# R&D and market size: who benefits from orphan drug regulation?

*preliminary version;*
*please do not cite/circulate*

Simona Gamba* Laura Magazzini† Paolo Pertile‡

December 15, 2018

**Abstract**

Orphan (rare) diseases occur infrequently in the population, thus providing little incentives for drug development. However, rare diseases affect a non-negligible part of the population, raising policy relevant issues. Since the early 80s, orphan specific legislations have addressed the lack of incentives to invest in innovation for rare diseases. We provide new evidence on the heterogeneous effect of these provisions on innovation targeting rare diseases belonging to different classes of prevalence. A theoretical model is developed that distinguishes market-related and input-related incentives to show how the different provisions are crucial in defining the relative convenience to invest across different classes of prevalence. Data on US orphan designations, as a proxy for R&D effort, are used to empirically test the model and to estimate the effect of orphan legislations. We show that, while the number of designations has increased over time for all orphan diseases, those in the class with the lowest prevalence have done so at a much slower pace. Our findings support the idea that the type of incentives in place may be responsible for this increase in inequality within orphan diseases.

*Department of Economics, University of Verona, Verona, Italy. gamba.simona@gmail.com
†Department of Economics, University of Verona, Verona, Italy. laura.magazzini@univr.it
‡Department of Economics, University of Verona, Verona, Italy. paolo.pertile@univr.it

# 1 Introduction

Orphan diseases are defined as those whose prevalence is below a given threshold, whose determination may vary from one institutional context to another. Despite the fact that each of these diseases often affects a very small number of individuals, the number of orphan diseases is such that it is estimated that 25 to 30 million US citizens and 27 to 36 million EU residents suffer from an orphan disease (Health and Safety, 2015). However, around 95% of the more than 7,000 orphan diseases currently known do not have a single approved drug treatment.

Given that the pharmaceutical industry is mainly responsible for R&D investments for new drugs, the allocation of resources across different therapeutic areas is affected by the amount of profits that each area can provide, in expected terms. The limited size of the market for drugs targeting orphan diseases may lead to underinvestment, in comparison with more common disease. Acemoglu and Linn (2004) consider the effect of the potential size of markets on pharmaceutical innovation and entry of new drugs. The authors derive an equilibrium condition for the levels of R&D effort and show that, the greater is the market size, the more profitable it is to supply the drug and so the greater will be the research effort required to gain market-leader position. Jobjörnsson et al. (2016) show how the interaction between the regulation of marketing approval by institutions such as FDA and EMA and reimbursement decisions by payers makes it less likely that an R&D project is undertaken for a rare vis-a-vis a common disease. The authors also discuss some adjustments of typical decision criteria that might contribute to closing this gap.

Policy makers have introduced a number of tools to incentivise R&D for orphan diseases. The main tools are tax credits on R&D expenditure, market exclusivity for new products, protocol assistance and reduced marketing authorization fees. The first special legislation was introduced in the United States, with the Orphan Drug Act (ODA), approved in 1983. Since then, several other countries have established regulations for the development of orphan drugs. The necessary formal step to access these incentives is obtaining an *orphan designation* from the competent regulatory authority.

The implementation of these policies raises a number of questions. A first question is whether such policies are effective in closing the gap between orphan and non-orphan diseases. Braun et al. (2010a), and Westermark et al. (2011a) report a positive impact, respectively for the United States and the European legislation. Overall, there seems to be a general consensus that special legislations adopted over the world have had a positive impact on R&D investments targeting orphan diseases. On the other hand, the sustainability of such policies has been questioned, given their substantial impact on pharmaceutical expenditure (Denis et al., 2010; Schey et al., 2011). It has also been argued that high prices may jeopardize access to innovation in some

countries (Detiček et al., 2018). It is therefore also important that incentive policies are efficiently designed, so to maximize the social return on expenditure.

While most of the literature has addressed the question whether special legislations are effective in reducing the gap between R&D for orphan and non orphan diseases, far less attention has been devoted to the possibly heterogeneous impact across different orphan diseases. However, this is an extremely relevant issue, given the huge number of orphan diseases and the large variability among these, along several dimensions. The dimension of heterogeneity on which this paper focuses is prevalence. Among orphan diseases, there are some that affect almost 100,000 individuals worldwide and others that are extremely rare. We believe that, if an equity argument provides the rationale for incentivising orphan versus non-orphan diseases, the equity implications of these incentives within the class of orphan diseases cannot be neglected either. Hence, our focus is on the heterogeneous impact of orphan drug legislations across different classes of prevalence.

Yin (2008) studies the impact of the ODA on R&D activity targeting rare diseases, proxied by the number of clinical trials, and shows that the rarest diseases have benefited less from the introduction of the special legislation in the United States. The theoretical interpretation of the results is focused on the mechanism through which tax credits provide an incentive to investment. Our analysis extends to the impact of special legislation introduced in other geographic areas. To account for the heterogeneity in the tool set used in different contexts, we extend the theoretical analysis to include *market-oriented* incentives (e.g., market exclusivity) in addition to *input-oriented* incentives (e.g., tax credits). Moreover, we use the number of orphan designations as a proxy of R&D effort, which allows us to extend the analysis to all orphan diseases.

In our theoretical model, a set of heterogeneous expected profit-maximizing firms decide: (i) whether to undertake an R&D project for a specific disease, (ii) conditional on undertaking a project, how much to invest in it. We study how different types of incentives may affect the relative convenience of investing in R&D targeting diseases with higher and lower levels of prevalence within of orphan diseases. For the empirical analysis, we use the number of orphan designations per disease as a proxy of R&D intensity. The empirical counterpart of our theoretical model is a zero-inflated negative binomial model. We adopt a difference-in-differences approach to exploit the fact that reforms have been introduced at different points in time in different geographic areas.

We find that, over time, R&D efforts have increased substantially less for the rarer diseases within the class of orphan diseases, thus increasing inequality between more and less rare diseases. According to the theory, *market-oriented* tools provide a stronger incentive than *input-*

*oriented* tools towards this type of allocation of R&D efforts. We argue that, the prominent role of market exclusivity as an incentive tool in the European legislation introduced since 2000, may be a key determinant of this widening gap.

In terms of policy implications, if providing an opportunity also to those who suffer from very rare diseases is a goal, than the composition of the incentive tool set should be revised by either reducing the relative weight of market exclusivity, or by introducing new tools that allow for prevalence weighted incentives.

The structure of the paper is as follows. Section 2 describes the different legislations that have been adopted over time. Section 3 describes the model, which is solved in section 4. Section 5 describes data and methodology for the empirical analysis whose results are presented in section 7. Section 8 concludes and discusses the policy implications.

# 2 Institutional context

Over the last 35 years, orphan drug legislations have been adopted in several countries around the world. The first country to develop a specific legislation were the US, where the Congress signed in 1983 the ODA, whose goal is to provide financial incentives for the development of drugs targeting orphan diseases. According to the ODA, a drug is considered as orphan if it treats a rare disease or condition affecting fewer than 200,000 persons in the US (6.25 in 10 thousand persons) or if it will not be profitable within seven years following approval by the FDA. The incentives for drugs designated as orphan are (1) assistance from the Office of Orphan Product Development during the development process; (2) tax credits (up to 50% of clinical development costs); (3) exemption or waiver of application (filing) fees; (4) seven years of marketing exclusivity[1] and (5) subsidies for clinical trials from the Orphan Products Grant Program.

Special legislations with the same objectives have subsequently been introduced in several countries, such as Singapore (1991), Japan (1993), Australia (1998), South Korea (1998), the EU (2000) and Taiwan (2000).[2] In what follows we only consider the introduction of special legislations in the three areas with the largest markets: US, Japan and the EU.

In April 1993, Japan substantially revised its orphan medicinal product system, introduced in 1985, so as to extend the tools used to incentivize research on orphan diseases. So, in addition to

[1]Market exclusivity represents a stronger protection for firms compared to patents. While patents prevent other companies from making, using, offering for sale, selling, and importing for these purposes the drug, market exclusivity implies that the regulatory agency cannot approve another drug for the same indication without the sponsor's consent.

[2]With the exception of Australia, all these countries provide (extra) market exclusivity for orphan drugs (Sharma et al., 2010).

the already existing (1) reductions in the required data for application, and (2) accelerated review process, the following incentives were introduced: (3) protocol assistance; (4) tax credits (up to 6% of clinical and non-clinical costs); (5) subsidies for clinical and nonclinical studies and (6) ten years of market exclusivity. In order to be designated as orphan, the drug, which has to be proved highly effective and safe, has to treat a rare and serious disease or condition affecting less than 50,000 persons in Japan (4 in 10 thousand persons), and such disease should not have any other available treatment. Since in Japan the incentive tools which are the main focus of our analysis were introduced in 1993, we refer to this as the date when the special legislation was introduced.

In December 1999, also the European Union approved a regulation on orphan medicinal products: the Regulation (EC) No 141/2000. The regulation establishes a procedure for designating orphan drugs and sets incentives for R&D. The incentives include (1) protocol assistance; (2) access to a centralised procedure allowing immediate marketing authorisation in all member states; (3) reduced fees for regulatory procedures and (4) ten years of market exclusivity. In order to benefit from the incentives, orphan drugs have to be designated as such before the marketing authorisation is granted. Moreover, the targeted drug has to treat a life threatening or chronically debilitating condition affecting no more than 5 in 10 thousand persons in the Community when the application is made, or it has to be unlikely, without incentives, to generate sufficient return to justify the necessary investment; finally, there should exist no satisfactory alternative methods authorised in the Community (article 3 of the Regulation). In addition to the incentives mentioned in the regulation, some member states have introduced other measures to support R&D, such as tax reductions (allowed in France and the Netherlands) (Health and Safety, 2015).

Incentives provided by the US, Japan and Europe are summarized in Table 1, together with requirements for drugs to be considered as orphan.

Since November 2007, the European Medicines Agency (EMA) and the FDA are collaborating to encourage joint applications to the orphan drug status both in Europe and the US. A common application form has been developed, in an effort to reduce the administrative burden on the orphan drug sponsor (Braun et al., 2010b; Mariz et al., 2016). Parallel applications in Japan are also encouraged, although a common application form is not in place yet, due to administrative differences between the two offices (Mariz et al., 2016).

| | US (1983) | Japan (1993) | Europe (2000) |
|---|---|---|---|
| Disease: | | | |
| Prevalence | 6.25 per 10,000 | 4 per 10,000 | 5 per 10,000 |
| Characteristics | Rare or not profitable | Rare<br>Serious<br><br>No other treatment available | Rare or not profitable<br>Life threatening or chronically debilitating<br>No other treatment available |
| Main incentives: | | | |
| Tax credit | Yes<br>(50% clinical costs) | Yes<br>(6% clinical and non-clinical costs) | Member state specific |
| Market exclusivity | Yes (7 years) | Yes (10 years) | Yes (10 years) |
| Reduced applic. fees | Yes (waved) | No | Yes (reduced) |
| Protocol assist. | Yes | Yes | Yes |
| Subsidies for clinical trials | Yes | Yes | No |

Table 1: Comparison of orphan drugs legislations in the US, Japan and EU.

# 3 The model

A representative profit-maximizing firm is free to decide on the size of an R&D investment, $I \geq 0$, targeting disease $j$. The number of individuals affected by disease $j$ is $n_j$, which is also assumed to be the size of the market for a product that obtains marketing authorization with an indication for that disease. For an orphan drug, there are two key regulatory steps in the development process. In the first step, the firm that has developed a molecular entity applies for "orphan drug designation" (ODD). If granted, the ODD makes the firm eligible for any incentive related to the development of an orphan drug. If the development process is successfully completed, the firm will approach the second regulatory stage: marketing authorization. From the perspective of the firm, both stages entail uncertainty. Let $p_j^d(I)$ be the probability that the firm obtains an ODD, given the R&D investment $I$. For the function $p_j^d(I)$ we introduce the standard assumptions $\frac{\partial p_j^d}{\partial I} > 0$ and $\frac{\partial^2 p_j^d}{\partial I^2} < 0$. Moreover, given that $p_j^d$ is a probability, $p_j^d(0) = 0$ and $\lim_{I\to\infty} p_j^d(I) = 1$.

Conditional on obtaining the ODD, the firm will carry on the development process. With probability $p_j^m$ this will lead to the marketing approval of the product.[3] Given the disease specific

[3]For the sake of simplicity, $p_j^m$ is assumed independent of $I$.

per patient net revenue $m_j$, conditional on obtaining an ODD, the expected net revenue is $p_j^m \cdot m_j$. To simplify notation, we define the individual level expected net revenue, conditional on having obtained the ODD, as $M_j(\Omega_j) = p_j^m m_j$. The parameter $\Omega_j$ is a vector of disease specific characteristics that may affect the probability $p_j^m$ and / or the net revenue $m_j$ . For example, some regulators grant a price premium to drugs targeting life threatening conditions.

The expected profit at the time when a decision on the R&D investment is made, can be written as:

$$E\Pi_j = p_j^d(I)[M_j \cdot n_j] - I + \delta_j \tag{1}$$

The term $\delta_j$ aims to capture any additional component of the expected profit that is only known to the firm. From the perspective of the researcher, $\delta_j$ is the realization of a random variable, with density $f(\delta_j)$.

The aim of our analysis is to study the impact of different forms of incentives among those that have been introduced as part of the special legislation on: *i)* the probability of having investment on a rare disease, *ii)* the probability of having an orphan designation. Our analysis is carried out *within* the class of orphan diseases. In other words, we do not contrast rare versus non-rare disease, but more versus less rare diseases within the class of orphan diseases. As a result, we assume that all diseases are eligible for incentives. Our focus is on how the impact of different types of incentives is affected by the prevalence of an orphan disease.

Incentives can be distinguished into two categories: *market-related* and *input-related*. *Market-related* incentives are those that aim to increase the net market revenue of investments made on orphan diseases. The best known instance of such instrument is *market exclusivity*, to which all products with an orphan designation are entitled. This is part of the incentive package provided, for example, by the US, Japan and Europe. We model this as a mark-up, $z$ ($z \geq 0$), on net revenues. This way of modelling *market-related* incentives is sufficiently flexible to also account for other types of incentives such as a price premium to which all orphan drugs are equally entitled.

*Input-related* incentives reduce the cost of R&D investment for rare diseases. Examples of such incentives include tax credits related to R&D expenditure, reduced fees for market authorization applications and protocol assistance. We model this type of incentive as an allowance on investment costs, such that the investment cost borne by the firm is $I_j(1-\gamma)$, with $0 \geq \gamma \geq 1$.

To take the role of these incentives into account, the expected profit function can be written as:

$$E\Pi_j = p_j^d(I)[M_j \cdot n_j](1+z) - (1-\gamma)I + \delta_j. \tag{2}$$

# 4 Optimal investment policy

We start by characterizing the first order condition (FOC) for optimal investment from the firms' perspective:

$$\frac{\partial p_j^d(I)}{\partial I} = \frac{1-\gamma}{M_j \cdot n_j(1+z)}. \tag{3}$$

The second order conditions are satisfied under the assumptions on the functional form of $p_j^d(I)$ that were introduced above. Eq. 3 implicitly defines the optimal investment level $I^*(M_j, n_j)$ and highlights the well known role of market size as an incentive for R&D investments: with $n_j$ small, other things being equal, the optimal investment level is lower.

We can use the implicit function theorem, to study the impact of an increase in $z$ on the optimal level of investment:

$$\frac{dI^*}{dz} = -\frac{1-\gamma}{(\partial^2 p_j^d/\partial I^2) M_j n_j (1+z)^2}. \tag{4}$$

According to eq. 4, an increase in $z$ provides an incentive to invest more, by reducing the value on the right hand side of eq. 3. From the perspective of our analysis, it is also interesting to investigate how the marginal impact on $I^*$ of an increase in $z$ varies with $n_j$. Differentiating the right hand side of eq. 4 with respect to $n_j$ obtains:

$$\frac{\partial^2 I^*}{\partial z \partial n_j} = \frac{(\partial p_j^d/\partial I)(\partial^3 p_j^d/\partial^3 I) - (\partial^2 p_j^d/\partial^2 I)^2}{(\partial^2 p_j^d/\partial^2 I)^2 (1+z)^2}(1+z)\frac{\partial I^*}{\partial n_j}. \tag{5}$$

Given that $\partial I^*/\partial n_j > 0$ (see eq. 3), the sign of eq. 5 is the same as the sign of its first term. Since $\partial^3 p_j^d/\partial I^3$ may be positive,[4] the expression cannot be unambiguously signed. This means that, conditional on $I^* > 0$, we cannot unambiguously say whether the impact on the probability of obtaining a designation of strengthening a *market-related* incentive is greater for a more or a less rare disease.

Given $I^*(M_j, n_j)$, the firm will only invest if the expected profit at the time of investment is non-negative, i.e.:

$$p_j^d(I^*(M_j, n_j, z, \gamma))[M_j \cdot n_j](1+z) - (1-\gamma)I^*(M_j, n_j, z, \gamma) + \delta_j \geq 0. \tag{6}$$

It is then possible to define a minimum value of the stochastic variable, $\hat{\delta_j}$, such that the firm

[4] Indeed, the sign is positive for the increasing and concave functional forms typically employed in economics.

makes any investment in R&D for disease $j$:

$$\hat{\delta_j} = (1-\gamma)I^*(M_j, n_j, z, \gamma) - p_j^d(I^*(M_j, n_j, z, \gamma))[M_j \cdot n_j](1+z). \tag{7}$$

The indicator function $\mathcal{I}_j$ denotes the decision to invest a strictly positive amount in R&D for disease $j$:

$$\mathcal{I}_j(M_j, n_j, z, \gamma) = \begin{cases} 1 & \text{if} \quad \delta_j \geq \hat{\delta_j}(M_j, n_j, z, \gamma) \\ 0 & \text{if} \quad \delta_j < \hat{\delta_j}(M_j, n_j, z, \gamma). \end{cases} \tag{8}$$

The probability of having investment for disease $j$ is therefore

$$\mathcal{P}(\mathcal{I} = 1 \mid M_j, n_j, z, \gamma) = \int \mathcal{I}(M_j, n_j, z, \gamma) f(\delta_j) d\delta_j. \tag{9}$$

To investigate the impact of $n_j$ on the decision whether to invest or not, we study the dependency of $\hat{\delta_j}$ on $n_j$. Noticing that

$$\hat{\delta_j} = -E\Pi_j(I^*) + \delta_j, \tag{10}$$

which allows to simplify calculations through the application of the Envelope Theorem, the following expression obtains:

$$\frac{\partial \hat{\delta_j}}{\partial n_j} = -p_j^d(I^*(M_j, n_j, z, \gamma)) M_j (1+z) \quad < 0. \tag{11}$$

Hence, other things being equal, for a comparatively rare disease the value of $\delta_j$ must be larger for the firm to decide to undertake any investment (eq. 11). Thus, it is less likely to observe R&D investment in comparatively rare diseases.

Using a similar approach, we can study the impact of an increase in $z$ on the value of the stochastic variable above which a positive amount is invested. This leads to

$$\frac{\partial \hat{\delta_j}}{\partial z} = -p_j^d(I^*(M_j, n_j, z, \gamma)) \cdot M_j \cdot n_j < 0, \tag{12}$$

which shows the role of $z$ in making it more likely that there is investment for disease $j$, by reducing the value of $\hat{\delta_j}$. Also in this case, we are interested in differences in the impact of this incentive tool that can exist across different classes of prevalence. By differentiating the right

hand side of eq. 12 with respect to $n_j$, we obtain:

$$\frac{\partial^2 \hat{\delta_j}}{\partial z \partial n_j} = -M_j \left[ \frac{\partial p_j^d}{\partial I} \frac{\partial I^*}{\partial n_j} n_j + p_j^d(I^*) \right] < 0. \tag{13}$$

The negative sign of the expression means that the impact on the probability that the firm undertakes any investment of an increase in $z$ is larger for comparatively large values of $n_j$.[5]

We can now move to the analysis of the impact of the *input-oriented* incentive, associated with the parameter $\gamma$. As for $z$, the impact on $I^*$ of an increase in $\gamma$ is positive (see eq. 3). Concerning the question on how this impact changes across different classes of prevalence, using the same approach as above, we find that:

$$\frac{\partial^2 I^*}{\partial \gamma \partial n_j} = \frac{(\partial^3 p_j^d / \partial^3 I)(\partial I^* / \partial n_j) n_j + \partial (p_j^d)^2 / \partial I^2}{(\partial (p_j^d)^2 / \partial I^2)^2 M_j \cdot n_j^2 (1+z)}. \tag{14}$$

As for $z$, this term cannot be unambiguously signed, meaning that the impact of an increase in $\gamma$ on $I^*$ and hence on the probability of having an ODD, conditional on investing, may be increasing or decreasing in the disease prevalence.

Concerning the impact on the probability of having any investment for disease $j$, we have that,

$$\frac{\partial \hat{\delta_j}}{\partial \gamma} = -I^*(M_j, n_j, z, \gamma) < 0 \tag{15}$$

and

$$\frac{\partial^2 \hat{\delta_j}}{\partial \gamma \partial n} = -\frac{\partial I^*}{\partial n_j} < 0. \tag{16}$$

Also in this case, the impact on the probability of having an investment is greater for the more common diseases. Having noted that the sign of the impact is the same, as a final step, we can try to compare the magnitude of a marginal increase in $z$ versus $\gamma$ on the impact on the probability of having investment for different values of $n_j$. A marginal increase in $z$ can be interpreted as an increase by, e.g. $1\%$, in expected revenues from commercialisation. Similarly, for $\gamma$, it can be seen as a $1\%$ reduction in the investment cost faced by the firm.

The comparison between eq. 16 and eq. 13 shows that in the former there is one additional effect. As a result of this, it can be shown that for $z$ sufficiently small, the difference in the marginal impact of an increase in $z$ on the probability of having any investment between a less

[5] Notice that the negative sign is related to a larger reduction in $\hat{\delta_j}$.

and a more common disease is greater than the difference for an increase in $\gamma$. In other words, under reasonable assumptions, we would expect a *market oriented* incentive to favour more the comparatively common diseases within the class of orphan in terms of probability of having any investment. Since, this probability is lower for the rarest diseases even without incentives (eq. 11) this mechanism is potentially increasing the inequality in terms of opportunity to have a therapeutic option available within the class of orphan diseases.

Combining the two decisions made by the firm – whether to invest and the level of investment – we can write the unconditional probability of observing an ODD for disease $j$:

$$\mathcal{P}\Big(\mathcal{I} = 1 \mid M_j, n_j, z, \gamma\Big) \cdot p_j^d\Big(I^*(M_j, n_j, z, \gamma) \mid \delta_j > \hat{\delta}_j\Big). \tag{17}$$

The analysis carried out in this section has shown that $z$ and $\gamma$ have a positive impact on both terms, but the size of these impacts depends on the disease prevalence. Both the *market-related* and the *input-related* incentive have a larger impact on the probability of having any incentive (first term) for diseases with comparatively higher prevalence. However, under reasonable assumptions, the magnitude of this impact is greater for a *market-oriented* incentive. On the other hand, whether the impact on the second term of eq. 17 is greater or smaller for comparatively rare (common) diseases cannot be said without ambiguity.

# 5 Data and measures

To identify the full list of orphan diseases, we rely on the Orphanet data base, which is the standard reference for information on rare diseases.[6] The list of rare diseases is systematically updated, as approximately 250 new diseases are described each year (Westermark et al., 2011b; Wästfelt et al., 2006): the version used in the empirical analysis was downloaded in October 2017. The full list downloaded counts 9,530 diseases, some of which are excluded from the analysis following the criteria described later.

Our proxy of R&D efforts targeting rare diseases is the number of ODD granted by FDA between 1983 and 2016. We focus on designations in the US as it is the largest pharmaceutical market in the World; moreover, the ODA establishment in 1983 allows us to study the dynamics in the number of applications before and after the year 1993, when Japan reviewed its orphan provisions, and the year 2000, when orphan legislation was introduced in Europe. Since the

[6] Orphanet was established in 1997 in France in order to gather knowledge on rare diseases so as to improve their diagnosis, care and treatment; from 2000 the initiative become a European endeavor. For additional information, please visit www.orpha.net.

pharmaceutical industry is a global one, it is convenient for the inventor to apply for the orphan drug status in several countries to benefit from additional incentives. Hence, looking at ODD granted in the US, we should be able to cover the large majority, if not all of them, worldwide. Having an ODD is a necessary condition for the project, and eventually for the drug, to be eligible for the incentives provided under the special legislation. One advantage of this measure is that all designations can be retrieved from a single administrative dataset, meaning that all orphan diseases can be included in the analysis. This is not the case, for example for clinical trials, which have been used in some previous studies (e.g. Yin, 2008). On the other hand, being specific to orphan diseases, the choice of the number of ODD as a proxy prevents from using a group of non-rare diseases as a control group.

We exclude designations referring to products for surgery methods, prevention, transplant, diagnostics and imaging procedures. The analysis starts from year 1983 (when the ODA was passed) as orphan designation was introduced as a regulatory step at that time.[7] For each drug, the FDA provides the date of orphan designation, marketing approval (if any), the designated indication, and the company sponsoring the request.

A major effort was undertaken to match the FDA designations to the list of orphan diseases in Orphanet. Out of 3,588 ODD granted by the FDA between 1983 to 2016, 199 records are dropped because information on the treated disease cannot be retrieved. The Orphanet list provides different levels of aggregation, including also "particular clinical situation", "histopathological subtype" and "group of phenomes". We refer to the level "disease". However, in some cases we were only able to match an ODD from the FDA dataset to a "group of phenomes", i.e. a broader category.[8] In this case, in order to be consistent in the definition of the therapeutic market, we rely on the hierarchical classification of orphan diseases provided by Orphanet to link the group with all relevant diseases belonging to it and match the FDA designation at the disease level. If more than one disease is included in the group, one orphan drug designation is attributed to each disease, i.e. a non-fractional count is adopted.

Orphanet also provides information on the class of prevalence of each disease.[9] In particular, we refer to worldwide prevalence. When this information is missing, we consider prevalence in Europe or, if missing, in the US. Diseases belonging to the following prevalence classes are

[7] During the period 1967-1983, only 34 drugs were approved by the FDA for rare conditions (Kesselheim, 2010).

[8] For example, some applications were designated for drugs treating the hypereosinophilic syndrome, which is classified as a "group of phenomes" in Orphanet and comprises different diseases included in the Orphanet list (i.e., idiopathic hypereosinophilic syndrome, primary hypereosinophilic syndrome, and secondary hypereosinophilic syndrome).

[9] In few cases (6.7% of diseases), a numeric value for prevalence is also provided. However, the availability of this information is unevenly distributed among classes of prevalence. Given these limitations, the point estimate of prevalence is not used in the empirical analysis.

| Prevalence | number of diseases | % total | average number of yearly ODD per disease |
|---|---|---|---|
| $N1$: <1/1,000,000 | 2,629 | 42.14 | 0.02 |
| $N2$: 1-9/1,000,000 | 208 | 3.33 | 0.12 |
| $N3$: 1-9/100,000 | 307 | 4.92 | 0.15 |
| $N4$: 1-5/10,000 | 161 | 2.58 | 0.20 |
| $N0$: Missing prev. | 2,933 | 47.02 | 0.11 |
| Total | 6,238 | 100 | – |

Table 2: Distribution of the diseases among prevalence classes

included in the analysis: "<1/1,000,000", "1-9/1,000,000", "1-9/100,000", and "1-5/10,000".[10] The distribution of diseases considered in the analysis among prevalence classes is reported in Table 2. Information on the prevalence is missing (or not yet documented) in Orphanet for a large share of the diseases: these are considered as a separate class. Among the classes with known prevalence, the large majority of diseases is classified with a prevalence lower to 1 in 1 million (42.14%), with the "less rare" diseases only accounting for 2.58% of the total. Table 2 also shows how the expected number of ODD changes from one class of prevalence to another. The numbers reported are calculated taking the average over years in the study period and over diseases in each class of prevalence. These descriptive statistics are coherent with our theoretical results and with the literature suggesting a positive correlation between market size and R&D effort (Acemoglu and Linn, 2004; Dubois et al., 2015).

Orphanet provides additional information at the disease level, including the therapeutic class(es) of each disease and information on the age of onset and age at death. These are classified in antenatal, neonatal, infancy, childhood, adolescence, adulthood and elderly. About 13% of diseases allows for a normal life expectancy. We exclude from the analysis those diseases emerging in the antenatal period or causing death before birth (323 diseases).[11]

All in all, our data comprise 212,092 observations (6,238 diseases over 34 years).

Finally, we complement information provided by Orphanet with an ad hoc search into PubMed,[12] in order to gather information on the number of publications related to each disease. An auto-

[10]Note that these diseases would be classified as orphan both in US and Europe, whereas diseases with prevalence higher than 4 per 10,000 would not benefit from orphan incentives in Japan (see Section 2).

[11]We also removed 568 diseases referring to surgical procedures, and 192 items representing an old nomenclature (these were moved to an updated item).

[12]Pubmed is a web-search service maintained by the US National Libraty of Medicine. It comprises more than 28 million citations for biomedical literature from MEDLINE, life science journals, and online books. For more information, please visit https://www.ncbi.nlm.nih.gov/pubmed/.

mated search was conducted on PubMed for each disease in our list, retrieving the number of articles published over the period 1970-2016 and containing the name of the disease in the title, abstract or content. We use this information to construct a measure for the stock of publications ($SP$), following the perpetual inventory method:

$$SP_{jt} = P_{jt} + (1 - \rho)SP_{j,t-1}$$

where $P_{jt}$ is the number of publications related to disease $j$ at time $t$ and $\rho = 0.1$ is the rate of obsolescence of knowledge typically applied in the empirical literature (Keller, 2002).

# 6 Empirical methods

We study the impact of the orphan legislation on the distribution across different classes of prevalence of pharmaceutical R&D effort, as proxied by the number of ODD granted by the FDA Office of Orphan Product Development. We adopt a difference-in-differences approach to exploit the natural experiment provided by the different time of introduction of orphan legislations in the three largest markets (US, Japan, Europe). Among the incentive tools that are considered most relevant, market exclusivity is present in all three legislations, whereas tax credits are used only in US and Japan (see section 2). The theoretical results of section 4 show that the impact of both market exclusivity and tax credits on the probability of undertaking an investment increases with the prevalence of the disease. We also show that this difference is greater for a *market-oriented* tool as market exclusivity than for an *input-oriented* incentive as tax-credit. This implies a different level of exposure to the treatment for more *versus* less rare diseases.

The empirical counterpart of our theoretical model is a Zero Inflated Negative Binomial (ZINB) model: the probability of observing a zero corresponds to the probability of (not) investing in R&D; the count data model is related to the level of investment, $I^*$, i.e. to the probability of obtaining a designation, conditional on $I^* > 0$. In our theory, when $I^* > 0$, a firm can either obtain one designation or no designation. However, in reality there are several firms and several projects, so that for each disease we can have more than one designation. However, since an increase (decrease) in the probability of obtaining a designation for each firm also implies an increase (decrease) in the expected number of designations observed, the comparative statics results of section 4 can be straightforwardly extended to the case where the number of designations is a non-negative integer. The ZINB model allows us to understand the determinants of the two different processes determining a zero outcome (Lambert, 1992): choice (the decision not to

invest in R&D) and nature (the lack of innovative output, probably related to the level of effort and to the availability of information on the disease) (Winkelmann, 2008). In this set up, it is important also to take into account the large overdispersion of the data. Hence, the choice of the ZINB model.

R&D effort, proxied by the number of ODD targeting disease $j$ in year $t$, is therefore modelled as:

$$y_{jt} = \begin{cases} 0, & \text{if } \mathcal{I}_j = 0 \\ y^*_{jt}, & \text{if } \mathcal{I}_j = 1 \end{cases} \tag{18}$$

where:

- from section 4, $\mathcal{I}_j$ is a binary variable related to the decision whether to invest in R&D or not.[13] If $\mathcal{I}_j$=0, the outcome is a "certain zero", also referred to as "strategic" or "structural" zero (Staub and Winkelmann, 2013);
- $y^*$ is an overdispersed count variable, representing the number of orphan designation applications filed at the FDA Office of Orphan Product Development. When $y^* = 0$, zeros in the outcome are due to nature.

Define $f_1(.)$ as the density of the binary process and $f_2(.)$ as the count density. The model taking into account both choice and nature processes has a density:

$$f(y_{jt}) = \begin{cases} f_1(\mathcal{I}_j = 0) + [1 - f_1(\mathcal{I}_j = 0)]f_2(0) & \text{if } y_{jt}\text{=0} \\ [1 - f_1(\mathcal{I}_j = 0)]f_2(y) & \text{if } y_{jt} \geq 1. \end{cases} \tag{19}$$

The two components of zero inflated models are estimated simultaneously. The Logit part estimates the probability to be in the "certain zero" case ($\mathcal{I}_j = 0$):

$$f_1(\mathcal{I}_j = 0) = \frac{\exp(x'_{jt}\beta_1)}{1 + \exp(x'_{jt}\beta_1)}$$

while the count part of the model explains the determinants of innovation output for diseases not included in the "certain zero" group with expected value:

$$E_{f_2}(y_{jt}|x_{jt}) = \exp(x'_{jt}\beta_2) \tag{20}$$

[13] We do not make explicit reference to time here, as there might be lags between R&D investments and ODD. However, this does not affect our empirical strategy.

The expected number of ODD is expressed as a combination of the two processes as:

$$E(y_{jt}|x_{jt}) = (1 - f_1(\mathcal{I}_j = 0))E(_{f_2}(y_{jt}|x_{jt})) = \frac{\exp(x'_{jt}\beta_2)}{1 + \exp(x'_{jt}\beta_1)} \tag{21}$$

where

$$x'_{jt}\beta = \alpha + \sum_{i=0}^{4} \zeta_i Ni_d + \sum_{t=1}^{4} \tau_t Dt + \sum_{i=0}^{4}\sum_{t=1}^{4} \delta_{it}(Ni_d \times Dt) + \theta C_{jt}. \tag{22}$$

Note that we use the same set of variables in the Logit and in the count part of the model. $Ni$ represents the class of prevalence, from the most rare (N1:"<1/1,000,000"; see Table 2) to the most common diseases (N4:"1-5/10,000"). The variables $Dt$ are binary variables indicating relevant periods of time, related to the introduction of special legislation in US, Japan and EU: 1983-1992; 1993-1999; 2000-2006 and 2007-2016. $\eta_{it}$ are the main parameters of interest, both in the Logit and count part of the model, representing the different effect of each reform for diseases belonging to the class of prevalence $Ni$, with respect to those in the class with the lowest prevalence (taken as reference category). $C_{jt}$ includes additional control variables, which, according to the analysis of section 4 may have an impact on R&D policies. The following variables are included in the specifications that we consider:

- the stock of publications, $SP$. This may be one of the determinants of the expected profit ($\Omega_j$). This variable is meant to account for the fact that advances in scientific knowledge in one therapeutic area may increase the probability of obtaining an ODD.

- a set of variables that might play a role, by affecting the disease specific per patient net revenue, $m_j$. In particular, there are disease characteristics for which some payers are willing to consider a price premium. We include a dummy variable indicating whether the disease allows for a normal life expectancy ($NLExp$, which takes value 1 if the disease allows for the same life expectancy as that of healthy individuals) and two dummy variables ($Ped1$ and $Ped2$) that aim at identifying paediatric diseases. $Ped1$ identifies those diseases that occur at a paediatric age (up to adolescence included), but death comes during adulthood or elderhood. The variable $Ped2$ instead identifies those diseases for which both the age of onset and death occur before adulthood. The reference category consists of diseases characterized by an adult or older age of onset. As some regulators grant a price premium to drugs targeting life threatening conditions, and paediatric drugs are granted additional

market exclusivity,[14] these characteristics of the drug might affect whether any investment is undertaken and the optimal level of investment.

- A proxy for the probability of obtaining marketing authorization, $p_j^m$ (see section 4). This variable is constructed as the ratio between the sum of marketing authorizations granted in the previous 5 years and the sum of designations received in the previous 5 to 9 years. We consider a time lag of 4 years as, from FDA data, about 50% of all approvals take place within 4 years from designation.[15]

# 7 Results

Results of the estimation of the ZINB model are presented in Table 3. In Column (1) we present a simplified model where we do not account for the heterogeneity in the effect of the regulations: we omit the interaction terms from Equation 22. In Column (2) results from the baseline estimate presented in Equation 22 are reported, whereas in Columns (3)-(5) additional control variables are included in the analysis. For each estimated model we present both the results of the Logit part of the model (probability of being a "certain zero" outcome), and the "count" part (modeling the determinants of innovation output for diseases not included in the "certain zero" group). Therapeutic class dummy variables, along with a dummy variable identifying genetic diseases, are included in all the specifications.[16]

Results in Column (1; Logit) show that it is more likely to have innovation in comparatively large diseases. Similarly, Column (1; count) shows that the level of innovation is higher for these diseases, and that over time the introduction of special legislations is associated with an increase in the number of ODD.

Interactions between the classes of prevalences and the time periods are added both in the Logit and the count parts of the model presented in Column (2), in order to account for any

[14]The extra market exclusivity for paediatric drugs lasts 2 years in Europe (Article 37 of the Orphan Regulation), and 6 months in US (Section 505(A) of the Food and Drug Administration Modernization Act of 1997). It is important to notice that the extension is granted when a Paediatric Investigation Plan is completed, while age of onset and age at death are not considered in the decision to grant the extension. However, information regarding the Paediatric Investigation Plan cannot be used to retrieve whether a disease is paediatric or not, since they are available only for those diseases for which there is at least one drug.

[15]This statistics is obtained by taking into account the designation-approval lag for designations that received a marketing authorization. We only considered designations obtained before the year 2005, as the designation-approval lag for more recent ODD would be censored. If also more recent designations are taken into account, we find that 60% of all approvals take place within 4 years from designation.

[16]The 26 therapeutic class dummy variables are not mutually exclusive, as a disease may belong to more than one classification. As an example, craniopharyngioma is classified as neurological, endocrine, and neoplastic disease. It is also a genetic disease.

heterogeneity in the impact of reforms across classes of prevalence. Results from Column (2; Logit) point out that it is more likely to observe zero R&D investments for more rare diseases for all periods but the first (reference period): the positive and significant coefficients for $N2$ and $N3$ indicates that, in the first period, the probability of having a "certain zero" is higher for more common diseases with respect to those belonging to the class of prevalence "<1/1,000,000" ($N1$), taken as the reference. The coefficient for $N4$ is positive but not statistically significant. In order to analyze the effect in subsequent time periods, we test the sum of coefficients for $Ni$ and the interaction term $Ni \times Dt$: these sums are all negative and statistically significant, pointing out that in all periods but the first the probability of a "certain zero" is lower for higher classes of prevalence. As for the dynamic over time, the coefficients associated to the interaction terms are all negative and statistically significant, confirming that market exclusivity reduces more the probability of zero investments for comparatively larger diseases, as predicted by the theoretical model. Figure 1(a) plots the predicted probability of having a "certain zero" as a function of time for each class of prevalence.[17] From the second to the third period, when market exclusivity is introduced also in Europe, the larger variation in probability is detected for the largest class of prevalence ($N4$). Moreover, it can be noticed that, for all periods but the first, the probability of having a "certain zero" decreases as prevalence increases.

The negative interaction terms in the count part of the model (Column 2; count) suggest that, for diseases not in the "certain zero" group, the extension of market exclusivity favour more diseases belonging to the lowest class of prevalence ($N1$): conditional on having any R&D investment, the extension of market exclusivity reduces the gap, in terms of ODD, between more and less rare diseases. Recall that this outcome is related to the intensity of investment in the theoretical model ($I^*$) and that the results for the comparative statics of $\gamma$ and $z$ are ambiguous. This trend is visible in Figure 1(b), plotting the dynamics in the linear combination $x_{jt}\hat{\beta}_2$ over time for the different classes of prevalence. In Figure 1(c) we plot the exponential value of the linear combination presented in 1(b). We can see that, although the gap reduces in percentage term, the differences in the expected number of ODD (conditional on a positive level of R&D investments) increases over time.

In order to take into account both parts of the model, graph (d) of Figure 1 shows the predicted number of annual ODD by disease, distinguishing for classes of prevalence and time periods. Already in years 1983-1999 (when orphan regulation was in force only in the US), the predicted

[17] As pointed out by Greene (2010), in the context of nonlinear models, statistical tests about partial effects and interaction terms are not necessarily informative: rather, it is better to analyze the dynamics in predicted values and interaction effects.

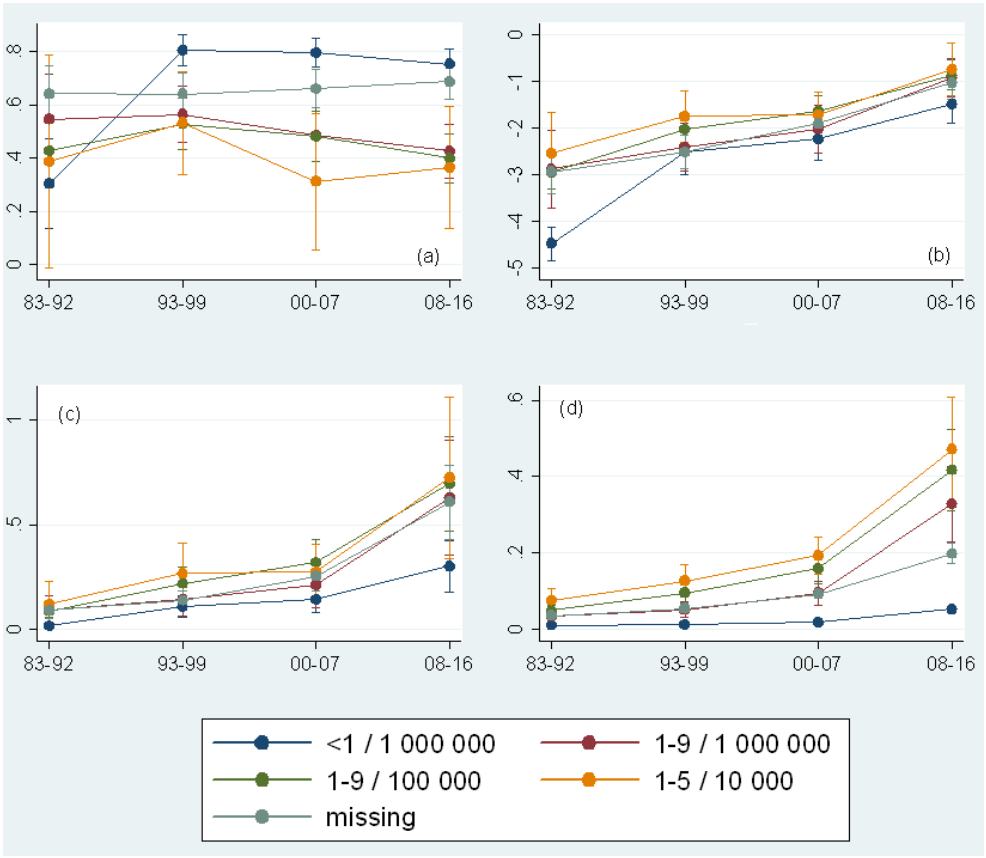


Figure 1: (a) Predicted probability that $\mathcal{I}_j = 0$; (b) linear combination $x_{jt}\hat{\beta}_2$; (c) predicted number of ODD, conditional on $\mathcal{I}_j > 0$; (d) predicted number of ODD

number of ODD was lower for diseases belonging to $N1$ compared to less rare diseases.[18] Over time, there has been an increase in the number of ODD for all classes of prevalence, but this has been greater for the less rare diseases. This means that the magnitude of the heterogeneous impact on the probability of undertaking any investment (Logit) part outweighs the effects on research intensity (count), which goes in the opposite direction.

In Column (3) we control for the level of scientific research related to disease $j$, measured as the (log) of the stock of publication at time $t-5$. This for two reasons. Firstly, pharmaceutical research is the leading example of a science-based sector (Pavitt, 1984): a large part of innovations builds on academic research (Mansfield, 1995). As a result, inputs from science can play a relevant role in stimulating R&D efforts at the market level. Secondly, since the list of orphan drugs is continuously updated, it is possible that some of the diseases considered in our analysis

[18] A test on the predicted number of designations for diseases having a prevalence of "<1/1,000,000" ($N1$) and for those having a prevalence of "1-9 /1,000,000" ($N2$) rejects the null hypothesis of no difference (p-value=0.006).

| | (1) | | (2) | | (3) | | (4) | | (5) | |
|---|---|---|---|---|---|---|---|---|---|---|
| | logit | count | logit | count | logit | count | logit | count | logit | count |
| $N2$ | -1.043** | 0.334 | 3.056*** | 1.412*** | 3.165*** | 1.358*** | 3.155*** | 1.339*** | -1.036* | -0.165 |
| | (0.523) | (0.241) | (0.798) | (0.460) | (0.838) | (0.459) | (0.838) | (0.449) | (0.595) | (0.328) |
| $N3$ | -1.406*** | 0.459*** | 2.000*** | 1.237*** | 2.233*** | 1.217*** | 2.117*** | 1.196*** | -1.145** | 0.217 |
| | (0.392) | (0.176) | (0.550) | (0.224) | (0.626) | (0.242) | (0.681) | (0.253) | (0.472) | (0.219) |
| $N4$ | -2.037*** | 0.489** | 1.592 | 1.511*** | 1.679 | 1.424*** | 1.573 | 1.400*** | -1.565* | 0.345 |
| | (0.672) | (0.227) | (1.325) | (0.489) | (1.382) | (0.488) | (1.381) | (0.487) | (0.814) | (0.348) |
| $N0$ | -0.086 | 0.112 | 3.015*** | 1.163*** | 3.097*** | 1.196*** | 3.042*** | 1.180*** | -1.182*** | -0.307* |
| | (0.209) | (0.104) | (0.493) | (0.264) | (0.583) | (0.283) | (0.632) | (0.299) | (0.328) | (0.158) |
| $D2$ | 1.623** | 0.855*** | 4.242*** | 1.966*** | 4.376*** | 1.969*** | 4.380*** | 1.928*** | | |
| | (0.635) | (0.093) | (0.530) | (0.255) | (0.592) | (0.275) | (0.607) | (0.287) | | |
| $D3$ | 1.710** | 1.400*** | 4.074*** | 2.243*** | 4.217*** | 2.236*** | 4.246*** | 2.209*** | -0.170 | 0.255* |
| | (0.758) | (0.130) | (0.491) | (0.233) | (0.559) | (0.258) | (0.573) | (0.260) | (0.286) | (0.135) |
| $D4$ | 1.677** | 2.277*** | 3.395*** | 2.996*** | 3.563*** | 2.985*** | 3.606*** | 2.967*** | -0.781*** | 1.047*** |
| | (0.736) | (0.132) | (0.462) | (0.213) | (0.549) | (0.238) | (0.568) | (0.242) | (0.278) | (0.135) |
| $N2 \times D2$ | | | -4.079*** | -1.493*** | -4.106*** | -1.509*** | -4.166*** | -1.491*** | | |
| | | | (0.839) | (0.446) | (0.927) | (0.473) | (0.922) | (0.454) | | |
| $N2 \times D3$ | | | -4.554*** | -1.379*** | -4.624*** | -1.416*** | -4.699*** | -1.396*** | -0.409 | 0.148 |
| | | | (0.978) | (0.509) | (0.993) | (0.511) | (1.006) | (0.497) | (0.654) | (0.277) |
| $N2 \times D4$ | | | -4.325*** | -1.033** | -4.316*** | -1.066** | -4.388*** | -1.051** | -0.175 | 0.450* |
| | | | (0.895) | (0.523) | (0.979) | (0.537) | (0.967) | (0.514) | (0.512) | (0.253) |
| $N3 \times D2$ | | | -3.392*** | -1.033*** | -3.500*** | -1.075*** | -3.453*** | -1.025*** | | |
| | | | (0.561) | (0.265) | (0.627) | (0.282) | (0.668) | (0.297) | | |
| $N3 \times D3$ | | | -3.629*** | -0.926*** | -3.706*** | -0.998*** | -3.660*** | -0.950*** | -0.263 | 0.040 |
| | | | (0.544) | (0.256) | (0.604) | (0.276) | (0.637) | (0.279) | (0.388) | (0.182) |
| $N3 \times D4$ | | | -3.624*** | -0.898*** | -3.704*** | -1.008*** | -3.684*** | -0.970*** | -0.326 | 0.023 |
| | | | (0.549) | (0.270) | (0.606) | (0.284) | (0.641) | (0.287) | (0.444) | (0.204) |
| $N4 \times D2$ | | | -3.327*** | -1.173** | -3.347*** | -1.191** | -3.322*** | -1.137** | | |
| | | | (1.144) | (0.456) | (1.219) | (0.468) | (1.227) | (0.467) | | |
| $N4 \times D3$ | | | -4.554*** | -1.416*** | -4.459*** | -1.435*** | -4.444*** | -1.394*** | -1.530 | -0.326 |
| | | | (1.180) | (0.478) | (1.225) | (0.486) | (1.233) | (0.480) | (1.095) | (0.313) |
| $N4 \times D4$ | | | -3.535*** | -1.195** | -3.441** | -1.245** | -3.435*** | -1.206** | -0.417 | -0.178 |
| | | | (1.282) | (0.531) | (1.337) | (0.530) | (1.320) | (0.522) | (1.171) | (0.401) |
| $N0 \times D2$ | | | -4.245*** | -1.548*** | -4.336*** | -1.571*** | -4.324*** | -1.525*** | | |
| | | | (0.558) | (0.307) | (0.627) | (0.329) | (0.662) | (0.344) | | |
| $N0 \times D3$ | | | -3.911*** | -1.215*** | -3.975*** | -1.256*** | -3.984*** | -1.221*** | 0.423 | 0.329** |
| | | | (0.542) | (0.297) | (0.613) | (0.320) | (0.649) | (0.330) | (0.281) | (0.139) |
| $N0 \times D4$ | | | -3.006*** | -1.093*** | -3.077*** | -1.155*** | -3.104*** | -1.133*** | 1.186*** | 0.359** |
| | | | (0.523) | (0.270) | (0.604) | (0.288) | (0.642) | (0.298) | (0.272) | (0.142) |
| $\ln(SP_{d,t-5})$ | | | | | -0.060** | 0.043*** | -0.057** | 0.045*** | -0.069*** | 0.047*** |
| | | | | | (0.025) | (0.015) | (0.026) | (0.016) | (0.027) | (0.017) |
| $Ped1$ | | | | | | | -0.299 | -0.077 | -0.161 | -0.016 |
| | | | | | | | (0.224) | (0.119) | (0.236) | (0.130) |
| $Ped2$ | | | | | | | -0.382 | 0.239 | -0.132 | 0.401** |
| | | | | | | | (0.454) | (0.194) | (0.444) | (0.200) |
| NLExp | | | | | | | 0.310 | 0.024 | 0.284 | -0.073 |
| | | | | | | | (0.242) | (0.156) | (0.257) | (0.150) |
| $p_j^m$ | | | | | | | | | -0.935** | -0.168 |
| | | | | | | | | | (0.470) | (0.233) |
| Constant | 0.179 | -3.499*** | -1.399** | -3.945*** | -1.324** | -4.018*** | -1.255* | -3.984*** | 2.738*** | -2.449*** |
| | (0.754) | (0.258) | (0.614) | (0.299) | (0.649) | (0.284) | (0.687) | (0.305) | (0.666) | (0.393) |
| $\ln(\alpha)$ | | 0.802*** | | 0.535*** | | 0.496*** | | 0.510*** | | 0.389* |
| | | (0.263) | | (0.190) | | (0.188) | | (0.194) | | (0.215) |
| $N$ | | 212092 | | 212092 | | 212092 | | 212092 | | 149712 |
| *AIC* | | 72104.53 | | 71883.30 | | 71681.89 | | 71629.41 | | 59397.70 |
| *BIC* | | 72833.33 | | 72858.46 | | 72677.58 | | 72686.68 | | 60339.77 |

Robust (clustered across pathologies) standard errors in parentheses.

Therapeutic class and genetic dummy variables included in all specifications.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$



Table 3: Results – baseline regression

were discovered only in recent years, explaining some of the zero outcomes. As expected, results highlight that the stock of publications decreases the probability of having a “certain zero", while it increases the number of ODD for any positive level of investment.

In Column (4) we take into account also the characteristics of the disease in terms of life expectancy and affected population. The coefficients for these variables, however, are never statistically significant.

Finally, in Column (5), we control for the average probability of receiving a marketing authorization for drugs belonging to each therapeutic class. This, however, causes a sample reduction, since the probability of success cannot be computed for years 1983-1990. Given the large reduction in the number of observations for the first time period (comprising years 1983 to 1992), we omit this time period from the estimation. For this reason, in Column (5) the reference category is the second time period (1993-1999). As can be seen in the Logit part of the model, a higher probability of success in the past is associated with a lower probability of belonging to the “certain zero" group. This might reflect a fixed effect at the disease level. The variable is not statistically significant in the count part of the model.

Importantly, results about the heterogeneous effect of Orphan Regulations across classes of prevalence reported in Column (3)-(5) confirm results of the baseline specification reported in Column (2). This also holds for the expected number of ODD predicted from the Logit and count models.

## 7.1 Robustness checks

In this section we perform a set of additional regressions in order to check the robustness of our specification. Results are reported in Table 4.

In the count part of the model presented in Column (1), therapeutic class dummies have been interacted with period dummies. In principle, the presence of technological reforms, taking place in the same years as the orphan regulations and positively affecting innovation in therapeutic classes characterized by a higher prevalence, might bias our results. As highlighted by the robustness of coefficients for the interaction terms, however, this is not the case.

Drugs may receive an orphan designation also after the marketing approval. This happens when the drug is not initially developed for the orphan disease, but its efficacy for the orphan disease is subsequently discovered. Since these designations do not really represent a new innovative effort, in Column (2) we exclude them when counting the number of ODD: the dependent variable takes into account only those designations referring to drugs for which the marketing approval was subsequent to the designation.

| | (1) | | (2) | | (3) | | (4) | | (5) |
|---|---|---|---|---|---|---|---|---|---|
| Dep. Var. | $y_{dt}$ | | $\tilde{y}_{dt}$ | | $y_{dt}$ | | $y_{d,t+5}$ | | $y_{dt}$ |
| Sample | Full | | Full | | $\sum_t P_{dt} > 0$ | | $t \geq 1988$ | | Full |
| | logit | count | logit | count | logit | count | logit | count | poisson |
| $N2$ | 3.936** | 2.493*** | 3.137*** | 1.397*** | 2.647** | 1.533*** | 2.880*** | 1.472*** | 0.309 |
| | (1.670) | (0.672) | (0.827) | (0.458) | (1.188) | (0.554) | (0.863) | (0.405) | (0.236) |
| $N3$ | 3.002** | 2.265*** | 1.925*** | 1.182*** | 1.703** | 1.404*** | 2.440*** | 1.601*** | 0.405** |
| | (1.373) | (0.387) | (0.574) | (0.236) | (0.730) | (0.270) | (0.693) | (0.271) | (0.166) |
| $N4$ | 2.974** | 2.503*** | 1.591 | 1.289** | 1.612 | 1.812*** | 1.907* | 1.819*** | 0.714*** |
| | (1.411) | (0.541) | (1.482) | (0.522) | (1.177) | (0.523) | (1.041) | (0.415) | (0.222) |
| $N0$ | 3.445** | 1.862*** | 3.052*** | 1.133*** | 2.714*** | 1.369*** | 2.658*** | 1.109*** | 0.041 |
| | (1.388) | (0.398) | (0.578) | (0.314) | (0.591) | (0.376) | (0.541) | (0.235) | (0.125) |
| $D2$ | 3.642** | 1.556*** | 4.309*** | 1.942*** | 3.434*** | 1.907*** | 4.498*** | 2.538*** | 0.105 |
| | (1.605) | (0.493) | (0.506) | (0.294) | (0.973) | (0.345) | (0.568) | (0.245) | (0.097) |
| $D3$ | 3.885** | 2.021*** | 4.195*** | 2.236*** | 3.420*** | 2.294*** | 4.263*** | 2.577*** | 0.484*** |
| | (1.573) | (0.464) | (0.485) | (0.289) | (0.996) | (0.323) | (0.553) | (0.247) | (0.113) |
| $D4$ | 3.369** | 2.598*** | 3.419*** | 2.957*** | 2.957*** | 3.110*** | 2.645*** | 2.733*** | 1.528*** |
| | (1.410) | (0.390) | (0.468) | (0.246) | (0.859) | (0.282) | (0.514) | (0.205) | (0.098) |
| $N2 \times D2$ | -4.562*** | -2.351*** | -4.151*** | -1.461*** | -3.429** | -1.468*** | -4.521*** | -2.067*** | 0.340 |
| | (1.658) | (0.606) | (0.995) | (0.538) | (1.413) | (0.539) | (1.059) | (0.426) | (0.208) |
| $N2 \times D3$ | -5.264** | -2.384*** | -4.721*** | -1.400*** | -3.896** | -1.407** | -4.627*** | -1.206*** | 0.582** |
| | (2.098) | (0.810) | (0.992) | (0.538) | (1.584) | (0.621) | (0.887) | (0.426) | (0.253) |
| $N2 \times D4$ | -5.379*** | -2.188*** | -4.342*** | -1.009* | -4.189*** | -1.245** | -3.625*** | -0.819** | 0.729*** |
| | (1.525) | (0.632) | (1.028) | (0.577) | (1.016) | (0.613) | (0.782) | (0.376) | (0.240) |
| $N3 \times D2$ | -3.954** | -1.803*** | -3.304*** | -0.980*** | -2.777*** | -1.030*** | -4.419*** | -1.879*** | 0.572*** |
| | (1.587) | (0.466) | (0.587) | (0.298) | (0.850) | (0.341) | (0.637) | (0.283) | (0.151) |
| $N3 \times D3$ | -4.595*** | -2.153*** | -3.642*** | -0.935*** | -3.179*** | -1.045*** | -4.397*** | -1.384*** | 0.629*** |
| | (1.584) | (0.459) | (0.567) | (0.299) | (0.819) | (0.310) | (0.655) | (0.330) | (0.167) |
| $N3 \times D4$ | -4.671*** | -1.956*** | -3.639*** | -0.861*** | -3.343*** | -1.051*** | -3.910*** | -1.012*** | 0.460*** |
| | (1.421) | (0.423) | (0.575) | (0.292) | (0.769) | (0.313) | (0.712) | (0.311) | (0.159) |
| $N4 \times D2$ | -3.938** | -1.709*** | -3.198** | -0.928* | -2.967*** | -1.245** | -4.278*** | -2.058*** | 0.409** |
| | (1.550) | (0.544) | (1.334) | (0.527) | (1.019) | (0.560) | (0.869) | (0.418) | (0.166) |
| $N4 \times D3$ | -5.979*** | -2.508*** | -4.460*** | -1.220** | -4.466*** | -1.619*** | -5.059*** | -1.767*** | 0.560*** |
| | (1.637) | (0.551) | (1.294) | (0.543) | (1.032) | (0.532) | (1.054) | (0.486) | (0.210) |
| $N4 \times D4$ | -4.969*** | -2.168*** | -3.336** | -0.909 | -3.887** | -1.549** | -3.342*** | -1.240** | 0.323 |
| | (1.326) | (0.574) | (1.432) | (0.580) | (1.575) | (0.702) | (1.198) | (0.495) | (0.204) |
| $N0 \times D2$ | -4.267*** | -2.072*** | -4.293*** | -1.498*** | -3.586*** | -1.550*** | -3.984*** | -1.735*** | 0.395*** |
| | (1.629) | (0.481) | (0.579) | (0.362) | (0.780) | (0.412) | (0.580) | (0.273) | (0.114) |
| $N0 \times D3$ | -4.222*** | -2.009*** | -4.003*** | -1.189*** | -3.425*** | -1.349*** | -3.492*** | -1.182*** | 0.549*** |
| | (1.618) | (0.481) | (0.582) | (0.363) | (0.682) | (0.387) | (0.577) | (0.262) | (0.136) |
| $N0 \times D4$ | -3.439** | -1.754*** | -3.040*** | -1.063*** | -2.683*** | -1.286*** | -1.979*** | -0.739*** | 0.363*** |
| | (1.464) | (0.404) | (0.576) | (0.311) | (0.610) | (0.343) | (0.545) | (0.223) | (0.127) |
| Constant | -1.618 | -3.938*** | -1.251** | -3.880*** | -1.339** | -4.291*** | -1.381** | -3.804*** | -4.377*** |
| | (1.099) | (0.325) | (0.620) | (0.280) | (0.599) | (0.415) | (0.646) | (0.280) | (0.147) |
| Ther. dummies | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| Ther. dummies $\times Dt$ | N | Y | N | N | N | N | N | N | N |
| $\ln(\alpha)$ | | 0.352 | | 0.609*** | | 0.579* | | 0.462** | |
| | | (0.243) | | (0.208) | | (0.344) | | (0.213) | |
| $N$ | | 212092 | | 212092 | | 183328 | | 180902 | 212092 |
| *AIC* | | 71072.68 | | 69989.72 | | 64799.53 | | 67753.11 | 89469.51 |
| *BIC* | | 72879.28 | | 70964.87 | | 65760.84 | | 68713.15 | 89951.95 |

Robust (clustered across pathologies) standard errors in parentheses.

Therapeutic class and genetic dummy variables included in all specifications.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$



Table 4: Results – Robustness checks

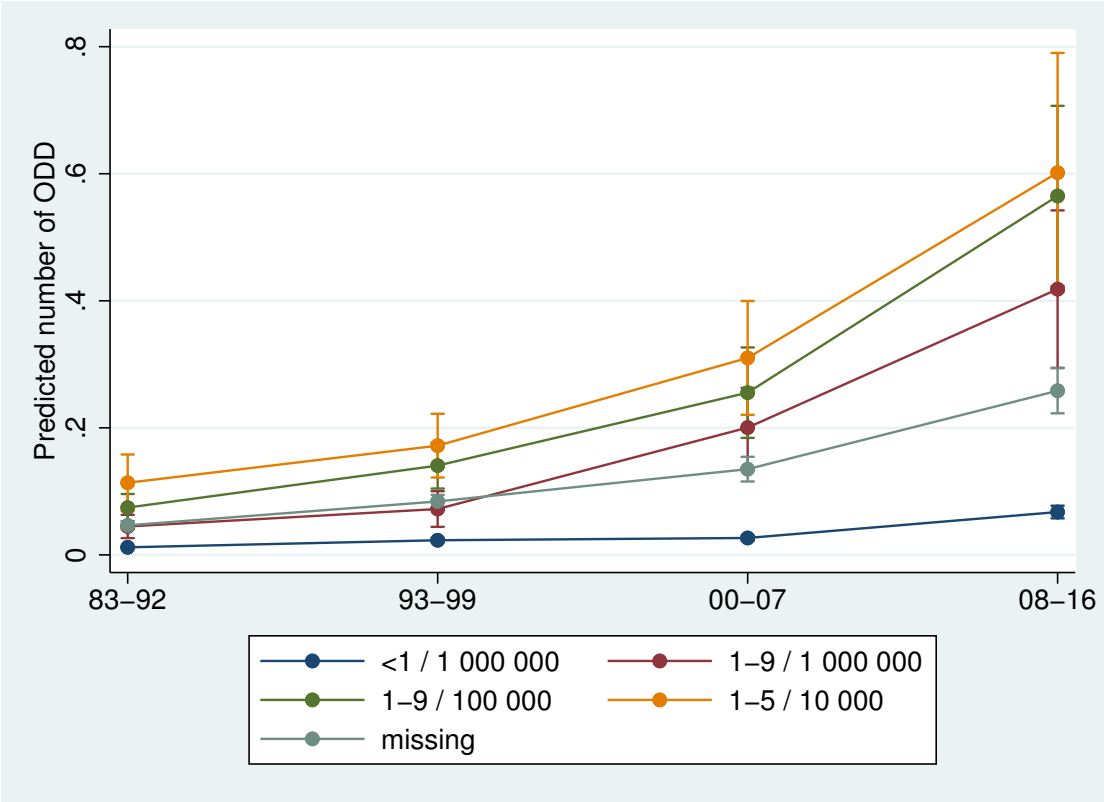


Figure 2: Predicted number of ODD when considering a five years lag in the regressors (see Column (4)).

As one of the main issues in our analysis is the large share of zero designations that characterizes our sample, in Column (3) we take into account only those diseases for which at least one scientific publication is observed over the entire sample period.

Results reported in the above mentioned columns are qualitatively similar to those reported in the baseline specification.

The reforms may have an immediate effect on R&D, but few years are needed for innovation input (R&D) to transform in innovation output (such as ODD). To take this into account, in Column (4) we consider the effect of independent variables at time $t$ on the number of ODD in $t + 5$. The five years window has been selected as it is the average time span from the beginning of clinical trials to the request for ODD.[19] When the time lag is taken into account, the estimated effect of the reforms is larger, as can be seen from the comparison of Figure 2 and Figure 1(d). As an example, consider the number of ODD in year 2001. When the lag is not taken into account, the increase in designations in 2001 is attributed to the extension of market exclusivity also to Europe. If there is a time lag, however, this increasing trend should be attributed to the extension of market exclusivity to Japan. As Japan is a smaller market compared to Europe, we expect the value of $z$ to be smaller. Therefore, the results that do not take into account the research designation lag may be downward biased.

Finally, a Poisson quasi maximum likelihood estimators is adopted, as this estimator is robust to distributional misspecification (Gourieroux et al., 1984; Gouriéroux et al., 1981). The Poisson specification, however, does not allow to disentangle the two processes that may determine a zero outcome, and only allows to analyse the combined effect. Results, reported in Column (5), confirm the positive effect of the reforms on R&D effort and its negative effect in terms of equity among diseases belonging to different classes of prevalence.

# 8 Concluding remarks

Since the early 80s regulators have started to address the lack of incentives to invest in innovation for rare diseases by means of specific provisions. As the pharmaceutical market is a global one,

[19]The five years window is estimated by combining our own computation on FDA data and data on drug development length provided by DiMasi et al. (2016). According to our computation, the *average* time lag between designation and marketing approval for drugs designated before 2005 is 68 months (again, we consider the 2005 limit to avoid data censoring that characterizes more recent years). DiMasi et al. (2016) reports a time period of 126 months from synthesis to approval. By taking the difference between these two numbers, we find that designations take place on average five years after synthesis of the compound. This result is in line with Hay et al. (2014), who find that ODD are most often received when a drug is in phase 2, that is roughly five years from synthesis (according to DiMasi et al., 2016 estimates).

these incentives for the development of *orphan drugs* have cumulated over time as new countries have introduced them. There is ample evidence that this has led to more investments in projects targeting rare diseases, meaning a potential reduction in inequality between orphan and common diseases. In this paper we study the distribution of R&D efforts within the class of orphan diseases, with a focus on heterogeneity with respect to prevalence.

We developed a theoretical model to show that the type of incentive that is used may be crucial to define the relative convenience to invest across different classes of prevalence. In particular, we compare *market-oriented* incentives, such as market exclusivity with *input-oriented* incentives, such as tax credits. We show that both increase more the probability of observing investment for a less rare disease, but this difference is greater for *market-oriented* incentives. The difference in the impact on the level of investment across different classes of prevalence is ambiguous in both cases.

We then use the number of orphan designations, a condition to become eligible for incentives, as a proxy of R&D effort, to investigate the impact of the introduction of incentives in different geographic areas over time. The estimated impacts on the probability of having any investment, and the intensity of investment, are in line with the theoretical predictions. We find that while the number of designations has increased over time for all orphan diseases, those in the class with the lowest prevalence have done so at a much slower pace.

The gap seems to have widened after 2000, when the orphan legislation has been introduced in the EU. We argue that the large weight of *market-oriented* incentives embodied in this legislation, when compared for example with the US legislation, combined with the large size of the EU market, may have contributed substantially to the increase in inequality among orphan diseases. If the reduction of this form of inequality is an objective of European policy makers, then the weight of input-related incentives should be increased. However, the adoption of some of these incentives, such as tax credits, may be more challenging than in other regulatory frameworks, due to the fact that single EU member countries are still responsible for the definition of fiscal policies. Hence, an extension of the incentive tool set to include provisions that can be tailored to the prevalence of a disease, should also be considered.

# References

Acemoglu, D. and Linn, J. (2004). Market size in innovation: theory and evidence from the pharmaceutical industry. *The Quarterly journal of economics*, 119(3):1049–1090.

Braun, M. M., Farag-El-Massah, S., Xu, K., and Coté, T. R. (2010a). Emergence of orphan drugs

in the united states: a quantitative assessment of the first 25 years. *Nature Reviews Drug Discovery*, 9(7):519.

Braun, M. M., Farag-El-Massah, S., Xu, K., and Coté, T. R. (2010b). Emergence of orphan drugs in the United States: a quantitative assessment of the first 25 years. *Nature Reviews Drug Discovery*, 9(7):519.

Denis, A., Mergaert, L., Fostier, C., Cleemput, I., and Simoens, S. (2010). A comparative study of european rare disease and orphan drug markets. *health Policy*, 97(2-3):173–179.

Detiček, A., Locatelli, I., and Kos, M. (2018). Patient access to medicines for rare diseases in european countries. *Value in Health*, 21(5):553–560.

DiMasi, J. A., Grabowski, H. G., and Hansen, R. W. (2016). Innovation in the pharmaceutical industry: new estimates of R&D costs. *Journal of health economics*, 47:20–33.

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

Gourieroux, C., Monfort, A., and Trognon, A. (1984). Pseudo maximum likelihood methods: Theory. *Econometrica: Journal of the Econometric Society*, pages 681–700.

Gouriéroux, C., Monfort, A., Trognon, A., and d'études prospectives d'économie mathématique appliquées à la planification (Paris), C. (1981). *Pseudo Maximum Likelihood, Methods: Applications to Poisson Models*. Centre d'études prospectives d'économie mathématique appliquées à la planification.

Greene, W. (2010). Testing hypotheses about interaction terms in nonlinear models. *Economics Letters*, 107(2):291–296.

Hay, M., Thomas, D. W., Craighead, J. L., Economides, C., and Rosenthal, J. (2014). Clinical development success rates for investigational drugs. *Nature biotechnology*, 32(1):40.

Health and Safety, F. (2015). *Inventory of Union and Member State incentives to support research into, and the development and availability of, orphan medicinal products*. European Commission.

Jobjörnsson, S., Forster, M., Pertile, P., and Burman, C.-F. (2016). Late-stage pharmaceutical r&d and pricing policies under two-stage regulation. *Journal of health economics*, 50:298–311.

Keller, W. (2002). Geographic localization of international technology diffusion. *American Economic Review*, 92(1):120–142.

Kesselheim, A. S. (2010). Innovation and the Orphan Drug Act, 1983–2009: the regulatory and clinical characteristics of approved orphan drugs. *Rare Diseases and Orphan Products: Accelerating Research and Development*, pages 291–308.

Lambert, D. (1992). Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, 34(1):1–14.

Mansfield, E. (1995). Academic research underlying industrial innovations: sources, characteristics, and financing. *The review of Economics and Statistics*, pages 55–65.

Mariz, S., Reese, J. H., Westermark, K., Greene, L., Goto, T., Hoshino, T., Llinares-Garcia, J., and Sepodes, B. (2016). Worldwide collaboration for orphan drug designation. *Nature reviews Drug discovery*, 15(6):440.

Pavitt, K. (1984). Sectoral patterns of technical change: towards a taxonomy and a theory. *Research policy*, 13(6):343–373.

Schey, C., Milanova, T., and Hutchings, A. (2011). Estimating the budget impact of orphan medicines in europe: 2010-2020. *Orphanet journal of rare diseases*, 6(1):62.

Sharma, A., Jacob, A., Tandon, M., and Kumar, D. (2010). Orphan drug: development trends and strategies. *Journal of Pharmacy and Bioallied Sciences*, 2(4):290.

Staub, K. E. and Winkelmann, R. (2013). Consistent estimation of zero-inflated count models. *Health economics*, 22(6):673–686.

Wästfelt, M., Fadeel, B., and Henter, J.-I. (2006). A journey of hope: lessons learned from studies on rare diseases and orphan drugs. *Journal of internal medicine*, 260(1):1–10.

Westermark, K. et al. (2011a). European regulation on orphan medicinal products: 10 years of experience and future perspectives. *Nature Reviews Drug Discovery*, 10(5):341.

Westermark, K. et al. (2011b). European regulation on orphan medicinal products: 10 years of experience and future perspectives. *Nature Reviews Drug Discovery*, 10(5):341.

Winkelmann, R. (2008). *Econometric analysis of count data*. Springer Science & Business Media.

Yin, W. (2008). Market incentives and pharmaceutical innovation. *Journal of Health Economics*, 27(4):1060–1077.