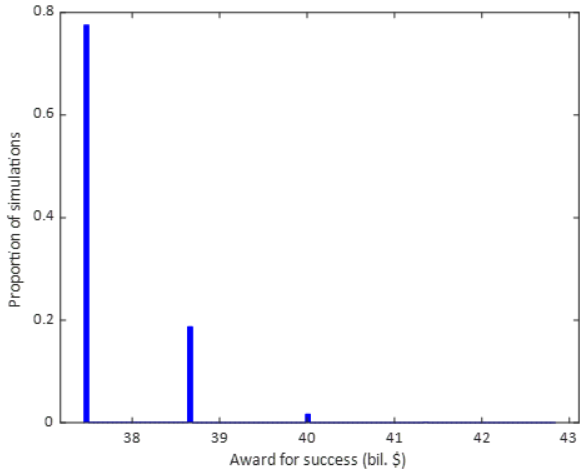
SOURCE Article text.

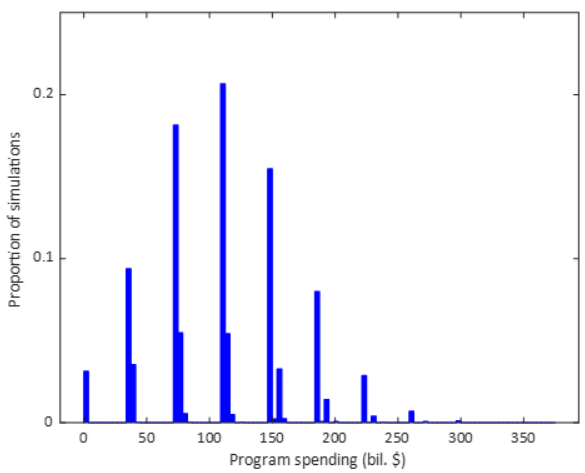
EXHIBIT A7

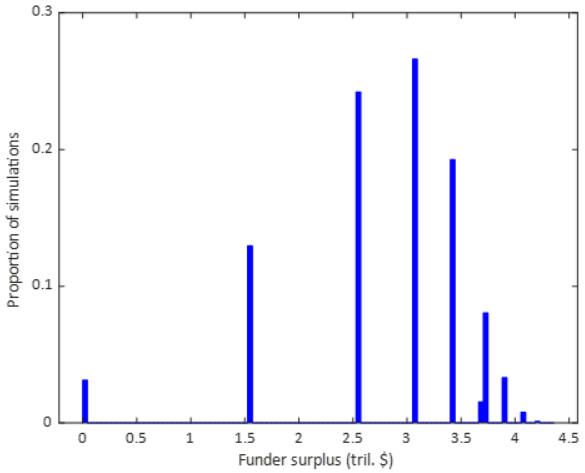
Histograms for outcome variables across simulations in baseline scenario











SOURCE Authors' calculations based on one million simulations. **NOTES** Complements results provided in Exhibit 3. That exhibit provides means only for outcome variables. This exhibit provides histograms showing entire distribution of outcome variables across the million simulations. For brevity, this exhibit only displays results for baseline scenario.