Data Privacy Regulation and Innovation*

Sukhun Kang
UCSB

Jennifer Kao
UCLA

March 25, 2026

Abstract

We investigate how data privacy regulations shape research and development. Such regulations are designed to protect individuals, but the innovation process that develops valuable new products for those same individuals depends on collecting, sharing, and analyzing their data. We examine this tension in the context of the pharmaceutical industry, where data is a key input to R&D, and the European Union's (EU) General Data Protection Regulation (GDPR), the most comprehensive data privacy regulation to date. Leveraging firm-level variation in exposure to the GDPR, we find that the regulation lowers clinical trials by 18 percent. Firms also shift the types of projects they pursue: trials are less likely to include sites in the EU or to be conducted across multiple countries, and they target a narrower set of diseases. Research collaborations decline overall and shift away from new partners towards established relationships. When looking at research outcomes, we observe that trials take longer to complete, are less likely to succeed, and report results with greater delay. The decline in clinical trial activity is greatest among young firms. 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; Clinical Trials

*We are grateful for feedback from seminar audiences at the Austin Technology & Entrepreneurship Conference, the Utah Strategy Conference, the ASSA Annual Meetings, UCLA, the SoCal Strategy Workshop, the Harvard Medical School and Brigham & Women's Hospital PORTAL Seminar, and Korea University, and to our discussant, Jie Fang. We are especially thankful to Nicholas DeVito for providing access to the EU Clinical Trials Register data. Emily Zhao, Brian Lu, and Sean Wang 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). Yet because data is often proprietary and sensitive, the growth of data-driven research and development (R&D) has raised public concerns regarding privacy 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. As of 2025, data privacy laws have been implemented in 172 countries (Greenleaf, 2025).

How do data privacy regulations shape research and development (R&D)? Prior work has shown that institutional arrangements governing data access (e.g., data repositories) shape the rate and direction of innovation (Furman and Stern, 2011; Furman et al., 2021; Nagaraj and Tranchero, 2023). However, relatively little is known about how the direct *regulation* of data shapes R&D. Unlike other commonly studied forms of regulation, such as entry requirements (Stern, 2017; Grennan and Town, 2020; Gupta and Kao, 2025), product-specific policies (Yin, 2008; Chandra et al., 2026), intellectual property protection (Williams, 2013; Sampat and Williams, 2019; Kyle and McGahan, 2012) or liability regimes (Galasso and Luo, 2017, 2022), data privacy regulations shape how data—a key input in the R&D process—is collected, analyzed, and shared.

This creates a tension: Data privacy regulations are designed to protect individuals, but the innovation process that generates valuable products for those same individuals depends on collecting, sharing, and analyzing their data. These regulations raise research costs by imposing additional regulatory requirements on data collection, sharing, and management. As a result, regulations intended to protect individuals may also delay access to the novel products they need. This paper sheds light on how data privacy regulations shape the level of research investment, the processes through which research is conducted, and the outcomes it produces.

In particular, we focus on the European Union (EU)'s General Data Protection Regulation (GDPR), the world's most comprehensive data privacy regulation to date (European Parliament and Council of the European Union, 2016). The GDPR was adopted in 2016 and became enforceable in 2018. It expanded consent and documentation requirements, imposed limitations on how data can be used and repurposed. restricted data transfers between organizations and across borders, and mandated timely reporting of breaches. These requirements were backed by strong enforcement mechanisms: fines for violations could reach up to 20 million euros or 4 percent of a firm's annual global revenue.

The implications of the GDPR are particularly salient in the pharmaceutical industry, where data is not only a critical input to the drug development process but is also particularly sensitive (Vakili and McGahan, 2016; DiMasi et al., 2016). Further, the GDPR's penalties exceed those under prior health data regulations (e.g., HIPAA). Unlike in consumer data, where the GDPR could help filter out lower-value data (Aridor et al., 2023) or improve the quality of consumer consent (Godinho De Matos and Adjerid, 2022), pharmaceutical data is produced through costly, regulated experimentation and cannot be easily substituted when access becomes more restricted. Drug development often relies on exploratory research, secondary uses of data (e.g., to identify new indications, serve as controls for follow-on studies), and iterative research designs (e.g., adaptive trials). As a result, limitations on secondary uses of data and requirements to outline ex-ante data uses have meaningful effects. Further, the drug development process is often global and collaborative. As a result, the GDPR can affect where firms conduct trials and how they coordinate with research partners. Taken together, data privacy regulations can impact drug development by shaping the level and types of new drug projects that firms pursue, as well as the likelihood and speed with which those drugs ultimately come to market. Figure 1 documents differences in Phase II trials across EU countries before and after the GDPR and suggests a decline in trial levels.

-Insert Figure 1 about here-

While the decline in Figure 1 may be driven by the GDPR, other factors such as firm-level

differences in economic or scientific opportunities may be driving the observed pattern. To motivate a systematic investigation into the decline, we develop a conceptual framework in which data privacy regulations affect firms' R&D decisions through three channels. First, such regulations increase compliance costs by requiring investments in legal, administrative, and data infrastructure, reducing R&D investment and changing which projects firms pursue. Second, because the GDPR specifically targets the collection and sharing of data between organizations, it raises the cost of inter-organizational exchange. This may impact collaborative relationships: as regulatory complexity increases, the informational, contractual, and infrastructural advantages that established partners possess become more valuable (Gulati, 1995; Dyer and Singh, 1998; Gulati and Nickerson, 2008). Third, data privacy regulations may reduce perceived risks for external stakeholders, though existing evidence suggests this channel is limited (Acquisti et al., 2016; Beresford et al., 2012). Under this framework, we expect higher GDPR exposure to reduce R&D investment, shift the composition of research and collaborative relationships, and slow trial progression, though the trust channel may partially offset these effects on specific margins.

We use a comprehensive dataset of clinical trials initiated between 2010 and 2023, with sites in the EU, US, and other countries. A key empirical challenge in examining the effect of the GDPR is that it is inherently a global shock, making it difficult to identify suitable treatment and control groups (Johnson, 2022). To tackle this challenge, we construct a continuous measure of exposure to the GDPR that is motivated by the idea that firms with more EU-based research before 2018 likely faced greater exposure to the regulation. We then use this firm-level exposure measure in a difference-in-differences design.

Our analysis yields three main findings. First, we find that firms more exposed to GDPR reduce their number of clinical trials by 18 percent relative to less exposed firms. This effect is immediate, persistent, and holds even after accounting for shifts in disease composition. Second, firms adjust the types of projects they pursue: trials are less likely to be conducted in the EU, they are less likely to operate across multiple countries, and they target a narrower set

of diseases. Overall research collaborations decline, particularly among new collaborators and among young firms. At the same time, firms modestly increase the number of collaborations with existing research partners. Finally, we show that these effects have meaningful impacts on research outcomes: trials take longer to complete, are less likely to successfully complete, and firms delay the reporting of trial results.

Our findings offer several insights for managers and policymakers. For managers, our results highlight two considerations. First, investment in data protection, which is often seen as an ex-post response to regulation rather than an ex-ante capability, may be an important source of competitive advantage. Consistent with this view, Cao and Kretschmer (2026) find that GDPR compliance investments by large US financial firms catalyzed improvements in data governance and downstream innovation that increased revenue and operating income. Second, established research partnerships are important assets when faced with increased regulatory requirements. While overall collaborations decline following the GDPR, firms shift toward existing partners who offer shared compliance infrastructure and pre-existing data agreements. Younger firms show the largest declines in collaborative research. These firms may benefit from investing in partnerships early or working with established firms who can share regulatory expertise.

For policymakers, our results point to potential unintended consequences of data privacy regulation. First, the 18 percent decline in clinical trials, combined with longer durations and lower completion rates, implies delays in the development of new treatments. Second, the shift of trials away from the EU raises concerns about whether patient populations in clinical trials will reflect the patient populations that ultimately use approved therapies. If EU patients are underrepresented in trials, the resulting evidence base may be less informative for EU prescribers and patients. Third, the disproportionate effects on young firms and new collaborations suggest that privacy regulation may contribute to market concentration, which can have implications for the types of drugs produced and patient access to those drugs. These trade-offs warrant consideration as policymakers design and refine data

privacy frameworks (e.g., explicit safe harbors for pseudonymized research data, streamlined procedures for secondary analyses).

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 short-run performance effects of data privacy regulations on digital services and consumer-facing firms (e.g., Goldfarb and Tucker, 2011; Miller and Tucker, 2009, 2018; Peukert et al., 2022; Johnson et al., 2023; Goldberg et al., 2024; Jia et al., 2021; Janßen et al., 2022; Jin et al., 2024). We build on this work by examining a setting where data is a core input to innovation and show that privacy regulation affects not only how much research firms undertake, but also how research is organized and its outcomes (cf. Vakili and McGahan, 2016). This complements a growing body of work on the economic effects of data privacy regulation. Conceptually, Goldfarb and Tucker (2012) argues that privacy regulation can inhibit innovation by raising the cost of using data as an input, a view supported by broader reviews of the field (Acquisti et al., 2016; Johnson, 2022). Empirically, Demirer et al. (2026) estimate that the GDPR increased the variable cost of data inputs by 22 percent, while Aridor et al. (2020) document how the regulation reshaped the collection and use of consumer data in digital markets. We extend this work by showing that when data is a core input to R&D, the consequences extend beyond costs to reshape how firms organize and conduct research.

Second, we contribute to the literature on R&D collaborations. Prior work has shown that collaborations depend on existing relationships and network position (Gulati, 1995; Gulati and Gargiulo, 1999), trust and governance mechanisms (Gulati and Nickerson, 2008; Reuer and Ariño, 2007), and relationship-specific investments (Dyer and Singh, 1998; Kale et al., 2002). This literature has established that prior ties reduce uncertainty about a partner's capabilities, repeated interaction generates relationship-specific assets, and trust lowers governance costs in exchange relationships. We show that data privacy regulation amplifies each of these dynamics: by raising the cost of inter-organizational data exchange, the

GDPR increases the relative value of established partnerships where informational familiarity, shared compliance infrastructure, and reusable governance structures already exist. The result is not a uniform reduction in collaboration but a reallocation away from new partners and toward established ones. This prediction is specific to data privacy regulation: because it targets inter-organizational data exchange, it shifts the composition of collaborative networks in ways that other cost-increasing regulations, such as tort liability (Galasso and Luo, 2022), do not. This finding extends the alliances literature by identifying regulatory complexity as a force that reshapes the structure of collaborative networks, with implications for the long-run concentration of research.

Finally, we contribute to the broader literature on regulation and firm strategy by showing how data privacy regulation has heterogeneous effects across firms. Consistent with prior work showing that larger firms benefit from economies of scale and scope in pharmaceutical R&D (Cockburn and Henderson, 2001), we find that data privacy regulations disproportionately negatively impacts younger firms. This complements prior work on regulatory capture and incumbent advantage by highlighting a distinct mechanism by which regulations create advantages for incumbents, who can navigate costly regulatory requirements and maintain collaborative relationships essential for data-driven innovation.

The paper proceeds as follows: Section 2 describes the setting and conceptual framework. Section 3 introduces the data and presents descriptive facts. Section 4 describes the empirical strategy and Section 5 presents the results. Finally, Section 6 concludes.

2 Institutional Background and Conceptual Framework

This section presents an overview of the GDPR's institutional context, its implications for research in the pharmaceutical industry, and a conceptual framework that connects data privacy regulation, more broadly, to R&D changes.

2.1 The European Union's General Data Protection Regulation

The EU's GDPR, which was adopted in April 2016 and took effect in May 2018, represents the world's most comprehensive and stringent regulation for governing data collection, storage, usage, and dissemination (European Parliament and Council of the European Union, 2016). The GDPR governs personal data, which ranges from an individual's name and location to health and genetic data. It applies to any firm using the data of EU residents or offering them goods or services. The GDPR replaced the 1995 Data Protection Directive, which had set minimum privacy protections but varied in its implementation across member states (European Union, 1995). In contrast, the GDPR established a unified, enforceable legal framework that was applied consistently across all EU member states.

The GDPR introduced several important changes that strengthened the rights of individuals providing personal data while increasing regulatory requirements for firms. First, it expanded consent and documentation requirements. Under the GDPR, valid consent must be affirmative (pre-checked boxes are prohibited), freely given, granular to each purpose of processing, and must list all third parties who will process the data (Goldberg et al., 2024). For example, a firm seeking to reuse patient data from an earlier trial to support a new therapeutic indication must now obtain explicit, purpose-specific consent for that secondary use; under the prior regime, broader consent language could cover multiple future uses without requiring re-consent.

Second, the GDPR imposed restrictions on data use, reuse, and purpose. Relatedly, individuals gained the right to access, correct, and delete their personal data. Although the 1995 Data Protection Directive included rights to access and correct data, these were implemented inconsistently across member states and lacked uniform enforcement (Peukert et al., 2022). The GDPR's goal was to harmonize and strengthen these protections across all member states, and added new rights, including the right to data portability. However, implementation has varied across member states, and this heterogeneity has in some cases

increased rather than reduced regulatory uncertainty for firms operating across jurisdictions (CIPL et al., 2018).

Third, the GDPR imposed restrictions on data transfers between organizations and across borders, and where multiple organizations jointly process personal data, firms must explicitly allocate compliance responsibilities through a formal arrangement (European Commission, 2025). Finally, the GDPR increased breach reporting and regulatory oversight, and introduced individuals' right to request deletion of their personal data. For example, following a data breach, firms would need to notify the appropriate regulatory authority within 72 hours, and in some cases, inform affected participants. The GDPR also introduced substantial fines for non-compliance: firms can be fined up to 20 million euros or 4 percent of a firm's global annual revenue for the previous year, whichever is greater (European Parliament and Council of the European Union, 2016). These penalties were significantly higher than most previous data privacy regulations (Intersoft Consulting, 2025). As shown in Figure 2, 279 GDPR fines have been imposed on healthcare organizations between 2018 and 2026, with enforcement intensifying over time.[1] Taken together, these features increased the cost of R&D. Indeed, Demirer et al. (2026) find that the GDPR increased the variable cost of data inputs by 22 percent on average for affected firms.

-Insert Figure 2 about here-

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

[1]Data from `https://www.enforcementtracker.com/`, filtered to "Health Care."

2.2 Pharmaceutical R&D and GDPR

2.2.1 Pharmaceutical R&D

Drug development relies heavily on data to evaluate the safety and efficacy of experimental therapies (DiMasi et al., 2016; Lakdawalla, 2018). The data include patient demographics, medical histories, clinical outcomes, adverse events, laboratory results, and biomarker measurements.

The drug development process 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 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 examining efficacy and safety in a larger patient population.

Clinical trial design varies along several dimensions. First, trials vary in geographic scale: some operate at a single site, while others span multiple sites across several countries. Multinational trials can accelerate recruitment of patients and generate evidence from heterogeneous patient populations, but they also require coordination across regulatory jurisdictions and data transfers between organizations. Second, trials vary in therapeutic scope: some target a single disease, while others study a compound's efficacy across several indications. Third, trials vary in their outcomes: some reach successful completion, while others are terminated early due to safety concerns, lack of efficacy, or operational challenges. Trial duration and the timely reporting of results also vary, with implications for how quickly evidence reaches regulators, clinicians, and subsequent researchers.

Within the EU, the conduct of clinical trials is governed by the Clinical Trials Regulation (Regulation EU No 536/2014), which established a single application portal and standardized authorization and safety reporting requirements across member states (European Commission, 2014). Developers can also conduct trials outside the EU with the intention of applying

for EU marketing authorization, provided those trials adhere to equivalent standards. The EU offers several pathways for marketing authorization, including a centralized procedure through the European Medicines Agency that grants approval across all member states.[2]

Given the operational complexity, high costs, and regulatory demands associated with drug development, pharmaceutical firms increasingly rely on collaboration to share resources and expertise. Many trials are jointly conducted by several organizations (Keezhupalat et al., 2016), and trials conducted with partners tend to have higher success rates (Smietana et al., 2016). The choice of whether to collaborate, and with whom, depends on knowledge requirements of the development process and governance challenges of coordinating across organizational boundaries (Macher and Boerner, 2012) . Research collaborations can help firms navigate regulatory environments, pool patient recruitment networks, and share compliance costs (Gulati, 1995; Cockburn and Henderson, 2001).

At the same time, collaboration introduces its own costs and risks. Partners must align on study protocols, data practices, and compliance responsibilities. Coordination across organizations can slow decision-making, and disagreements over intellectual property or data access can strain relationships. When trials involve sensitive patient data, partners share responsibility for data protection, meaning that a compliance failure by one party can expose others to liability. Because drug development depends extensively on the collection and exchange of sensitive patient data, changes to data protection laws have the potential to substantially impact R&D in such settings.

Qualitative reports have highlighted these challenges. The European Federation of Pharmaceutical Industries and Associations (EFPIA) emphasizes that the implementation of the GDPR has complicated clinical research without delivering the anticipated harmonization of regulatory requirements (EFPIA, 2022). Heterogeneity in approaches to the implementation of the GDPR across member states has led to greater regulatory uncertainty than under the prior privacy regime (CIPL et al., 2018). Lalova-Spinks et al. (2024) document that over

[2]See Appendix B for additional details on regulatory approval processes in the EU and US.

40 EU research sites were unable to enroll in NIH-sponsored COVID-19 therapeutic trials due to GDPR data transfer restrictions. Similarly, large-scale research consortia, such as the International Genomics of Alzheimer's Project, were unable to share data across sites in real time and instead conducted separate analyses at each location, limiting statistical power and slowing the pace of discovery. These examples illustrate how GDPR's regulatory complexity can delay clinical trial enrollment, increase research costs, and reduce the feasibility of collaborative research.

These industry-specific challenges motivate a more general framework for understanding how data privacy regulations shape R&D decisions.

2.3 Conceptual Framework

This section develops a framework for understanding how data privacy regulations shape R&D. We organize the framework around three channels. First, regulations raise the cost of conducting research, which reduces R&D investment and changes which projects firms pursue. Second, regulations reshape the relative value of different types of collaborative relationships, favoring established partnerships over new ones. Third, regulations may increase external stakeholders' willingness to participate in research, though existing evidence suggests this channel is limited. These channels jointly predict changes in research levels, processes, and outcomes.

2.3.1 Compliance Costs

Data privacy regulation increases the cost of conducting research through several ways. Firms must invest in consent management systems and data protection audits (Goldberg et al., 2024). Regulatory requirements also raise the marginal cost of data acquisition and management, particularly for projects requiring large or diverse datasets (Aridor et al., 2020). Further, ongoing costs arise from monitoring, reporting, and responding to data subject requests (Demirer et al., 2026).

Higher compliance costs systematically change which projects firms pursue. 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). These constraints raise the relative cost of certain types of research. Projects requiring extensive data collection or cross-border data transfers face higher compliance costs and may not be initiated. Restrictions on cross-border data transfers may lead firms to reduce the geographic scope of their projects—e.g., by concentrating research in fewer jurisdictions or shifting away from regions with stricter requirements. Similarly, requirements for explicit consent may discourage secondary analyses and exploratory research directions, leading firms to narrow the set of questions they investigate within a given project. In the clinical trial context, these dynamics could manifest as fewer multinational trials, reduced trial in EU countries, and trials targeting fewer therapeutic indications.

Given the asymmetry between compliance costs and the limited evidence for offsetting participation benefits (discussed below), we expect the cost mechanism to dominate, yielding a net reduction in R&D investment.

2.3.2 Research Collaboration

Data privacy regulations affect not only how much research firms conduct but also with whom they collaborate. A large literature on alliance formation has shown that prior ties reduce uncertainty about a partner's capabilities and reliability (Gulati, 1995), that relational governance built through repeated interaction can substitute for formal contracting (Gulati and Nickerson, 2008), and that repeated interaction generates relationship-specific assets that are costly to replicate with a new partner (Dyer and Singh, 1998; Kale et al., 2002). In drug development specifically, Macher and Boerner (2012) show that the nature of knowledge at organizational boundaries shapes whether firms conduct R&D internally or through external partners. We argue that data privacy regulation amplifies each of these dynamics, shifting the relative value of established partnerships over new ones.

This shift operates through three reinforcing mechanisms. First, the GDPR's joint controller framework requires collaborating organizations to formally allocate compliance responsibilities before sharing personal data (European Commission, 2025). For new partners, firms must assess compliance capabilities under substantial uncertainty: a prospective collaborator's data-handling practices, IT infrastructure, and legal preparedness are largely unobservable prior to engagement. With established partners, firms have direct experience with these capabilities from prior projects, reducing the informational cost of partner assessment. This is consistent with Gulati (1995)'s argument that prior ties serve as informational conduits that mitigate the hazards of new exchange relationships.

Second, partners who have previously collaborated on data-intensive projects will have already invested in compatible compliance infrastructure, including shared consent protocols, aligned data-processing systems, and joint controller agreements. These are relationship-specific investments in the sense of Dyer and Singh (1998). They have value within the existing relationship but must be recreated at full cost with a new partner. The GDPR raises the magnitude of these required investments because data-sharing arrangements must satisfy its detailed requirements for consent documentation, breach-reporting procedures, and cross-border transfer safeguards. As a result, the cost advantage of continuing an existing relationship over initiating a new one increases.

Third, established partners can extend existing governance structures, such as data-processing agreements, compliance audit procedures, and templates for consent protocols, to cover new projects. Gulati and Nickerson (2008) show that relational governance developed through repeated exchange can substitute for formal contracting. In this context, the relevant governance costs are those associated with negotiating, drafting, and monitoring GDPR-compliant data-sharing agreements. These costs are largely fixed per relationship and can be amortized across multiple projects with the same partner, whereas each new partner requires negotiation from scratch.

These three mechanisms are reinforcing: informational familiarity, co-specialized compli-

ance investments, and reusable governance structures all increase the relative attractiveness of established partners when regulatory complexity rises. We therefore expect overall collaborations to decline, with the reduction concentrated among new partnerships. Established collaborators should prove more resilient and may even increase modestly as firms consolidate research within relationships where compliance infrastructure already exists. This prediction is consistent with evidence that the GDPR reduced new relationship formation in other contexts, such as venture capital investment (Jia et al., 2021), and that it created incentives for within-boundary data processing over cross-boundary data sharing (Peukert et al., 2022).

The magnitude of this shift should vary across firms. Younger firms have had less time to accumulate a portfolio of vetted partners and to build relationship-specific compliance assets, so they should face the largest declines in collaborative research. This generates a secondary prediction: data privacy regulation may concentrate collaborative research among incumbents with established networks, with implications for the long-run structure of R&D (Johnson et al., 2023).

2.3.3 Stakeholder Participation

Data privacy regulations may also affect external stakeholders' willingness to participate in research. By strengthening privacy protections and implementing substantial penalties for non-compliance, such regulations may reduce perceived risks for trial participants, whose enrollment decisions can be influenced by trust in researchers and institutions (Alsan et al., 2025; Buckman et al., 2023).

However, the empirical basis for this channel is limited. Research on privacy valuation reveals a persistent gap between stated and revealed preferences: individuals report high concern for privacy but consistently accept small compensation to disclose sensitive information (Acquisti and Grossklags, 2005; Acquisti et al., 2016). Field experiments confirm that consumers are largely unwilling to pay a premium for privacy-protective options (Beresford et al., 2012). In the clinical trial context, enrollment decisions depend on many factors

beyond data privacy, including perceived health benefits, physician recommendation, and disease severity (Alsan et al., 2025). Although we cannot rule out that the GDPR's privacy protections modestly increased willingness to participate in trials that are initiated, this channel is unlikely to offset the compliance costs documented above.

2.3.4 Research Outcomes

The compliance burden introduced by GDPR may affect research outcomes. Administrative requirements for consent management, data subject requests, and breach reporting may delay trial completion. Cross-border data transfer restrictions may delay data aggregation across multinational sites, further slowing progress. These requirements could also reduce the likelihood that trials reach successful completion: firms may terminate trials that become too costly or complex to manage under increased regulatory requirements. Additionally, the GDPR's restrictions on data sharing and the complexity of ensuring compliance before releasing information could delay when firms report trial results. Delays in results reporting have implications beyond individual firms, as timely dissemination of clinical evidence is essential for informing subsequent research and clinical practice.

Taken together, this framework predicts variation in overall R&D levels, the composition of firms' research portfolios, and the speed and success with which projects reach completion. The collaboration channel provides a more specific prediction that distinguishes data privacy regulation from other cost-increasing regulations: because the GDPR targets the collection and sharing of data between organizations, it raises the cost of inter-organizational exchange in a way that disproportionately penalizes new partnerships relative to established ones. This reallocation of collaborative research is the framework's central prediction, and the one most distinctive to data privacy as a regulatory instrument.

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 trial registries. First, ClinicalTrials.gov (via the Aggregate Analysis of ClinicalTrials.gov Database) provides data on trials initiated in the US and elsewhere.[3] Second, EU Clinical Trials Register contains data on trials initiated in the EU and those registered with the European Medicines Agency (EMA). Both registries provide detailed trial information, 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).

Our main analyses focus on Phase II trials as they represent a firm's first major investment for a drug and require substantial data collection and analysis. They typically enroll several hundred patients, last more than two years, and require coordination across several organizations (e.g., hospitals, contract research organizations) (Sertkaya et al., 2024). Further, unlike earlier stages (e.g., Phase I trials), Phase II trials are required to be reported in public registries.

We aggregate the clinical trials data at the firm-year level, where each observation represents a unique sponsoring firm active in a given year. Because trial registries record sponsors as listed on individual trial records, a single firm may appear under multiple names corresponding to subsidiaries or country-level affiliates. To construct the firm-year panel, we harmonized sponsor names by matching subsidiaries to their parent organizations. For example, country-level entities of the same multinational were consolidated under the parent firm. We restrict the sample to firms that initiated at least one Phase II trial before 2018, ensuring that all firms have observable pre-GDPR research activity. We supplement this

[3]Our analysis period (2010-2023) takes place after major changes in trial registration for industry-sponsored trials, the focus of our analysis. For more details, see https://clinicaltrials.gov/about-site/about-ctg.

trial-level data with firm-level information, including firm headquarters location and founding year, using data from Crunchbase. The panel for each firm begins in its founding year or the year of its first trial, whichever is later, and extends through 2023.

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 73,535 firm-year observations representing 7,508 unique firms.

Measuring Firm-Level Exposure: Our empirical strategy is motivated by the idea that firms with extensive EU-based operations prior to the GDPR's implementation face greater exposure to the regulation compared to those with minimal EU-based research activities before the reform. As a proxy for EU-based activity, we examine the share of firms' clinical trial operations.[4] For each firm, we calculate the share of its Phase II trial sites located in the EU before 2018. Figure 3 shows these shares across firms. Firms are classified as having high exposure to the GDPR if their shares are above the median; others are classified as having low exposure to the GDPR. For simplicity and give the bimodal distribution of the share measure, our main analyses compare firms that have high vs. low exposure. However, we also show that our results are similar when using a continuous measure of exposure.

-Insert Figure 3 about here-

Measuring Research Processes: To characterize how GDPR shifts the types of projects that firms pursue, we examine changes in trial design and trial collaborations. For trial design, we measure several characteristics recorded in the trial registries: (i) whether the trial has a pre-specified plan to share individual-level participant data (a binary indicator); (ii) whether the trial has sites in multiple countries; (iii) whether the trial includes sites inside vs. outside the EU; (iv) the number of patients enrolled; and (v) the number of distinct diseases targeted, measured using unique 3-digit Medical Subject Headings (MeSH)

[4]In the context of the pharmaceutical industry, a firm's headquarter location is a poor proxy for exposure to the GDPR because firms often conduct clinical trials across different countries (e.g., targeting countries where they intend to seek regulatory approval).

codes (e.g., musculoskeletal diseases, respiratory tract diseases).

To measure research collaborations, we identify Phase II trials that are conducted in partnership 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 collaborative trials into trials with new and trials with established collaborators based on their relationship with the sponsoring firm. A collaborator is classified as *new* if the sponsoring firm has either never collaborated with that partner on a Phase II trial or if all prior collaborations occurred within the 2 years immediately preceding the focal trial. Collaborators with longer histories of collaboration are classified as *established*.[5]

Measuring Research Outcomes: To characterize research outcomes, we examine trial progression and the public availability of results. We measure trial duration as the number of years between the trial start date and trial completion date. We then measure the share of trials that are completed (as opposed to being terminated or suspended). To characterize data disclosure, we examine the share of trials that publicly report results within one year of trial completion (the timeline as mandated by the Food and Drug Administration Amendments Act of 2007).

Measuring Firm Characteristics: In several analyses, we examine heterogeneity by firm age and size. We classify firms as "Older" if their age in 2018 exceeds the median across all sponsors and "Younger" otherwise. We measure firm size by research intensity: firms are classified as "Larger" if they conducted an above-median number of Phase II trials prior to

[5] Appendix E shows that our results are robust to alternative measures of collaboration.

2018 and "Smaller" otherwise.

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 percent (0.36) are multicountry trials. Firms conduct 0.12 trials with collaborators each year and firms are more likely to have new, rather than established, collaborators (0.09 vs. 0.04, respectively). Appendix Table C2 further compares pre-GDPR characteristics of firms with and without collaborators. On average, trials last 2.7 years and 56 percent successfully complete. Timely trial results reporting is rare: within our sample, about one percent of trials report their results within one year. Approximately 40 percent of firm-years are associated with old firms and 20 percent are associated with large firms.

-Insert Table 1 about here-

4 Empirical Strategy

We measure the causal effect of the GDPR on innovation using a DID design. Our baseline specification compares outcomes for firms with higher vs. lower exposure to the GDPR, before and after the regulation's implementation. Our dynamic specification is:

$$y_{it} = \sum_{t} \beta_t \mathbf{1}\left\{year = t\right\} \times \mathbf{1}\{EUSiteShare_i\} + \mathbf{1}\{EUSiteShare_i\} + \delta_i + \delta_t + \epsilon_{it} \quad (1)$$

where y_{it} denotes the firm-level outcome (e.g., the number of clinical trials initiated) for firm i in year t. The indicator $\mathbf{1}\{EUSiteShare_i\}$ identifies whether the firm has an above-median share of pre-2018 Phase II trial sites based in the EU. The coefficients β_t capture changes in the outcome for firms with high vs. low GDPR exposure over time. Appendix Table C1 shows that firms with high exposure to the GDPR 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 δ_t to control for common annual shocks across firms. Standard errors are clustered at the firm level.

In regression specifications, we modify equation (1) to include an indicator for years after the GDPR's implementation ($\mathbf{1}\{GDPR_t\}$) interacted with $\mathbf{1}\{EUSiteShare_i\}$. In particular, we estimate:

$$y_{it} = \beta \mathbf{1}\{GDPR_t\} \times \mathbf{1}\{EUSiteShare_i\} + \mathbf{1}\{EUSiteShare_i\} + \delta_i + \delta_t + \epsilon_{it} \tag{2}$$

In addition to a binary measure of exposure to the GDPR, we also present our main estimates using a continuous measure of firms' pre-2018 EU trial site share. This yields similar results.

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 Phase II trials among firms more exposed to the regulation. Figure 4 shows the raw trends in the total (Panel A) and average (Panel B) number of trials across firms with below vs. above median share of Phase II trial sites in the EU before 2018. Trial counts across both groups are relatively stable before 2018. However, after 2018, firms with an above median share of Phase II trial sites (i.e., those facing relatively higher exposure to the GDPR) reduce their trials. In contrast, firms with a below median share

of Phase II trial sites (i.e., those facing lower exposure to the GDPR) increase their trials.[6] While this descriptive evidence suggests that the GDPR led to a relative decline in research levels among firms more exposed to the regulation, these patterns could also be driven by firm-level differences (e.g., in resources, R&D ability) or by changing scientific and economic opportunities. As a result, we now turn to a formal estimation of Equation (1) to investigate the causal impact of the GDPR.

-Insert Figure 4 about here-

To begin, we use the number of Phase II trials (with the inverse hyperbolic sine transformation) and a binary measure of $EUSiteShare$. Panel A of Figure 5 shows that the difference in the inverse hyperbolic sine-transformed number of Phase II 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). This timing is also consistent with the onset of GDPR enforcement in the healthcare sector, where fines have been imposed since 2018 (Figure 2). The decline in the level of Phase II trials remains large and statistically significant through 2023. Panel B of Figure 5 shows similar results using the direct (non-transformed) number of Phase II trials.

-Insert Figure 5 about here-

Column 1 of Panel A in Table 2 presents the corresponding estimates from Equation (2). 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. Given the skewed distribution of trial counts and the

[6] As described in 3, we harmonized sponsor names by matching subsidiaries to their parent organizations. This consolidation means the observed decline is unlikely to reflect firms shifting trials from one subsidiary to another within the same corporate structure.

large number of zeroes, we use the inverse hyperbolic sine transformation as our preferred measure of research levels. However, by way of comparison, when looking at the direct (non-transformed) number of trials, we continue to observe that the GDPR leads firms to reduce their research levels (Column 2 of Panel A in Table 2 and Panel B of Figure 5). Panels C and D of Figure 5 and Panel B of Table 2 show similar patterns when using a continuous measure of $EUSiteShare_i$. Together, these results suggest that the GDPR led to declines in overall research levels.

-Insert Table 2 about here-

5.1.2 Robustness Checks and Heterogeneity Analyses

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

Accounting for Changing Disease Composition. 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. For example, firms may prioritize diseases that are more challenging to investigate, leading to a mechanical decline in research levels. To address this, Table D1 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. Table D1 shows that 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 relative decline in clinical trial levels among firms more exposed to the GDPR. In addition, Table D5 shows that the decline in trials is concentrated among diseases with high levels of prior research activity, consistent with the view that.

Alternative Outcome Specifications. We confirm that our findings hold under alternative outcome measures. Table D2 shows that the estimated impact of GDPR remains nega-

tive and statistically significant when the outcome variable is an indicator for any phase II trial (Column 1) or transformed using a log transformation (Column 2).

Confounding Events. One potential concern may be that our results are driven by concurrent events, such as Brexit or the onset of Covid-19. Brexit, which took effect in January 2020, could independently affect clinical trial research through regulatory uncertainty. Although the UK subsequently incorporated the EU GDPR into national law (with minor modifications), firms anticipating Brexit may have reduced EU-based research for reasons unrelated to the GDPR. Given the global nature of R&D, cleanly separating firms exposed to Brexit from those that were not is challenging. As an approximation, we exclude firms likely to be most affected by Brexit (i.e., those conducting trials in the UK) from our treatment group by reconstructing our exposure measure using only non-UK trial sites. Note that this approach effectively allocates firms whose GDPR exposure was driven primarily by UK-based research into the control group, which would dampen our estimated effects if Brexit did indeed significantly dampen subsequent research levels. Despite this conservative specification, Column 3 of Table D2 shows that our results remain negative and statistically significant. To address concerns about Covid-19, we rerun our main analysis using only firm-year observations with years before 2020. Column 4 of Table D2 shows that the effect of GDPR remains negative and statistically significant. While we cannot fully rule out a role for Brexit or Covid-19, these results suggest these events are unlikely to be the primary driver of our findings

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 and therefore better able to adjust their research activities (e.g., scale back R&D programs that would have taken place in the EU absent the GDPR). To address these concerns, we restrict the sample to only include multinational firms, which we define as those with at least one clinical trial 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. Column 5 of Table D2 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.

Heterogeneity Across Firms. The effects of the GDPR may vary across firm age. Older firms may have more experience with regulators and established processes for complying with regulations. In contrast, younger firms may have less expertise in dealing with regulators and changing regulatory requirements. The effects of GDPR may also vary by firm size, which may reflect a firm's ability to meet regulatory requirements. Larger firms often have in-house legal teams and data protection resources. In contrast, smaller firms may not be able to afford the infrastructure needed to meet GDPR standards.

Table D3 shows that the level of trials declines among both younger and older firms, but the decline is significantly larger among younger firms. We observe that the decline is greater among smaller firms than larger firms, though the difference between the two is not statistically significant at conventional levels, likely due to small sample sizes. Taken together, these findings provide suggestive evidence that regulatory experience and resources play an important role in shaping the types of firms most affected by the GDPR.

Differences Across Phases. Table D4 examines the effect of the GDPR across different trial Phases. We find a small, but statistically significant decline in Phase I trials and no significant decline in Phase III trials. The larger effect in Phase II trials relative to Phase I may be explained by two factors: First, Phase II trials are substantially more data-intensive relative to Phase I trials and thus may be particularly impacted by the regulatory requirements imposed by the GDPR. Second, as discussed in Section 3.1, Phase II trials are better reported in clinical trial registries compared to Phase I trials. The absence of a measurable effect on Phase III trials is consistent with the view that GDPR's costs operate primarily on the exploratory margin of R&D. Phase III trials represent commercially committed investments with substantial sunk costs, making them more resilient to incremental regulatory cost increases. In contrast, Phase II trials are the stage at which firms decide whether to

advance a compound, and it is this decision margin that is most sensitive to changes in the cost of collecting and sharing data.

5.2 Impact on Research Processes

5.2.1 Trial Design

To further shed light on how the GDPR shaped the types of projects firms pursue, we examine Phase II trial design. Table 3 shows firms with high exposure to the GDPR are significantly more likely to conduct trials with pre-specified plans for sharing individual participant data. This is consistent with the view that the GDPR shapes how firms plan and manage trial data.

Columns 2 through 4 shed light on the location in which trial levels decline. Consistent with the GDPR's geographic focus, we find that high exposure firms significantly reduce the level of trials taking place in the EU (i.e., trials with at least one EU site). In contrast, the reduction in trials in non-EU countries is notably lower. Among trials that are conducted, we observe little change in the average number of enrolled patients. The lack of increase is consistent with the privacy valuation literature, which finds that individuals place low value on privacy in actual decisions (Acquisti et al., 2016; Acquisti and Grossklags, 2005; Beresford et al., 2012). As a result, the participation channel is unlikely to make a large offsetting role. Finally, we observe a significant decline in the average number of diseases studied in a given trial, suggesting that in response to the GDPR, firms reduce their research scope. Taken together, these findings are consistent with the view that data privacy regulations cause firms to shift their research design.

Taken together, the increase in pre-specified data sharing plans is consistent with the trust channel operating on a limited margin: firms respond to the GDPR's transparency requirements by committing to data sharing protocols ex ante. The absence of a decline in per-trial enrollment reflects a different logic: conditional on initiating a trial, firms must

enroll enough patients to achieve adequate statistical power, so the GDPR's cost effects manifest in fewer trials being initiated rather than in smaller trials. Neither pattern is sufficient to counteract the overall decline in trial activity, consistent with our expectation that compliance costs dominate.

-Insert Table 3 about here-

5.2.2 Trial 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 1.9 ($\approx e^{-0.0190} - 1$) 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 resilience of collaborative trials relative to overall trial levels is consistent with the framework developed in Section 2.3.2: collaboration offers a way to share compliance costs, and established partnerships provide informational, contractual, and infrastructural advantages that partially insulate them from regulatory shocks. However, the framework predicts that this resilience should be concentrated among established partners, with new partnerships bearing the brunt of the decline. To test this, we compare collaborations with relatively new partners (those with no prior collaboration or whose collaboration began within 2 years of the focal trial) to those with established partners (those with longer collaborative histories).

Columns 2 and 3 in Table 4 show 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). This pattern is consistent with the framework's prediction that the GDPR shifts the relative value of different types of partnerships. With new partners, firms face informational uncertainty about compliance capabilities, must negotiate governance structures from scratch, and lack the relationship-specific compliance investments that reduce the marginal cost of additional joint projects. With established partners, these costs have already been incurred.

5.2.3 Robustness Checks and Heterogeneity Analyses

The preceding analysis shows that firms reduce collaborative trials following the GDPR, with the decline concentrated among new rather than established partners. Several questions arise from these findings. We probe the robustness of these findings and examine heterogeneity below.

Alternative Measures of Collaboration Novelty. The distinction between new and established collaborators depends on how we define prior collaborative experience. Table E1 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 for prior collaborative experience with Phase II trials. Table E2 shows that the results are similar when measuring new vs. existing collaborators based on prior collaborative experience with Phase III trials.

Heterogeneity Across Firms. The framework in Section 2.3.2 predicts that the shift toward established partners should be largest among firms that have had less time to accumulate a portfolio of vetted collaborators and to build relationship-specific compliance assets (e.g., compatible compliance infrastructure). Firm age proxies for the opportunity to develop such partnerships, while firm size proxies for the ability to meet regulatory requirements independently.

Table E3 shows that the decline in collaborative trials is driven by younger firms (column 1), while older firms (column 2) show no significant change. This is consistent with the prediction that older firms have accumulated relational assets that insulates their col-

laborative research from regulatory shocks. Columns 3 and 4 show results by firm size. While the estimated effect is negative for small firms and positive for large firms, the differences between small and large firms are not statistically significant. The significant effect of firm age but not firm size suggests that the relevant advantage is relational (accumulated through repeated interaction with partners) rather than resource-based (available through scale). This pattern is more consistent with the informational and governance mechanisms in Section 2.3.2 than with a pure cost-sharing explanation.

Deal-based Measures of Collaboration. 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 E4 shows that the GDPR led to a significant decline across both types of deal-based collaborative research.

Geographic Distance Between Collaborators' Headquarters. If GDPR's data transfer restrictions drive the decline in collaboration, effects may vary with geographic distance between partners. To the extent that firms' research activities take place closer to their headquarters, greater geographic distance could increase the costs of compliance with the GDPR (e.g., by requiring data transfers across borders) and make it more difficult to establish trust between partners (Catalini et al., 2020). Table E5 examines whether the effects of GDPR vary depending on the geographic distance between collaborators' headquarters. 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.

This null result is informative for distinguishing among the mechanisms proposed in Section 2.3.2. If spatial search costs were the primary driver of the decline in new collaborations, we would expect geographic distance to matter, since locating and screening distant

partners is costlier. The absence of a geographic gradient suggests that the mechanism operates through accumulated relationship-specific assets rather than through spatial frictions in partner search. A secondary explanation is that headquarters location is an imperfect proxy for where research occurs: multinational firms often operate through subsidiaries, so headquarters distance may not reflect the true regulatory exposure of a given collaboration.

5.3 Impact on Research Outcomes

While the prior results indicate that the GDPR led to an overall decline in research levels, the effect on trial outcomes remains unclear. On the one hand, faced with increased costs, firms may be more selective about the projects they initiate and only conduct trials with a greater likelihood of successful completion. We would expect the share of trials that complete successfully to increase following the GDPR. On the other hand, increased costs associated with collecting, analyzing, and sharing data may increase trial durations and cause firms to increase trial terminations. Table 5 provides evidence consistent with this latter channel. Column 1 shows that following the GDPR, trials take 12 percent longer to complete. Column 2 shows that 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. Column 3 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 5 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 largely focused on examining how regulation shapes immediate firm performance outcomes,

raising open questions about how such 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 clinical trials by 18 percent and led to meaningful shifts in their research processes and outcomes.

Our analysis has several limitations that suggest opportunities for future work. First, our data measures 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 work could explore whether patterns emerge in other data-intensive settings. Third, our firm-level approach restricts our analysis to examining responses among existing firms. A natural extension is to examine how the GDPR affected firm entry and exit and how these dynamics redirected clinical trials to other jurisdictions (e.g., Asia) (Liu et al., 2025).

These findings carry broader implications. First, the GDPR was designed to protect EU individuals' privacy, but our results suggest it has the unintended consequence of slowing medical innovation that serves those same individuals (Mahoney et al., 2009). This is consistent with a broader pattern in which regulatory interventions designed to achieve important social objectives generate unintended costs for innovation. Galasso and Luo (2022) document a similar dynamic in the medical implants industry, where liability protections intended to compensate victims of defective products had a chilling effect on downstream innovation. In both cases, the policy achieves its proximate objective while imposing costs on the innovation process that are not immediately visible to policymakers. An 18 percent reduction in clinical trials, combined with longer trial durations and lower completion rates, suggests that some valuable treatments may be delayed or never reach patients. Given the growth in the number of privacy regulations modeled after the GDPR (Greenleaf, 2025), our findings suggest that policymakers should consider the effects of regulations in research-intensive industries, for instance, through explicit safe harbors for pseudonymized research

data or streamlined consent procedures for secondary analyses. Second, the shift from new to established collaborators suggests that data privacy regulation may concentrate research towards established collaborative networks. Our framework identifies a specific mechanism for this: when regulatory complexity raises the cost of forming new data-sharing relationships, firms with accumulated relationship-specific assets, including vetted partners, compatible compliance infrastructure, and reusable governance structures, face lower marginal costs of collaboration. Younger firms, which have had less time to build such assets, bear disproportionate costs. Combined with evidence that the GDPR reduced venture investment in technology firms (Jia et al., 2021), this pattern suggests that data privacy regulation may generate persistent advantages for incumbents with established research networks, with implications for the long-run distribution of innovation activity and for the types of products that ultimately reach individuals.

Taken together, our study shows that data privacy regulation shapes innovation through multiple channels. Beyond reducing the aggregate level of R&D, the GDPR reshapes the relational structure through which R&D is pursued. By documenting how established partnerships become more valuable as regulatory complexity increases, we contribute a mechanism, grounded in the alliances literature, that distinguishes data privacy regulation from other cost-increasing shocks. These findings advance understanding of how regulation shapes firms' R&D decisions in data-intensive industries and underscore the importance of relational assets in navigating regulatory environments.

References

Acquisti, A. and Grossklags, J. (2005). Privacy and Rationality in Individual Decision Making. *IEEE Security & Privacy*, 3(1):26–33.

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

Adjerid, I., Acquisti, A., Telang, R., Padman, R., and Adler-Milstein, J. (2016). The Impact of Privacy Regulation and Technology Incentives: The Case of Health Information Exchanges. *Management Science*, 62(4):1042–1063.

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

Alsan, M., Campbell, R. A., Leister, L., and Ojo, A. (2025). Investigator racial diversity and clinical trial participation. *Journal of Health Economics*, 100:102968.

Aridor, G., Che, Y.-K., and Salz, T. (2020). The economic consequences of data privacy regulation: Empirical evidence from GDPR. *National Bureau of Economic Research Working Paper No. 26900*.

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):695–730.

Beresford, A. R., Kübler, D., and Preibusch, S. (2012). Unwillingness to Pay for Privacy: A Field Experiment. *Economics Letters*, 117(1):25–27.

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

Cao, R. and Kretschmer, T. (2026). Regulation as Opportunity: Proactive GDPR Compliance in the US Financial Services Industry. *Industrial and Corporate Change*, 35(1):150–170.

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. (2026). Regulatory Incentives for Innovation: The FDA's Breakthrough Therapy Designation. *Review of Economics and Statistics*, 108(2):470–484.

CIPL, EFPIA, and FPF (2018). Can GDPR Work for Health Scientific Research? Accessed: 2025-05-28.

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., Jiménez Hernández, D. J., Li, D., and Peng, S. (2026). Data, Privacy Laws and Firm Production: Evidence from the GDPR. *National Bureau of Economic Research Working Paper No. 32146*.

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.

Dyer, J. H. and Singh, H. (1998). The Relational View: Cooperative Strategy and Sources of Interorganizational Competitive Advantage. *Academy of Management Review*, 23(4):660–679.

EFPIA (2022). EFPIA Statement on a GDPR Code of Conduct. https://www.efpia.eu/news-events/the-efpia-view/statements-press-releases/efpia-statement-on-a-gdpr-code-of-conduct/. Accessed: 2025.

European Commission (2014). Clinical Trials Regulation (Regulation EU No 536/2014). Retrieved from https://health.ec.europa.eu/medicinal-products/clinical-trials/clinical-trials-regulation-eu-no-5362014$_en$.

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

European Parliament and Council of the European Union (2001). Directive 2001/83/EC on the Community Code Relating to Medicinal Products for Human Use. Official Journal of the European Union L 311/67. As amended by Directive 2004/27/EC.

European Parliament and Council of the European Union (2004). Regulation (EC) No 726/2004 Laying Down Community Procedures for the Authorisation and Supervision of Medicinal Products for Human and Veterinary Use. Official Journal of the European Union L 136/1.

European Parliament and Council of the European Union (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Official Journal of the European Union L 119/1. https://eur-lex.europa.eu/eli/reg/2016/679/oj.

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.

Furman, J. L., Nagler, M., and Watzinger, M. (2021). Disclosure and Subsequent Innovation: Evidence from the Patent Depository Library Program. *American Economic Journal: Economic Policy*, 13(4):239–270.

Furman, J. L. and Stern, S. (2011). Climbing atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research. *American Economic Review*, 101(5):1933–1963.

Galasso, A. and Luo, H. (2017). Tort Reform and Innovation. *Journal of Law & Economics*, 60(3):385–412.

Galasso, A. and Luo, H. (2022). When Does Product Liability Risk Chill Innovation? Evidence from Medical Implants. *American Economic Journal: Economic Policy*, 14(2):366–401.

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.

Greenleaf, G. (2025). Global data privacy laws 2025: 172 countries, 12 new in 2023/24. *Privacy Laws & Business: International Report*.

Grennan, M. and Town, R. J. (2020). Regulating innovation with uncertain quality: information, risk, and access in medical devices. *American Economic Review*, 110(1):120–161.

Gulati, R. (1995). Social structure and alliance formation patterns: A longitudinal analysis. *Administrative Science Quarterly*, 40(4):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.

Gupta, C. N. and Kao, J. (2025). Beyond the Label: Regulatory Arbitrage in Pharmaceutical Market Entry.

Intersoft Consulting (2025). Fines / Penalties - General Data Protection Regulation (GDPR). https://gdpr-info.eu/issues/fines-penalties/. Accessed: 2025-05-30.

Janßen, R., Kesler, R., Kummer, M. E., and Waldfogel, J. (2022). GDPR and the Lost Generation of Innovative Apps. *National Bureau of Economic Research Working Paper No. 30028*.

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.

Jin, G. Z., Liu, Z., and Wagman, L. (2024). The GDPR and SDK Usage in Android Mobile Apps. *National Bureau of Economic Research Working Paper No. 33099*.

Johnson, G. (2022). Economic research on privacy regulation: Lessons from the GDPR and beyond. *National Bureau of Economic Research Working Paper No. 30705*.

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.

Kale, P., Dyer, J. H., and Singh, H. (2002). Alliance capability, stock market response, and long-term alliance success: the role of the alliance function. *Strategic Management Journal*, 23(8):747–767.

Keezhupalat, S. M., Naik, A., Gupta, S., Srivatsan, R., and Saberwal, G. (2016). An Analysis of Sponsors/Collaborators of 69,160 Drug Trials Registered with ClinicalTrials.gov. *PLOS ONE*, 11(2):e0149416.

Kyle, M. K. and McGahan, A. M. (2012). Investments in Pharmaceuticals Before and After TRIPS. *Review of Economics and Statistics*, 94(4):1157–1172.

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.

Liu, S., Hu, H., Ge, C., Yuan, S., Jiang, J., and Chen, X. (2025). The rise of China's pharmaceutical industry from 2015-2024: a decade of innovation. *Nature reviews. Drug Discovery*, 24(10):738–739.

Macher, J. T. and Boerner, C. (2012). Technological Development at the Boundaries of the Firm: A Knowledge-Based Examination in Drug Development. *Strategic Management Journal*, 33(9):1016–1036.

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. (2009). Privacy Protection and Technology Diffusion: The Case of Electronic Medical Records. *Management Science*, 55(7):1077–1093.

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

Nagaraj, A. and Tranchero, M. (2023). How Does Data Access Shape Science? Evidence from the Impact of US Census's Research Data Centers on Economics Research. *National Bureau of Economic Research*, (31372).

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.

Reuer, J. J. and Ariño, A. (2007). Strategic alliance contracts: dimensions and determinants of contractual complexity. *Strategic Management Journal*, 28(3):313–330.

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.

Smietana, K., Siatkowski, M., and Møller, M. (2016). Trends in Clinical Success Rates. *Nature Reviews Drug Discovery*, 15:379–380.

Stern, A. D. (2017). Innovation under regulatory uncertainty: Evidence from medical technology. *Journal of public economics*, 145:181–200.

U.S. Food and Drug Administration (2023). New Drug Application (NDA). https://www.fda.gov/drugs/types-applications/new-drug-application-nda. Accessed: 2025.

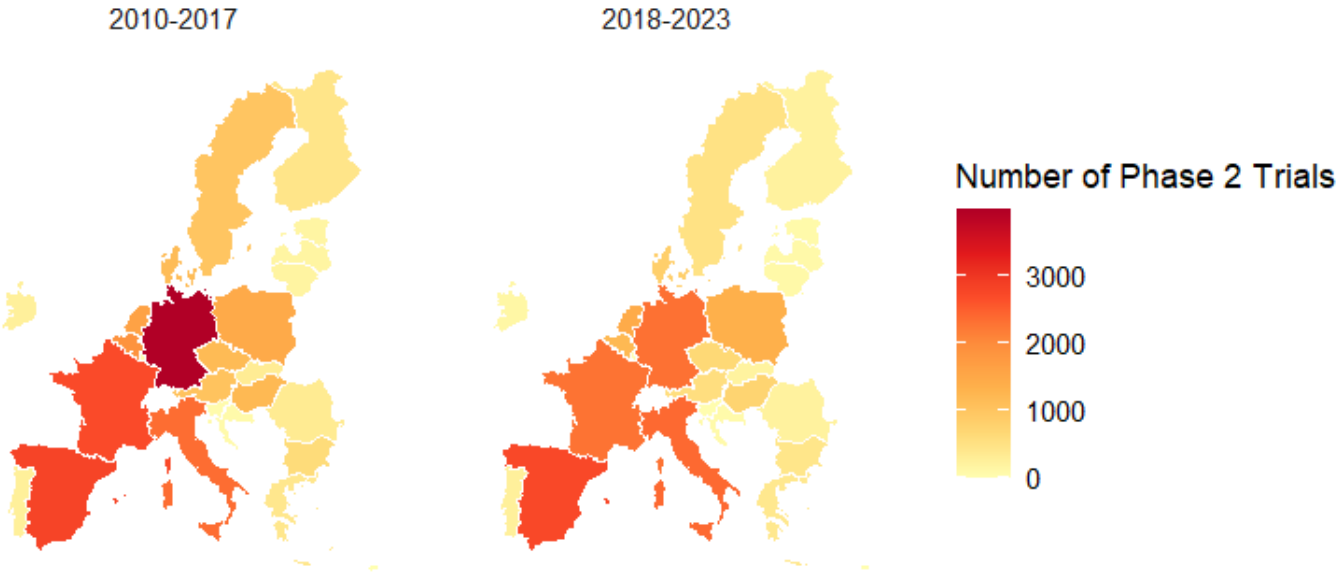
Vakili, K. and McGahan, A. M. (2016). Health Care's Grand Challenge: Stimulating Basic Science on Diseases that Primarily Afflict the Poor. *Academy of Management Journal*, 59(6):1917–1939.

Williams, H. L. (2013). Intellectual property rights and innovation: Evidence from the human genome. *Journal of Political Economy*, 121(1):1–27.

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

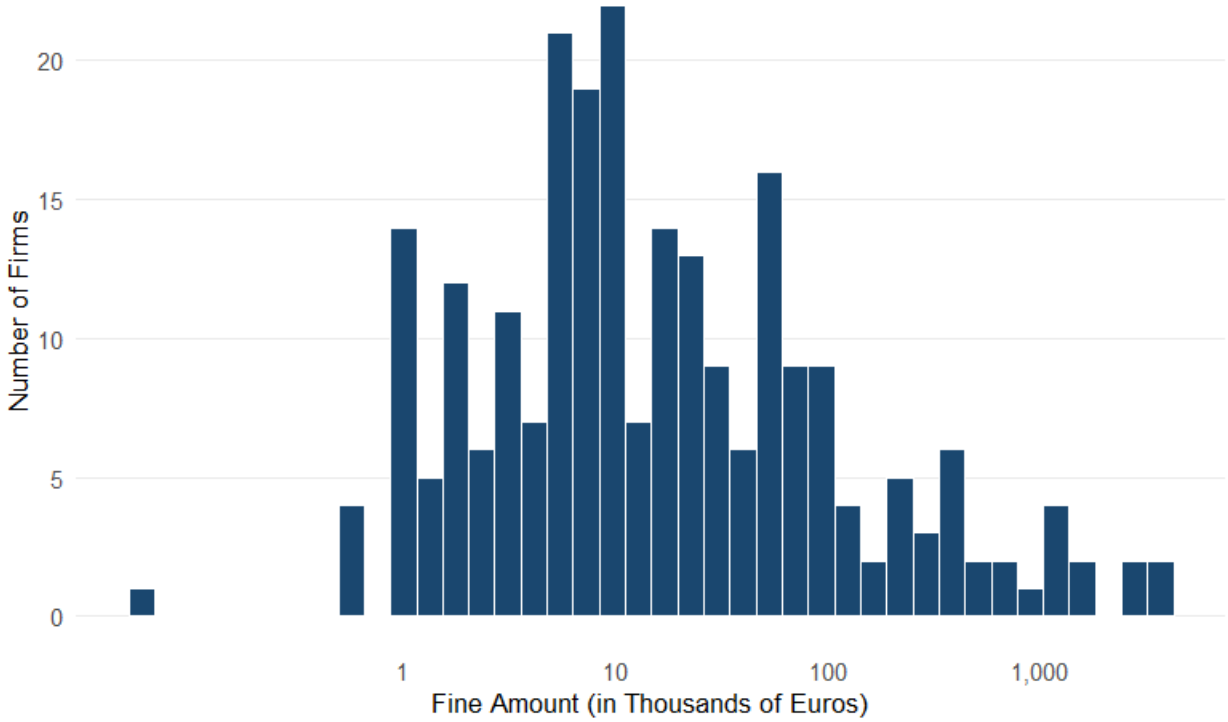
7 Main Figures and Tables

Figure 1: Phase II Clinical Trials in the EU Before and After GDPR



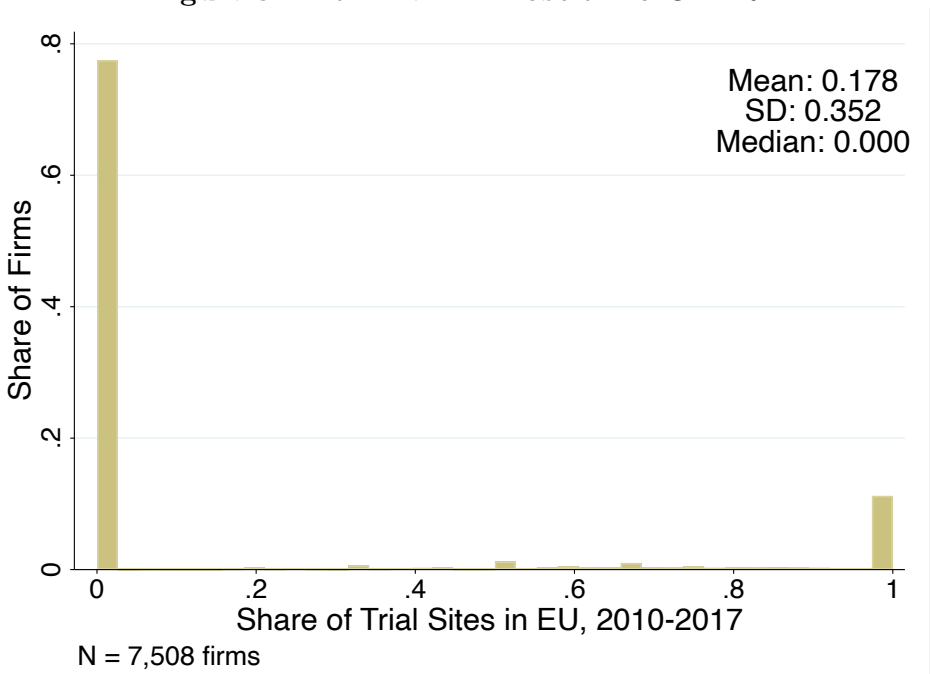
Notes. This figure shows the number of Phase II clinical trials initiated in EU countries before the GDPR (2010–2017) and after its implementation (2018–2023). Darker shading indicates more trials.

Figure 2: Distribution of GDPR Fines in Healthcare (2018–2025)



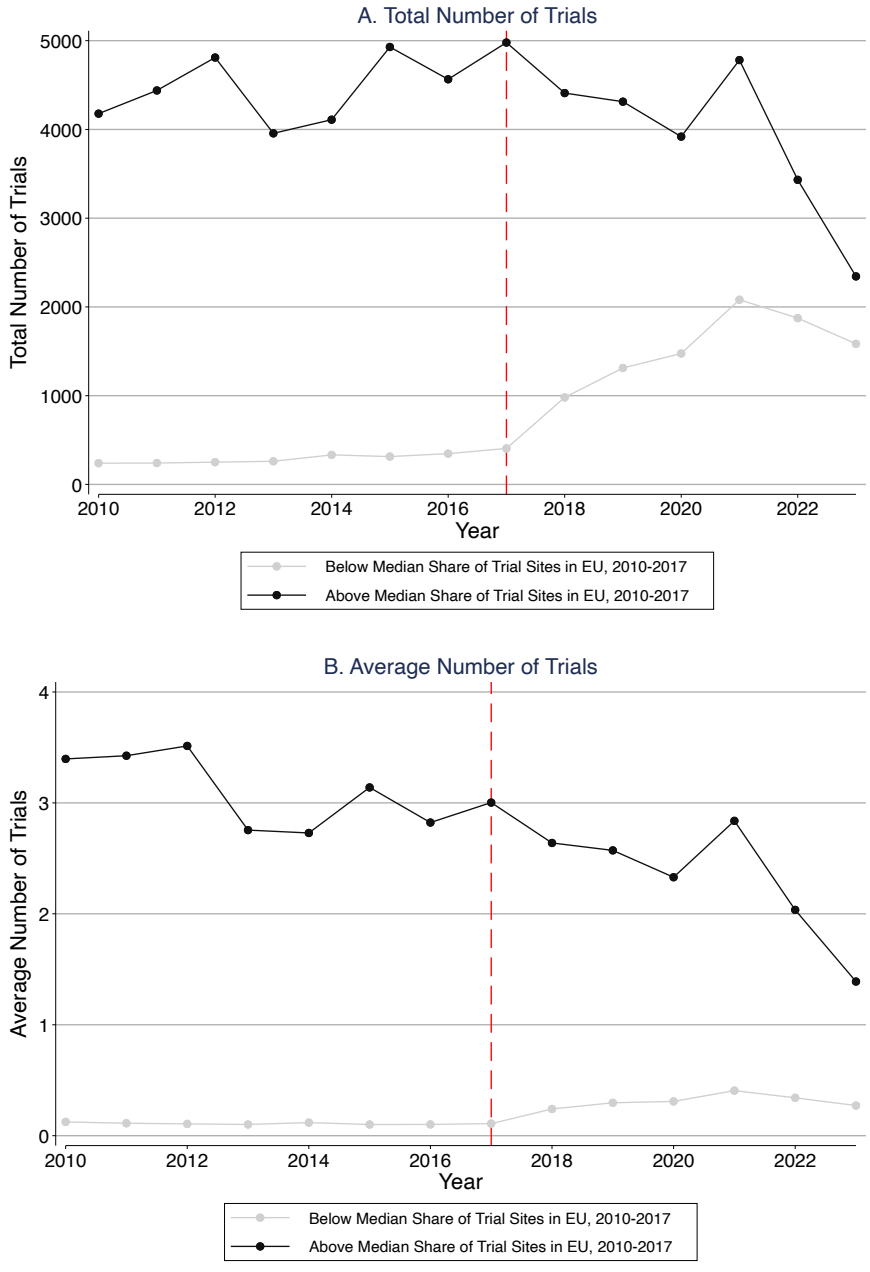
Notes. This figure shows the distribution of publicly reported GDPR fines imposed on healthcare organizations. The horizontal axis shows the fine amount in thousands of euros on a logarithmic scale. The vertical axis shows the number of firms fined. Data from the GDPR Enforcement Tracker (https://www.enforcementtracker.com/), filtered to "Health Care." $N = 279$.

Figure 3: Firm-level Exposure to GDPR



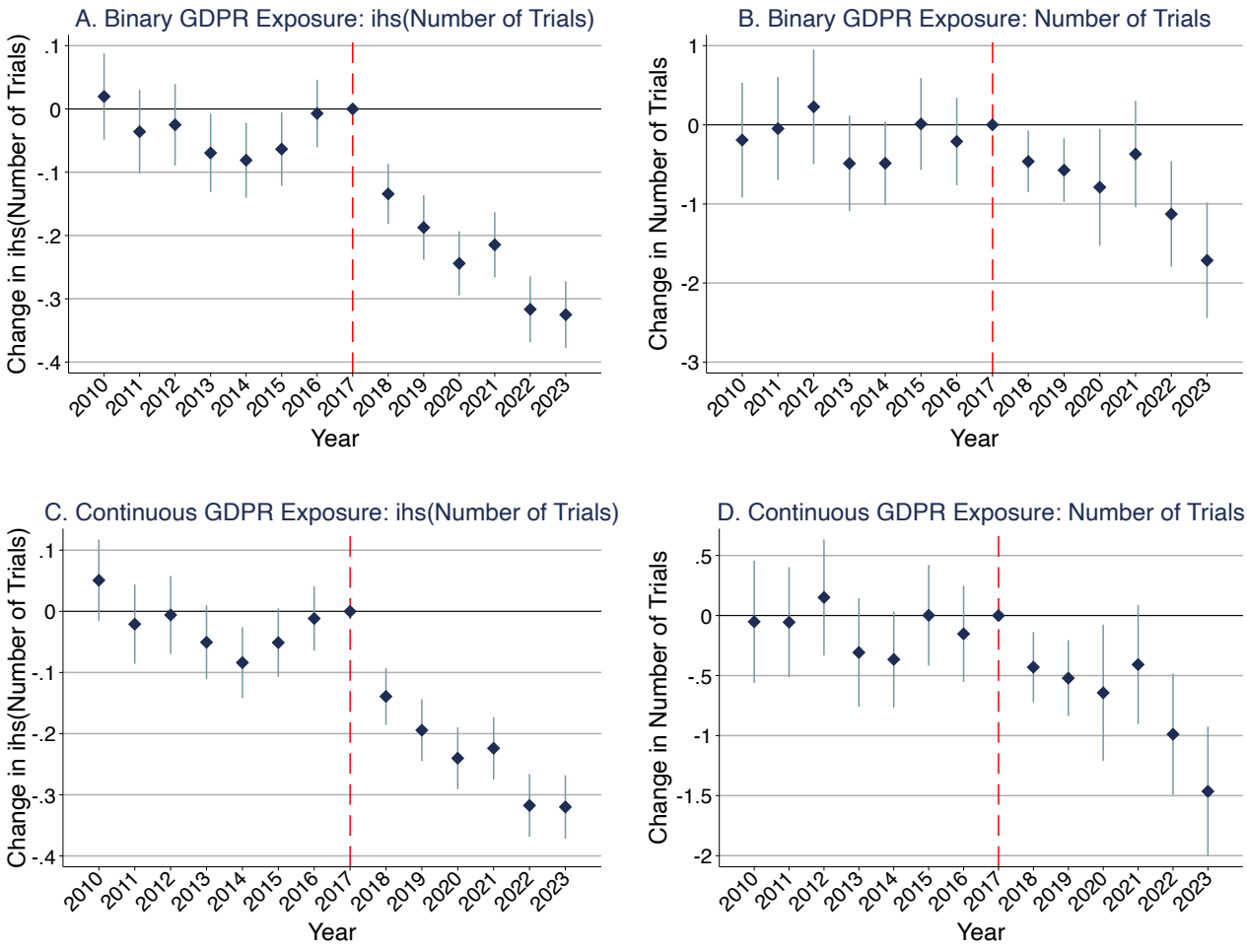
Notes. This figure shows the distribution of firms' pre-2018 share of Phase II trial sites in the EU.

Figure 4: Trends in Trial Levels



Notes. This figure plots the total number of clinical trials across (i) firms with a below median share of trial sites in the EU before 2018 (light gray) and (ii) firms with an above median share of trial sites in the EU before 2018 (dark gray). Panel A shows the total number of trials across firms. Panel B shows the average number of trials across firms. The dotted vertical line represents the year before the GDPR.

Figure 5: Impact of GDPR on Trial Levels



Note: Panels A and B present coefficients from firm-year regressions that regress the outcome on an indicator for after the GDPR interacted with an indicator for whether a firm has an above median share of Phase II trial sites in the EU prior to 2018. Panels C and D present coefficients from firm-year regressions where an indicator for after the GDPR interacted with a continuous rather than binary measure of share of Phase II trial sites in the EU prior to 2018. In all regressions, we include year fixed effects and firm fixed effects and cluster standard errors at the firm level. In Panels A and C, the outcome is the inverse hyperbolic sine transformed number of Phase II trials in a year. In Panels B and D, the outcome is the number of Phase II trials in a year. Solid vertical lines are 95 percent confidence intervals. 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	73,535	0.96	10	0	559
Research Process					
# Multicountry Trials	73,535	0.36	2	0	97
# EU Trials	73,535	0.83	10	0	554
# Non-EU Trials	73,535	0.13	1	0	30
# Diseases	9,969	1.36	1	0	18
Share of Trials with Pre-specified Plans for Sharing Participant Data	9,969	0.05	0	0	1
# Patients Per Trial	9,969	117.24	166	0	5,000
# Trials with Collab.	73,535	0.12	1	0	149
# Trials with New Collab.	73,535	0.09	1	0	95
# Trials with Established Collab.	73,535	0.04	1	0	67
Research Outcomes					
Duration Per Trial (Years)	7,643	2.70	2	0	16
Share of Trials Completed	9,969	0.56	0	0	1
Share of Trials Reporting Results Within 1 Year	9,969	0.01	0	0	1
Firm Characteristics					
0/1: Old Firm	69,118	0.39	0	0	1
0/1: Large Firm	73,535	0.19	0	0	1

Notes. This table reports firm-year level summary statistics of key research characteristics. There are 73,535 firm-year observations, representing 7,508 unique firms. Certain variables (e.g., "# Patients Per Trial") may have fewer observations since they are only reported for the firm-year observations with at least one clinical trial.

Table 2: Impact of GDPR on Trial Levels

	ihs(# Trials) (1)	# Trials (2)
A. Binary GDPR Exposure		
$\mathbf{1}\{GDPR\} \times \mathbf{1}\{EUSiteShare\}$	-0.204***	-0.692***
	(0.0147)	(0.195)
Mean of Dep. Var.	0.219	0.965
Observations	73,197	73,197
Firm FEs	Yes	Yes
Year FEs	Yes	Yes
B. Continuous GDPR Exposure		
$\mathbf{1}\{GDPR\} \times EUSiteShare$	-0.216***	-0.643***
	(0.0145)	(0.143)
Mean of Dep. Var.	0.219	0.965
Observations	73,197	73,197
Firm FEs	Yes	Yes
Year FEs	Yes	Yes

Notes. Panel A presents coefficients from firm-year regressions that regress the outcome on an indicator for after the GDPR interacted with an indicator for whether a firm has an above median share of Phase II trial sites in the EU prior to 2018. Panel B presents coefficients from firm-year regressions where an indicator for after the GDPR interacted with a continuous rather than binary measure of share of Phase II trial sites in the EU prior to 2018. In both, we include year fixed effects and firm fixed effects. 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 Trial Design

	Share with Pre-specified Plans for Sharing Individual Participant Data (1)	ihs(# Multicountry Trials) (2)	ihs(# EU Trials) (3)	ihs(# Non-EU Trials) (4)	ihs(# Diseases Per Trial) (5)	ihs(# Patients Per Trial) (6)
$\mathbf{1}\{GDPR\} \times \mathbf{1}\{EUSiteShare\}$	0.0822***	-0.0617***	-0.211***	-0.0163***	-0.0944***	0.0305
	(0.0191)	(0.0145)	(0.0141)	(0.00606)	(0.0254)	(0.0854)
Observations	8,102	73,197	73,197	73,197	8,102	8,102
R-squared	0.594	0.714	0.544	0.429	0.681	0.496
Firm FEs	Yes	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Mean of Dep. Var.	0.0606	0.196	0.153	0.0891	1.024	4.749

Notes. This table presents coefficients from firm-year regressions that regress the outcome on an indicator for after the GDPR interacted with an indicator for whether a firm has an above median share of Phase II trial sites in the EU prior to 2018. Columns 1, 5 and 6 contain fewer than 73,535 observations because their outcomes are only defined when there is a non-zero number of trials. Estimates are based on OLS regressions with year fixed effects and firm fixed effects. 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 Trial Collaboration

	Dependent variable: ihs(# Trials)		
	With Any Collaborator	With New Collaborators	With Established Collaborators
	(1)	(2)	(3)
$\mathbf{1}\{GDPR\} \times \mathbf{1}\{EUSiteShare\}$	-0.0190**	-0.0250***	0.00923*
	(0.00721)	(0.00685)	(0.00429)
Mean of Dep. Var.	0.047	0.040	0.011
Diff. Wald Test p-value		0.00	
Observations	73,197	73,535	73,535
Firm FEs	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes

Notes. This table presents coefficients from firm-year regressions that regress the outcome on an indicator for after the GDPR interacted with an indicator for whether a firm has an above median share of Phase II trial sites in the EU prior to 2018. The outcome in column 1 is the number of collaborative Phase II trials. The outcome in column 2 is the number of trials conducted by research partners with no prior Phase II collaboration or Phase II collaborations that began within two years of the focal trial. The outcome in column 3 is the number of trials conducted by other (established) research partners. Outcomes are transformed with the inverse hyperbolic sine transformation. 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. 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 Trial Outcomes

	ihs(Duration Per Trial (Years))	Share of Trials Completed	Share of Trials Reporting Results Within 1 Year
	(1)	(2)	(3)
$\mathbf{1}\{GDPR\} \times \mathbf{1}\{EUSiteShare\}$	0.111***	-0.0416*	-0.0134*
	(0.0407)	(0.0247)	(0.00694)
Mean of Dep. Var.	1.584	0.579	0.0127
Observations	6,304	8,102	8,102
Firm FEs	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes

Notes. This table presents coefficients from firm-year regressions that regress the outcome on an indicator for after the GDPR interacted with an indicator for whether a firm has an above median share of Phase II trial sites in the EU prior to 2018. Estimates are based on OLS regressions with year fixed effects and firm fixed effects. 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 Comparison of Data Privacy Regulations

The GDPR is part of a broader global trend toward strengthening data protections. This appendix compares the GDPR to two prominent U.S. regulations: the California Consumer Privacy Act (CCPA) and the Health Insurance Portability and Accountability Act (HIPAA).

California Consumer Privacy Act (CCPA)

The CCPA was passed on June 28, 2018, and became effective on January 1, 2020. Its subsequent amendment, the California Privacy Rights Act (CPRA), took effect on January 1, 2023. The CCPA grants California residents the right to know what personal data is collected about them, request deletion of their data, and opt out of the sale of their data (Jia et al., 2021; Peukert et al., 2022).

The CCPA differs from the GDPR in several important ways. First, the CCPA follows an opt-out approach: firms can set the default to allow data collection and then provide consumers an option to change this default. In contrast, the GDPR requires affirmative opt-in consent. Second, the CCPA applies only to for-profit businesses that meet at least one of the following criteria: gross annual revenue exceeding $25 million, collecting or sharing the data of more than 50,000 consumers, or deriving at least 50 percent of annual revenue from selling consumer data. These thresholds exclude most small businesses and startups (Jia et al., 2021). The GDPR, by contrast, applies broadly to any organization processing the personal data of EU residents, regardless of firm size or revenue.

Third, the CCPA's geographic scope is narrower. While it can apply to firms located outside California if they serve California residents, the GDPR's doctrine extends European privacy standards to any firm that offers goods or services to EU residents or monitors their behavior, regardless of where the firm is located (Peukert et al., 2022). Finally, the CCPA's penalties are lower than the GDPR's. CCPA penalties can reach up to $2,500 per unintentional violation and $7,500 per intentional violation. In contrast, the GDPR permits fines of up to €20 million or 4 percent of global annual revenue, whichever is greater.

Health Insurance Portability and Accountability Act (HIPAA)

HIPAA was enacted in 1996 and amended by the Health Information Technology for Economic and Clinical Health (HITECH) Act in 2009, which added breach notification requirements for covered entities (Adjerid et al., 2016; Acquisti et al., 2016). HIPAA governs the use and disclosure of protected health information by covered entities, which include healthcare providers, health plans, and healthcare clearinghouses.

HIPAA differs from the GDPR in scope, consent requirements, and enforcement. First, HIPAA is sector-specific, applying only to healthcare-related entities and their business associates. The GDPR applies broadly across all industries. Second, HIPAA permits certain disclosures of health information without explicit patient consent. For example, healthcare

providers may share protected health information for treatment, payment, and healthcare operations without obtaining patient authorization (Adjerid et al., 2016). In contrast, the GDPR requires explicit consent for processing personal data, including health data, which it classifies as a special category warranting additional protections.

Third, HIPAA's penalties are substantially lower than the GDPR's. HIPAA penalties are tiered based on the level of culpability, ranging from $100 to $50,000 per violation, with an annual maximum of $1.5 million per violation category. The GDPR's maximum fines can reach up to €20 million or 4 percent of global revenue, whichever is higher.

Fourth, enforcement under HIPAA has historically been less aggressive. The HITECH Act strengthened enforcement by requiring the Department of Health and Human Services to conduct periodic audits and by enabling state attorneys general to bring civil actions. Nonetheless, HIPAA enforcement has been described as relatively limited compared to the active enforcement regime that has emerged under the GDPR, where data protection authorities across EU member states have levied over €3 billion in fines in the five years following implementation (Demirer et al., 2026).

Summary

Table A1 summarizes the key differences across these three regulations. The GDPR is the most comprehensive along several dimensions: it applies across industries, requires affirmative consent, imposes substantial penalties, and extends to any firm serving EU residents regardless of location. These features help explain why the GDPR has attracted significant attention from firms operating internationally and why compliance costs have been substantial.

Table A1: Comparison of Data Privacy Regulations

	GDPR	**CCPA**	**HIPAA**
Effective Date	May 2018	January 2020	1996
Scope	All industries	All industries (with thresholds)	Healthcare sector
Consent Approach	Opt-in	Opt-out	Permits disclosure without consent for treatment, payment, operations
Size Thresholds	None	Revenue >$25M, >$50K users, or >$50% revenue from data sales	Applies to covered entities
Maximum Penalties	€20M or 4% global revenue	$7,500 per intentional violation	$1.5M per violation category per year
Geographic Reach	Effects-based (applies to firms serving EU residents)	Applies to firms serving CA residents	US only

Appendix B Background on Drug Development

This appendix provides additional detail on the regulatory approval processes in the European Union and United States.

Regulatory Approval in the European Union

In the European Union, the centralized procedure is managed by the European Medicines Agency (EMA) and yields a single marketing authorization valid across all EU and EEA member states (European Parliament and Council of the European Union, 2004). This procedure is mandatory for biotechnology-derived products, orphan drugs, advanced therapy medicinal products, and new active substances for treating cancer, HIV/AIDS, neurodegenerative disorders, diabetes, autoimmune diseases, and viral diseases. The EMA's Committee for Medicinal Products for Human Use (CHMP) evaluates applications within 210 active days; the European Commission then issues the final authorization within 67 days.

Drugs not required to use the centralized procedure may pursue the decentralized procedure, where drug developers can seek authorization of a drug for more than one member state at the same time (European Parliament and Council of the European Union, 2001). 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 (a drug approved in one EU country can be approved for another country) and national authorization procedures (approval is sought within a single member state for use in that country only).

Clinical trial conduct in the EU is governed by the Clinical Trials Regulation (Regulation EU No 536/2014), which harmonizes procedures across member states (European Commission, 2014). Sponsors may also conduct trials outside the EU for subsequent EU marketing authorization as long as those trials meet equivalent standards.

Regulatory Approval in the United States

In the United States, the Food and Drug Administration (FDA) governs the drug approval process (U.S. Food and Drug Administration, 2023). Before conducting clinical trials, sponsors must file an Investigational New Drug (IND) application containing preclinical data demonstrating sufficient safety for human testing. Following successful trials, sponsors submit a New Drug Application (NDA) for small-molecule drugs or a Biologics License Application (BLA) for biologics. The FDA typically reviews both application types within 10 months (or 6 months under priority review).

Appendix C Additional Descriptives

Table C1: 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	20.09	0.00***
Research Process					
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 Trials	0.04	0.54	3.10	19.26	0.00***
# Non-EU Trials	0.11	0.53	0.22	1.16	0.00***
# Diseases	1.04	0.60	1.67	1.99	0.00***
# Patients Per Trial	102.12	161.08	127.56	145.15	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***
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 C2: Pre-GDPR 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	22.20	0.00***
Research Process					
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	4.16	0.00***
# EU Trials	0.20	1.47	3.49	21.46	0.00***
# Non-EU Trials	0.06	0.33	0.40	1.32	0.00***
# Diseases	1.07	0.59	1.71	1.96	0.00***
# Patients Per Trial	105.28	140.85	131.45	191.18	0.00***
Research 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	0.06*
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 D Robustness Checks and Heterogeneity Analyses

Table D1: Impact of GDPR on Trial Levels, Firm-Disease-Year Analysis

	Dependent variable: ihs(# Trials)			
	(1)	(2)	(3)	(4)
$\mathbf{1}\{GDPR\} \times \mathbf{1}\{EUSiteShare\}$	-0.00975***	-0.00823***	-0.00823***	-0.00823***
	(0.00161)	(0.00159)	(0.00160)	(0.00160)
Mean of Dep. Var.	0.00694	0.00694	0.00694	0.00694
Observations	807,375	807,375	807,375	807,375
Firm FEs	No	Yes	Yes	Yes
Year FEs	No	No	Yes	Yes
Disease FEs	No	No	No	Yes

Notes. This table presents coefficients from a firm-disease-year level regression that regresses the outcome on an indicator for after the GDPR interacted with an indicator for whether a firm has an above median share of Phase II trial sites in the EU prior to 2018. 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 D2: Impact of GDPR: Alternative Outcome, Exposure Measures, and Samples

	Dep. Var: $\mathbf{1}\{AnyTrial\}$	Dep. Var: ln(# Trials + 1)	Dep. Var: ihs(# Trials) Exposure: Excluding UK Sites	Dep. Var: ihs(# Trials) Sample: Years before 2020	Dep. Var: ihs(# Trials) Sample: Multinational firms
	(1)	(2)	(3)	(4)	(5)
$\mathbf{1}\{GDPR\} \times \mathbf{1}\{EUSiteShare\}$	-0.111***	-0.161***	-0.233***	-0.139***	-0.212***
	(0.00700)	(0.0119)	(0.0147)	(0.0173)	(0.0289)
Mean of Dep. Var.	0.134	0.174	0.219	0.230	0.494
Observations	73,197	73,197	73,197	45,233	18,666
Firm FEs	Yes	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes	Yes

Notes. This table reports coefficients from firm-year regressions that regresses the outcome on an indicator for after the GDPR interacted with an indicator for whether a firm has an above median share of phase II trial sites in the EU prior to 2018. The outcome in column 1 is an indicator for whether there is any phase II trial conducted in the firm-year. The outcome in column 2 is the log number of phase II trials conducted in the firm-year. In Column 3, the *EUSiteShare* measure excludes UK-based sites. Column 4 restricts firm-years to those before 2020. Column 5 restricts firm-years to multinational firms. Multinational firms are defined as those whose pre-2018 share of EU trials is between 0 and 1. In all regressions, we include year fixed effects and firm fixed effects. 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 D3: Impact of GDPR on Trial Levels by Firm Age and Size

	Dependent variable: ihs(# Trials)			
	Firm Age		Firm Size	
	Younger Firm (1)	Older Firm (2)	Smaller Firm (3)	Larger Firm (4)
$\mathbf{1}\{GDPR\} \times \mathbf{1}\{EUSiteShare\}$	-0.237***	-0.173***	-0.104***	-0.0595
	(0.0206)	(0.0230)	(0.0111)	(0.0349)
Mean of Dep. Var.	0.206	0.267	0.116	0.656
Diff. Wald Test $p-value$	0.04		0.22	
Observations	42,196	26,922	59,271	14,264
Firm FEs	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes

Notes. This table presents coefficients from firm-year regressions that regress the outcome on an indicator for after the GDPR interacted with an indicator for whether a firm has an above median share of Phase II trial sites in the EU prior to 2018. Estimates are from seemingly unrelated regressions. Outcomes are transformed with the inverse hyperbolic sine transformation. 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. 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 Across Trial Phases

	ihs(# Phase I Trials) (1)	ihs(# Phase II Trials) (2)	ihs(# Phase III Trials) (3)
$\mathbf{1}\{GDPR\} \times \mathbf{1}\{EUSiteShare\}$	-0.0290***	-0.204***	-0.0162
	(0.00862)	(0.0147)	(0.0164)
Mean of Dep. Var.	0.131	0.219	0.191
Observations	73,197	73,197	73,197
Firm FEs	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes

Notes. This table reports coefficients from firm-year regressions that regress the outcome on an indicator for after the GDPR interacted with an indicator for whether a firm has an above median share of Phase II trial sites in the EU prior to 2018. In all regressions, we include year fixed effects and firm fixed effects. 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 D5: Impact of GDPR on 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)
$\mathbf{1}\{GDPR\} \times \mathbf{1}\{EUSiteShare\}$	-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 E Additional Collaboration Analyses

Table E1: Impact of GDPR on Trial Collaboration: Alternative Time-Horizon for Measuring New and Existing Collaborations

	Dependent variable: ihs(# Trials)		
	With New Collab.		With Established Collab.
	(1)		(2)
$\mathbf{1}\{GDPR\} \times \mathbf{1}\{EUSiteShare\}$	-0.0237**		0.00809*
	(0.00735)		(0.00370)
Mean of Dep. Var.	0.044		0.006
Diff. Wald Test p-value		0.00	
Observations	73,535		73,535
Firm FEs	Yes		Yes
Year FEs	Yes		Yes

Notes. This table presents coefficients from firm-year regressions that regress the outcome on an indicator for after the GDPR interacted with an indicator for whether a firm has an above median share of Phase II trial sites in the EU prior to 2018. The outcome in column 1 is the number of trials conducted by research partners with no prior collaboration or collaborations that began within five years of the focal trial. The outcome in column 2 is the number of other trials conducted by research partners. Outcomes are transformed with the inverse hyperbolic sine transformation. Estimates are from seemingly unrelated regressions. The "Wald Test p-value" is the p-value comparing the $GDPR \times HighExposure$ coefficient. Standard errors are clustered at the firm level and reported in parentheses, with *** p<0.01, ** p<0.05, * p<0.1.

Table E2: Impact of GDPR on Trial Collaboration: Alternative Phases for Measuring New and Existing Collaborations

	Dependent variable: ihs(# Trials)	
	With New Collab.	With Established Collab.
	(1)	(2)
$\mathbf{1}\{GDPR\} \times \mathbf{1}\{EUSiteShare\}$	-0.0230**	0.00619
	(0.00730)	(0.00373)
Mean of Dep. Var.	0.044	0.005
Diff. Wald Test p-value	0.00	
Observations	73,535	73,535
Firm FEs	Yes	Yes
Year FEs	Yes	Yes

Notes. This table presents coefficients from firm-year regressions that regress the outcome on an indicator for after the GDPR interacted with an indicator for whether a firm has an above median share of Phase II trial sites in the EU prior to 2018. The outcome in column 1 is the number of trials conducted by research partners with no prior Phase III collaboration or Phase III collaborations that began within three years of the focal trial. The outcome in column 2 is the number of other trials conducted by research partners. Outcomes are transformed with the inverse hyperbolic sine transformation. Estimates are from seemingly unrelated regressions. The "Wald Test p-value" is the p-value comparing the $GDPR \times HighExposure$ coefficient. Standard errors are clustered at the firm level and reported in parentheses, with *** p<0.01, ** p<0.05, * p<0.1.

Table E3: Impact of GDPR on Trial 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)
$\mathbf{1}\{GDPR\} \times \mathbf{1}\{EUSiteShare\}$	-0.0372***	-0.00213	-0.00649	0.0229
	(0.00805)	(0.0129)	(0.00446)	(0.0168)
Mean of Dep. Var.	0.041	0.061	0.025	0.139
Diff. Wald Test p-value	0.02		0.09	
Observations	42,196	26,922	59,271	14,264
Firm FEs	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes

Notes. This table presents coefficients from firm-year regressions that regress the outcome on an indicator for after the GDPR interacted with an indicator for whether a firm has an above median share of Phase II trial sites in the EU prior to 2018. Estimates are from seemingly unrelated regressions. Outcomes are transformed with the inverse hyperbolic sine transformation. 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. Standard errors are clustered at the firm level and reported in parentheses, with *** p<0.01, ** p<0.05, * p<0.1.

Table E4: Impact of GDPR on Deal-Based Research Collaboration

	Dependent variable: ihs(# Deals)	
	R&D Collab. Deals	Asset Transfer Deals
	(1)	(2)
$\mathbf{1}\{GDPR\} \times \mathbf{1}\{EUSiteShare\}$	-0.0135***	-0.0211***
	(0.00434)	(0.00461)
Mean of Dep. Var.	0.0547	0.0581
Observations	73,197	73,197
Firm FEs	Yes	Yes
Year FEs	Yes	Yes

Notes. This table presents coefficients from firm-year regressions that regress the outcome on an indicator for after the GDPR interacted with an indicator for whether a firm has an above median share of Phase II trial sites in the EU prior to 2018. The outcome in column 1 is the number of co-development deals. The outcome in column 2 is the number of asset transfer deals. Outcomes are transformed with the inverse hyperbolic sine transformation. 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 E5: Impact of GDPR on Trial Levels, by Geographic Distance of Collaborator Headquarters

	Dependent variable: ihs(# Trials)			
	Firm with EU HQ		Firm with Non-EU HQ	
	Collab. Firm with EU HQ	Collab. Firm with Non-EU HQ	Collab. Firm with EU HQ	Collab. Firm with Non-EU HQ
	(1)	(2)	(3)	(4)
$\mathbf{1}\{GDPR\} \times \mathbf{1}\{EUSiteShare\}$	-0.00301	0.00295	-0.00240	-0.00917
	(0.00457)	(0.00687)	(0.00502)	(0.00753)
Mean of Dep. Var.	0.00802	0.0180	0.00613	0.0139
Observations	21,321	21,321	51,332	51,332
Firm FEs	Yes	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes	Yes

Notes. This table presents coefficients from firm-year regressions that regress the outcome on an indicator for after the GDPR interacted with an indicator for whether a firm has an above median share of Phase II trial sites in the EU prior to 2018. Columns vary the outcome by the geographic location of the collaborator's headquarters relative to the focal firm. Outcomes are transformed with the inverse hyperbolic sine transformation. 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.

Appendix F GDPR Provisions and Empirical Findings

This appendix summarizes the mapping between specific GDPR provisions, the mechanisms through which they affect pharmaceutical R&D, and the corresponding empirical findings reported in the main text. As discussed in earlier sections, the GDPR comprises multiple distinct provisions that generate different predictions across outcomes. The geographic shift in trial activity away from the EU is consistent with Article 44's restrictions on cross-border data transfers. The narrowing of disease scope within trials is consistent with Article 5's data minimization principle. The decline in new collaborations, alongside increased reliance on established partners, is consistent with the joint controller framework under Article 26. The increase in pre-specified data sharing plans is consistent with the expanded consent and documentation requirements under Articles 6 and 9. Longer trial durations and delayed results reporting are consistent with the administrative burdens introduced by Articles 17 and 33. Table F1 provides a consolidated reference for this mapping.

Table F1: GDPR Provisions and Their Implications for Pharmaceutical R&D

GDPR Provision	**Plausible Mechanism**	**Empirical Finding**
Article 5 (Data minimization and purpose limitation)	Restricts data collection to what is necessary for the stated purpose at the time of collection, foreclosing speculative collection across multiple indications	Decline in the number of diseases studied per trial (Table 3, Column 5)
Articles 6 & 9 (Consent requirements)	Requires explicit, granular consent for each data use, increasing the cost of secondary analyses and multi-indication research	Increase in the share of trials with pre-specified data sharing plans (Table 3, Column 1)
Article 17 (Right to erasure)	Creates ongoing administrative burden when participants request deletion of personal data mid-study	Longer trial durations and lower completion rates (Table 5, Columns 1–2)
Article 26 (Joint controllers)	Requires collaborating organizations to formally allocate compliance responsibilities before sharing patient data, raising the cost of new data-sharing arrangements	Decline in new research collaborations; shift toward established partners (Table 4, Columns 2–3)
Article 33 (Breach reporting)	Creates recurring compliance obligations that slow data aggregation and release of trial results	Delays in results reporting within one year of trial completion (Table 5, Column 3)
Article 44 (International data transfers)	Requires firms to demonstrate adequate data protection standards before transferring patient data across borders, raising the cost of multinational coordination	Decline in EU trial activity and multicountry trials (Table 3, Columns 2–4)

Notes. This table maps specific GDPR provisions (European Parliament and Council of the European Union, 2016) to the mechanisms through which they affect pharmaceutical R&D and the corresponding empirical findings. Provision descriptions follow the regulatory text. Empirical findings refer to the main results tables in the paper.