Data Privacy Regulation and Innovation*

Sukhun Kang
UCSB

Jennifer Kao
UCLA

PRELIMINARY. PLEASE DO NOT CIRCULATE.

January 2, 2026

Abstract

We investigate how data privacy regulations shape firms' research and development decisions. While such regulations aim to increase stakeholder privacy and data security, they can also introduce compliance costs in settings where collecting, analyzing, and sharing sensitive data is central to innovation. We examine this tension in the context of the European Union's General Data Protection Regulation (GDPR) and the pharmaceutical industry. Using detailed clinical trial data and firm-level variation in exposure to the GDPR, we find that GDPR reduces clinical trials by 18 percent, and declines persist even after accounting for shifts in disease composition. GDPR also shifts research processes: firms conduct trials across fewer sites and countries. Research collaborations decline overall, particularly among young firms, and shift away from new partners towards established relationships. Finally, the GDPR shifts innovation outcomes: trials take longer to complete, are less likely to complete successfully, and results are reported with a greater delay. These findings reveal how data privacy regulations shape not only the rate of innovation, but also the process by which it is pursued and its outcomes.

JEL Classifications: O31, O38, I18
Keywords: Innovation; Privacy; Regulation; Collaboration; Experimentation

*We are grateful for feedback from seminar audiences at UCLA, the SoCal Strategy Workshop, and Korea University. Emily Zhao and Brian Lu provided excellent research assistance.

1 Introduction

Firms increasingly rely on collecting, analyzing, and sharing data to develop novel products and services (Goldfarb and Que, 2023). However, the expansion of data-driven research and development (R&D) has raised public concerns regarding privacy protection and data security (Acquisti et al., 2016; Miller and Tucker, 2018). In response, regulatory authorities have implemented data privacy regulations designed to safeguard individual privacy rights and maintain public trust. Most notably, the European Union (EU)'s General Data Protection Regulation (GDPR), which was implemented in 2018 and is widely considered the world's most comprehensive data privacy regulation to date.

How do data privacy regulations shape firms' research and development (R&D) decisions? Unlike other forms of commonly studied regulation, such as entry requirements or intellectual property protection, data privacy regulations shape how data—a key input in the R&D process—is collected, analyzed, and shared. On the one hand, these policies may raise the costs of data-intensive research by implementing additional regulatory requirements. On the other hand, since such regulations strengthen stakeholder privacy and improve data security, they may ultimately increase external stakeholders' willingness to participate in R&D activities that require the use of sensitive data. When data is central to innovation, this may have consequences for not only how much firms innovate, but also how research is conducted and its outcomes (Mahoney et al., 2009; Vakili and McGahan, 2016). In this paper, we utilize the unique features of the pharmaceutical R&D setting to empirically evaluate the causal effect of the GDPR on innovation.

To motivate our empirical analysis, we develop a conceptual framework in which data privacy regulations impact firms' R&D decisions through two mechanisms. First, data privacy regulations increase compliance costs by requiring upfront investments in legal, administrative, and data infrastructure. When compliance costs increase, firms are less likely to invest in projects with marginal expected returns. Second, the R&D process can involve substantial

uncertainty and risks for external stakeholders. By implementing enforceable requirements, data privacy regulations may lower participation costs by reducing uncertainty and risks. The net effect depends on the relative size of these effects. This can affect not only the level of firms' R&D investments but also how research is organized and its outcomes.

Despite the key role of data in the R&D process, there is limited empirical evidence on how data privacy regulations shape innovation. Existing studies have largely examined the immediate impacts of data privacy regulations on immediate performance outcomes (Blind, 2012; Aridor et al., 2020; Jia et al., 2021; Peukert et al., 2022; Johnson et al., 2023; Goldberg et al., 2024). These studies show that such regulations can lead to meaningful effects on online advertising, website traffic and revenues, and venture capital investment. Clarifying the effects of GDPR on innovation is important because well-intentioned regulatory policies might unintentionally constrain firms' R&D decisions with long-term consequences. This paper aims to fill this gap.

Specifically, we examine the causal impact of the EU's GDPR on firms' R&D decisions in the pharmaceutical industry. The GDPR harmonizes data protection standards across EU Member States while granting individuals control over their personal information. In the pharmaceutical industry, data is not only a critical input to R&D but is also particularly sensitive and tightly regulated (Vakili and McGahan, 2016; Buckman et al., 2023). Unlike in consumer data, where these regulations helped filter out lower-value data (Aridor et al., 2023) or even increase the effectiveness of user targeting (Godinho De Matos and Adjerid, 2022), the data required for pharmaceutical research is produced through costly, regulated experimentation and cannot be readily substituted when access and governance become more difficult. Further, because pharmaceutical development and trial operations are global, the GDPR is also relevant beyond Europe: it can affect where firms run trials, how they coordinate across organizations, and how quickly evidence becomes available.

We utilize a comprehensive dataset of clinical trials initiated between 2010 and 2023, with sites spanning the EU, US, and other countries. We then leverage a difference-in-differences

design that exploits cross-firm variation in exposure to GDPR based on firms' pre-regulation geographical distribution of clinical trial activity. Firms with extensive EU-based clinical trial activity prior to GDPR implementation faced greater regulatory exposure than those with minimal EU activity, providing natural variation in exposure to the regulation.

Our analysis yields three main findings. First, we find that firms more exposed to GDPR reduced their number of clinical trials by 18 percent relative to less exposed firms. This effect was immediate, persistent, and holds even after accounting for shifts in disease composition. Second, firms adjust their research processes. They increase the share of trials with pre-specified plans for sharing individual participant data and narrow the scope of their research activity: trials are less likely to be conducted in multiple countries and to involve an EU site. We also observe meaningful shifts in collaboration: overall research collaborations decline, with effects driven by young firms. Further, firms significantly reduce collaborations with new partners while modestly increasing research activity with established collaborators. Finally, we show that the GDPR shapes research outcomes: trials take longer to complete, are less likely to reach successful completion, and firms delay the reporting of trial results.

Our findings offer several important insights for managers navigating data privacy regulations. First, firms should develop greater flexibility in managing their R&D investments. This includes developing expertise in navigating complex consent procedures. Second, our results highlight the value of investing in and maintaining research partnerships. Rather than viewing collaboration solely through the lens of scientific benefits, managers should recognize established partnerships as regulatory assets that provide compliance expertise, shared infrastructure, and reduced coordination costs in data-intensive environments. Firms (particularly young firms) should therefore invest in deepening existing collaborative relationships while building the institutional capabilities needed to establish trust with new partners when opportunities arise. As data privacy regulations continue to spread, the firms best positioned for success may be those that can balance regulatory compliance with sustained innovation through strategic flexibility and collaboration.

This paper contributes to several literatures. First, we contribute to the emerging literature on data privacy regulation by showing how data-related policies shape firms' R&D decisions. Prior research has primarily examined the immediate performance outcomes of data privacy regulations on digital services and consumer-facing firms (e.g., Blind, 2012; Aridor et al., 2020; Jia et al., 2021; Johnson et al., 2023; Goldberg et al., 2024; Goldfarb and Tucker, 2011). We show that regulations targeting data, a key input in the innovation process, impact not only the level of research, but also how research is conducted and its outcomes. This contributes to a growing recognition that data privacy regulation differs from other forms of regulation by directly targeting key inputs essential for R&D (Goldfarb and Tucker, 2012; Acquisti et al., 2016; Aridor et al., 2020; Jia et al., 2021). Recent work on the pharmaceutical industry shows how intellectual property (Sampat and Williams, 2019) and product-specific policies (Yin, 2008; Chandra et al., 2024) can shape pharmaceutical R&D investments, but has not yet examined how data-specific regulations impact decisions throughout the R&D process. By documenting an 18 percent decline in clinical trial along with changes in research processes and outcomes, we demonstrate that data privacy regulations may act as an important constraint on a key input for innovation with meaningful consequences for the level, type, and outcomes of research pursued.

Second, we shed light on the drivers of R&D collaborations. Our findings reveal that firms respond to regulation not by reducing collaboration uniformly, but by reallocating their collaborative efforts toward established partners while reducing the creation of new partnerships. This extends existing work on alliance formation (Gulati, 1995; Gulati and Gargiulo, 1999; Gulati and Nickerson, 2008) by showing how regulatory environments create differential costs that favor established relationships over network expansion. We provide empirical evidence that established partnerships serve as valuable assets that enable firms to share regulatory compliance expertise. This suggests that collaborations may become more concentrated under stringent regulatory regimes.

Finally, we contribute to the broader literature on regulation and firm strategy by showing

how data privacy regulation can have uneven effects among firms. Consistent with prior work that has shown that large firms possess advantages in navigating regulatory requirements (Cockburn and Henderson, 2001), we find that data privacy regulations disproportionately affect younger firms and new partnerships. This complements prior work on regulatory capture and incumbent advantage (Mahoney et al., 2009) by highlighting a distinct mechanism by which regulations create advantages for incumbents, who have the ability to navigate costly regulatory requirements and maintain collaborative relationships essential for data-driven innovation.

2 Institutional Background and Conceptual Framework

This section provides a detailed overview of GDPR's institutional context, highlighting its implications for the pharmaceutical industry, and outlines a conceptual framework linking regulatory changes to firm-level strategic responses and innovation dynamics.

2.1 GDPR

The GDPR was adopted by the European Union in April 2016 and took effect in May 2018 (European Union, 1995). It represents the world's most comprehensive and stringent regulation for governing the collection, storage, usage, and dissemination of personal data. Under the GDPR, personal data is defined to include any information that can identify an individual and ranges from names and location to health and genetic data. While GDPR has strengthened data protection and privacy rights, its provisions have introduced significant regulatory challenges for industries that depend heavily on the collection, processing, and sharing of personal data. In addition, while the GDPR was enacted within the EU, its implications extend globally: it applies not only to organizations based within the EU but also to any organization worldwide that processes the data of EU residents or offers goods and services to individuals in the EU. Thus, the GDPR has established a global regulatory standard for data privacy, affecting firms regardless of geographic boundaries.

The GDPR introduced several important changes. First, it replaced the 1995 Data Protection Directive, which had set minimum privacy protections but varied in its implementation across member states. In contrast, the GDPR established a unified, enforceable legal framework that was applied consistently across all member states.

Second, it strengthened the rights of individuals and increased requirements for firms. Individuals gained the right to access, correct, and delete their personal data. At the same time, firms faced a new set of regulatory expectations that increased costs. These included implementing organizational changes for data security, requirements for internal oversight and documentation, obligations to respond to consumer requests and data breach requests in a timely manner, and expand its geographic reach to apply not only to firms based in the EU but to any firms that offered goods or services to EU residents (European Commission, 2025b).[1] Demirer et al. (2024) find that the GDPR is responsible for the 20% increase in the variable cost of data inputs for affected firms.

Third, GDPR also imposes strict requirements on data transfers, both between organizations and across international borders. Specifically, personal data can only be transferred to entities that meet GDPR-defined standards for "adequate data protection" (European Commission, 2025a). This requirement significantly complicates international data exchanges, making GDPR a regulatory framework with profound implications for global data-driven industries, such as pharmaceuticals, where cross-border data collaboration is integral to innovation.

The GDPR is part of a broader global trend toward strengthening data protections, reflecting rising concerns about privacy in a digital economy. Comparable frameworks include the United States' Health Insurance Portability and Accountability Act (HIPAA) and California's Consumer Privacy Act (CCPA). Although these frameworks share some principles such as informed consent, data minimization, and breach notification requirements, GDPR

[1]In particular, the GDPR mandated that data breaches be reported within 72 hours of discovery and introduced substantial fines for non-compliance: firms can be fined up to 20 million euros or 4% of a firm's global annual revenue for the previous year, whichever was greater—a level significantly higher than most previous data privacy regulations. (https://gdpr-info.eu/issues/fines-penalties/, Accessed: 09/23/2025)

is notably more stringent. Unlike HIPAA, which permits certain healthcare disclosures without explicit consent, GDPR applies broadly across industries. Similarly, while the CCPA focuses primarily on granting California residents control over data sales, GDPR imposes uniform standards for data processing and international transfers. Given the pharmaceutical sector's reliance on large-scale, sensitive personal data, GDPR's expansive coverage and strict requirements impose more substantial operational constraints, making it a particularly critical context for examining the relationship between privacy regulation and innovation.

2.2 Pharmaceutical R&D and GDPR

2.2.1 Pharmaceutical R&D

The pharmaceutical industry is highly data-intensive, with drug development representing a high-risk, high-reward process that relies heavily on clinical trial data to evaluate the safety and efficacy of experimental therapies (DiMasi et al., 2016; Lakdawalla, 2018). This process typically begins with preclinical research, where compounds are tested on cells and animals to assess their potential. Promising candidates then move into human clinical trials, which are divided into three main phases: Phase I focuses on assessing safety and dosage in a small group of healthy volunteers or patients. Phase II evaluates the drug's efficacy and further assesses safety in a slightly larger group. Phase III, often the costliest and most extensive phase, involves large-scale trials to confirm efficacy, monitor side effects, and compare the new treatment to standard therapies. The conduct of clinical trials is governed by the Clinical Trials Regulation (Regulation EU No 536/2014) (European Medicines Agency, 2014), which harmonizes procedures across member states and ensures high standards of safety and transparency.

Following the completion of clinical trials, drug developers can pursue one of several pathways for marketing authorization (Khanna et al., 2018; Van Norman, 2018). For example, under the centralized procedure, drug developers submit a single application to the

European Medicines Agency (EMA). The EMA conducts a scientific evaluation and provides a recommendation to the European Commission (EC). If approved, the drug is authorized for marketing across all EU member states. This pathway is mandatory for drugs targeting certain life-threatening conditions (e.g., cancer, HIV/AIDS, diabetes, neurodegenerative and rare diseases) and is also available for drugs that demonstrate substantial patient benefit. For more details, see Appendix A.

Drug developers located outside the EU frequently engage in clinical trials with the intention of applying for EU marketing authorization. In these cases, developers must ensure that trials adhere not only to local regulatory standards but also to principles equivalent to those outlined in the EU's Clinical Trials Regulation. This underscores the inherently global nature of pharmaceutical research and drug development. As a result, clinical data and regulatory practices often cross international borders, compelling firms to manage and harmonize diverse regulatory expectations while conducting research on a global scale.

Given the significant operational complexity, high costs, and regulatory demands associated with global drug development, pharmaceutical firms increasingly rely on collaboration among firms to share resources, expertise, and risks (Moss et al., 2011). Such collaborations, involving partnerships across multiple jurisdictions, help firms navigate varying regulatory environments, manage compliance costs, and efficiently share sensitive clinical data. Such partnerships are especially critical for costly or risky projects such as those targeting rare or complex diseases, where collective resources and expertise are essential for successful development. Because clinical development depends so extensively on the collection and exchange of sensitive patient data, changes to data protection law have the potential to substantially alter the industry's operating environment.

2.2.2 Pharmaceutical R&D and GDPR

The GDPR has significantly altered the regulatory landscape for pharmaceutical firms, which rely extensively on collecting, processing, and sharing sensitive patient data. Industry groups

have argued that the GDPR substantially increases the complexity and cost of clinical research in the EU. The European Federation of Pharmaceutical Industries and Associations (EFPIA), for example, emphasizes that GDPR implementation has complicated clinical research without delivering anticipated regulatory harmonization (European Federation of Pharmaceutical Industries and Associations, 2022). Compared to less stringent privacy laws such as HIPAA, GDPR imposes significantly broader and stricter demands on pharmaceutical firms. For instance, GDPR mandates explicit and informed consent for each specific use of personal data, restricts the collection of data to only what is strictly necessary, and allows patients to request deletion of their data under certain conditions. Such provisions are not typically present in regulatory frameworks like HIPAA, creating unique operational challenges for multinational pharmaceutical firms.

Several GDPR provisions have direct implications for clinical trial practices. First, the GDPR's requirement for explicit consent complicates secondary analyses and longitudinal studies, where multiple uses of data are common. Second, the data minimization principle limits exploratory research, reducing the capacity for firms to pursue innovative but initially uncertain research directions. Third, participants' right to request deletion of personal data poses substantial operational challenges, especially if deletion occurs mid-study. Finally, GDPR restricts data transfers between organizations and across national borders, requiring firms to demonstrate adequate data protection standards and creating significant uncertainty about compliance, especially for multinational trials and collaborations (Eiss, 2020).

Qualitative reports have highlighted the impact of the GDPR: For instance, Lalova-Spinks et al. (2024) document that over 40 EU research sites could not enroll in NIH-sponsored COVID-19 therapeutic trials due to data transfer restrictions. Similarly, large-scale research consortia, such as the International Genomics of Alzheimer's Project, have been compelled to conduct fragmented analyses, impeding real-time data sharing and collaborative insights. These examples illustrate how GDPR's regulatory complexity can delay clinical trial enrollment, increase research costs, and reduce the feasibility of collaborative clinical research,

thus reshaping the landscape in which pharmaceutical firms innovate and collaborate. Consequently, GDPR may introduce specific innovation frictions for pharmaceutical R&D, including higher compliance costs, increased complexity in data sharing, and constraints on research collaborations .

2.3 Conceptual Framework

This section outlines the conceptual framework that connects the institutional dynamics of GDPR to firms' R&D decisions. We propose that GDPR impacts R&D primarily through elevated compliance costs, which then cascade into two strategic adjustments: shifts in research composition and changes in collaboration patterns. While these mechanisms apply broadly across research-intensive industries, we illustrate them using pharmaceutical R&D.

GDPR increases both fixed and variable costs of conducting research. Firms must invest upfront in consent management systems, data protection audits, and contractual alignment with partners (Goldberg et al., 2024). These fixed investments create barriers to project initiation, making it more costly to begin new research. Variable costs arise from ongoing monitoring, reporting, and responding to data subject requests. Regulatory requirements significantly raise the marginal cost of data acquisition and management, particularly for projects requiring large, diverse datasets (Aridor et al., 2020). Demirer et al. (2024) estimate that GDPR increased the variable cost of data inputs by 20 percent for affected firms. These costs compound when projects involve multiple partners or cross-border data transfers, as each additional party introduces coordination complexity and shared liability exposure. As project costs rise, those with marginal expected returns become unviable. For heavily exposed firms—those conducting substantial research in jurisdictions where GDPR applies—these pressures reduce overall R&D investment by making previously viable projects economically infeasible.

Elevated compliance costs systematically alter which projects firms pursue. GDPR's data minimization requirements constrain how much data firms can collect, while restrictions on

data reuse limit the extent to which existing datasets can be repurposed for new questions (Allen and Crawford, 2019). When project diversity increases compliance burden—for instance, by requiring consent protocols for multiple data uses or transfer agreements across jurisdictions—firms respond by narrowing scope. They limit the range of questions addressed, reduce project scale, and concentrate resources in areas with clearer regulatory pathways. Projects requiring extensive data collection, diverse participant populations, or complex data-sharing arrangements incur higher compliance costs and are more likely to be scaled back or abandoned. Projects with simpler data requirements or standardized protocols face lower relative burden and may proceed largely unaffected. These adjustments lower costs and reduce regulatory risk but constrain the breadth of evidence generated (Buckman et al., 2023). Over time, such changes influence not only the quantity but also the composition of firms' research portfolios, potentially steering innovation away from questions that require diverse or complex datasets.

Heightened compliance costs also reshape collaboration incentives, though the net effect remains ambiguous. GDPR makes all participating organizations jointly responsible for compliance, creating new costs for partner screening, contract negotiation, and oversight (Blind, 2012; Kircher and Foerderer, 2022). This joint liability increases coordination complexity: partners must align data governance practices, agree on responsibility allocation, and establish mechanisms for responding to data subject requests or breaches. Such coordination costs may discourage collaboration, particularly with unfamiliar partners whose data practices are uncertain and whose compliance standards require extensive verification (Jia et al., 2021). In this scenario, firms retreat from collaboration or limit partnerships to a narrow set of proven partners. Conversely, firms may intensify collaboration to share compliance expertise and infrastructure. Established partners with proven compliance practices can pool resources for data protection audits, share templates for consent protocols, and jointly navigate regulatory uncertainty. In this scenario, collaboration becomes a strategic response to elevated costs rather than a casualty of them. The net effect depends critically on whether coordination

costs or resource-sharing benefits dominate. Coordination difficulty likely varies systematically: established partnerships face lower costs because partners have aligned practices and trust, while new partnerships require extensive due diligence, creating barriers to network expansion. If regulation standardizes compliance practices across firms, coordination may become easier, encouraging collaboration. If instead regulations introduce partner-specific requirements or create uncertainty about joint liability, coordination costs rise, discouraging collaboration.

Collectively, these mechanisms position compliance costs as the critical friction, with composition and collaboration adjustments representing firms' responses to manage increased costs. The framework thus predicts variation not only in overall R&D levels but also in the types of projects firms pursue and the collaborative structures through which innovation occurs.

3 Data and Summary Statistics

3.1 Data

To empirically investigate the impact of data privacy regulation on innovation, we utilize data on clinical trials from two primary sources: ClinicalTrials.gov (US data) via the Aggregate Analysis of ClinicalTrials.gov Database, and the EU Clinical Trials Register (EU data) from the European Medicines Agency. These databases provide detailed and accurate information on trial sponsors, phases, patient enrollment, locations, duration, and completion status.

Our dataset consists of detailed information for each clinical trial, including the sponsoring firm (e.g., AstraZeneca), trial dates (start and completion), trial phase, patient enrollment numbers, trial duration, and locations of trial sites (e.g., EU or non-EU). We focus on Phase II trials because they represent a substantial investment by firms (averaging $21 million per trial) and are publicly reported (Sertkaya et al., 2024).

We aggregate the clinical trials data at the firm-year level, where each observation represents a unique sponsoring firm active in a given year. We supplement this trial-level data

with comprehensive firm-level information, including firm headquarter location and founding year, sourced from Crunchbase.

The analysis is restricted to clinical trials initiated by industry sponsors between 2010 (or the sponsor's founding year if it was later than 2010) and 2023. The final sample is an unbalanced panel of 73,591 firm-year observations representing 7,508 unique firms.

Measuring High Exposure Firms. Our empirical strategy is motivated by the idea that firms with extensive EU-based operations prior to the GDPR's implementation face higher exposure to the regulation compared to those with minimal EU activities. We categorize firms as being "High Exposure" if they have an above median share of Phase II trial sites in the EU prior to 2018. Figure 1 presents the distribution of pre-2018 EU trial site shares across firms. The average firm's share of trial sites in the EU is 0.18 prior to 2018.

Measuring Research Levels. To measure research level, we count the total number of Phase II clinical trials initiated by each firm annually. Phase II trials represent substantial investment—averaging $21 million per trial (Sertkaya et al., 2024)—and require data collection, consent management, and multi-party coordination subject to GDPR regulation. We focus on Phase II trials because they involve significant patient enrollment, generate extensive clinical data, and are publicly reported, making them particularly sensitive to compliance cost increases. This measure captures whether firms reduce overall research activity in response to elevated costs.

Measuring Research Processes. To examine shifts in research processes, we focus on two dimensions: trial design and research collaboration. For trial design, we measure different dimensions associated with data planning, scope, and scale. As a proxy for data planning, we identify whether the trial includes a pre-specified plan for sharing individual data. To characterize a trial's scope, we examine whether it takes place in multiple countries and the number of trial sites located inside and outside of the EU. Finally, to measure trial scale, we examine the number of patients enrolled in the trial.

For research collaboration, we identify Phase II trials that are conducted in collaboration with other organizations. Following the definition used by ClinicalTrials.gov, we define a collaborative trial as one involving additional organizations that provide substantial support to a trial, including funding, study design, implementation, data analysis, or reporting. In the US trial data, collaborators are explicitly listed separately from sponsors. We flag a trial as a collaborative trial if at least one collaborator is listed. In contrast, the EU trial data does not distinguish between sponsors and collaborators. Instead, when a trial is associated with multiple sponsors, we classify it as a collaborative trial.

We further characterize the nature of collaboration by classifying collaborating trials into two mutually exclusive categories: new and established collaborators based on their relationship with the sponsoring firm. If two firms have no prior collaboration or the collaboration began within 2 years of the focal trial, the collaborator is coded as new. All other collaborations are coded as established.

In subsequent analyses, we examine heterogeneity across firm types. First, we examine differences by firm age and size. We classify firms as "old" if they are older than the median sponsor age in 2018 and "young" otherwise. We also examine variation by firm location and categorize firms based on the location of their headquarters (EU or non-EU). Similarly, we classify firms as "large" if they have an above median number of phase II trials prior to 2018 and "small" otherwise.

Measuring Research Outcomes. For research outcomes, we construct measures that characterize trial progress and the timing in which research results become publicly available. To measure trial progress, we focus on trial duration (the number of years between the trial start date and trial completion date) and trial completion (the share of trials completed). To characterize data disclosure, we examine the share of trials reporting results within one year of trial completion.

3.2 Summary Statistics

Table 1 shows that from 2010 to 2023, the average firm-year observation has 0.96 Phase II trials. Of these trials, 37.5% (0.36) are multicountry trials. Trials typically involve five sites, with approximately three sites in the EU and two outside the EU. Firms conduct 0.12 trials with collaborators each year and the average number of new and established collaborators is similar (0.09 vs. 0.10, respectively). On average, trials typically last 2.7 years and 56% successfully reach completion. Timely trial results reporting is rare: about 1% of trials report their results within 1 year. Approximately 40 percent of firm-years are associated with an old firm and 20 percent are associated with a large firm.

4 Empirical Strategy

To estimate the impact of GDPR on innovation and collaboration, we leverage the differential exposure of firms to GDPR based on their geographical distribution of clinical trials. Our baseline specification compares outcomes of firms with high vs. low exposure to the GDPR, before and after the implementation of the GDPR:

$$Y_{it} = \alpha + \beta(\text{GDPR}_t \times \text{High Exposure}_i) + \delta_i + \tau_t + \epsilon_{it} \tag{1}$$

where Y_{it} represents firm-level outcomes, such as the number of clinical trials initiated, for firm i in year t. The variable GDPR_t is an indicator that equals 1 in the years following GDPR implementation (2018 and onward) and 0 otherwise. The High Exposure_i indicator identifies firms with substantial exposure to GDPR, as determined by the share of pre-2018 EU-based clinical trials.

The interaction term $\text{GDPR}_t \times \text{High Exposure}_i$ captures the causal impact of GDPR. Appendix Table B1 shows that High Exposure firms systematically differ from low exposure firms. For example, they are more research-intensive and their trials are more likely to involve sites in multiple countries and to involve more patients. To account for time-invariant firm

differences, we include firm fixed effects δ_i. Further, we include year fixed effects τ_t to control for common annual shocks across firms. Standard errors are clustered at the firm level.

5 Results

Building on the discussion in the conceptual framework, we provide empirical evidence on how the GDPR affects firms' R&D decisions along three dimensions: research levels, research processes, and research outcomes.

5.1 Impact on Research Levels

5.1.1 Baseline Results

To understand how the GDPR shaped research levels, we begin by exploring whether the regulation changed the level of research in clinical trials among firms more exposed to the regulation. We begin by estimating Equation (1) where the outcome is the number of Phase II trials (with the inverse hyperbolic sine transformation) in a firm and year. The estimates in column 1 of Panel A in Table 2 show that the GDPR is associated with an 18 percent ($\approx e^{-0.2} - 1$) decline in the relative level of clinical trials in High Exposure firms. Column 2 shows a similar, though smaller, decline when using the log number of trials. Looking to the extensive margin, we observe a similar effect: the GDPR leads to a 0.11 decline in the likelihood of a trial, an 84 percent decline relative to the sample mean.

-Insert Table 2 about here-

To assess the timing of this baseline effect, we conduct an event study using indicators for each year relative to GDPR implementation. Figure 2 shows that the difference in the inverse hyperbolic sine-transformed number of clinical trials among firms with low and high exposure prior to the GDPR is, on average, close to zero. After the implementation of the GDPR, we observe a sharp and persistent decline in the number of trials conducted by High Exposure firms. The relatively quick response is consistent with the idea that the GDPR

may have prevented further investment in previously developed drugs, which is common as drugs typically have multiple therapeutic uses (Greenblatt et al., 2023). The decline in the level of clinical trials remains large and statistically significant through 2023.

-Insert Figure 2 about here-

One potential concern is that the observed decline in clinical trial levels following the GDPR may instead reflect a shift in the types of diseases that firms pursue (e.g., firms may prioritize diseases where clinical trials are more costly) rather than a reduction in research investment. To address this, Panel B presents results from a firm-disease-year analysis, where diseases are measured at the 3-digit MeSH level (e.g., musculoskeletal diseases, respiratory tract diseases). This approach allows us to directly control for shifts in disease composition by including disease fixed effects. As shown in Panel B in Table 2, the GDPR has a consistent, negative, and statistically significant effect on clinical trial levels even with the addition of disease fixed effects. These findings suggest that shifts in disease composition are unlikely to be the primary driver of the observed change in clinical trial levels. In the subsequent sections, given the skewed distribution of trial counts and the large number of zeroes, we use the inverse hyperbolic sine transformation as our preferred measure of research levels.

5.1.2 Robustness Checks and Heterogeneity Analyses

We probe the robustness of our baseline findings and examine heterogeneity in several ways.

Confounding Events. A potential concern may be that our results are driven by concurrent events, such as the onset of Covid-19. If firms with greater exposure to the GDPR (e.g., multinational firms) are more likely to be affected by the pandemic, this could bias our estimates downwards. To address this, we rerun our main analysis using only data on years before 2020. Column 1 of Table C1 shows that the effect of GDPR remains negative and statistically significant, suggesting that our findings are not primarily driven by pandemic-related disruptions.

Alternative Firm Samples. A related concern is that firms with high exposure to the GDPR may be systematically different from other firms as they are more likely to be multinational. To address these concerns, we restrict the sample to multinational firms, which we define as those with at least some clinical trial activity in both EU and non-EU countries before the implementation of the GDPR. Note that this also limits the sample to firms with some pre-GDPR prior research activity. Column 2 of Table C1 shows that the effect of GDPR remains negative and statistically significant when restricting the sample to multinational firms, suggesting that our results are not driven by differences between multinational and single-country firms.

Differences Across Phases. Table C2 examines the effect of the GDPR across different trial phases. We find a small, but statistically significant decline among Phase I trials and no significant decline among Phase III trials. The larger effect among Phase II trials is consistent with this phase being the first phase in which firms make significant financial investments. The lack of measurable effect among Phase III trials may be partially explained by the longer lag between the GDPR and the onset of Phase III trials.

5.2 Impact on Research Processes

5.2.1 Trial Design

To further shed light on how the GDPR shaped firms' R&D decisions, we examine Phase II trial design. Table 3 shows High Exposure firms are significantly more likely to conduct trials with pre-specified plans for sharing individual participant data. At the same time, trials narrow in geographic scope: they are less likely to take place in multiple countries and the number of trial sites declines overall. Consistent with the GDPR's geographic focus, the decline in clinical trial sites is largely driven by the reduction in EU-based sites, with little effect on non-EU-based sites. In contrast, we observe little change in the average number of participants per trial. This is consistent with the view that declines associated

with increased costs related to regulatory compliance may be partially offset by the benefits associated with external stakeholders' (i.e., trial participants) perceptions of increased data privacy and security. In turn, this may increase their willingness to participate in the R&D process.

-Insert Table 3 about here-

5.2.2 Research Collaboration

The previous results suggest that regulatory requirements associated with the GDPR increased the cost of research. Given the importance of data sharing in the research process and the GDPR's focus on data sharing, this should also affect certain forms of collaboration between firms. Column 1 in Table 4 shows a significant decline in the level of collaboration between firms: in particular, column 1 shows a 1.8 ($\approx e^{-0.0189} - 1$) percent decline in the level of Phase II collaborative trials in a given year. This decline is modest compared to the overall 18 percent reduction in total trials, suggesting research collaborations showed some resilience to the broader effects of the GDPR.

-Insert Table 4 about here-

The smaller decline in collaborative trials compared to overall trial activity is notable and raises important questions about why firms adjust their R&D collaborations. Rather than stepping away from collaboration entirely, firms may be responding to new data privacy requirements by being more selective about who they work with. One explanation is that collaboration offers a way to share the rising costs of compliance. To explore these possibilities, we next examine how the effects of GDPR vary by collaborator type and firm characteristics.

Heterogeneity by New and Existing Collaborations. To better understand whether firms respond to privacy regulations by adjusting collaboration, we examine differences in collaboration patterns based on the prior relationship between firms and their partners.

Specifically, we compare collaborations with relatively new partners (i.e., those that have no prior collaboration or the collaboration began within 2 years of the focal trial) to those with relatively established partners (i.e., those with whom the firm has had long-standing relationships).

Columns 2 and 3 in Table 4 suggests that the overall decline in collaborative trials is driven by a significant drop in collaborations with relatively new partners (column 2). At the same time, collaborations with established partners modestly increase (column 3). These results are echoed when looking at the unique number of firm collaborators (columns 5 and 6).

We conduct several robustness checks. Table D1 shows that our results hold when we use an alternative definition of new vs. existing collaborators using a 5-year threshold instead of the baseline 2-year threshold.

Heterogeneity by Firm Age and Size. Next, we examine whether the effects vary by firm age and firm size. We use firm age to capture the extent to which a firm has had time to develop relationships with collaborators. Older firms are more likely to have worked with the same partners repeatedly and may have established shared practices that ease compliance under new regulations. For example, an older firm may already have data-sharing agreements in place, making it easier to continue joint research without added legal or administrative burdens. In contrast, firm size reflects a firm's ability to absorb the costs of compliance, such as hiring legal staff, updating IT systems, or managing documentation requirements. Larger firms often have in-house legal teams and data protection resources. In contrast, smaller firms may find it more difficult to take on new collaborations if they cannot afford the infrastructure needed to meet GDPR standards.

Table 5 shows that the decline in collaborative trials is driven by younger firms (column 1), while older firms (column 2) show no significant change. The difference in effects is statistically significant. This suggests that older firms are better able to maintain collaborations despite the added requirements introduced by GDPR.

-Insert Table 5 about here-

Columns 3 and 4 show results by firm size. While the estimated effect is slightly more negative for small firms, the differences between small and large firms are not statistically significant. This is consistent with the view that the ability to share compliance costs across collaborators is less important than having established relationships in shaping firms' collaboration decisions.

5.2.3 Additional Heterogeneity Analyses

Appendix D presents additional extensions of the collaboration results.

Appendix Table D3 examines whether the effects of GDPR vary depending on the geographic distance between collaborators. Longer geographic distance could increase the costs of compliance with the GDPR (e.g., data must be transferred across borders) but it might also make it harder to establish trust between partners (Catalini et al., 2020). However, the results show that GDPR does not significantly affect collaborations with either EU-based or non-EU-based partners, regardless of whether the focal firm is located within or outside the EU. These results suggest that distance-related frictions (which may increase coordination costs or raise difficulties in forming trust) are not the primary drivers of the observed decline in collaboration. Instead, firms appear to respond more broadly to the regulatory environment, regardless of where their partners are located.

While earlier analyses focused on one measure of collaboration (collaborative clinical trials), collaboration can also take other forms, such as R&D-related licensing deals between firms. We examine two types of deals: R&D collaboration deals, where firms jointly work on developing a drug, and asset transfer deals, where one firm acquires the rights to a compound or technology developed by another. Consistent with the findings outlined in Table 4, Table D4 shows that the GDPR led to a significant decline across both types of deal-based collaborative activity.

5.3 Impact on Research Outcomes

Having examined the impact on research levels and processes, we next examine how the GDPR shaped research outcomes. Table 6 shows that following the GDPR, trials take 12 percent longer to complete and the share of trials that successfully complete declines. These results, combined with the previous findings on trial levels, are consistent with the view that the GDPR shapes both the number of projects that firms initiate and the number that successfully come to market. Finally, Table 6 shows delays in the availability of information about trial outcomes: following the GDPR, the share of trials reporting results within one year of completion declines.

-Insert Table 6 about here-

6 Conclusion

Understanding how data privacy regulation shapes innovation is of substantial interest to strategy and innovation scholars, managers, and policymakers. Yet prior empirical work has been largely focused on examining how regulation shapes immediate firm performance outcomes, raising open questions about how regulation alters firms' R&D decisions. Using variation in firms' pre-GDPR geographic distribution of clinical trials, we show that the GDPR led firms to reduce their clinical trial levels by 18 percent and led to meaningful shifts in firms' research processes and outcomes.

Our analysis has several limitations that suggest opportunities for future work. First, our data capture the investment in clinical trials but not the downstream outcomes, such as drug quality or patient health. Future research could examine whether GDPR exposure ultimately affects the quality of therapies developed or consumer access to treatment. Second, while we study the pharmaceutical industry, future research could explore whether similar mechanisms operate in other data-intensive, research-driven sectors. Such work would help assess the external validity of our framework and clarify whether the effects we document

are general features of data privacy regulation. Third, although our analysis emphasizes firm-level responses, future work could examine heterogeneity within organizations (e.g., across therapeutic portfolios, project teams, or geographic subsidiaries) to shed light on how managers reallocate resources in response to regulatory constraints.

Taken together, our study highlights how data privacy regulation and innovation are closely linked. By documenting how GDPR reshaped firms' research levels, processes, and outcomes, we provide evidence that privacy regulation not only shapes short-term performance but also leads to meaningful shifts in firms' R&D decisions. These findings advance understanding of how regulation shapes firms' R&D decisions in data-intensive industries and its broader effects for the rate and direction of subsequent innovation.

References

Acquisti, A., Taylor, C., and Wagman, L. (2016). The economics of privacy. *Journal of Economic Literature*, 54(2):442–492.

Allen, F. S. and Crawford, G. (2019). Clinical trials under the GDPR: What should sponsors consider? *Privacy Laws & Business*.

Aridor, G., Che, Y.-K., and Salz, T. (2020). The economic consequences of data privacy regulation: Empirical evidence from GDPR. Technical report, National Bureau of Economic Research, Cambridge, MA, USA.

Aridor, G., Che, Y.-K., and Salz, T. (2023). The effect of privacy regulation on the data industry: empirical evidence from GDPR. *RAND Journal of Economics*, 54(4).

Blind, K. (2012). The influence of regulations on innovation: A quantitative assessment for OECD countries. *Research Policy*, 41(2):391–400.

Buckman, J. R., Adjerid, I., and Tucker, C. (2023). Privacy Regulation and Barriers to Public Health. *Management Science*, 69(1):342–350.

Catalini, C., Fons-Rosen, C., and Gaulé, P. (2020). How do travel costs shape collaboration? *Management Science*, 66(8):3340–3360.

Chandra, A., Kao, J., Miller, K. L., and Stern, A. D. (2024). Regulatory Incentives for Innovation: The FDA's Breakthrough Therapy Designation. *Review of Economics and Statistics*, pages 1–46.

Cockburn, I. M. and Henderson, R. M. (2001). Scale and scope in drug development: unpacking the advantages of size in pharmaceutical research. *Journal of health economics*, 20(6):1033–1057.

Demirer, M., Hernández, D. J. J., Li, D., and Peng, S. (2024). Data, privacy laws and firm production: Evidence from the GDPR. Technical report, National Bureau of Economic Research.

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.

Eiss, R. (2020). Confusion over Europe's data-protection law is stalling scientific progress. *Nature*, 584(7822):498–499.

European Commission (2025a). Adequacy decisions: how EU personal data flows to non-EU countries. Retrieved from https://commission.europa.eu/law/law-topic/data-protection/international-dimension-data-protection/adequacy-decisions_en.

European Commission (2025b). Data protection in the EU. Retrieved from https://commission.europa.eu/law/law-topic/data-protection/data-protection-eu_en.

European Federation of Pharmaceutical Industries and Associations (2022). GDPR implementation has complicated clinical research without delivering regulatory harmonization. EFPIA Report.

European Medicines Agency (2014). Clinical Trials Regulation (Regulation EU No 536/2014). Retrieved from https://www.ema.europa.eu.

European Union (1995). Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. https://eur-lex.europa.eu/eli/dir/1995/46/oj/eng. Official Journal L 281, 23/11/1995, p. 31–50.

Godinho De Matos, M. and Adjerid, I. (2022). Consumer Consent and Firm Targeting After GDPR: The Case of a Large Telecom Provider. *Management Science*, 68(5):3330–3378.

Goldberg, S. G., Johnson, G. A., and Shriver, S. K. (2024). Regulating privacy online: An economic evaluation of the GDPR. *American Economic Journal: Economic Policy*, 16(1):325–358.

Goldfarb, A. and Que, V. F. (2023). The economics of digital privacy. *Annual Review of Economics*, 15(1):267–286.

Goldfarb, A. and Tucker, C. (2012). Privacy and Innovation. *Innovation Policy and the Economy*, 12:65–90.

Goldfarb, A. and Tucker, C. E. (2011). Privacy regulation and online advertising. *Management science*, 57(1):57–71.

Greenblatt, W., Gupta, C., and Kao, J. (2023). Drug Repurposing During The COVID-19 Pandemic: Lessons For Expediting Drug Development And Access. *Health Affairs*, 42(3):424–432.

Gulati, R. (1995). Social structure and alliance formation patterns: A longitudinal analysis. *Administrative Science Quarterly*, pages 619–652.

Gulati, R. and Gargiulo, M. (1999). Where do interorganizational networks come from? *American Journal of Sociology*, 104(5):1439–1493.

Gulati, R. and Nickerson, J. A. (2008). Interorganizational Trust, Governance Choice, and Exchange Performance. *Organization Science*, 19(5):688–708.

Jia, J., Jin, G. Z., and Wagman, L. (2021). The Short-Run Effects of the General Data Protection Regulation on Technology Venture Investment. *Marketing Science*, 40(4):661–684.

Johnson, G. A., Shriver, S. K., and Goldberg, S. G. (2023). Privacy and Market Concentration: Intended and Unintended Consequences of the GDPR. *Management Science*, 69(10):5695–5721.

Khanna, R., Guler, I., and Nerkar, A. (2018). Entangled decisions: Knowledge interdependencies and terminations of patented inventions in the pharmaceutical industry. *Strategic Management Journal*, 39(9):2439–2465.

Kircher, T. and Foerderer, J. (2022). Ban targeted advertising in apps? An empirical investigation of the consequences for app development. *Management Science*. Forthcoming.

Lakdawalla, D. N. (2018). Economics of the pharmaceutical industry. *Journal of Economic Literature*, 56(2):397–449.

Lalova-Spinks, T., Valcke, P., Ioannidis, J. P., and Huys, I. (2024). EU-US data transfers: an enduring challenge for health research collaborations. *npj Digital Medicine*, 7(1):215.

Mahoney, J. T., McGahan, A. M., and Pitelis, C. N. (2009). Perspective—The Interdependence of Private and Public Interests. *Organization Science*, 20(6):1034–1052.

Miller, A. R. and Tucker, C. (2018). Privacy Protection, Personalized Medicine, and Genetic Testing. *Management Science*, 64(10):4648–4668.

Moss, A. J., Francis, C. W., and Ryan, D. (2011). Collaborative clinical trials. *The New England journal of medicine*, 364(9):789–791.

Peukert, C., Bechtold, S., Batikas, M., and Kretschmer, T. (2022). Regulatory Spillovers and Data Governance: Evidence from the GDPR. *Marketing Science*, 41(4):746–768.

Sampat, B. and Williams, H. L. (2019). How do patents affect follow-on innovation? Evidence from the human genome. *American Economic Review*, 109(1):203–236.

Sertkaya, A., Beleche, T., Jessup, A., and Sommers, B. D. (2024). Costs of drug development and research and development intensity in the US, 2000-2018. *JAMA network open*, 7(6):e2415445–e2415445.

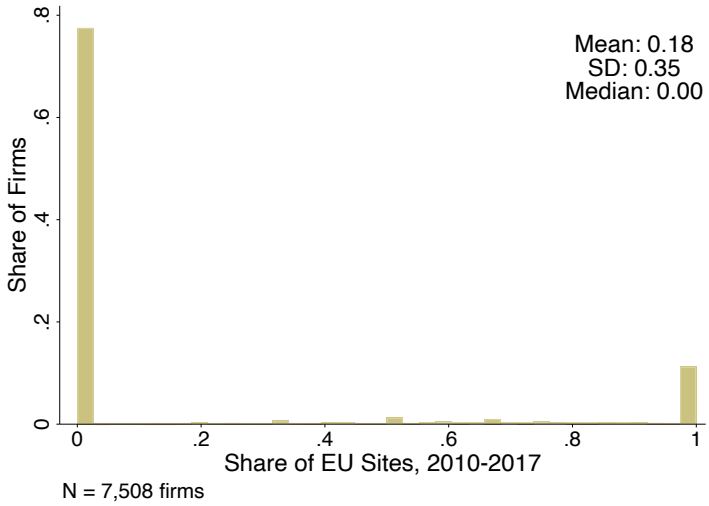
Vakili, K. and McGahan, A. M. (2016). Health Care's Grand Challenge: Stimulating Basic Science on Diseases that Primarily Afflict the Poor. *Academy of Management Journal*, 59(6):1917–1939.

Van Norman, G. A. (2018). Expanding Patient Access to Investigational New Drugs: Overview of Intermediate and Widespread Treatment Investigational New Drugs, and Emergency Authorization in Public Health Emergencies. *JACC: Basic to Translational Science*, 3(3):403–414.

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

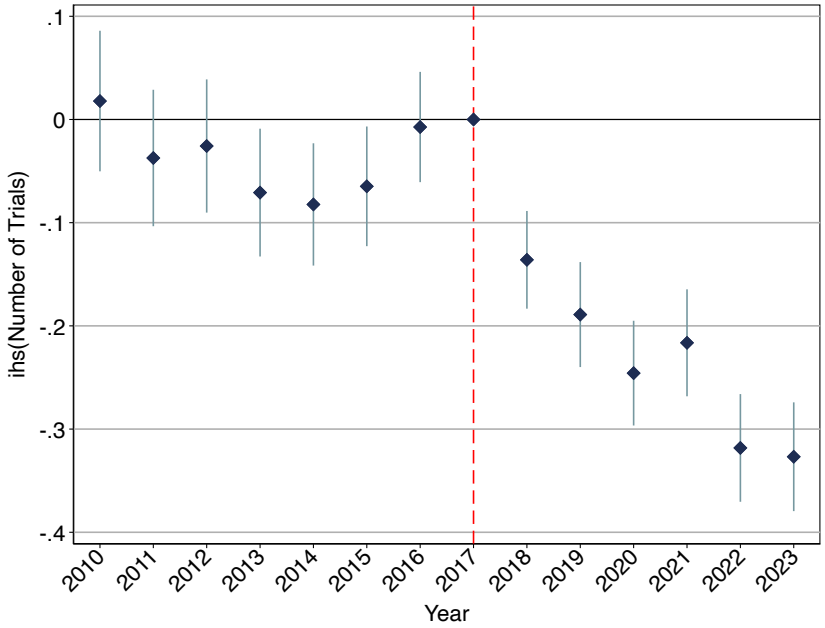
7 Main Figures and Tables

Figure 1: Firm-level Exposure to GDPR



Notes. This figure shows the distribution of firms' pre-2018 EU trial site share.

Figure 2: Impact of GDPR on Phase II Trials



Note: This figure presents event study coefficients corresponding to a dynamic version of Equation (1). Observations are at the firm-year level. We plot the coefficients on the interaction between year and an indicator for whether the firm is a high exposure firm. The outcome is the inverse hyperbolic sine transformed number of trials in a year. The dotted vertical line represents the year before the GDPR's implementation.

Table 1: Firm-Year Level Summary Statistics

	Count (1)	Mean (2)	SD (3)	Min (4)	Max (5)
Research Levels					
Trials	73,591	0.96	10	0	559
Research Process					
Share of Trials with Pre-specified Plans for Sharing Participant Data	9,973	0.05	0	0	1
# Multicountry Trials	73,591	0.36	2	0	97
# EU Sites Per Trial	9,973	2.93	5	0	27
# Non-EU Sites Per Trial	9,973	2.27	4	0	39
# Patients Per Trial	9,973	117.22	166	0	5,000
# Trials with Collab.	73,591	0.12	1	0	149
# Trials with New Collab.	73,591	0.09	1	0	95
# Trials with Established Collab.	73,591	0.04	1	0	67
# Collab.	73,591	0.18	1	0	87
# New Collab.	73,591	0.09	1	0	35
# Established Collab.	73,591	0.10	1	0	77
Research Outcomes					
Duration Per Trial (Years)	7,647	2.70	2	0	16
Share of Trials Completed	9,973	0.56	0	0	1
Share of Trials Reporting Results Within 1 Year	9,973	0.01	0	0	1
Firm Characteristics					
0/1: Old Firm	69,174	0.39	0	0	1
0/1: Large Firm	73,591	0.19	0	0	1

Notes. This table reports firm-year level summary statistics of key research characteristics. There are 73,591 firm-year observations, representing 7,508 unique firms. Variables may have fewer observations due to missing data. The variables "# EU Sites Per Trial," "# Non-EU Sites Per Trial," "Duration Per Trial (Years) ," "Patients Per Trial," are only reported for the firm-year observations with at least one clinical trial.

Table 2: Impact of GDPR on Research Levels

	ihs(# Trials) (1)	log(1 + # Trials) (2)	1(Any Trial) (3)
A. Firm-year level			
GDPR × High Exposure	-0.205***	-0.162***	-0.112***
	(0.0147)	(0.0119)	(0.00700)
Mean of Dep. Var.	0.219	0.174	0.134
Observations	73,253	73,253	73,253
R-squared	0.523	0.544	0.334
Firm FEs	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes
B. Firm-disease-year level			
GDPR × High Exposure	-0.00823***	-0.00640***	-0.00547***
	(0.00160)	(0.00131)	(0.000679)
Mean of Dep. Var.	0.00694	0.00546	0.00473
Observations	807,375	807,375	807,375
Firm FEs	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes
Disease FEs	Yes	Yes	Yes

Notes. This table reports DID estimates of the effect of GDPR on phase II trials. Panel A reports estimates from regressions where the unit of analysis is the firm-year. Panel B reports estimates from a regression where the unit of analysis is the firm-disease-year. Diseases are measured at the 3-digit MeSH code level. Estimates are based on OLS regressions. Standard errors are clustered at the firm level and reported in parentheses, with *** p<0.01, ** p<0.05, * p<0.1.

Table 3: Impact of GDPR on Research Processes: Phase II Trial Design

	Share with Pre-specified Plans for Sharing Individual Participant Data	ihs(# Multicountry Trials)	ihs(# EU Sites Per Trial)	ihs(# Non-EU Sites Per Trial)	ihs(# Patients Per Trial)
	(1)	(2)	(3)	(4)	(5)
GDPR × High Exposure	0.0821***	-0.0616***	-0.109**	0.0426	0.0282
	(0.0191)	(0.0145)	(0.0545)	(0.0503)	(0.0853)
Mean of Dep. Var.	0.0605	0.196	1.235	1.171	4.749
Observations	8,106	73,253	8,106	8,106	8,106
Firm FEs	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes

Notes. This table reports DID estimates of the effect of GDPR on the design of phase II trials. The level of observation is the firm-year. Estimates are from OLS regressions. Standard errors are clustered at the firm level and reported in parentheses, with *** p<0.01, ** p<0.05, * p<0.1.

Table 4: Impact of GDPR on Research Processes: Phase II Research Collaboration

	Dependent variable: ihs(# Trials)			Dependent variable: ihs(# Collaborators)		
	All Collab.	With New Collab.	With Established Collab.	All Collab.	With New Collab.	With Established Collab.
	(1)	(2)	(3)	(4)	(5)	(6)
GDPR × High Exposure	-0.0189**	-0.0249***	0.00922*	-0.00867	-0.0245***	0.0193*
	(0.00722)	(0.00686)	(0.00428)	(0.0101)	(0.00683)	(0.00773)
Mean of Dep. Var.	0.047	0.040	0.011	0.103	0.040	0.011
Diff. Wald Test p-value		0.00			0.00	
Observations	73,253	73,591	73,591	73,253	73,591	73,591
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes

Notes. This table reports DID estimates of the effect of GDPR on the number of phase II trials with collaborators. The level of observation is the firm-year. Column 1 reports OLS estimates. Columns 2 and 3 report estimates from seemingly unrelated regressions. The "Wald Test p-value" is the p-value comparing the $GDPR \times HighExposure$ coefficient in columns 2 vs. 3. Outcomes are transformed with the inverse hyperbolic sine transformation. Standard errors are clustered at the firm level and reported in parentheses, with *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Impact of GDPR on Research Processes: Phase II Research Collaboration by Firm Age and Size

	Dependent variable: ihs(# Collab. Trials)			
	Firm Age		Firm Size	
	Young Firm	Old Firm	Small Firm	Large Firm
	(1)	(2)	(3)	(4)
GDPR × High Exposure	-0.0377***	-0.00129	-0.00601	0.0229
	(0.00805)	(0.0129)	(0.00453)	(0.0168)
Mean of Dep. Var.	0.041	0.061	0.025	0.139
Diff. Wald Test p-value	0.02		0.10	
Observations	42,236	26,938	59,326	14,265
Firm FEs	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes

Notes. This table reports DID estimates of the effect of GDPR, by firm age and size. Estimates are from seemingly unrelated regressions. The "Wald Test p-value" is the p-value comparing the $GDPR \times HighExposure$ coefficient in columns 1 vs. 2 and columns 3 vs. 4. Outcomes are transformed with the inverse hyperbolic sine transformation. Standard errors are clustered at the firm level and reported in parentheses, with *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Impact of GDPR on Research Outcomes

	ihs(Duration Per Trial (Years))	Share of Trials Completed	Share of Trials Reporting Results Within 1 Year
	(1)	(2)	(3)
GDPR × High Exposure	0.111***	-0.0422*	-0.0133*
	(0.0407)	(0.0246)	(0.00693)
Mean of Dep. Var.	1.584	0.579	0.0127
Observations	6,308	8,106	8,106
Firm FEs	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes

Notes. This table reports DID estimates of the effect of GDPR on the outcomes of phase II trials. The level of observation is the firm-year. Estimates are from OLS regressions. Standard errors are clustered at the firm level and reported in parentheses, with *** p<0.01, ** p<0.05, * p<0.1.

Data Privacy Regulation and Innovation: Evidence from GDPR

Online Appendix

Appendix A Additional Background on Drug Development

Drugs not required to use the centralized procedure may pursue the decentralized procedure, where drug developers can seek that a drug be authorized in more than one member state at the same time. In this case, a Reference Member State conducts the primary review and notifies the other member states to which the application has been sent.

Additional pathways include the mutual recognition procedure, where a drug approved in one EU country can be approved for another, and national authorization procedures, where approval is sought within a single member state for use in that country only.

Appendix B Additional Figures and Tables

Table B1: Pre-GDPR Differences Across Low and High Exposure Firms

	Low Exposure		High Exposure		
	Mean	SD	Mean	SD	p-value
	(1)	(2)	(3)	(4)	(5)
Research Levels					
# Trials	0.15	0.80	3.31	20.08	0.00***
Research Processes					
Share of Trials with Pre-specified Plans for Sharing Participant Data	0.03	0.16	0.05	0.18	0.00***
# Multicountry Trials	0.05	0.30	1.02	3.71	0.00***
# EU Sites Per Trial	0.18	0.71	5.13	5.69	0.00***
# Non-EU Sites Per Trial	1.39	1.33	2.83	5.32	0.00***
# Patients Per Trial	102.12	161.08	127.46	145.08	0.00***
# Trials with Collab.	0.03	0.39	0.32	2.45	0.00***
# Trials with New Collab.	0.03	0.26	0.24	1.86	0.00***
# Trials with Established Collab.	0.01	0.26	0.11	1.50	0.00***
# Collab.	0.13	1.32	0.29	1.26	0.00***
# New Collab.	0.07	0.57	0.15	0.78	0.00***
# Established Collab.	0.06	0.96	0.14	0.76	0.00***
Research Outcomes					
Duration Per Trial (Years)	2.31	1.80	3.10	2.12	0.00***
Share of Trials Completed	0.77	0.40	0.75	0.38	0.08*
Share of Trials Reporting Results Within 1 Year	0.01	0.11	0.03	0.13	0.00***
Firm Characteristics					
0/1: Old Firm	0.42	0.49	0.51	0.50	0.00***
0/1: Large Firm	0.10	0.30	0.52	0.50	0.00***

Notes. This table compares summary statistics of key research characteristics before the GDPR across firms with low and high exposure. Column 5 presents p-values from t-tests comparing the difference of means. *** p<0.01, ** p<0.05, * p<0.1

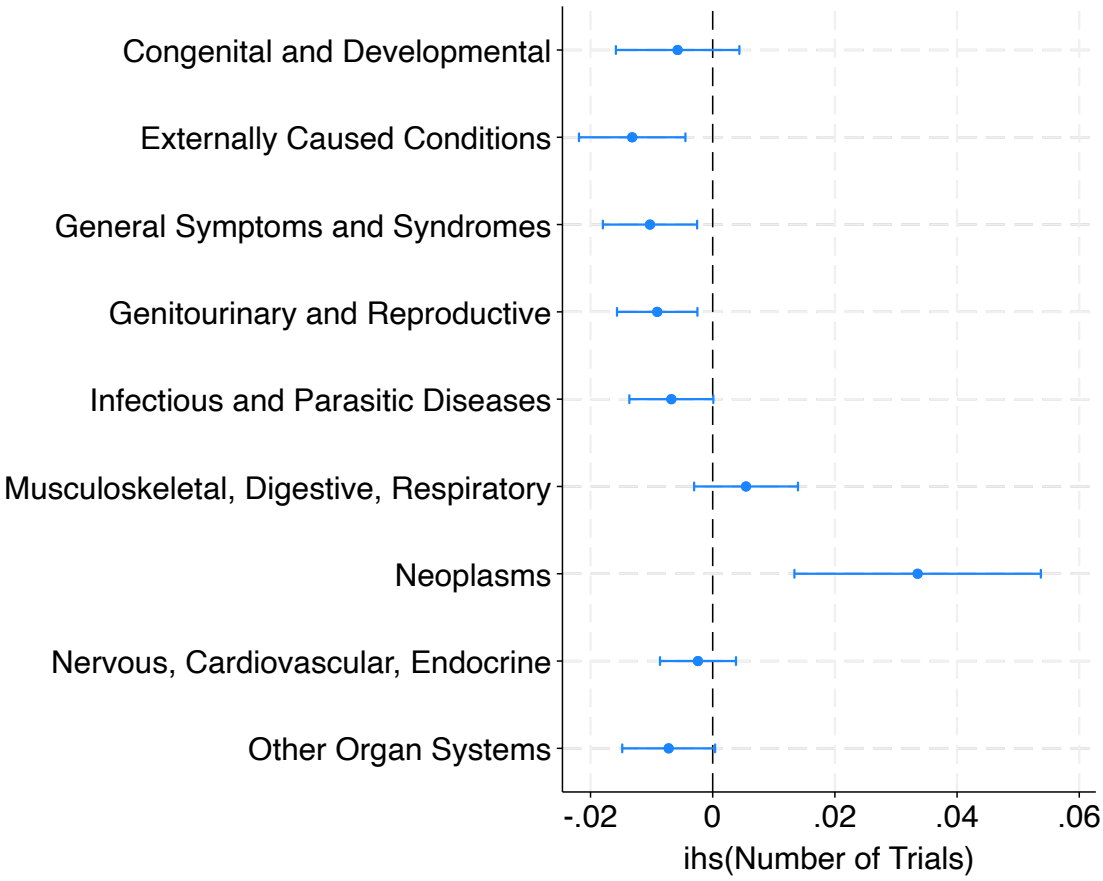
Table B2: Differences Across Firms Without and With Collaborators

	No Collaboration		Any Collaboration		
	Mean	SD	Mean	SD	p-value
	(1)	(2)	(3)	(4)	(5)
Research Levels					
# Trials	0.27	1.54	3.89	22.19	0.00***
Research Processes					
Share of Trials with Pre-specified Plans for Sharing Participant Data	0.03	0.17	0.08	0.25	0.00***
# Multicountry Trials	0.21	0.70	0.98	4.15	0.00***
# EU Sites Per Trial	1.94	3.05	4.11	6.10	0.00***
# Non-EU Sites Per Trial	1.47	2.23	3.21	5.12	0.00***
# Patients Per Trial	105.26	140.83	131.42	191.15	0.00***
Trial Outcomes					
Duration Per Trial (Years)	2.34	1.68	3.10	2.35	0.00***
Share of Trials Completed	0.55	0.47	0.57	0.45	0.05**
Share of Trials Reporting Results Within 1 Year	0.01	0.09	0.01	0.10	0.00***
Firm Characteristics					
0/1: Old Firm	0.39	0.49	0.38	0.48	0.00***
0/1: Large Firm	0.15	0.35	0.39	0.49	0.00***

Notes. This table compares summary statistics of key research characteristics across firms with and without any collaborators. Column 5 presents p-values from t-tests comparing the difference of means. *** p<0.01, ** p<0.05, * p<0.1

Appendix C Robustness Checks and Heterogeneity Analyses

Figure C1: Differences in the Effect of GDPR Across Diseases Groups



Notes. This figure shows the impact of GDPR across different types of disease groups. We plot the coefficients from a firm-disease-year level version of Equation 1 where $\text{GDPR}_t \times \text{High Exposure}_i$ is interacted with an indicator for different disease groups: (1) Blood, Immune, and Metabolic; (2) Congenital and Developmental; (3) Externally Caused Conditions; (4) General Symptoms and Syndromes; (5) Genitourinary and Reproductive; (6) Infectious and Parasitic; (7) Musculoskeletal, Digestive, and Respiratory; (8) Neoplasms; (9) Nervous, Cardiovascular, and Endocrine; and (10) Other Organ Systems. The omitted group in the figure is "Blood, Immune, and Metabolic." The outcome is the inverse hyperbolic sine transformed number of trials in a year.

Table C1: Impact of GDPR on Trial Levels, Alternative Samples

	Dependent variable: ihs(# Trials)	
	Years before 2020	Multinational firms
	(1)	(2)
GDPR × High Exposure	-0.140***	-0.212***
	(0.0173)	(0.0289)
Mean of Dep. Var.	0.230	0.494
Observations	45,289	18,675
Firm FEs	Yes	Yes
Year FEs	Yes	Yes

Notes. This table reports DID estimates of the effect of GDPR using alternative samples. Column 1 restricts firm-years to those before 2020. Column 2 restricts firm-years to multinational firms. Multinational firms are defined as those whose pre-2018 share of EU trials is between 0 and 1. Estimates are from OLS regressions. Standard errors are clustered at the firm level and reported in parentheses, with *** p<0.01, ** p<0.05, * p<0.1.

Table C2: Impact of GDPR on Trial Levels, Across Phases

	ihs(# Phase I Trials)	ihs(# Phase II Trials)	ihs(# Phase III Trials)
	(1)	(2)	(3)
GDPR × High Exposure	-0.0298***	-0.205***	-0.0163
	(0.00861)	(0.0147)	(0.0163)
Mean of Dep. Var.	0.131	0.219	0.190
Observations	73,253	73,253	73,253
Firm FEs	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes

Notes. This table reports DID estimates of the effect of GDPR across different trial phases. Estimates are from OLS regressions. Standard errors are clustered at the firm level and reported in parentheses, with *** p<0.01, ** p<0.05, * p<0.1.

Table C3: Impact of GDPR on Phase II Trial Levels Across Diseases with Varying Levels of Prior Research Activity

	Dependent variable: ihs(# Trials)		
	Diseases with Low Research Activity (1)		Diseases with High Research Activity (2)
GDPR × High Exposure	-0.00195		-0.0150***
	(0.00106)		(0.00308)
Mean of Dep. Var.	0.007		0.007
Diff. Wald Test p-value		0.00	
Observations	807,375		807,375
Firm FEs	Yes		Yes
Year FEs	Yes		Yes
Disease FEs	Yes		Yes

Notes. This table reports DID estimates of the effect of GDPR across diseases with low vs. high level of prior research activity. A disease is considered as having a low (high) level of prior research activity if the total number of clinical trials before 2010 is below (above) the median. The unit of observation is the firm-disease-year level. Diseases are measured at the 3-digit MeSH code level. Estimates are from seemingly unrelated regressions. The "Wald Test p-value" is the p-value comparing the $GDPR \times HighExposure$ coefficient in columns 1 vs. 2. Standard errors are clustered at the firm level and reported in parentheses, with *** p<0.01, ** p<0.05, * p<0.1.

Appendix D Additional Collaboration Analyses

Table D1: Impact of GDPR on Phase II Research Collaboration (Alternative Definition of New and Existing Collaborations)

	Dependent variable: ihs(# Trials)	
	With New Collab.	With Established Collab.
	(1)	(2)
GDPR × High Exposure	-0.0236**	0.00808*
	(0.00736)	(0.00370)
Mean of Dep. Var.	0.044	0.006
Diff. Wald Test p-value	0.00	
Observations	73,591	73,591
Firm FEs	Yes	Yes
Year FEs	Yes	Yes

Notes. This table reports DID estimates of the effect of GDPR on the number of phase II trials with collaborators, across new and existing collaborators, where if two firms have no prior collaboration or the collaboration began within five years of the focal trial, the collaborator is coded as new. All other collaborations are coded as established. The level of observation is the firm-year. Estimates are from seemingly unrelated regressions. The "Wald Test p-value" is the p-value comparing the $GDPR \times HighExposure$ coefficient in columns 1 vs. 2. Outcomes are transformed with the inverse hyperbolic sine transformation. Standard errors are clustered at the firm level and reported in parentheses, with *** p<0.01, ** p<0.05, * p<0.1.

Table D2: Impact of GDPR on Phase II Trial Levels, by Industry vs. Non-Industry Collaborator

	Dependent variable: ihs(# Trials)			
	Novel Collab.		Established Collab.	
	Industry Collab.	Non-Industry Collab.	Industry Collab.	Non-Industry Collab.
	(1)	(2)	(3)	(4)
GDPR × High Exposure	-0.0136*	-0.00963*	0.00769	0.00204
	(0.00610)	(0.00410)	(0.00425)	(0.00113)
Mean of Dep. Var.	0.024	0.019	0.024	0.019
Diff. Wald Test p-value	0.54		0.20	
Observations	73,591	73,591	73,591	73,591
Firm FEs	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes

Notes. This table reports DID estimates of the effect of GDPR, by industry vs. non-industry collaborator. Estimates are from seemingly unrelated regressions. The "Wald Test p-value" is the p-value comparing the $GDPR \times HighExposure$ coefficient in columns 1 vs. 2; 3 vs. 4. Outcomes are transformed with the inverse hyperbolic sine transformation. Standard errors are clustered at the firm level and reported in parentheses, with *** p<0.01, ** p<0.05, * p<0.1.

Table D3: Impact of GDPR on Phase II Trial Levels, by Geographic Distance of Collaborator

	Dependent variable: ihs(# Trials)			
	Firm with EU HQ		Firm with Non-EU HQ	
	Collab. Firm with EU HQ (1)	Collab. Firm with Non-EU HQ (2)	Collab. Firm with EU HQ (3)	Collab. Firm with Non-EU HQ (4)
GDPR × High Exposure	-0.00321	0.00207	-0.00301	-0.00755
	(0.00494)	(0.00712)	(0.00480)	(0.00740)
Mean of Dep. Var.	0.00711	0.0155	0.00650	0.0152
Observations	17,309	17,309	54,430	54,430
Firm FEs	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes

Notes. This table reports DID estimates of the effect of GDPR, by geographic distance of collaborator. Estimates are from OLS regressions. Outcomes are transformed with the inverse hyperbolic sine transformation. Standard errors are clustered at the firm level and reported in parentheses, with *** p<0.01, ** p<0.05, * p<0.1.

Table D4: Impact of GDPR on Deal-Based Research Collaboration

	Dependent variable: ihs(# Deals)	
	R&D Collab. Deals	Asset Transfer Deals
	(1)	(2)
GDPR × High Exposure	-0.0136***	-0.0212***
	(0.00433)	(0.00461)
Mean of Dep. Var.	0.0547	0.0581
Observations	73,253	73,253
Firm FEs	Yes	Yes
Year FEs	Yes	Yes

Notes. This table reports DID estimates of the effect of GDPR on the number of deals with other firms. The level of observation is the firm-year. Estimates are from OLS regressions. Standard errors are clustered at the firm level and reported in parentheses, with *** p<0.01, ** p<0.05, * p<0.1.