Data Privacy Regulation and Innovation*

Sukhun Kang
UCSB

Jennifer Kao
UCLA

December 19, 2025

Abstract

This paper examines how data privacy regulations shape pharmaceutical firms' research and development decisions. While such regulations are designed to increase data security, they also introduce compliance costs and coordination challenges in settings where collecting, analyzing, and sharing sensitive data is central to innovation. We investigate this tension in the context of the European Union's General Data Protection Regulation (GDPR) and its impact on the medical innovation. Using detailed clinical trial data and firm-level variation in exposure to the GDPR, we find that GDPR reduces clinical trials by 22 percent, and declines persist even after accounting for shifts in disease composition. GDPR also shifts research processes: firms conduct trials across fewer sites and countries. Collaborative activity declines overall, particularly among young firms, and shifts away from new partners towards established relationships. Finally, the GDPR shifts innovation outcomes: trials take longer to complete, are less likely to complete successfully, and results are reported with a greater delay. These findings reveal how data privacy regulations shape not only the rate of medical innovation, but also the process by which it is pursued and its outcomes.

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

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

Extended Abstract

Firms increasingly rely on collecting, analyzing, and sharing data to develop novel products and services (Goldfarb and Que, 2023). However, the expansion of data-driven research and development (R&D) has raised public concerns regarding privacy protection and data security (Acquisti et al., 2016; Miller and Tucker, 2018). In response, regulatory authorities have implemented data privacy regulations designed to safeguard individual privacy rights and maintain public trust. Most notably is the European Union (EU)'s General Data Protection Regulation (GDPR), which was implemented in 2018 and is widely considered the world's most comprehensive data privacy regulation to date. For firms, data privacy regulations create a tension: such policies aimed at protecting privacy and increasing public trust can also raise the costs of collecting, analyzing, and sharing data. When data is central to innovation, this may have consequences for not only how much firms innovate, but also how research is conducted and its outcomes (Mahoney et al., 2009; Vakili and McGahan, 2016).

Recent work documents the immediate impacts of data privacy regulations on immediate performance outcomes (Blind, 2012; Aridor et al., 2020; Jia et al., 2021; Peukert et al., 2022; Johnson et al., 2023; Goldberg et al., 2024). These studies show that such regulations can lead to meaningful effects on online advertising, website traffic and revenues, and venture capital investment. However, we know little about how data privacy regulations shape firms' decisions in a setting where data is a key input: innovation. Addressing this gap is important because well-intentioned regulatory policies might unintentionally constrain firms' innovation activities with long-term consequences. This paper addresses that gap by studying how the GDPR changes pharmaceutical firms' clinical trial activity, research processes, and research outcomes.

We begin by proposing a conceptual framework in which data privacy regulations impact R&D through increased compliance and coordination costs. Data privacy regulations increase compliance costs by requiring upfront investments in legal, administrative, and data

infrastructure that are required before R&D can proceed. Higher costs disproportionately affect data-intensive innovations by raising barriers to project initiation. When compliance costs rise, projects with low expected returns become less attractive and reduce R&D investment levels.

In response to data privacy regulation, firms also adjust research processes. Firms may concentrate resources in geographic areas where the cost of compliance is lower. Data privacy regulation also affects the scale of research activity: test users are a key component of R&D costs. While data privacy regulations (which add consent, documentation, and governance requirements) increase per user costs, they also increase user trust and willingness to participate, thus minimizing costs. Further, regulation can shift collaboration incentives due to varying coordination costs. The effects are ambiguous: while regulations discourage collaboration due to increased coordination challenges, they encourage firms to collaborate to share compliance expertise and infrastructure. This is likely to be particularly salient among established partners whose data practices are familiar.

Finally, our conceptual framework highlights that these adjustments carry consequences for research outcomes. Compliance procedures add administrative burden at each stage of data collection, extending the time required to complete trials. Further, if firms narrow geographic scope to avoid regulatory exposure, they may struggle to enroll sufficient participants, increasing the likelihood that trials fail to reach completion. We therefore expect GDPR exposure to lengthen trial duration and reduce completion rates.

To empirically test this framework, we examine the causal impact of the EU's GDPR on firms' R&D decisions in the pharmaceutical industry. The GDPR harmonizes data protection standards across EU Member States while granting individuals control over their personal information. In the pharmaceutical industry, data is not only a critical input to R&D but is also particularly sensitive and tightly regulated (Vakili and McGahan, 2016; Buckman et al., 2023). Unlike many consumer data settings, the data required for pharmaceutical research is produced through costly, regulated experimentation and cannot be readily substituted when

access and governance become more difficult. Further, because pharmaceutical development and trial operations are global, the GDPR is also relevant beyond Europe: it can affect where firms run trials, how they coordinate across organizations, and how quickly evidence becomes available. These central concerns for assessing the functioning of medical innovation system.

We utilize a comprehensive dataset of clinical trials initiated between 2010 and 2023, with sites spanning the EU, US, and other countries (Table 1). We then leverage a difference-in-differences design that exploits cross-firm variation in exposure to GDPR based on firms' pre-regulation geographical distribution of clinical trial activity (Figure 1). Firms with extensive EU-based clinical trial activity prior to GDPR implementation faced greater regulatory exposure than those with minimal EU activity, providing natural variation in exposure to the regulation.

Our analysis yields three main findings. First, we find that firms more exposed to GDPR reduced their number of clinical trials by 22 percent relative to less exposed firms (Table 2). This effect is immediate, persistent, and holds even after accounting for shifts in disease composition (Figure 2). In examining heterogeneity across disease areas, we find that the GDPR led to declines across nearly all disease groups. Cancer is a notable exception, which is consistent that oncology trials may be considered high priority and have substantial investment as they are supported by large, specialized clinical networks (e.g., the National Cancer Institute's Cooperative Groups Program) and strong financial incentives.

Second, firms adjust their research processes. They increase the share of trials with pre-specified plans for sharing individual participant data and narrow the scope of their research activity: trials are less likely to be conducted in multiple countries and to involve an EU site (Table 3). We also observe meaningful shifts in collaboration (Tables 4 and 5): overall collaborative activity declines, with effects driven by young firms. Further, firms significantly reduce collaborations with new partners while modestly increasing research activity with established collaborators. Finally, we show that the GDPR shapes research outcomes: trials

take longer to complete, are less likely to reach successful completion, and firms delay the reporting of trial results (Table 6).

This paper contributes to several literatures. First, we contribute to the emerging literature on data privacy regulation by showing how data-related policies shape firms' R&D decisions. Prior research has primarily examined the immediate performance outcomes of data privacy regulations on digital services and consumer-facing firms (e.g., Blind, 2012; Aridor et al., 2020; Jia et al., 2021; Johnson et al., 2023; Goldberg et al., 2024; Goldfarb and Tucker, 2011). We show that regulations targeting data, a key input in the innovation process, impact not only the level of research, but also how research is conducted and its outcomes. This contributes to a growing recognition that data privacy regulation differs from other forms of regulation by directly targeting key inputs essential for R&D (Goldfarb and Tucker, 2012; Acquisti et al., 2016; Aridor et al., 2020; Jia et al., 2021). Recent work on the pharmaceutical industry shows how intellectual property (Sampat and Williams, 2019) and product-specific policies (Yin, 2008; Chandra et al., 2024) can shape pharmaceutical R&D investments, but has not yet examined how data-specific regulations impact decisions throughout the R&D process. By documenting a 22% decline in clinical trial along with changes in research processes and outcomes, we demonstrate that data privacy regulations may act as an important constraint on a key input for innovation with meaningful consequences for the level, type, and outcomes of research pursued.

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

concentrated under stringent regulatory regimes.

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

References

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

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

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

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

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

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

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.

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

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

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

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

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

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

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

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

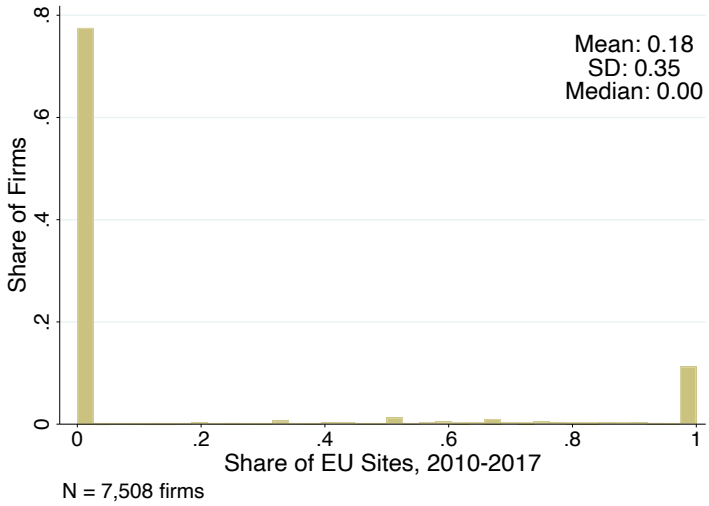
Sampat, B. and Williams, H. L. (2019). How do patents affect follow-on innovation? Evidence from the human genome. *American Economic Review*, 109(1):203–236.

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.

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

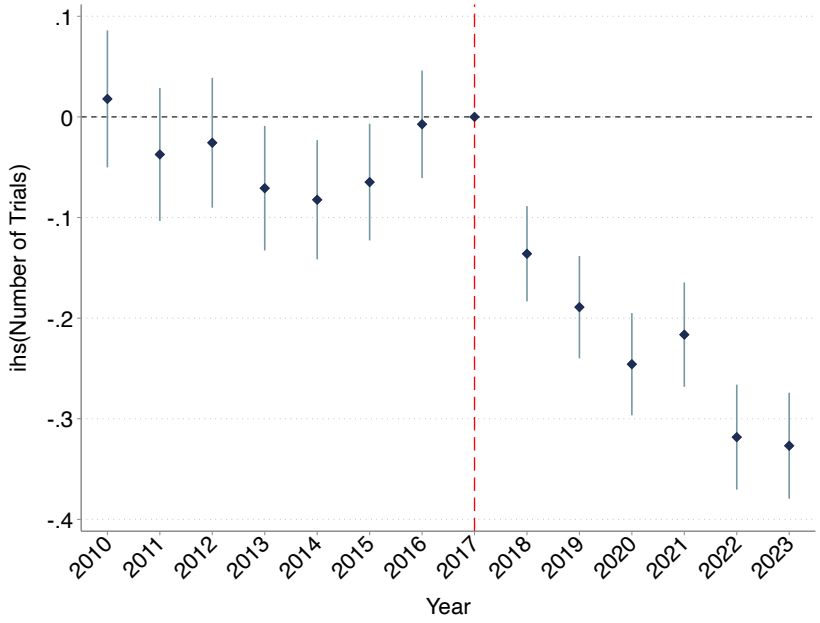
Main Figures and Tables

Figure 1: Firm-level Exposure to GDPR



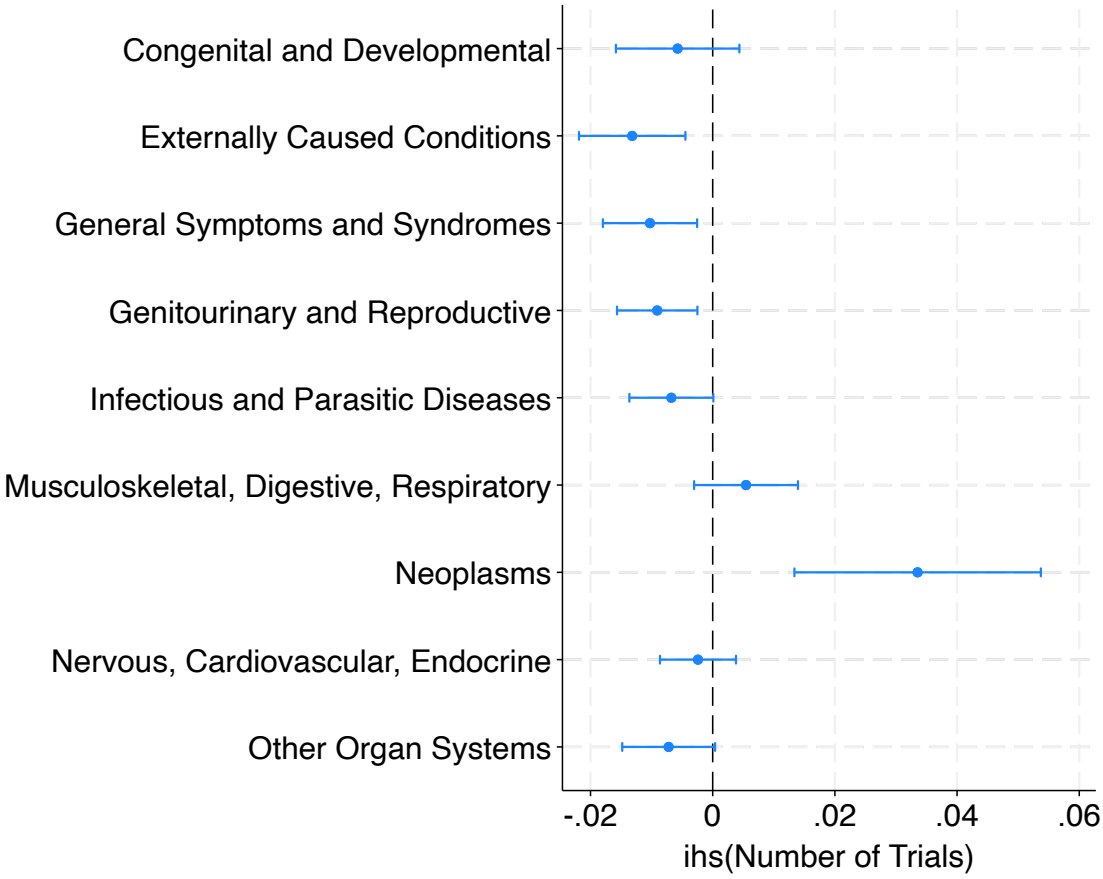
Notes. This figure shows the distribution of firms' pre-2018 EU trial site share.

Figure 2: Impact of GDPR on Research Levels



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

Figure 3: Differences in the Effect of GDPR Across Diseases Groups



Notes. This figure shows the impact of GDPR across for different disease groups: (1) Blood, Immune, and Metabolic; (2) Congenital and Developmental; (3) Externally Caused Conditions; (4) General Symptoms and Syndromes; (5) Genitourinary and Reproductive; (6) Infectious and Parasitic; (7) Musculoskeletal, Digestive, and Respiratory; (8) Neoplasms; (9) Nervous, Cardiovascular, and Endocrine; and (10) Other Organ Systems. The omitted group in the figure is "Blood, Immune, and Metabolic." The outcome is the inverse hyperbolic sine transformed number of trials in a year.

Table 1: Firm-Year Level Summary Statistics

	Count (1)	Mean (2)	SD (3)	Min (4)	Max (5)
Research Levels					
Trials	73,591	0.96	10	0	559
Research Process					
Share of Trials with Pre-specified Plans for Sharing Participant Data	9,973	0.05	0	0	1
# Multicountry Trials	73,591	0.36	2	0	97
# EU Sites Per Trial	9,973	2.93	5	0	27
# Non-EU Sites Per Trial	9,973	2.27	4	0	39
# Patients Per Trial	9,973	117.22	166	0	5,000
# Trials with Collab.	73,591	0.12	1	0	149
# Trials with New Collab.	73,591	0.09	1	0	95
# Trials with Established Collab.	73,591	0.04	1	0	67
# Collab.	73,591	0.18	1	0	87
# New Collab.	73,591	0.09	1	0	35
# Established Collab.	73,591	0.10	1	0	77
Research Outcomes					
Duration Per Trial (Years)	7,647	2.70	2	0	16
Share of Trials Completed	9,973	0.56	0	0	1
Share of Trials Reporting Results Within 1 Year	9,973	0.01	0	0	1
Other Trial Characteristics					
Firm Characteristics					
Pre-2018 EU Trial Site Share	73,591	0.23	0	0	1
0/1: Old Firm	69,174	0.39	0	0	1
0/1: Large Firm	73,591	0.19	0	0	1

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

Table 2: Impact of GDPR on Research Levels

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

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

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

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

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

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

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

Notes. This table reports DID estimates of the effect of GDPR on the number of phase II trials with collaborators. The level of observation is the firm-year. Columns 1 and 4 report OLS estimates. Columns 2, 3, 5, and 6 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 and columns 5 vs. 6. Outcomes are transformed with the inverse hyperbolic sine transformation. Standard errors are clustered at the firm level and reported in parentheses, with *** p<0.01, ** p<0.05, * p<0.1.

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

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

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

Table 6: Impact of GDPR on Research Outcomes

	ihs(Duration Per Trial (Years))	Share of Trials Completed	Share of Trials Reporting Results Within 1 Year
	(1)	(2)	(3)
GDPR × High Exposure	0.111***	-0.0422*	-0.0133*
	(0.0407)	(0.0246)	(0.00693)
Observations	6,308	8,106	8,106
R-squared	0.793	0.595	0.325
Firm FEs	Yes	Yes	Yes
Year FEs	Yes	Yes	Yes
Mean of Dep. Var.	1.584	0.579	0.0127

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