# DATA-DRIVEN HEALTH INNOVATION AND PRIVACY REGULATION

Chiara Conti,[a] Anna D'Annunzio,[b] Pierfrancesco Reverberi[a]

(Preliminary version)

## Abstract

Data-driven health innovation may lead to develop targeted treatments using health data. We consider privacy-sensitive patients who may decide to share personal health data if compensated. Each patient does not internalize the impact of sharing data on drug innovation. We show that investment incentives in targeted treatments are too weak due to the costs for collecting health data. Then, privacy protection measures reducing data sharing risks can promote pharmaceutical R&D and social welfare. We also investigate the effects of a policy allowing firms to access health data for medical research without patients' consent.

**JEL classification**: L12, L51, I10

**Keywords:** Health data; Data-driven innovation; Precision medicine; Privacy regulation

[a] Department of Computer, Control, and Management Engineering, Sapienza University of Rome

[b] TBS Business School

## 1. Introduction

In recent years, the availability of extensive health-related datasets, combined with the use of artificial intelligence techniques, are shaping the way research on new drugs and treatments is accomplished. Data-driven health innovation can lead to the early intervention of diseases, as well as enable the design of tailored treatments for patients. These features have enormous potential for tackling important health care problems, in particular oncological and genetic illnesses.

The motivating example for this paper is precision medicine (PM), which aims at finding the most effective cure depending on patients' genetic, environmental, and lifestyle factors, thereby reducing side effects and improving survival rates. In doing so, it is candidate to outperform traditional medical treatments adopting a one-size-fits-all therapy designed for the 'average patient'. The potential value of PM influenced the $215-million allocation of public funds to the US Initiative.[1]

If, on the one hand, easy access to large amounts of health data is essential to R&D activities, on the other hand, there are significant concerns about potential misuses of sensitive data. Privacy regulation aims at addressing these concerns by providing patients with control over personal data. Nonetheless, pharmaceutical R&D may be hindered when it is costly to obtain consent from each individual patient (EC, 2022).[2]

We build a theoretical model to study the interplay between pharmaceutical innovation, price regulation of drugs and privacy regulation. We analyze a monopolist's incentives to develop a targeted treatment for an eligible patient group, rather than offer a standard treatment for all patients.[3] Due to perceived risks of data sharing, the firm must compensate patients to obtain consent to collect

[1] See https://obamawhitehouse.archives.gov/precision-medicine.

[2] In Europe, personal data collection for R&D purposes may be subject to special rules (Article 89 of GDPR). The appropriate use of personal data for improving health care is a pillar of the EU Data Strategy (EC, 2020).

[3] Gonzalez *et al*. (2016) define (radical) 'horizontal' drug innovations as advances benefiting a given patient group because of lower side effects. They provide several examples of such innovations in the market for statins.

personal health data for research. Drug prices are negotiated with the government, which fully reimburses patients.[4]

We investigate the following issues. What determines incentives to invest in targeted treatments? Are private incentives to invest strong enough from a social viewpoint? Does consent-based privacy regulation ultimately improve patient surplus and social welfare? The answers depend on demand-side factors such as the eligible patient group size and the incremental benefit of the targeted treatment, and supply-side factors such as data collection and research costs, and R&D uncertainty.

Our paper connects two related strands of literature. The first one considers the interaction between drug price regulation and investment incentives in pharmaceutical R&D (Bardey *et al*., 2010; Gonzalez *et al*., 2016). The second strand studies how consent-based privacy regulation affects product innovation (Lefouili and Toh, 2019; Conti and Reverberi, 2021).

Despite the widespread interest in PM (see e.g. Stern *et al*., 2017), the theoretical literature on this topic is limited. Antoñanzas *et al*. (2015) study the incentives of health authorities to use predictive biomarkers to inform treatment choices. Brekke *et al*. (2022) examine how biomarker tests affect competition between existing drugs and the design of health plans. Mougeot and Naegelen (2022) assess the impact of price regulation on the viability of PM.

These papers ignore personal data sharing and privacy issues, which are key to our analysis. Miller and Tucker (2018) find empirically that patients' control over data redisclosure, but not an informed consent policy, boosts the spread of genetic testing.

This paper is organized as follows. Section 2 presents the model. Section 3 derives the equilibrium. Section 4 analyzes welfare. Section 5 discusses policy implications. Section 6 introduces an extension of the baseline model to allow for higher health benefit from data sharing. Section t concludes.

## 2. The model

[4] This assumption fits oncological, genetic, and degenerative diseases in the national health systems of many EU countries.

We consider three types of players: patients, a pharmaceutical firm, and the government. There is a unit mass of heterogeneous patients in the therapeutic market. Patients differ across two dimensions, genetic features and privacy attitudes.

Patients' genetic features entail diverse therapy responses. From this perspective, we distinguish a (homogeneous) patient group of size $\alpha$ $(0 < \alpha < 1)$ from the rest of patients of size $(1 - \alpha)$, where $\alpha$ is common knowledge (e.g., from previous studies). Moreover, patients incur idiosyncratic privacy costs $\beta$ due to their perceived risks of misuse of personal health data. Let $\beta$ be uniformly distributed in $[0,\ \bar{\beta}]$.

A monopolist producing at zero cost offers a 'one-size-fits-all' treatment with an average health benefit of $q_0$. Using patients' health data, the firm may undertake project $H$. This may yield a targeted treatment (e.g., PM) to the eligible group of size $\alpha$, whose patients receive a health benefit of $q_H$, with $q_H > q_0$ (e.g., due to lower side effects). Thus, $\alpha(q_H - q_0)$ is the total incremental benefit of the targeted treatment. We assume that project $H$ is stochastic, and the probability of success $\phi(d)\colon [0,1] \to [0,1]$ increases with the amount $d$ of patients' health data (i.e., $\phi'(d) > 0$, with $\phi(0) = 0$ and $\phi(1) = 1$). For simplicity, we assume that $\phi(d) = d$.

The total cost for undertaking project $H$ is $F + c(\delta)$, where $F > 0$ is the R&D cost[5] and $c(\delta) = \delta d$ is the data collection cost, with $\delta$ being the incentive offered to each patient to gather personal health data. Given the perceived risks of sharing health data, the firm must compensate patients (through monetary payments or health benefits from participation in medical trials) to obtain their consent to use personal data for R&D.[6]

[5] This may include, for instance, the cost of developing biomarker tests.

[6] For each patient, the probability of receiving $q_H$ is independent of personal data sharing. Indeed, genetic profiles (i.e., patients' types) are identified at the time when research is performed for those sharing data, or *ex post* for those not sharing data (e.g., through biomarker tests).

Let $p_i$ be the price of treatment $i$ $(i = 0, H)$. The government negotiates $p_H$ with the monopolist through Nash bargaining, and fully reimburses patients (so that the market is covered). We consider $p_0$ to be exogenous (e.g., negotiated in the past). The timing is as follows:

1. The monopolist decides whether and how much health data to collect for undertaking project $H$.
2. Patients decide whether or not to share personal health data.
3. If treatment $H$ is achieved, then the firm and the government negotiate the price.

## 3. Equilibrium

As usual, we solve the game backwards for subgame perfect Nash equilibria.

*Stage 3. Price negotiation*

Let the government care about patients' health benefits net of social expenditure (see e.g. Bardey *et al*., 2010), namely, $\widetilde{W}_H = \alpha q_H + (1-\alpha)q_0 - \alpha p_H - (1-\alpha)p_0$ with treatment $H$ for the eligible patient group, and $\widetilde{W}_0 = q_0 - p_0$ with the standard treatment for all patients. Let $\lambda > 0$ be the government's bargaining power in price negotiation.

Let the firm (with bargaining power $1-\lambda > 0$) care about profit, namely, $\Pi_H = \alpha p_H + (1-\alpha)p_0$ with treatment $H$ and $\Pi_0 = p_0$ otherwise. R&D and data collection costs are sunk at the price negotiation stage.

Then, the Nash bargaining problem is:

$$\max_{p_H} (\Delta\widetilde{W})^{\lambda} (\Delta\Pi)^{1-\lambda}$$

where $\Delta\widetilde{W} = \widetilde{W}_H - \widetilde{W}_0 = \alpha(q_H - q_0 + p_0) - \alpha p_H$ and $\Delta\Pi = \Pi_H - \Pi_0 = \alpha(p_H - p_0)$ respectively are the incremental surplus and the incremental profit with the targeted treatment relative to the status quo. Thus, the negotiated price is:

$$p_H = p_0 + (1-\lambda)(q_H - q_0).$$

*Stage 2. Data sharing*

Given that the firm undertakes project $H$, a patient with privacy cost $\beta$ accepts compensation $\delta$ and shares personal health data if the (net) individual privacy surplus from sharing data is non-negative, that is, $(\delta - \beta) \geq 0$.

Let $\hat{\beta}$ be the indifferent patient between sharing data or not, who gains zero privacy surplus. Then $\hat{\beta} = \delta$, and patients with privacy costs $\beta$ such that $\beta \leq \hat{\beta}$ do share data. Note that personal data sharing creates an externality, since each patient does not internalize the impact on drug innovation.

*Stage 1. Data collection and investment*

The firm may decide to collect health data and undertake project $H$, with an expected profit of $E(\Pi_H) = \phi(\alpha p_H + (1-\alpha)p_0) + (1-\phi)p_0 - c(\delta) - F$. For a given value of $\delta$, such that $\delta \leq \bar{\beta}$,[7] there are $d = \delta/\bar{\beta} \leq 1$ patients sharing data, so that $c(\delta) = \delta^2/\bar{\beta}$.

Under project $H$, the firm chooses the optimal compensation to induce patients to share data by solving:

$$\max_{\delta} E(\Pi_H) = \frac{\delta}{\bar{\beta}}\alpha(1-\lambda)(q_H - q_0) + p_0 - \frac{\delta^2}{\bar{\beta}} - F. \quad (1)$$

Given that the second-order condition (SOC) holds, the level of $\delta$ maximizing expected profit is:

$$\delta^* = \frac{\alpha}{2}(1-\lambda)(q_H - q_0). \quad (2)$$

Henceforth, we focus on interior solutions where patients with high privacy costs are not compensated and thereby do not share data. Intuitively, this occurs when the highest privacy cost is high enough, namely, $\bar{\beta} > \alpha(q_H - q_0)$ -see equation (4).[8]

By comparing $E(\Pi_H(\delta^*)) = \frac{\alpha^2(1-\lambda)^2(q_H-q_0)^2}{4\bar{\beta}} + p_0 - F$ with $\Pi_0 = p_0$, we find the condition under which the firm prefers to collect health data and undertake project $H$.

[7] A higher value of $\delta$ than the value for which all patients share data is neither profitable nor socially optimal.

[8] Corner solutions where all patients share data add no further insight (results are available from the authors upon request).

***Result* 1.** *If* $\alpha_H = \frac{2}{(1-\lambda)}\sqrt{\overline{\beta}F} \leq \alpha(q_H - q_0) < \overline{\beta}$*, then the firm collects health data and undertakes project* $H$.

For simplicity, we only consider cases where investment in treatment $H$ is not precluded in equilibrium. This requires $\bar{\beta} \geq \alpha_H$, i.e., $\bar{\beta} \geq \frac{4F}{(1-\lambda)^2}$.

From Result 1, data sharing risks reduce the likelihood for targeted treatments being developed, since the critical value of $\alpha(q_H - q_0)$ increases with the highest privacy cost $\overline{\beta}$. Indeed, health data are essential to R&D investment, but are costly to collect.

In Section 4, we investigate whether private and social incentives for pharmaceutical innovation are aligned by comparing private investment decisions with those implemented by a regulator choosing optimal compensation for data sharing while taking into account privacy costs .

## 4. Welfare

Let $W_i = \widetilde{W}_i + \Pi_i + PS_i$ be the social welfare with treatment $i$ ($i = 0, H$),[9] where $PS_H$ is the privacy surplus under $H$, namely, the total compensation for personal data sharing net of aggregate privacy costs (with $PS_0 = 0$). Since $\delta \leq \bar{\beta}$, then $PS_H = \left(\delta\frac{\widehat{\beta}}{\overline{\beta}} - \int_0^{\widehat{\beta}} \frac{1}{\overline{\beta}} s ds\right) = \frac{\delta^2}{2\overline{\beta}}$. When the firm collects health data and undertakes project $H$, the expected social welfare is:

$$E(W_H) = \frac{\delta}{\overline{\beta}}\alpha(q_H - q_0) + q_0 - \frac{\delta^2}{2\overline{\beta}} - F. \tag{3}$$

Given that the SOC holds, the level of $\delta$ maximizing social welfare under project $H$ (with partial data sharing) is:

$$\delta^w = \alpha(q_H - q_0). \tag{4}$$

Then, Result 2 follows from comparing $E(W_H(\delta^w)) = \frac{\alpha^2(q_H - q_0)^2}{2\overline{\beta}} + q_0 - F$ with $W_0 = q_0$.

[9] Qualitative results are not affected when profit is excluded.

***Result* 2**. *If $\alpha(q_H - q_0) \geq \alpha_W = \sqrt{2\bar{\beta}F}$, then (expected) social welfare is higher under treatment H than in the status quo.*

We now compare private and social incentives to collect health data and undertake project $H$.

***Result* 3**. *Since $\delta^* < \delta^W$ and $\alpha_W < \alpha_H$ then private incentives to collect health data and undertake project H are too weak.*

Result 3 indicates that the pharmaceutical firm underinvests in targeted treatments. Specifically, when $\alpha_W \leq \alpha(q_H - q_0) < \alpha_H$ the firm does not undertake project $H$ (since $E(\Pi_H) < \Pi_0$), but social welfare would be higher under treatment $H$ ($E(W_H) > W_0$). Moreover, when $\alpha_H \leq \alpha(q_H - q_0) \leq \bar{\beta}$ the firm does undertake project $H$, but the amount of health data collected, and thereby the probability of success for $H$, are lower than socially optimal.

We also find that the parameter region for which there is underinvestment in the targeted treatment widens with the value of $\overline{\beta}$ (Corollary 1). Indeed, we obtain that $\partial\left[\bar{\beta} - \sqrt{2\overline{\beta}F}\right]/\partial\overline{\beta} > 0$.

***Corollary* 1**. *Underinvestment in project H is more likely as $\overline{\beta}$ increases.*

## 5 Policy implications

Results discussed in the previous section highlight the hurdle the firms may face in gathering health data for medical research. In particular, the need to compensate each patient to obtain consent produces underinvestment in targeted treatments. Our result is in line with the issues raised by the European Commission concerning the limited secondary use of health data (EC 2022).

In the following, we discuss policy interventions aiming at facilitating access to health data for medical research.

### 5.1 Reducing risks of data sharing

Corollary 1 shows that privacy protection measures aimed at reducing perceived risks of sharing personal health data can promote pharmaceutical R&D and social welfare by aligning private and

social incentives for drug innovation. Policy makers may try to reduce the privacy cost that patients perceive by decreasing the perceived risks of potential misuse of data. Policy makers may, for example, foster the provision of more transparent information to consumers, limit the timespan of data retention or require data to be anonymized and/or aggregated. Presumably, these measures can be implemented without reducing the ability to use this data for medical innovation.

**5.2 Restricting patients control over health data**

Recently, the European Commission proposed the creation of a European Health Data Space (EHDS). One of the purposes of this proposal is to provide an efficient set-up for the use of health data for medical research and innovation, thus facilitating data collection. The Communication of the European Commission pertaining the creation of the EHDS envisages the possibility to access health data simply by applying for "a permit from a health data access body" (EC, 2022).

Based on the model discussed above, we analyze how the possibility for the firm to obtain patients' health data without offering compensation for privacy costs affects the incentives to invest in targeted treatments and the overall welfare.

When patients' data are freely available, the monopolist's profit under project *H* is

$$\mathrm{E}(\Pi_H) = \phi\alpha(q_H - q_o)(1-\lambda) + p_o - F$$

Given $\phi = d$, $\Pi_H$ is maximized using all patients' data. Then, the monopolist invests in *H* when $\Pi_H > \Pi_0$, that is

$$\alpha(q_H - q_o) > \frac{F}{(1-\lambda)}$$

As expected, the private threshold for investment is lower than that when the firm needs to compensate patients, meaning that a policy that gives free access to data increases the likelihood to observe investment. However, this does not always result in welfare improvement. Since, under this scenario, $\mathrm{E}(\mathrm{W}_H) > W_0$ when

$$\alpha(q_H - q_o) > \frac{\bar{\beta}}{2} + F$$

and given

$$\frac{F}{(1-\lambda)} < \frac{\bar{\beta}}{2} + F$$

it turns out that overprovision of targeted treatments may arise.

Indeed, this policy may lead to excessive data collection, therefore overall privacy costs are too high and at the same time they are not compensated.

Private and public incentives to invest could be better aligned in presence of a central authority that gives for free an *optimal amount* of patients' data, when socially desirable. In this case, the optimal amount of data $d$ is retrieved by solving the following problem

$$\max_d E(\mathrm{W}_H) = d\alpha(q_H - q_0) + q_0 - F - d^2\frac{\bar{\beta}}{2}$$

This problem is equivalent to the case where we derived of the optimal compensation for health data, because cost of data collection and compensation received by patients cancel out in welfare function. Hence, we conclude that

$$d^w = \frac{\alpha(q_H - q_0)}{\bar{\beta}} = \frac{\delta^w}{\bar{\beta}}$$

Then the threshold for public investment is the same as that given in Result 2, that is $E(\mathrm{W}_H) > \mathrm{W}_0$ when $\alpha(q_H - q_0) \geq \alpha_W = \sqrt{2\bar{\beta}F}$. When the firm receives $d^w$ for free, investment in H occurs when

$$\mathrm{E}(\Pi_H) = d^w\alpha(q_H - q_o)(1-\lambda) + p_o - F > \Pi_0$$

that is for

$$\alpha(q_H - q_0) > \sqrt{\frac{\bar{\beta}F}{(1-\lambda)}}$$

Since this threshold is larger than $\frac{F}{(1-\lambda)}$, the problem of overprovision discuss above is reduced. We find that overprovision (respectively, under-provision) occurs when $\lambda < 1/2$ (respectively, $\lambda > 1/2$). Note that if $\lambda = 1/2$ public and private incentives are perfectly aligned. In this case, allowing firms to use the optimal amount of data for free is always welfare improving.

Nevertheless, patients could be hurt by this policy. Indeed, on the one hand the probability of successful investment increases, leading to higher expected health benefit, on the other hand they receive no more compensation for their data. We compare consumer surplus given that the investment occurs. Let

$$CS_{\delta} = \frac{\delta^*}{\bar{\beta}}\alpha(q_H - q_0) + q_0 + \frac{(\delta^*)^2}{2\bar{\beta}}$$

be the patients' surplus when they receive compensation for their data, and

$$CS_f = d^w\alpha(q_H - q_0) + q_0 - \frac{(d^w)^2\bar{\beta}}{2}$$

be the patients' surplus when the policy is implemented. Then, inserting for $\delta^*$ and $d^w$, we find that the policy increases patients' surplus if $\lambda > 3 - 2\sqrt{2}$~0,17.

We conclude that this policy can increase not only the overall welfare but also patients' surplus.

## 6 Extensions

We are currently working to a promising extension of the model. In the current version of the model we assume that patients that do not share data can be identified with the same probability as those sharing, hence they can have the same expected health benefit from the investment in precision medicine. In this extension, the utility of consumers is the following

$$EU = \begin{cases} \phi[\varepsilon\alpha q_H + (1-\varepsilon\alpha)q_0] + (1-\phi)q_0 + (\delta - \beta), if\ no\ data\ sharing \\ \phi[\alpha q_H + (1-\alpha)q_0] + (1-\phi)q_0 + (\delta - \beta), if\ data\ sharing \end{cases}$$

where parameter $0 \leq \varepsilon < 1$, means that patients not sharing data receive a lower benefit when precision medicine is developed. Hence, this parameter increases the incentives of patients to share the data, ceteris paribus. However, the parameter also reduces the share of the patients to which precision medicine can be provided, hence revenues of the firm. Moreover, differently from the baseline model, patients' decisions on data sharing may generate aggregate privacy costs that exceed firm's costs of data collection, potentially leading to excessive investment. Indeed, in this extension we find a region of parameters where the incentives of the firm to invest in precision medicine may be too high from a social welfare point of view.

## 7 Conclusions

We have studied the incentives to develop a targeted treatment for an eligible patient group identified through personal health data. We have found that the pharmaceutical firm underinvests in the targeted treatment. This is because the firm does not develop the treatment when it is welfare improving, or because she collects a smaller amount of data than socially optimal, thereby reducing the probability of success for the new treatment. Indeed, to collect health data, the firm must compensate privacy-sensitive patients who do not internalize the impact of sharing personal data on pharmaceutical innovation. Therefore, privacy protection measures that reduce patients' perceived risks of sharing health data can promote pharmaceutical R&D and social welfare. Moreover, we have shown that, when patients' health data are freely available for research purposes after permission of a public agency, the likelihood to observe investment in innovative treatments increases. Thus, delegating control over patients' data to a central authority may reduce the externalities generated by individual decisions on data sharing, improving welfare and patients' surplus.

There are several directions for future work. First, as anticipated in Section 6, we aim at extending our analysis to the case in which patients sharing health data benefit more from innovative treatments than patients who refuse to share (e.g., because the latter can be identified via biomarker tests with

limited precision). Thus, for a given compensation, patients may have stronger incentives to share personal data, and this reduces the importance of coordination issues for successful R&D. However, stronger individual incentives to share data could possibly generate excessive data collection from a social point of view.

Second, we may consider that health authorities and privacy agencies interact to incorporate the privacy surplus in price negotiations. As a result, we may have higher negotiated prices for targeted treatments which, in turn, may increase R&D incentives. Finally, we may depart from full reimbursement of drugs and/or drug price negotiation to investigate whether pharmaceutical R&D (e.g., PM) widens social disparities when it implies high treatment prices for patients.[10]

**Acknowledgements**

We acknowledge financial support from Agence Nationale de la Recherche (ANR) – TEPREME project (ANR–21–CE26–0014–01) and Sapienza University of Rome (grant: RP120172B96CC6F4).

**References**

Antoñanzas, F., Juárez-Castelló, C.A., Rodríguez-Ibeas, R., 2015. Some economics on personalized and predictive medicine. *European Journal of Health Economics*, 16, 985–994.

Bardey, D., Bommier, A., Jullien, B., 2010. Retail price regulation and innovation: Reference pricing in the pharmaceuticals industry. *Journal of Health Economics*, 29, 303–316.

Brekke, K.R., Dalen, D.M., Straume, O.R., 2022. Competing with precision: Incentives for developing predictive biomarker tests. NHH Discussion Paper, 06/22.

Conti, C., Reverberi, P., 2021. Price discrimination and product quality under opt-in privacy regulation. *Information Economics and Policy*, 55, 100912.

[10] MIT Technology Review, 2018. *The Precision Medicine Issue*, https://www.technologyreview.com/magazines/the-precision-medicine-issue/.

EC–European Commission, 2020. A European strategy for data. COM(2020), 66 final.

EC–European Commission, 2022. A European health data space: Harnessing the power of health data for people, patients and innovation. COM(2022), 196 final.

Gonzalez, P., Macho-Stadler, I., Perez-Castrillo, D., 2016. Private versus social incentives for pharmaceutical innovation. *Journal of Health Economics*, 50, 286–297.

Lefouili, Y., Toh, Y.L., 2019. Privacy regulation and quality investment, https://ideas.repec.org/p/tse/wpaper/31623.html

Miller, A.R., Tucker, C., 2018. Privacy protection, personalized medicine and genetic testing. *Management Science*, 64, 4648–4668.

Mougeot, M., Naegelen, F., 2022. Incentives to implement personalized medicine under second-best pricing. *Health Economics*, DOI: 10.1002/hec.4588.

Stern, A.D., Alexander, B.M., Chandra, A., 2017. How economics can shape precision medicines. *Science*, 355, 1131–1133.