Market Size, Product Recalls, and Pharmaceutical Innovation

Federico Nutarelli*[1] and Massimo Riccaboni†[2]

[1]AXES Group, IMT School for Advanced Studies
[2]IUSS Pavia, Pavia, Italy

Abstract

We examine the causal effect of market size on pharmaceutical innovation in a context where observed revenues are endogenous equilibrium outcomes jointly determined by demand, competition, and innovation. To address this endogeneity, we use major drug recalls as a novel instrument for expected market rewards, relying on their sharp and unanticipated nature at the therapeutic-class level. Combining U.S. sales, recall, and clinical trial data at the relevant submarket level, we estimate the effect of market size on innovation within a control-function IV Poisson framework that addresses both idiosyncratic and heterogeneity-driven endogeneity. The identifying variation comes from recall-driven contractions in class-level revenues that persist beyond the recall year. We find that a 10% increase in market size leads to approximately a 6% increase in clinical trial activity, an elasticity at the upper range of existing estimates. Importantly, when innovation is restricted to trials that result in marketed products, the estimated elasticity falls substantially. This finding suggests that standard measures of innovation effort overstate the responsiveness of commercially successful innovation to market-size incentives and underscores the importance of distinguishing between innovation effort and realized outcomes for both empirical measurement and policy design.

Keywords: Product recalls; Innovation; Empirical Industrial organisation; Pharmaceutical industry.

JEL Codes: O31, J10, J20, I11, L11

Acknowledgments: The authors are grateful to the Crisis Lab project for the PHID database. We thank Prof. Jeff Wooldridge for his kind support in applying and interpreting his innovative methodology, and Prof. Eric von Hippel for his insights. We also thank the FDA for providing data on recalls and for clarifying any questions about recall timing and procedures. We are grateful to Evaluate for providing up-to-date and detailed checks of clinical trial data. Furthermore, we thank Springer AdInsight for enabling detailed classification at the ATC-3 level of rare clinical trials. Finally, we thank the Young Economists of Tuscan Institutions (YETI), the AXES unit of the IMT Lucca for Advanced Studies, Armando Rungi, Francesco Serti, Laura Magazzini, participants to the ASSA 2022, DRUID 2022, Unicatt 2023, International Panel Data Conference 2025, and Research Policy Online Conference for Early Career Researchers participants for their valuable comments on the work.

*Corresponding author: federico.nutarelli@imtlucca.it.
†massimo.riccaboni@imtlucca.it

1 Introduction

That larger markets induce more innovation is a cornerstone of innovation economics (Scherer, 1982; Schmookler, 2013; Klepper and Malerba, 2010; Ayerst, 2025), yet estimating this relationship causally remains elusive when market size and innovation are jointly determined equilibrium outcomes. This identification challenge has concrete policy stakes: the premise that larger expected rewards induce more research activity underlies the design of public demand incentives such as orphan drug legislation (Yin, 2008), making it important to establish whether and how strongly market size drives innovation. Schmookler's demand-pull hypothesis has been challenged on precisely these grounds. As early as the 1990s, Kleinknecht and Verspagen (1990) noted that the direction of causality between market size and innovation was far from obvious, suspecting a simultaneous relationship that their methods could not address. More recently, Stoneman (2010) and Ball et al. (2018b) have developed more rigorous approaches to detect this endogeneity, yet a credible causal estimate remains elusive

At the market level, Acemoglu and Linn (2004) proposed a strategy to overcome endogeneity bias by exploiting changes in market size across drug categories driven by demographic trends. After Acemoglu and Linn (2004), the literature shifted from establishing the existence of reverse causality between market size and innovation to identifying the optimal identification strategy. The instrument used in Acemoglu and Linn (2004) was later criticized by Cerda (2007) as being endogenous. As detailed in Cerda (2007), pharmaceutical innovation may increase patient age, while an increase in the average age of patients raises demand for innovative products. Demographic trends may therefore affect market size, which in turn affects innovation, while innovation also influences demographic trends. To the best of our knowledge, the identification strategy for assessing the impact of market size on innovation remains an open issue.

Inspired by Acemoglu and Linn (2004), most authors have examined the relationship between market size and innovation in the pharmaceutical industry (Janssen, 2023), which offers an ideal setting for this question. In this sector, consumer needs are diverse and remain almost stable over time, so the market can be divided into relatively independent submarkets based on medical needs (Sutton, 2001; Bertoni et al., 2010). Moreover, investment in innovation is essential to the industry's existence, and innovation is relatively easy to measure. In the pharmaceutical industry, relevant submarkets are well-defined using the Anatomical Therapeutic Chemical (ATC) classification system, which identifies submarkets based on *"the active ingredients of drugs according to the organ or system on which they act and their therapeutic, pharmacological, and chemical properties"* (WHO, 2009). Our contribution also considers the pharmaceutical industry as an ideal setting where we provide a novel identification strategy to estimate the causal effect of market size on innovation by exploiting product recalls as plausibly exogenous shocks to expected market rewards.

Therefore, our first contribution is methodological. We employ product recalls as an instrument to address the endogeneity of market size, an issue often overlooked in previous studies. Unlike the traditional Poisson approach, we introduce a Control Function (CF) IV strategy that leverages the unpredictability and abrupt nature of product recalls. The main improvement is using recalls as an instrument for market size, directly addressing the endogeneity challenge. This approach relies on the exogeneity of recalls, assuming that firms cannot anticipate competitors' recalls at the market level. The sudden and unexpected nature of major recalls, which are not necessarily more common in "riskier" ATC markets, offers a unique perspective. As detailed in Section 4, our arguments support the validity of treating major recalls as sharp and unanticipated, thereby strengthening the robustness of our methodology.

Our second contribution concerns the measurement of innovation. We measure innovation using the total number of active trials at the ATC-3 level, rather than traditional indicators such as cumulative R&D expenditures or the count of New Molecular Entities (NMEs). R&D expenditures are an input-based proxy for innovation and, in the pharmaceutical industry, they are known to be noisy and strongly shaped by accounting conventions and firms' financial and strategic choices. In particular, aggregate R&D at the market level tends to scale mechanically with the size and composition of the firms operating in that market,[1] so that using cumulative R&D spending as a proxy for innovation risks capturing market size itself rather than innovation effort. More importantly, R&D decisions are inherently forward-looking: they respond to current cash flows and to expectations about future market rewards. This intertemporal decision structure may induce

[1] For instance, Stoneman (2010) and others have posited that smaller, newer market entrants may be more disposed toward R&D investment than their larger, established counterparts. Consequently, defining innovation merely in terms of a market's cumulative R&D spending could inadvertently reflect the market's size, as gauged by the aggregate sales of the participating firms.

persistence in unobservables and dynamic endogeneity, that is, shocks affecting investment at time $t-1$ can remain correlated with investment at time t through omitted expectations and financing conditions. To mitigate these concerns, we measure innovation through the number of active (pre-clinical and clinical) trials at the ATC-3 level, which captures real experimentation activity, including failures, rather than ex post commercial success. Unlike expenditure-based measures, counts of active trials are less mechanically tied to contemporaneous sales and are less affected by accounting practices. In addition, our robustness checks explicitly address potential temporal dependence by including lagged innovation measures and testing for serial correlation in the error structure.

The choice not to rely on NMEs is also important. As in Acemoglu and Linn (2004), NMEs are innovative products containing active moieties not previously approved by the FDA. Despite being comprehensive, NMEs do not fully capture firms' intentions to innovate because they are observed only at the end of a long and uncertain development and regulatory process. Pharmaceutical drug approval begins with a pre-clinical phase, followed by clinical trials.[2] Thus, NMEs are recorded only when products successfully complete trials and receive FDA approval (NDA classification). Because approval occurs several years after early research and clinical development, the timing of the effect of market size on NMEs is inherently lagged and potentially heterogeneous across therapeutic areas (Sertkaya et al., 2024), implying that NME-based specifications would require strong assumptions on the lag structure that are difficult to discipline in our ATC-3-by-year panel specification. Accordingly, our metric measures innovation upstream by counting active (pre-clinical and clinical) trials, thereby capturing both successful and unsuccessful projects and providing a timelier measure of innovation effort. This approach recognizes even unsuccessful trials as indicative of firms' commitment to innovation.[3]

A further contribution is the use of a more refined segmentation of market categories, namely ATC-3 classifications, rather than the broader classifications commonly used in the literature. Data categorized at the ATC-3 level more precisely reflect the structure of submarkets, which are often synthetically constructed or overlooked in the scholarly discourse. Traditionally, the literature has focused predominantly on ATC-1 and ATC-2 levels, with ATC-3 serving mainly as a basis for comparison with these broader classifications. Importantly, adopting the ATC-3 level in our analysis is motivated not only by its finer granularity but also by its alignment with the categorizations used by antitrust agencies. This alignment underscores the relevance of ATC-3 for deriving policy implications and offers a framework that mirrors the operational parameters considered in regulatory assessments.[4]

Finally, we address concerns in the recent literature that market size effects may be overestimated when innovation is measured by outcomes that do not necessarily result in innovative products. Previous studies often used NMEs as dependent variables, but these metrics may not fully capture the timing and scope of market-driven innovation incentives because they focus only on long-term successful outcomes and ignore earlier-stage efforts and failures, which are also shaped by market size. In one of our preferred specifications, we examine only those trials that led to marketed innovations and find evidence of overestimation in previous estimates of the elasticity of market size to innovation. The importance of including impact-based innovations when measuring technological change has been highlighted in several works (see, among others, Arrow, 1962; Cohen and Levinthal, 1989; Mowery and Rosenberg, 1999). These contributions advocate greater attention to empirical studies that trace innovation from conception to commercialization and measure the economic impact of innovations that culminate in successfully marketed products, thereby shaping industry

[2] In Phase 1, the company performs clinical trials on healthy volunteers to establish the drug's fundamental characteristics and safety profile in humans. According to DiMasi et al. (2003), *"typically, the drug remains in this stage for one to two years"*. Once the drug is dispensed to volunteers drawn from the targeted population, Phase 2 begins, and Phase 3 subsequently compares the new drug against a standard-of-care drug, thereby establishing efficacy and safety.

[3] The length of the approval process is also an argument in favor of the exogeneity of recalls. In this analysis, we work at the ATC level and argue that major recalls occur randomly, in the sense that it is impossible to predict in advance whether a significant issue will arise with a given drug. This randomness is underscored by the fact that major recalls typically occur post-marketing; if the regulator had had prior suspicions, the drug would not have been approved in the first place. Acemoglu and Linn (2004), whose theoretical framework informs our study, approach the issue from a firm-level perspective. At the firm level, exogeneity of major recalls may not hold. While it may be valid if the recall occurs at time t or $t+1$, it is conceivable that if a recall occurs at $t-1$, firms might reconsider innovating in that sector. In essence, they may delay innovation due to the recall. One possible resolution to this issue lies in the fact that pharmaceutical innovation is a lengthy process, often taking many years. Given that major recalls are quite rare at the ATC level, it would likely take a significant number of recalls to induce a firm to halt ongoing clinical trials, especially given the substantial sunk costs associated with such trials. Subsequent analyses validate this argument.

[4] For more details, see Section 3.

standards and driving economic growth.

Our results show that market size has a positive and economically significant effect on pharmaceutical innovation. In our preferred specification, we estimate an elasticity of clinical trial activity with respect to market size of about 0.63. Since our IV design exploits recall-driven contractions in class-level revenues, this estimate is identified from negative shocks to expected market rewards. This estimate aligns with the upper range of the existing literature, but it depends on how innovation is measured. When we focus only on trials that lead to marketed products, the estimated elasticity is much lower, suggesting that broader measures of innovation may overstate the responsiveness of commercially successful innovation to market size. This finding underscores the importance of distinguishing between innovation effort and realized innovation outcomes. We also find that major product recalls cause persistent contractions in class-level revenues, supporting their validity as an instrument for market size and reinforcing the role of expected market rewards in shaping firms' innovation decisions.

The remainder of the paper is structured as follows. Section 2 reviews the related literature. Section 3 presents the data and the underlying theoretical framework. Section 4 describes the empirical strategy and identification. Section 5 reports the main results and a series of robustness checks. Section 6 concludes.

2 Literature Review

The importance of market size in explaining the rate of innovation has long been recognized in the literature. In 1942, Schumpeter indicated that larger firms are more innovative than smaller ones. In the early 1960s, the focus shifted to a broader consideration of the possible effects of demand on market size (see, e.g., Scherer (1982)). At that time, it was unclear whether reverse causality between demand and innovation was at play. Scherer (1982), for instance, argued that causality ran primarily from sales to innovation. However, this approach has been criticized in several respects. The definition of demand remained too broad and did not conclusively determine the unique direction of the relationship between demand and innovation; that is, reverse causality (see, e.g., Mowery and Rosenberg, 1979). At that time, the research did not focus specifically on selected R&D intensive sectors nor did it consider the aggregate market level (Pakes and Schankerman, 1984). More recently, Kleinknecht and Verspagen (1990) found clear reverse causality between demand and innovation, thereby invalidating prior studies. Soon after, Geroski and Walters (1995) empirically verified these conclusions and showed that innovations increase market size by creating their own demand.

It was clear that heterogeneous shifts of demand played a prominent role in determining technological development (see, e.g., Malerba (2007)). Between 1980 and 1990, and most recently in 2002, several studies showed, for instance, how innovation reacted elastically to energy prices. Nowadays, much of the research exploring the relationship between market size and innovation is concentrated in the pharmaceutical industry, where innovation is a key driver. The literature primarily considers two levels of aggregation: the firm level and the market level. Past research has aimed to determine the impact of firm size on R&D investments and outputs. However, this issue remains a topic of ongoing debate (refer to Mellahi and Wilkinson (2010), Kolluru and Mukhopadhaya (2017), among others). Specifically, the discussion has produced conflicting results, largely due to challenges in fully eliminating unobservable sources of endogeneity that change over time. These unobservable factors may result from strategic decisions within firms, which could be influenced by firm size. For instance, smaller pharmaceutical firms may opt for riskier decisions compared to their larger, more established counterparts (Hall and Rosenberg, 2010). Shifting focus to market-level aggregation effectively addresses these issues. Unobservables related to market size can generally be viewed as inherent characteristics of the markets and, therefore, constant over time. Consequently, fixed effect techniques enable researchers to account for unobservable heterogeneity, thereby eliminating the specific endogeneity associated with market size. As a result, the market level has appeared to be a more appropriate level of analysis, leading many researchers to adopt this broader level of aggregation.

The body of literature concerning the pharmaceutical sector is extensive, and its diversity can be attributed to the various measures of innovation that have been employed. Some researchers have relied on accounting data, with a particular focus on R&D spending. While R&D data is reliable in the context of perfect capital markets, its conclusions become less clear in imperfect markets where current investment decisions are influenced by expectations of future choices. In the presence of market imperfections, current revenues (which indicate market size) serve as a practical proxy

for estimating future market sizes. Initially, this approach may introduce endogeneity issues due to the correlation between current revenues and unobserved factors, such as the firm management's propensity for risk. Moreover, given that present R&D may be responding both to present and future sales opportunities, the coefficient of market size might incorporate two effects that are difficult to separate. In light of this issue, scholars have included lagged proxies of the market size (see, e.g., Giaccotto et al., 2005, , who estimated that a 1% increase in price leads to a 0.58% increase in R&D spending). Other problems related to R&D measure are reported in Hall and Rosenberg (2010). To give an idea, several authors (see e.g. Pammolli et al., 2011, among others) showed that although pharmaceutical firms made substantial investments in R&D, they did not produce innovation[5]. Other measures of innovation include clinical trials (Kyle and McGahan, 2012) and changes in Medicare part D. To the best of our knowledge, however, among innovation measures, no work has used Investigational New Drugs (INDs) and early-stage clinical trials together with late-stage trials. The use of Medicare part-D as an innovation measure might affect both present and future market size (Blume-Kohout and Sood, 2013; Dubois et al., 2015). As noticed in Blume-Kohout and Sood (2013): *"Medicare Part D could have affected firms' R&D expenditures both to its expansion of expected future markets for products still in pipeline, and also via two supply side mechanisms."*[6]. Therefore, Medicare Part D may have increased not only current sales but also, importantly, future sales through R&D expenditures. Researchers have observed that innovation responds positively to changes in market size. However, a persistent issue has been the simultaneous response of innovation to both immediate and anticipated cash flows resulting from changes in market size. As a result, it has been difficult to distinguish between the immediate and long-term effects within the estimated coefficient for market size. Furthermore, as highlighted in Section 1, it is important to note that changes in Medicare Part D could be subject to endogeneity due to lobbying activities by pharmaceutical firms.

In this work, we state that our measure of innovation did not influence current or future sales, as it includes both successful and unsuccessful innovation-generating projects across all clinical and pre-clinical phases. This methodological improvement addresses the endogeneity issue by focusing on the directness of innovation efforts rather than their market success (output). The inclusion of both successful and unsuccessful trials as a proxy for innovation captures the entire spectrum of R&D activities, providing a more accurate and nuanced view of innovation dynamics. Crucially, this approach decouples the analysis from the feedback loop where market success potentially distorts future R&D expenditures. By reflecting on actual innovation efforts, our measure offers a clearer, albeit not flawless, insight into how market size influences innovation, unmarred by the biases tied to market outcomes. This direct focus on R&D activities, rather than the indirect implications of market success, enriches our understanding of the innovation process, enhancing the reliability and depth of our analysis.

To better account for the possibility that the size of future markets could be influenced by both ongoing and unsuccessful projects, we addressed this potential bias in our robustness checks. We did this by recognizing the impact of active and failed projects on future market size and including lagged innovation as a variable. This approach clearly outlines the causal relationships between market size and innovation, ensuring that previous innovation efforts are not confused with future research and development investment decisions.

Innovation has also been quantified by the number of relevant medical journal articles (Lichtenberg, 2006). However, these measures do not account for innovations that occur exclusively within industries and focus mainly on academic and scientific discoveries, thereby suffering from selection bias.

Additional metrics include the number of new drugs introduced, covering both generic and branded drugs (Acemoglu and Linn, 2004; Dubois et al., 2015), categorized as New Chemical Entities (NCEs), New Molecular Entities (NMEs) or new medicines approved by the FDA. Active trials offer a direct measure that contrasts with NCEs and NMEs. These are indeed lagging indicators of innovation, representing the culmination of years of R&D. Furthermore, by including unsuccessful trials, our measure recognizes that innovation is not only about successful outcomes but also about the process and learning from failures. This approach incorporates the concept of learning by doing,

[5]Thus undermining the validity of R&D investment as a proxy for innovation.

[6]Blume-Kohout and Sood (2013) leverages the increase in prescription drug usage among seniors enrolled in Medicare Part D as a natural experiment to assess the impact of market size on innovation. However, this approach is frequently criticized for its vulnerability to endogeneity issues, primarily due to the lobbying efforts by pharmaceutical firms to include their new drugs in Medicare Part D (Drotleff, 2006; Heath, 2013), and more recently (Hwang et al., 2022). Such potential biases complicate the ability to draw clear causal conclusions from the observed changes in innovation and market dynamics.

as observed in several studies in the industrial organization literature (e.g., Arrow, 1962; Hall and Rosenberg, 2010).

Likewise, various indicators of market size have been utilized in the literature. Acemoglu and Linn (2004) made the first significant contribution to the relation between market size and innovation. Their idea relies on adopting demographic shifts to instrument market size, while controlling for observables or unobservables arising from reverse causality. In particular, Acemoglu and Linn (2004) utilize variations in the expenditure shares of different U.S. age cohorts across various therapeutic classes from 1970 to 2000. They find that a 1% increase in expenditure shares results in a 4% increase in the number of new medicines. This demonstrates a significantly higher elasticity than the average elasticity reported in the broader literature (Dubois et al., 2015).

Cerda (2007) provided further insights on the results found in Acemoglu and Linn (2004). Employing US demographic data, in their critical analysis, Cerda (2007) underscored the significance of feedback effects previously overlooked by Acemoglu and Linn (2004). Specifically, Cerda (2007) highlighted that the introduction of new drugs could fundamentally alter market dynamics by affecting mortality rates. The advent of innovative medicines, with their potential to cure a broader range of diseases, invariably leads to an increase in the average age of the population. This demographic shift results in a higher proportion of older individuals who are more likely to require these new treatments, thereby adjusting demand patterns. This observation crucially brings to the forefront the issue of reverse causality, emphasizing the intricate interplay between market forces and pharmaceutical innovations.

Recent literature has predominantly focused on enhancing methodological approaches within the field, as evidenced by works such as Lichtenberg (2006); Civan and Maloney (2009); Dubois et al. (2015); Rake (2017), among others. This emphasis on methodology contrasts with the fundamental critique presented by Cerda (2007). Notably, Dubois et al. (2015) introduced a novel methodology by utilizing global pharmaceutical data at the ATC-1 and 2 levels, yet they persisted in using demographic shifts as an instrument for assessing market size.

It is worth mentioning also the study of Civan and Maloney (2009) where authors discovered that "the higher the prices of existing drugs in a therapeutic category, the larger the number of drugs in the development pipeline for that category", which underscores the dynamic interaction between drug pricing and pharmaceutical innovation. Importantly, Civan and Maloney (2009)'s research is primarily centered on estimating the elasticity of drug development in response to market prices of drugs, delving deeper into the economic mechanisms that drive pharmaceutical advancements. Their methodology involves using R&D as a measure of innovation, although their results are challenged by the endogeneity of prices in the elasticity equations utilized.

Furthermore, Rake (2017) employs a unique database alongside a Poisson Quasi Maximum Likelihood approach to derive his findings. His work notably utilizes NMEs and New Drug Approvals (NDA) as metrics of innovation, contributing valuable insights into the factors that influence pharmaceutical development.

Rake (2017) adopted several measures of innovation from NCE to clinical trials in Phase II and Phase III, finding no evidence of reverse causality when adopting NCE. One of his efforts was to account for changes in the industry's R&D process, from "random screening" to "guided drug development". He modeled technological opportunities and inserted them as regressors in the analysis, finding a positive relationship with Phases II and III trials. His results are in line with Cerda (2007) and Acemoglu and Linn (2004). Dubois et al. (2015) used NCE to measure innovation and defined market size as a measure of expected revenue. Their dataset included information about sales for 14 different countries. Specifically, Dubois et al. (2015) measured market size as the total revenue over the entire life cycle of a branded drug. Dubois et al. (2015) performed a control function approach and recovered an estimate of the relationship between market size and innovation for each therapeutic class at level 1. The average elasticity of innovation to market size reported by Dubois et al. (2015) was about 23%, which is relatively lower than typically expected averages. A potential explanation for this lower elasticity—aside from the use of an IV approach—can be found in Blume-Kohout and Sood (2013). This study notes that many of the countries analyzed in Dubois et al. (2015) regulate prescription drug prices and that such regulations can change rapidly. Consequently, given the lower expected profit per consumer and greater uncertainty regarding future profits and prices, a firm's R&D decisions in these countries are likely to be less responsive to a unit change in expected revenues compared to the same unit change in the US market.

Table Appx.1 in the Appendix provides a schematic literature review of previous estimates of the

relationship between innovation and market size[7]

The existing literature generally indicates that a 1% increase in market size results in a 0.4% to 0.7% increase in innovation. Our study aligns with these findings under our favored specification. However, a significant deviation emerges when we focus exclusively on trials that result in a marketed innovation. In this more specific context, we observed a notably lower estimate, which closely aligns with the findings of (Dubois et al., 2015). Our analysis, additionally, distinctly quantifies, the variance in elasticity of market size between successful and unsuccessful trials when market size is accurately instrumented. Furthermore, while previous research has typically focused on disease levels or, at most, at the ATC-1 or ATC-2 levels of aggregation, as noted by Dubois et al. (2015)), our study ventures into the less explored ATC-3 level. According to our analysis, there are several advantages to utilizing drug classes over disease classes in the context of pharmaceutical market studies. Primarily, since pharmaceutical firms are the entities that initiate requests for new clinical trials (NCTs) focusing excessively on the demand side might overlook critical supply-side dynamics. These dynamics are pivotal as they drive firms to pursue NCTs or NDAs. Specifically, the aggregate sales of drug classes are more indicative of supply-side dynamics, whereas sales based on disease classes—aggregated sales of products purchased by patients—tend to reflect demand-side dynamics more closely. Furthermore, while firms may respond to demand-side incentives to initiate an NDA or an NCT, they primarily consider their competitors within the same ATC class. This is particularly true in commercial trials sponsored by the pharmaceutical industry, rather than those related to academic or research institutions. Therefore, aggregating sales into disease classes groups them into a demand-driven category, which provides less insight into the supply-side dynamics, such as competitor behavior, that influence a firm's decision to undertake a clinical trial. Consequently, when estimating the impact of market size categorized into disease classes (demand-driven) versus innovation (supply-driven), there is a risk of underestimating the influence of supply-side factors. This underestimation can skew the understanding of the market forces at play, thereby impacting strategic decisions and policy formulations within the pharmaceutical industry.

Moreover, by categorizing by disease classes, the definition of innovation encompasses various chemical and therapeutic drug types, ranging from topical to systemic applications and from vaccines to ointments. This lack of differentiation could result in endogeneity through multiple channels, such as public expectations. For example, patients might distrust certain drugs, influencing the market size potential of the drug type in question. Other sources of endogeneity could include potential correlations between regressors and the error term (which incorporates "drug-type"). Regulations, for instance, might be specific to the product type (e.g., WHO regulations for vaccines do not apply to other drug types). Control variables such as knowledge stocks could also be problematic, potentially varying with the drug type. Furthermore, knowledge stocks may increase when new medications are developed in categories previously dominated by a single type of medicine. A pertinent example is in dermatology, where research focuses on adapting topical treatments for systemic use due to the adverse side effects of some systemic medications. Finally, the duration of clinical trials may vary with the type of medicine being tested, which could result in delayed effects on market size if disease class is used as a categorization method.

3 Data

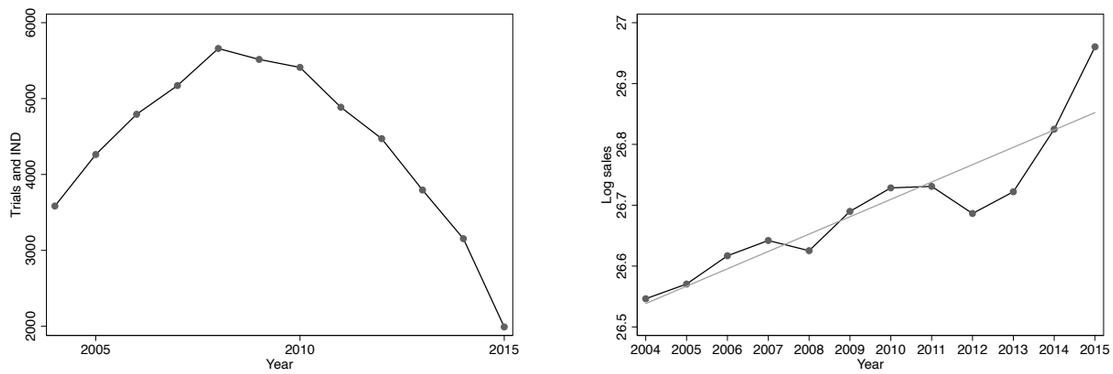
This study combines four main data sources. Sales data are drawn from the Evaluate Pharma dataset, which provides U.S. pharmaceutical revenues at the product and molecule levels from 2004 to 2015, subsequently aggregated to the ATC-3 level. Data on clinical trials and Investigational New Drugs (INDs) for the same period are sourced from ClinicalTrials.gov and a Pharmaceutical Industry Database (PHID), respectively.[8] Recall data are manually collected from multiple sources, including the FDA website, openFDA, Onakpoya et al. (2016), the WHOCC website, PubMed, and Siramshetty et al. (2016). Control variables are derived from Evaluate, PHID, and the FDA's website.

[7]The latter also underscores a notable gap in the existing literature concerning the relationship between market size and innovation: the general absence of an IV approach. An exception is found in the work of (Dubois et al., 2015). Previous studies have predominantly relied on direct measures of market size, typically derived from demographic variables. This prevalent methodology underscores the significance and novel insight offered by the current study, which introduces an IV approach to navigate the complexities and potential biases associated with directly measuring market size.

[8]PHID is a dataset maintained at IMT Lucca. For more information see Pammolli et al. (2011).

Our measure of market size is the logarithm of aggregated sales at the ATC-3 level, calculated as the total revenue of products competing within each therapeutic class in a given year, adjusted to 2015 dollars using the Consumer Price Index. In the ATC classification system, drugs are categorized into five levels, with higher levels indicating more detailed classification. While Acemoglu and Linn (2004) utilized ATC-1 and ATC-2 categories to define market size, we use the ATC-3 level for three reasons. First, ATC-1/2 classifications aggregate products targeting vastly different patient needs, potentially conflating distinct competitive dynamics. A firm investing within a specific ATC-2 category may operate across multiple ATC-3 submarkets, so that market size defined at broader levels may fail to capture the relevant competitive environment for firms' innovation decisions. Second, the ATC-3 level aligns with the market definitions adopted by antitrust authorities, which often rely on ATC-3 or narrower classifications.[9] This alignment is relevant for the policy implications of our analysis, as discussed in Thill-Tayara and Provost (2020); Siotis et al. (2020); Bonaime and Wang (2024); Feng et al. (2024), among others. Third, we do not use the ATC-4 level because products within the same ATC-4 class may not differ significantly from those in adjacent classes, potentially generating cross-group dependencies that could invalidate statistical inference and weaken the effectiveness of our instrumental variable.[10]

Our measure of innovation is the number of active clinical trials, encompassing both pre-clinical and clinical phases, at the ATC-3 level in a given year. Clinical trials are research studies assessing medical, surgical, or behavioral interventions in human subjects. An Investigational New Drug (IND) application allows a pharmaceutical company to begin human clinical trials before marketing approval is granted. Trials progress through Phase I to Phase IV, with our measure capturing all active phases. Figure 1 presents the annual number of trials and commercial INDs over the sample period, along with the trend in aggregate sales. Both series show an upward trend through 2013, followed by a notable decline. This decline reflects the exclusion of approaches and interventions from the trial count combined with the well-documented productivity crisis in pharmaceutical R&D (Pammolli et al., 2011). For this reason, and to avoid introducing systematic bias in the Poisson estimates, the analysis is restricted to the period 2004–2013.



(a) Number of trials and INDs per year

(b) Trend of sales by year

Figure 1: Overview of the trends for sales and trials

Recall data are manually collected and classified following the FDA's recall classification system, which distinguishes between Class I recalls, involving products that may cause serious adverse health consequences or death, Class II recalls, involving products that may cause temporary or medically reversible adverse health consequences, and Class III recalls, involving products unlikely to cause adverse health consequences. Our identification strategy relies on major recalls, defined as Class I recalls and withdrawals identified by critical keywords in their stated causes, including contamination, death, overdose, symptoms, particulate matter, and adverse reaction. This definition captures sharp and unanticipated events that are likely to generate meaningful contractions in class-level revenues, as discussed in Section 5.1. In total, 7.19% of the firm sample (697 firms) issued at least one Class II recall over the sample period. Among firms that experienced a recall, 51 underwent a Class I recall, 27 of which issued a single Class I recall, and only three firms issued

[9]This is the case for FTC market definition in pharmaceuticals, which is mainly context-specific and, especially for generic drugs, is often considerably narrower, commonly at the molecule/formulation/route level. See for instance `https://www.ftc.gov/system/files/documents/cases/1810017_amneal_impax_complaint_final.pdf`.

[10]We explored various other levels of analysis, including firm-specific, product-specific, and ATC-firm aggregations, in Nutarelli (2021). The results are consistent with those presented here.

more than nine Class I recalls. A practical challenge in constructing the recall dataset concerns the consistency of counts across sources. Table 1 documents the number of recalls recorded by each source in each year. Discrepancies across sources are substantial, particularly from 2009 onwards, and reflect differences in counting methodologies and in the treatment of recalls from compounding and repackaging firms.

Table 1: Number of recalls by source and year

Year	CNN	Reg. Focus	Hall et al. (2016)	FDA Enf.	Laguna (2019)	Avg.
2004		68				68
2005		140				140
2006	384	109				243
2007	391	56				189
2008	426	128		176		244
2009	1742	85		1660		890
2010		135		389		262
2011		236		1279	75	530
2012		381	499	1518		799
2013		1031	1283	848	60	805
2014		640	1344	893		959
2015				1584		1584

To address these discrepancies, we use the cross-source average as a benchmark and validate our collected recall data against it. The main discrepancies are concentrated in 2011, 2013, and 2015, and are largely explained by the high share of compounding firms in those years, exceeding 60% and 75% of total recalls in 2013 and 2015 respectively, which our dataset only partially captures due to data unavailability.[11] Once compounding recalls are excluded from both the benchmark and our dataset, Figure 2 shows close agreement between the two series throughout the sample period, supporting the validity of our recall data for the purposes of the analysis.

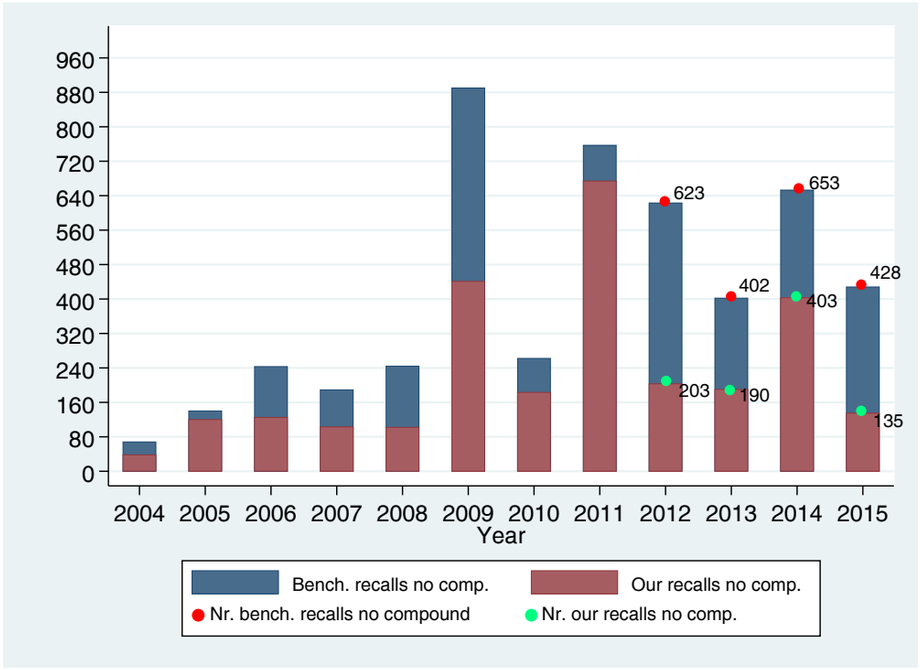


Figure 2: Comparison of collected recalls against the cross-source benchmark, excluding compounding recalls. Mint-colored points indicate the minimum number of recalls across sources; red points indicate the maximum. From 2011 onwards, the collected recalls closely track the benchmark once compounding recalls are excluded.

4 Empirical strategy

This section presents the empirical strategy for identifying the causal effect of market size on pharmaceutical innovation. We begin by outlining a conceptual framework that links product recalls to expected market rewards and innovation incentives. Next, we introduce the empirical specification and discuss the identification assumptions underlying our control-function IV approach.

[11] We confirmed that the representativeness of the recall sample is maintained as described in Nutarelli (2021).

4.1 Theoretical framework

This section outlines the conceptual framework underlying our identification strategy. We formalize how product recalls affect expected market rewards and, in turn, firms' innovation decisions. Our conceptual framework is intentionally parsimonious, yet tailored to pharmaceutical markets in two ways. First, innovation choices are forward-looking and respond to the discounted stream of expected market rewards, not merely contemporaneous sales.[12] Second, the observed market size is measured in terms of revenue; therefore, it is an equilibrium outcome that may change with both demand and supply. The goal is to formalize the conditions under which a major recall causes a persistent shift in expected market rewards at the therapeutic-class level, while remaining consistent with the exclusion restriction embedded in our control-function IV Poisson specification. To do this, we will partly draw from Jovanovic (2021).

Consider an ATC-3 therapeutic class indexed by i and discrete time t. Let $\mathscr{J}_{it}$ be the set of products available in class i at time t. Each product $j \in \mathscr{J}_{it}$ delivers an effectiveness/quality level $a_{jt} > 0$. To capture imperfect substitution across products within a therapeutic class, define an aggregate "treatment service" through a CES aggregator[13],

$$\mathscr{A}_{it} \equiv \left(\sum_{j \in \mathscr{J}_{it}} a_{jt}^{\sigma-1} \right)^{\frac{1}{\sigma-1}}, \qquad \sigma > 1, \tag{1}$$

where σ governs within-class substitutability, and Equation 1 summarizes the variety-adjusted effective quality in the class. Demand for the class-level treatment service is assumed proportional to $\mathscr{A}_{it}$,

$$Q_{it} = D_{it}\, \mathscr{A}_{it}, \tag{2}$$

where $D_{it} > 0$ is a demand fundamental capturing prevalence and diagnosis intensity, prescribing norms and guidelines, reimbursement generosity, and (importantly in this context) perceived safety risk and reputational considerations at the class level. Class-level sales are revenues,

$$\text{Sales}_{it} = P_{it}\, Q_{it}, \tag{3}$$

where P_{it} is a class-year price index. At annual frequency in U.S. pharmaceutical markets, the class-level price index P_{it} is largely mediated by institutional features—most directly, administered reimbursement benchmarks that rely on lagged average sales prices in Medicare Part B (Acquatella et al., 2026)—and by contracting frictions in coverage and rebates (e.g., Part D/PBM arrangements) (Duggan and Morton, 2011; Berndt and Newhouse, 2012). As a result, while P_{it} may move in equilibrium, it need not adjust quickly or strongly enough to offset quantity changes one-for-one within a year. Notice that we do not model P_{it} explicitly and the only role of this assumption is to rule out the case in which prices adjust within the year so strongly as to offset a recall-induced quantity contraction one-for-one in revenues (i.e., $\Delta \log P_{it} = -\Delta \log Q_{it}$), which would mechanically eliminate first-stage variation in $\log(\text{Sales}_{it})$.

Taking logs and combining (2)–(3) yields

$$M_{it} \equiv \log(\text{Sales}_{it}) = \log P_{it} + \log D_{it} + \log \mathscr{A}_{it}, \tag{4}$$

which motivates our interpretation of M_{it} as a reduced-form proxy for the market rewards relevant to innovation incentives. Innovation is measured by N_{it}, the number of active (activated) clinical trials in class i and year t. In the market-size tradition (e.g., Acemoglu and Linn, 2004; Dubois et al., 2015), innovative activity is increasing in the expected present value of profits from a marginal project, which scales with expected market rewards. We therefore define V_{it} as the incremental

[12] Notice that although R&D decisions are forward-looking, our empirical specification is a reduced-form intensity equation in which current innovation responds to the current state of expected market rewards. The model does not require state dependence in innovation counts beyond what is captured by market fixed effects, year effects, and the control function; adding lagged innovation would instead address serial correlation in the unobservables, which we handle through fixed effects and clustered inference.

[13] We introduce a CES aggregator over products within an ATC-3 class as a parsimonious way to capture imperfect substitution among therapeutically related drugs. In pharmaceutical markets, substitution is typically high but not perfect due to contraindications, adherence, clinical guidelines, and switching costs; hence the removal of a major product need not be fully offset by competitors, implying a class-level contraction in quantity and revenues. The CES form is not essential for our results; any aggregator with finite substitution delivers the same qualitative implication that product removal can reduce class-level effective variety/quality and thus market rewards.

expected present value of activating one additional (representative) innovative project targeted to class i at time t:

$$V_{it} \equiv \mathbb{E}_t \left[\sum_{\tau \geq 0} \beta^\tau \, \pi(\text{Sales}_{i,t+\tau}) \right], \qquad \beta \in (0,1), \tag{5}$$

where $\pi(\cdot)$ denotes the per-period operating profits attributable to that project and is increasing in market revenue (through demand, coverage, and appropriability). Under standard regularity conditions (in particular, π increasing), higher expected revenues $\{\text{Sales}_{i,t+\tau}\}_{\tau \geq 0}$ imply a higher project value V_{it}, formalizing the forward-looking nature of trial activation decisions.

To connect V_{it} to the observed count N_{it}, we adopt a standard "many potential projects" framework for innovation counts (Hausman et al., 1984; Dranove et al., 2014; Dubois et al., 2015). Fix an ATC-3 market i and year t. Suppose there is a large pool of potential trial opportunities indexed by k. A project is activated when its net expected value is non-negative:

$$\text{activate}_{k,it} = 1\Big\{ V_{it} - \chi_{it} + \varepsilon_{k,it} \geq 0 \Big\},$$

where χ_{it} captures market-year costs and constraints (regulatory burden, congestion in recruitment, scientific difficulty, etc.) and $\varepsilon_{k,it}$ is an i.i.d. idiosyncratic component across projects. Under the common type-I extreme value assumption for $\varepsilon_{k,it}$, the activation probability is increasing in V_{it} and proportional to $\exp(V_{it} - \chi_{it})$. With many independent opportunities, the number of activated projects in (i,t) is well approximated by a Poisson count with intensity proportional to that activation rate, so that $N_{it} \mid \mathscr{I}_{it} \sim \text{Poisson}(\lambda_{it})$ with $\lambda_{it} \propto \exp(V_{it} - \chi_{it})$.

In the data we do not observe V_{it} (nor firms' information set) directly. Instead we observe a revenue-based measure of market size $M_{it} \equiv \log(\text{Sales}_{it})$, which in (4) is a reduced-form function of demand fundamentals and effective variety/quality, and is therefore a natural empirical proxy for the profitability conditions that enter V_{it} at annual frequency (Acemoglu and Linn, 2004; Dubois et al., 2015). Importantly, we do not require that M_{it} be a sufficient statistic for the entire expectation vector; we only require that it be an informative (and persistent) proxy for the latent profitability state that shifts innovation incentives. Formally, let S_{it} denote the latent profitability state, i.e. the component of expected market rewards relevant for trial activation at the annual horizon. We assume that the value of activating one additional project is increasing in that state,

$$V_{it} = v(S_{it}), \qquad v'(\cdot) > 0,$$

so that higher S_{it} increases the incentive to activate trials at t, and that observed market size is a noisy signal of it,

$$M_{it} = S_{it} + \nu_{it},$$

where ν_{it} captures measurement error and other equilibrium noise in revenues.
Each ATC-3 market i is subject to a certain latent defect intensity at time t, denoted by d_{it}, capturing severe product-level safety or manufacturing problems that may trigger regulatory intervention in the style of Jovanovic (2021). The latter is represented by $d_{it} = \bar{d}_i + \eta_{it}$, where $\bar{d}_i$ is a time-invariant component reflecting the baseline risk profile of the therapeutic class and η_{it} is an idiosyncratic severe-defect shock. Modeling the full FDA decision process is beyond the scope of the present paper. Nevertheless, it is reasonable to assume that the regulator acts so as to protect patients' health, and therefore forces products out of the market when sufficiently severe adverse effects are detected. In this sense, a major recall can be viewed as the observable realization of a latent defect event once its seriousness crosses a regulatory threshold. Formally, one may write

$$R_{it} = \mathbf{1}\{d_{it} > \bar{d}\},$$

where R_{it} denotes a major recall event and $\bar{d}$ is the threshold above which the defect becomes severe enough to trigger regulatory action [14]. In the spirit of Jovanovic (2021), d_{it} cannot be predicted by neither the regulator nor the firm and $\bar{d}$ is a private information of the regulator. Furthermore in our context a major recall is interpreted as a public signal of a severe product-level problem that depresses expected market rewards at the ATC-3 level. We therefore summarize the effect of recalls on the latent profitability state as

$$S_{it} = \bar{S}_{it} - \psi_0 R_{it} - \psi_1 R_{i,t-1} + \xi_{it}, \qquad \psi_0, \psi_1 > 0,$$

[14] More generally, the model works even if one wants to model recalls as an intensity rather than a binary event. In this case, R_{it} can be interpreted as arising from a stochastic process with arrival rate increasing in d_{it}

where $\bar{S}_{it}$ is the underlying profitability state absent recall events, and ξ_{it} captures other shocks to expected market rewards. The contemporaneous term captures the immediate contraction in expected revenues induced by the recall, while the lagged term allows for gradual recovery.

Let κ_{it} denote an unobserved profitability component (e.g., scientific opportunity, guideline shifts) that affects trial activation through net costs/constraints and may be persistent over time [15]. If κ_{it} is persistent, then M_{it} and its lags can remain correlated with κ_{it} through the dependence of S_{it} on persistent unobservables, so replacing M_{it} with $M_{i,t-1}$ does not generally deliver causal identification. In other words, lagged market size may improve timing or forecasting, but it removes endogeneity only under the strong assumption that the unobservables generating endogeneity have no relevant temporal dynamics (Bellemare et al., 2017). This motivates instrumenting M_{it} in our empirical analysis. Basically, although trial activation is forward-looking, it is based on information available at time t: because the lag from clinical development to marketing is long and uncertainty about distant payoffs is substantial and heavily discounted (Keyhani et al., 2006; Glied and Smith, 2013; DiMasi et al., 2016), firms effectively condition on the current profitability state of the market. Empirically, contemporaneous market size M_{it} provides a noisy but informative signal of this persistent state, whereas lagging M_{it} does not in general remove endogeneity when unobserved profitability determinants are persistent.

In the data we do observe contemporaneous market size $M_{it} = \log(\text{Sales}_{it})$ and controls z_{it}, together with market fixed effects and year effects. Absorbing the proportionality constant and time-invariant heterogeneity into c_i, and writing the market-year cost shifter as $z_{it}'\gamma + \kappa_{it}$, yields the intensity specification that matches our empirical framework:

$$\mathbb{E}[N_{it} \mid M_{it}, z_{it}, c_i, \kappa_{it}] = c_i \exp\big(\beta_1 M_{it} + z_{it}'\gamma + \kappa_{it}\big). \tag{6}$$

Here c_i captures time-invariant ATC-3 heterogeneity (absorbed by ATC-3 fixed effects), while κ_{it} is an idiosyncratic profitability component potentially correlated with M_{it}, which motivates instrumenting M_{it} using major recalls within our control-function IV Poisson estimator. This implies that major recalls affect innovation incentives indirectly, by shifting the latent profitability state S_{it}, which in turn is reflected in the observed revenue-based measure of market size M_{it}.
Formally Equation 6 is recovered as follows. First we better specify the number of activated projects as a Poisson count with intensity proportional to the activation rate: $N_{it} \mid \mathscr{I}_{it} \sim \text{Poisson}(\lambda_{it}), \quad \lambda_{it} = \Lambda_i \exp(V_{it} - \chi_{it})$, where we added a time-invariant scale $\Lambda_i > 0$ (the size of the opportunity set) in market i.
Second, because S_{it} is not observed and M_{it} is its empirical proxy, in the spirit of Acemoglu and Linn (2004) we adopt the following reduced-form approximation:

$$V_{it} = \alpha_i + \beta_1 M_{it} + u_{it},$$

where α_i captures time-invariant ATC-3 heterogeneity and u_{it} is an idiosyncratic profitability component. Similarly, decompose activation costs as $\chi_{it} = b_i + z_{it}'\gamma + \omega_{it}$, where b_i is time-invariant, z_{it} are observed cost shifters/controls, and ω_{it} is an unobserved market-year cost shock. Substituting these decompositions into the intensity yields

$$\lambda_{it} = \Lambda_i \exp\big((\alpha_i - b_i) + \beta_1 M_{it} + z_{it}'\gamma + (u_{it} - \omega_{it})\big).$$

Define $c_i \equiv \Lambda_i \exp(\alpha_i - b_i)$ and $\kappa_{it} \equiv u_{it} - \omega_{it}$. Then the conditional mean of the Poisson count satisfies Equation 6 which is the mean function estimated in our fixed-effects Poisson framework. Note that product quality a_{jt} enters the analysis only through the effective-quality aggregator $\mathscr{A}_{it}$ and therefore through revenues $M_{it} = \log P_{it} + \log D_{it} + \log \mathscr{A}_{it}$; in the reduced-form approximation below, this channel is summarized by M_{it} rather than modeled separately.

4.2 Identification strategy

This subsection explains how we identify the causal effect of market size on innovation. We proceed in three steps. First, we present the empirical specification and the source of endogeneity that motivates our estimator. Second, we explain why major recalls provide relevant variation in

[15] Here ξ_{it} captures shocks to the latent profitability state, whereas κ_{it} below denotes the reduced-form unobserved component entering the empirical innovation equation.

class-level market size. Third, we discuss the identifying assumptions underlying the use of recalls as an instrument and summarize the evidence supporting their validity.

Our baseline empirical specification links innovation at the ATC-3-year level to market size through a Poisson model:

$$E[N_{it}|\mu_i,\zeta_t,X_{it},M_{it}] = \exp(\beta_1 \log M_{it} + \beta_2 X_{it} + \mu_i + \zeta_t), \qquad i = 1,\ldots,N,\ t = 1,\ldots,T, \quad (7)$$

where N_{it} is the number of active trials in ATC-3 class i and year t, M_{it} is class-level market size measured by sales, X_{it} is a vector of controls, μ_i are ATC-3 fixed effects, and ζ_t are year effects. In our application, X_{it} includes variables capturing product age, competition, generic penetration, and scientific production.

Estimating (7) by treating $\log M_{it}$ as exogenous may yield biased estimates for two reasons. First, market size is an equilibrium outcome and may therefore be correlated with unobserved profitability shocks that also affect trial activity. Second, in a nonlinear count model, this endogeneity must be handled jointly with fixed effects. For this reason, we adopt the control-function IV approach of Lin and Wooldridge (2019), which is specifically designed for nonlinear models with fixed effects and allows one to address both endogeneity arising from time-invariant heterogeneity and endogeneity arising from time-varying idiosyncratic shocks.

Formally, let c_i denote time-invariant unobserved heterogeneity and let κ_{it} denote an idiosyncratic market-year shock. The conditional mean can be written as

$$E[N_{it} \mid M_{it}, z_{it}, c_i, \kappa_{it}] = c_i \exp(x_{it}\beta_1 + \kappa_{it}), \quad (8)$$

where $x_{it} = (M_{it}, z_{it})$ and z_{it} contains the exogenous controls, including year effects. The key concern is that M_{it} may be correlated with κ_{it}. In the absence of such idiosyncratic endogeneity, a fixed-effects Poisson estimator would be consistent under the strict exogeneity condition

$$E[N_{it} \mid x_{it}, c_i] = c_i \exp(x_{it}\beta_1).$$

The control-function approach allows us to test this restriction directly.

To implement the procedure, we first estimate the reduced form for the endogenous regressor. Let R_{it} denote the excluded instruments, namely contemporaneous and lagged normalized major recalls. The reduced form for market size is

$$M_{it} = z_{it}'\Pi + R_{it}'\delta + c_{i2} + u_{it2}, \quad (9)$$

where u_{it2} is the reduced-form error. We then obtain the within residuals

$$\ddot{u}_{it2} = \ddot{M}_{it} - \ddot{z}_{it}'\hat{\Pi} - \ddot{R}_{it}'\hat{\delta}.$$

In the second step, we estimate a fixed-effects Poisson model augmented with the first-stage residual:

$$E[N_{it} \mid M_{it}, z_{it}, c_i, \hat{\hat{u}}_{it2}] = c_i \exp(x_{it}\beta_1 + \hat{\hat{u}}_{it2}\rho). \quad (10)$$

The coefficient ρ captures idiosyncratic endogeneity. A robust Wald test of $H_0 : \rho = 0$ therefore provides a direct test of whether market size can be treated as conditionally exogenous. When $\rho \neq 0$, the inclusion of $\hat{\hat{u}}_{it2}$ delivers consistent estimation of β_1 under the assumptions in Lin and Wooldridge (2019). Intuitively, the control function isolates the component of market size that is correlated with the unobserved idiosyncratic shock, so that the remaining variation in M_{it} can be interpreted causally. A more detailed derivation for our case, including the Mundlak representation that rationalizes this control-function correction in the presence of fixed effects, is reported in Appendix B.

A practical implication of the fixed-effects Poisson estimator is that the dependent variable must be nonzero for at least one time period within a panel unit. We therefore exclude ATC-3 classes that never record positive trial counts over the sample period. These account for about 10% of the ATC-3 categories in our data.

Recalls as an instrument for market size. To identify exogenous variation in market size, we instrument M_{it} with normalized major recalls. Let R_{it} denote the number of major recall events in class i during year t, and define

$$\tilde{R}_{it} \equiv \frac{R_{it}}{|\mathscr{J}_{it}|} \cdot 100,$$

that is, recalls normalized by the number of products in the ATC-3. This normalization is important because broader markets mechanically have more opportunities for recalls; using recall intensity rather than raw recall counts removes this source of spurious correlation between recalls and market size.

The relevance of the instrument follows directly from the conceptual framework in Section 4.1. A major recall affects class-level sales through two channels. The first is a product-removal channel: the recalled drug exits, or becomes partially unavailable in, the set of marketed products, thereby reducing effective class-level variety and quality. The second is an informational channel: major recalls convey negative information about safety and compliance and may reduce prescribing, continuation, and adoption at the therapeutic-class level. When within-class substitution is not perfect and/or the event carries reputational content, both channels reduce class-level revenues M_{it}.

This mechanism also implies persistence in the first stage. A recall may depress sales contemporaneously and continue to do so in subsequent periods if the recalled product is not immediately replaced or if the reputational shock dissipates only gradually. This is why both contemporaneous and lagged normalized recalls enter the first stage. Consistent with this logic, Table 7 shows that $\tilde{R}_{it}$ and $\tilde{R}_{i,t-1}$ are significant predictors of class-level log sales. The same evidence also shows that this effect is stronger for major recalls than for minor recalls, which is precisely what one would expect if severe recalls are the relevant source of exogenous variation in market rewards.

The fact that recalls originate from a supply-side disruption does not invalidate their use as an instrument in our setting. If substitution within the ATC-3 were effectively perfect and recalls had no informational content, the sales loss from one recalled product would be immediately offset by reallocation toward close substitutes, leaving class-level revenues largely unchanged. In that case the instrument would be weak. Our identifying strategy instead relies on the empirically relevant case emphasized in Section 4.1: imperfect substitution and/or a reputational demand response imply that recalls reduce class-level revenues rather than simply reshuffling sales across incumbent products. Since our endogenous regressor is class-level sales, this is exactly the variation in market rewards that matters for innovation incentives.

Identification assumptions. Our identification strategy relies on two conditions: instrument relevance and the exclusion restriction.

Relevance is supported both by the theory and by the first-stage estimates. Major recalls reduce class-level sales on impact and, because the resulting disruption is not immediately undone, they also predict lower sales with a lag. The first-stage results in Table 7 confirm this pattern. Additional supporting evidence comes from the abnormal-sales analysis in Figure 5, which shows a marked decline in ATC-3 sales around major recall events, and from the time-to-event analysis in Appendix C, which indicates that recalled products remain economically affected for a prolonged period.

The exclusion restriction requires that recalls affect innovation only through their effect on expected market rewards, not through a direct effect on trial activity. In our setting, the main threat would be that a recall creates a technological or commercial "gap" that competing firms immediately rush to fill by starting new trials. We believe this channel is unlikely to drive our results for three reasons.

First, we study major recalls of marketed products, whereas the time from trial initiation to commercialization is long. A competitor's recall therefore occurs late relative to the innovation cycle and does not mechanically generate an immediate short-run wave of new trials, since trial initiation is slow and responds to expected long-run profitability rather than to a contemporaneous gap in marketed products. Second, the sales shortfall created by a major recall is not rapidly eliminated. Our time-to-event evidence in Appendix C shows that recalled products often remain affected for several years, and the recall process itself typically unfolds over many months. This persistence is inconsistent with an immediate replacement mechanism through new trials. Third, our outcome construction explicitly removes the most direct mechanical links: we focus on active trials, exclude trials sponsored by the recalling firm(s), and, when a major recall can be linked to a specific active ingredient, exclude trials tied to that same ingredient. These restrictions ensure that recall exposure does not mechanically enter the dependent variable through the recalling firm's own trial pipeline.

A second concern is that recalls may be endogenous to market size itself, for example because larger markets generate more litigation or more regulatory scrutiny. We address this concern in two ways. First, the normalization of recalls by the number of products in the ATC-3 removes the most direct mechanical link between market breadth and recall incidence. Second, the institutional timing of the recall process makes reverse causality of this sort unlikely to be first-order: severe

recalls are typically triggered by product-specific safety or manufacturing problems, and litigation generally follows rather than causes the recall process. Since our specifications include ATC-3 fixed effects, year effects, and controls, identification comes from within-class variation in recall intensity over time rather than from permanent differences across therapeutic classes.

We also present several pieces of supporting evidence consistent with exogeneity. Figure 3 shows that ATC-3 classes experiencing major recalls are not systematically more innovative than those that do not. Figure 4 shows that trial activation does not display a discrete change immediately after the recall year, which is inconsistent with a direct short-run effect of recalls on innovation. Finally, the Monte Carlo exercise reported below indicates that the observed distribution of major recalls across ATC-3 classes is consistent with random assignment once the structure of the sample is taken into account.

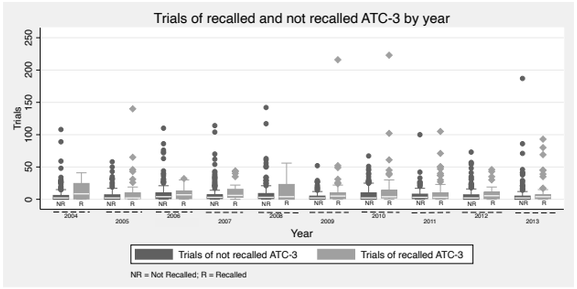


Figure 3: The box plots show how, on average, recalls do not necessarily happen in more innovative markets. The average number of trials amounts to 123 for ATC-3 markets having undergone a recall and to 118 for ATC-3 markets not having undergone a recall.
This graph reports a yearly analysis of the average number of trials in recalled and not recalled ATC-3 groups. The average number of trials is similar in both the ATC-3 markets, having undergone at least a recall (R) and ATC-3 markets not having undergone any recall (NR).

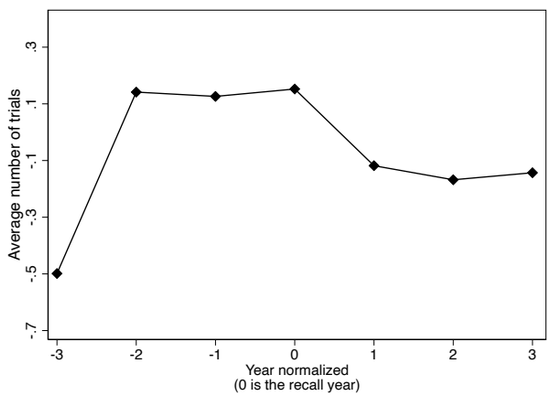


Figure 4: Change in the number of trials as measured by $ln(Trials_t) - ln(Trials_{t-1})$

Taken together, these arguments support the interpretation of normalized major recalls as a plausibly exogenous source of variation in class-level market size. Combined with the control-function IV Poisson estimator, this allows us to interpret the coefficient on market size as the causal semi-elasticity of clinical-trial activity with respect to ATC-3-level market rewards.

A final interpretive point is worth emphasizing. Because major recalls generate primarily contractionary variation in class-level revenues, our IV estimates are identified from negative shocks to expected market rewards. Accordingly, the estimated coefficient on market size should be interpreted as a local elasticity recovered from downside movements. Applying it symmetrically to market expansions relies on the maintained log-linear functional form (rather than on separate positive-shock variation in the data).

5 Results

The results section is organized into two main subsections. First, we analyze the impact of recalls on the endogenous market size, as measured by the total sales of ATC category i. Our goal here is

to offer compelling evidence supporting the relevance of the instrument we have adopted.
Second, we present the findings concerning the impact of the instrumented market size on innovation. This analysis seeks to elucidate how changes in market size, influenced by recalls, affect the rate and nature of innovation within the same ATC category.

5.1 The impact of recalls on sales

5.1.1 Summary statistics

This section presents the summary statistics for our sample. Table 2 displays the average values and standard deviations of the relevant variables for the entire sample and two distinct subsamples—those associated with recalls and those not associated with recalls. The data in this table are specific to the ATC-3 level and focus on major recalls. Table 2 also provides detailed information about all the relevant control variables that were used in Table 7.

Table 2: Summary statistics at the ATC-3 level for the full sample, the subset of ATC-3 having undergone a recall in the period considered, and the subset that has not undergone a recall. The database at the ATC-3 level is balanced.

		ATC-3			
Variable		Full Sample	Subs. recalls	Subs. no recalls	**Description**
Sales (log)	Overall mean	19.405	20.794	19.054	Log of sales at ATC-3 level.
	Overall Std. Dev.	2.233	1.475	2.256	
	Between Std. Dev.	2.152	1.441	2.163	
	Within Std. Dev.	.614	.378	.661	
Outflow rate ($\frac{K_{t+1}}{P_{-1}}$)	Overall mean	.086	.056	.093	This is defined as the number of lost products in an ATC-3 (K_{t+1} in regressions) over the total number of products in $t-1$ (P_{-1} in regressions).
	Overall Std. Dev.	.257	.064	.285	
	Between Std. Dev.	.109	.036	.120	
	Within Std. Dev.	.233	.053	.259	
Avg. age of firms within ATC	Overall mean	35.907	33.275	36.573	This is the average age of the firms competing within an ATC-3. The foundation year of the firms was present in the data.
	Overall Std. Dev.	7.631	5.420	7.960	
	Between Std. Dev.	6.767	4.452	7.093	
	Within Std. Dev.	3.556	3.160	3.650	
Herfindahl–Hirschman Index (hhi)	Overall mean	.431	.268	.434	The hhi measures the competition within a market. It can range from 0 to 1.0, moving from a huge number of very small firms to a single monopolistic producer.
	Overall Std. Dev.	.260	.159	.262	
	Between Std. Dev.	.236	.166	.240	
	Within Std. Dev.	.110	.020	.115	
Share generics by ATC	Overall mean	.746	.725	.752	This represents the percentage of generic products, among all products sold in an ATC-3 market
	Overall Std. Dev.	.255	.214	.264	
	Between Std. Dev.	.238	.210	.245	
	Within Std. Dev.	.092	.052	.099	
Avg. age prod. by ATC	Overall mean	13.159	12.043	13.441	This represents the average age of product within an ATC-3. The age of a product is based on the foundation year of the firm that produced it.
	Overall Std. Dev.	5.333	3.763	5.627	
	Between Std. Dev.	4.909	3.568	5.164	
	Within Std. Dev.	2.109	1.304	2.268	
Scientific knowledge within ATC	Overall mean	6.327	6.787	6.211	The number of papers and scientific publications for an ATC-3 present in PubMed and other sources.
	Overall Std. Dev.	1.718	1.623	1.724	
	Between Std. Dev.	1.705	1.627	1.709	
	Within Std. Dev.	.242	.177	.256	
Number of firms within ATC	Overall mean	21.054	32.802	18.082	Number of firms trading within an ATC-3
	Overall Std. Dev.	20.954	23.442	19.175	
	Between Std. Dev.	20.604	23.069	18.876	
	Within Std. Dev.	4.050	5.367	3.645	

Table 2 provides detailed statistics, including the overall, between, and within standard deviations for the main control variables (including sales). These statistics are presented for the entire sample, as well as for the sub-sample of ATC-3 categories that have experienced at least one recall [16], and the sub-sample of ATC-3 categories without any recalls. The sales panel in Table 2 illustrates that recalls typically occur in larger markets than the average.
Moreover, as expected, more competitive markets tend to experience more recalls, as indicated by the Herfindahl–Hirschman Index (*hhi*). This finding supports the notion that there are significant differences in terms of competition among ATC-3 groups. The variation in *hhi* values across these groups highlights the relationship between market structure and the frequency of recalls, suggesting that the dynamics of competition may influence the likelihood of product issues leading to recalls.

[16] To mitigate any potential issues of reverse causality with market size, recalls have been normalized by the number of products in the ATC market. These normalized recalls are referred to as $\tilde{recalls}$ in the subsequent sections of the analysis.

The finding that recalls more frequently occur in firms with a high share of generics supports the notion that there may be less stringent regulatory policies for the approval of generic drugs compared to branded drugs. This observation aligns with a well-documented concern in the literature regarding the safety of generic medications (see, e.g., Gallelli et al. (2013)).
Furthermore, within ATC markets, the outflow rate exhibits greater variance within groups than between them. This indicates that there is no significant difference in outflow rates among different ATC-3 groups. This outcome is expected at the ATC level, contrary to what might be observed at the firm level, where strategic product placement policies can influence outcomes. In the context of ATC aggregation, market dynamics, governed by demand and supply, tend to equalize outflow rates across even vastly different ATC markets.
Additionally, two other variables at the ATC-3 level appear to correlate with recalls: the first is the level of scientific knowledge within an ATC, and the second is the number of firms operating within that ATC. Specifically, recalls are more prevalent in ATC markets that not only involve a higher number of trading firms but also boast more advanced scientific knowledge compared to other markets. This suggests that these markets might be both more competitive and more dynamic, leading to higher recall rates.
Nutarelli (2021) provides further insights into the types of firms and products that typically experience recalls. According to Nutarelli (2021), there is a general tendency for recalls to occur in large, established firms and to involve relatively older products than the average product age.
In contrast, at the level of firm and product types (exploring the lines of productions within firms), recalls tend to occur in more dynamic ATC classifications, where the drugs that were recalled had previously been pioneers in their field. The use of fixed effects (at the ATC-3 level) techniques in the analysis helps to control for the time-invariant characteristics of these ATC groups, isolating the impact of other variables.
In summary, the recalls generally involve relatively old drugs produced by large, established firms. However, the major recalls occur in dynamic markets characterized by a high activity level, where, on average, many younger firms operate and trade relatively young products. One possible explanation for why markets with these characteristics are more prone to recalls is that they are under heightened scrutiny by regulators, who pay special attention to these rapidly evolving sectors.

5.1.2 Analysis of the determinants of drug recalls

This section presents the first stage results. The analysis employs a Fixed-Effects estimation approach. Table 7 displays the estimates for the first stage at the ATC-3 level. Additionally, a subsequent level, ATC-Firm, has been introduced to examine compensatory mechanisms within ATCs across different firms. To ensure alignment with the optimal model, the sample was truncated in 2013 for the initial stage as well. Results using a non-truncated sample are remarkably similar, as shown in Nutarelli (2021). The F-statistic is reported at 14.32. The standard errors reported in the tables for this and subsequent sections are robust and clustered at the ATC-3 level of aggregation.

Table 3: First stage results at different levels. ATC-3 aggregation represents the main specification.

	ATC-3-Firm Aggregation Log sales	**ATC-3 Aggregation** Log sales
$\tilde{recalls}$	−0.0053	−0.0283***
	(0.0033)	(0.0056)
$\tilde{recalls}_{t-1}$	−0.0226**	−0.0267***
	(0.0083)	(0.0070)
$\frac{K_{t+1}}{P_{-1}}$		0.1932**
		(0.0628)
average age firm		0.1576
		(0.0921)
average age firm^2		−0.0020
		(0.0013)
$hhi^{(a)}$		1.2405***
		(0.2590)
share generics in ATC		−0.1895
		(0.3373)
papers		−0.0260
		(0.0507)
# firms		0.0077
		(0.0071)
Year Dummies	Yes	Yes
Obs.	48915	1664
Groups	8634	208

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

$^{(a)}$ Relative Herfindahl–Hirschman index

Huber-White robust and clustered at ATC-3 level. (ATC-3-Firm Aggregation) fits an F.E. model at ATC-Firm level (i.e., ATC-3 lines of productions within firms). This level is introduced to check the possibility of compensations between sales of products belonging to the same ATC-3 within a firm. (ATC-3 Aggregation) fits an F.E. model at the ATC-3 level. $\tilde{recalls}$ represents normalised recalls. At the ATC-3 level, recalls are normalised for the number of products within an ATC.

We found a significant and negative impact of recalls on the logarithm of sales at the market level. At the same time, the production lines of medicines belonging to the same ATC-3 encounter a drop in sales due to recalls (ATC-Firm Aggregation in Table7). This evidence excludes the possibility of compensations between sales of products belonging to the same ATC inside a firm in line with Krieger et al. (2022). Therefore, the negative effect of recalls at the market level is enforced, whose lack is not filled by the same firms with other medicines of the same ATC-3.

The second column of Table 7 presents the first stage of the main specification. The analysis reveals that the effect of recalls at the ATC-3 level is substantial and significant, both for current and delayed recalls. Following an adequate number of bootstrap repetitions, it was determined that the t-statistic remains unchanged whether the model uses current recalls or lagged recalls [17]. This finding corresponds to a Sargan-Hansen test for over-identification in our analysis, indicating no over-identifying restrictions (Lin and Wooldridge (2019)).

We ascribe the robustness of these findings to the chosen level of aggregation. Firms characterized by high-quality management and a readiness for risk can promptly address severe recalls. Nevertheless, such recalls frequently surprise ATC-3 markets. Consequently, competitors fail to anticipate severe recalls impacting firms within the same ATC, an occurrence that is only recognizable at the market level.

To further verify the absence of compensatory effects at the market level, we investigated the impact of recalls on aggregated sales, excluding the sales of firms that had experienced recalls. Our analysis indicates that the observed decline in sales diminishes when firms that have undergone recalls are omitted from the sample (see Figure Appx.2 in the Appendix). The decline in sales observed at the ATC-3 level is not only apparent from the estimates provided in Table 7 but also from the examination of abnormal growth rates detailed in Section 5.1.3.

Furthermore, while it could be argued that recalled products, being highly innovative, lack direct substitutes within the same market, our analysis counters this notion. We have access to the generic names of products, both recalled and not recalled, as well as the active principles of medicines. Our findings indicate that, on average, there are ten products in the market that utilize the same active principle as the recalled products. This observation supports the hypothesis that market gaps created by recalls are typically filled by existing products, thereby challenging the notion that recalled

[17] The t-stat is obtained after 30000 repetitions and amounts to 2.438.

products are uniquely innovative and without substitutes.

5.1.3 Analysis of abnormal values

This section presents our analysis of the impact of drug recalls on sales. We define the effect of recalls by establishing a reference, which we call the "potential" value of a given economic indicator under "normal" economic conditions. Consequently, the Abnormal Value (AV) of an indicator y for unit i at time t is determined as the difference between the observed value and this potential value Thirumalai and Sinha (2011):

$$AV_{it} = y_{it} - E\left(y_{it}\right), \tag{11}$$

The potential value $E(y_{it})$ is estimated by running a Fixed-Effects regression on the following model:

$$y_{it} = \alpha + \beta y_{st} + \gamma X_{it} + \mu_i + \lambda_t + u_{it}, \tag{12}$$

where y_{st} is the aggregated value of y in year t at the sector level. The usual control variables (X) and year dummies are included as regressors. After obtaining estimates of AV_{it} for all i and t, referred to as $\widehat{AV}_{it}$, the time dimension is re-scaled. Specifically, the time dimension is centered on the year when the recall is issued for all units experiencing a recall in the time frame considered. Only these units are kept in the sample. The market-level Abnormal Value $\overline{AV}_t$ associated to recalls is then computed as the simple average of $\widehat{AV}_{it}$ for any $t \in \{-(T-1), ..., 0, ..., (T-1)\}$, as follows:

$$\overline{AV}_t = \sum_{i=1}^{N_t} \widehat{AV}_{it}, \tag{13}$$

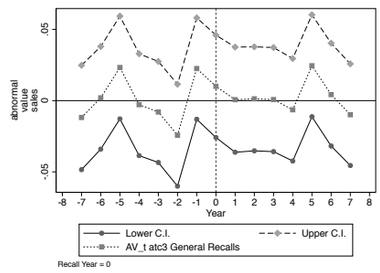
where N_t is the number of units with available data in t among those experiencing one recall. Confidence intervals for $\overline{AV}_t$ are constructed by calculating the variance of $\widehat{AV}_{it}$ as follows:

$$Var\left(\overline{AV}_t\right) = \frac{\sum_{i=1}^{N_t} Var\left(\widehat{AV}_{it}\right)}{N_t^2}, \tag{14}$$

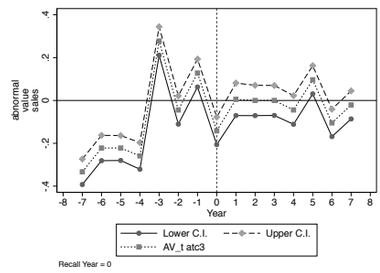
where $Var(\widehat{AV}_{it})$ is the variance of the forecast error derived from estimation of Equation 12. The focus of our analysis is on the growth rate of sales volumes. The exercise is replicated for three classifications of recalls (i.e., standard recall definition, major recalls, and type of recall) and three levels of analysis (i.e., product, firm, and sector level). The main text only reports the analysis at the ATC-3 level because it is the level at which the first and second stages are conducted. Abnormal values at the firm and product level can be found in the Appendix.

Note that in the model for the sector level, y_{st} is replaced with y_{mt} in Equation 12; that is, the value at the whole market level.

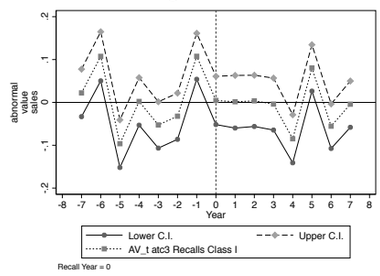
Figure 5 reports estimates of the effects of recalls on the AV of sales growth.



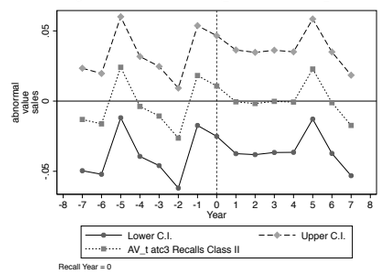
ATC-3: general recalls sales



ATC-3: Maj. recalls sales (preferred specification)



ATC-3: Class I recalls sales



ATC-3: Class II recalls sales

Figure 5: Abnormal values are analyzed at the ATC-3 level of aggregation, with years normalized such that year 0 corresponds to the year of the recall. The analysis encompasses four scenarios, delineating the sales trajectory before and after the recall year, according to four distinct recall classifications: major recalls, Class I recalls, general recalls, and Class II recalls. Diagrams clearly illustrate a decline in sales during the recall year across all recall types. Notably, major recalls exhibit a more significant drop. Additionally, the analysis shows that the lowest error bound is achieved when evaluating major recalls.

Figure 5 presents the abnormal values at the ATC-3 level, with confidence intervals constructed at the 95% level. As illustrated in Figure 5, sales decline sharply in the recall year and then partially recover over the following one to two years, particularly in the figure associated with major recalls. This pattern suggests that the effect of recalls is economically meaningful on impact and remains relevant beyond the recall year, even if it is not permanent. In this sense, recalls generate a medium-run contraction in class-level revenues, rather than the very persistent demographic shifts studied by Acemoglu and Linn (2004). This is sufficient for our purposes, since innovation decisions are forward-looking and respond to changes in expected market rewards over the relevant planning horizon.

The analysis of abnormal values corroborates the findings discussed in previous sections, particularly the significant impact of recalls on sales during the year of the recall. The error bound is notably lower for the ATC-3 level when analyzing major recalls, reinforcing the anticipated decline in sales coinciding with the recall event.

5.2 Relationship between innovation and market size

In this section, we report the findings related to the relationship between market size, denoted as M_{it}, and innovation, represented by N_{it}. The data used are already adjusted to 2015 dollars using the Consumer Price Index (CPI), allowing direct measurement of market size as the total sales across ATC market i at time t. Innovation is quantified by the number of activated trials in ATC i at time t. The analysis covers the period from 2004 to 2013. The final two years of the data set (i.e., 2014 and 2015) are excluded due to the minimal number of trials conducted during these years. Including 2014 and 2015 could introduce biases into the analysis, which utilizes Poisson estimates and does not accommodate a zero value for the dependent variable in most observations.

The dataset is a strongly balanced panel, as required for the estimation process, with data available for 208 therapeutic classes each year.

The preferred specification is estimated using Eq.(7).

We have incorporated a variety of regressors in our analysis, encompassing supply-side determinants, technological opportunities, and age factors. Some of these controls are drawn directly from existing literature, such as the knowledge stock, which is quantified by the number of scholarly papers associated with ATC category i. For this measurement, we have utilized the PubMed database, specifically gathering data on the number of scientific works related to a given ATC-3 in a particular year through the use of MeSH (Medical Subject Headings) Terms. According to the National Institutes of Health (NIH), MeSH terms are official designations that represent specific biomedical concepts. Indexers label articles using only official MeSH list terms, avoiding other spellings or

variations.
To determine whether a paper pertains to a specific ATC class, we first linked a MeSH Term to an ATC category i, primarily leveraging the official synthetic description provided for the ATC. If this initial attempt did not yield results or did not align with evidence from the literature, a secondary verification was conducted using level 3 indications as MeSH terms [18]. The NCBI MeSH database facilitated customized searches tailored to our specific needs. Moreover, observing an upward trend in the number of papers, we detrended this variable by first differentiating its logarithm to neutralize the effects of this trend, allowing for a clearer analysis of the underlying patterns.
Another essential control variable derived from the literature is the share of generics in the market. As noted by Dubois et al. (2015), the ease of market entry and the significant financial incentives associated with the use of generics can diminish the expected profitability of new innovations. Therefore, it is crucial to assess the degree of generics penetration within markets, as this factor may act as a deterrent to firms contemplating innovative endeavors. This analysis helps understand the broader economic and competitive dynamics influencing innovation rates in the pharmaceutical sector.
As highlighted in both Acemoglu and Linn (2004) and Dubois et al. (2015), an additional factor contributing to the declining margins of innovation is the increasing influx of young entrants within an ATC market. This heightened pharmaceutical competition might impede the productivity of innovation. Consequently, it is critical to accurately measure and account for competition levels. While Acemoglu and Linn (2004) addresses competition, much of the empirical literature, including Dubois et al. (2015), does not explicitly model it. In this study, we have developed two metrics to quantify pharmaceutical competition. The primary measure employed is the Herfindahl-Hirschman Index (HHI), utilized hereafter as an indicator of competitive intensity within an ATC-3 market. The major advantage of the HHI over other metrics, such as the concentration ratio, is that it assigns greater weight to larger firms. The index ranges from 0 to 1.0, indicating a spectrum from a highly fragmented market with many small firms to a monopolistic market dominated by a single producer.
The second measure we use to control competition is the average age of firms within a market. This metric complements the Herfindahl-Hirschman Index (*hhi*) by capturing different aspects of competition. While the *hhi* effectively quantifies the "degree of monopoly" within an ATC, it does not fully encapsulate the diversity of firm types within the market. Thus, we introduced the average age of firms as a further control. This measure primarily identifies the presence of small biotechnology firms, which are significant for two reasons: these firms are often at the forefront of competing for innovation, yet they typically possess fewer financial resources compared to more established companies (see, e.g., Hall and Rosenberg (2010) among others). This duality underscores the importance of considering firm age when analyzing the competitive dynamics and innovation potential within pharmaceutical markets. Because margins decline with the number of young entrants, we expect a negative sign for the firms' average age.
Table 4 displays the principal results from our analysis, representing the second stage of the methodological approach detailed earlier. Specifically, we designate z_{it2} as the excluded instruments, which include $(\text{recalls}_{it}$, and $\text{recalls}_{it-1})$[19]. In the first stage of the analysis, we compute the residuals, $\ddot{u}_{it2}$, from a linear fixed-effect model where market size serves as the dependent variable. The second stage then incorporates these residuals into a fixed-effect Poisson model for further estimation. For a detailed outline of these steps, please refer to steps 1 and 2 in the methodological section.
Contrary to previous literature, in this study, it is unnecessary to construct M_{it} based on demographic shifts, as our innovative instrument, recalls, already eliminates endogeneity from market size. Instead, M_{it} is simply the logarithm of aggregated sales at the ATC-3 level, calculated by multiplying the number of purchased drugs (expressed in standard units for comparability) by their price.
It is important to note a critical assumption of the model: the excluded exogenous variables, R_{it}, which are part of z_{it}, do not explicitly appear in the equation governing trials. Given the refined aggregation level at our disposal, ATC-3, it is reasonable to assume that the average elasticity

[18]For example, category C6B is officially designated as "PULMONARY ARTERIAL HYPERTENSION (PAH) PRODUCTS." Given the length and potential variability in abbreviations within the MeSH Terms list, our search for relevant MeSH Terms involved examining various specifications of the description, such as "PAH PRODUCTS" and "PULMONARY ARTERIAL HYPERTENSION PRODUCTS." If these searches did not yield satisfactory results or if the results did not align with existing literature findings, we then opted for the more concise Mesh indication at level 3, specifically "PAH." This method ensures that our data collection is both precise and aligned with standardized biomedical terminology

[19]We included two instruments following Hansen et al. (2008), which suggests that using more valid instruments leads to more accurate estimates.

is consistent across categories. This assumption simplifies the model without compromising the validity of our analyses.

Table 4: Impact of market size on innovation. Column (P) employs a simple Poisson model not considering fixed effects. Column (NB) adopts a Negative Binomial, due to the presence of overdispersion. Column (CF-IV) is the main specification (control function fixed effect IV Poisson). Column (CF-IV Marketed) is the main specification using as a dependent only the trials which resulted in a marketed innovation. Column (A-B) and Column (A-B linear) add the lag of the dependent following Acemoglu and Linn (2004). Column (NR) is the main specification eliminating all of the regressors

	(P) Trials	**(NB)** Trials	**(CF-IV)** Trials	**(CF-IV Marketed)** Trials (marketed)	**(A-B)** Trials	**(A-B linear)** *log Trials*	**(NR)** Trials
$trials_{t-1}$					−0.00741	0.0732*	
					(0.0005)	(0.0335)	
Log sales	0.1378***	0.122***	0.6362**	0.2713***	0.802**	0.1176***	0.8229**
	(0.0060)	(0.0224)	(0.2149)	(0.0756)	(0.266)	(0.0153)	(0.3174)
residuals			−0.8018***	−.3657***	−0.862**		−0.9711**
			(0.2157)	(0.0745)	(0.269)		(0.3177)
$\frac{K_{t+1}}{P_{-1}}$	−0.5378***	−0.423*	−0.0926	−.2635**	−0.484***	−0.0504	
	(0.0847)	(0.182)	(0.0909)	(0.0983)	(0.147)	(0.0914)	
average age firm	0.2890***	0.178***	−0.1332***	−0.0608***	−0.106*	0.0634**	
	(0.0139)	(0.0333)	(0.0377)	(0.0116)	(0.0398)	(0.0214)	
average age firm^2	−0.0038***	−0.00234***	0.0021***	0.0006***	0.00178***	−0.0008**	
	(0.0002)	(0.0004)	(0.0005)	(0.00015)	(0.0005)	(0.0003)	
$hhi^{(a)}$	0.2245***	0.548**	−0.3199	−0.457*	−0.145	0.1106	
	(0.0446)	(0.200)	(0.2903)	(0.2403)	(0.360)	(0.1153)	
share generics in ATC	−0.5571***	−0.526**	−0.3168**	−0.0782*	−0.898***	−0.2036	
	(0.0404)	(0.180)	(0.1124)	(0.0241)	(0.143)	(0.1068)	
average age product	−0.0658***	−0.0512*	−0.0592**	−0.1023***	−0.0928**	−0.0564***	
	(0.0061)	(0.0240)	(0.0190)	(0.0130)	(0.0323)	(0.0143)	
average age product^2	0.0010***	0.0005	0.0011	−0.0034***	0.0100	0.0010*	
	(0.0002)	(0.0007)	(0.0010)	(0.00053)	(1.64)	(0.0004)	
papers	0.5608***	0.175	0.1558*	0.0492***	0.101	−0.0443	
	(0.0728)	(0.237)	(0.0750)	(.00482)	(0.0013)	(0.1477)	
papers^2	−0.6672***	−0.316*	−0.0067	−0.0078	−0.0929	−0.0083	
	(0.1056)	(0.131)	(0.0838)	(0.0853)	(0.083)	(0.0883)	
# firms	0.0090***	0.0111***	0.0008	0.0120***	−0.0032	0.0048**	
	(0.0007)	(0.0028)	(0.0035)	(0.0026)	(0.0792)	(0.0016)	
Year Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Obs.	1664	1664	1664	1216	1664	1664	1872
Groups	208	208	208	152	208	208	208
Pseudo R^2	0.1346	0.027	.	.	.	.	.
Overdis.	Yes	.	.	.	.	.	.
Zero-inflated (Vuong)	.	No	.	.	.	.	.

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

(a) Relative Herfindahl–Hirschman index

Huber-White robust and clustered at ATC-3 level. The dependent variable is the count of active trials in ATC *i* at time *t* in (P), (NB), (CF-IV) and (NR). The time interval is 10 years.

Column (P) features a simple Poisson model that uses exogenous market size to explore the robustness of market size's positive effect, devoid of fixed effects and controls for endogeneity. Due to observed overdispersion in the data, Column (NB) introduces a naive Negative Binomial model [20] to accommodate this variance.

Column (CF-IV) represents our main specification, employing a fixed effect Poisson model that addresses the endogeneity of market size. This approach provides a more precise estimation by controlling for unobserved heterogeneity and potential biases that may affect the validity of the causal inference regarding market size and its impacts.

In Table 4, the coefficients of interest are Log sales and residuals: Log sales reflect market size, while residuals quantify the endogeneity of market size. A significant coefficient for residuals indicates a correlation between the error term (refer to the second stage regression specification in Table 4) and the error functions from the first stage market size model. Essentially, residuals help control for co-movements between sales and unobservable factors related to the number of trials, effectively "purging" market size of its potentially endogenous parts.

Endogeneity is tested using a Wald test on the coefficient of residuals, denoted ρ. If ρ is significantly different from zero, endogeneity is confirmed. This condition is met in our model, as expected (see Column (CF-IV)), where fully robust standard errors reveal strong idiosyncratic endogeneity. Utilizing Fixed Effects methodologies allows for unobserved heterogeneity to be correlated with all explanatory variables, including the excluded exogenous recalls. This suggests that, even after accounting for the possibility that market size may be correlated with ATC-specific heterogeneity, market size remains non-exogenous to idiosyncratic shocks.

The coefficient of market size is positively significant, aligning with previous studies. Our estimates indicate that a 10% increase in market size results in an approximate 6.3% increase in active trials, a magnitude that conforms with the existing literature. This finding underscores the significant role that market size plays in driving clinical trial activity within the pharmaceutical industry.

[20] This model assumes exogenous market size and does not control for fixed effects.

Previous research generally finds elasticities to be approximately 0.5, which is consistent with our estimates.
The recent literature has raised the possibility that, while clinical trials may be highly responsive to market size, the proportion of these trials that result in effective innovation may be declining (see, e.g., Dubois et al. (2015) among others). This suggests that previous studies might have overestimated the impact of market size on clinical trials, as the effect should ideally be assessed only on those trials that actually lead to innovation. In this paper, we address this issue by using active trials as our dependent variable, which we believe represents a more selective subset of trials with a higher potential for innovative contribution.
The higher effect observed in our analysis, compared to studies that use New Molecular Entities (NMEs) or New Chemical Entities (NCEs) as dependent variables, can be explained by the substantial costs associated with developing new pharmaceutical entities. Our approach assumes that active trials, by virtue of being ongoing and not yet concluded, include those most likely to contribute meaningfully to pharmaceutical innovation. This subset provides a more accurate reflection of the dynamic relationship between market size and the generation of valuable new treatments, thereby partially mitigating the issue of overestimation found in prior research. Drug development is, in fact, quite expensive, the cost ranging between $800 million to $2.5 billion (see, for instance the FDA programme MedWatch). Undertaking clinical trials is, instead, sensibly cheaper, amounting to an average of $20 million to $40 million (see Martin et al. (2017) as well as John Hopkins Bloomberg Health School, 2018). Thus, it is reasonable to suppose that, ceteris paribus, a 10% increase in market size stimulates more trials than NMEs or NCEs (on average). However, exceptions are still present (see Acemoglu and Linn (2004), Duggan and Scott Morton (2010), who estimated an higher elasticity than the one of the present work). In addressing the subsequent concern, we conducted an additional analysis, utilizing solely the trials that culminated in a marketed innovation within the specified time frame (Column (CF-IV Marketed)[21]). This rigorous examination revealed an overestimation bias concerning the elasticity of market size to innovation as presented in existing literature. Previously, this elasticity has been estimated at an average value of 0.5, indicating a potential overestimation bias due to endogeneity. We notice that the latter may also have a further explanation. Namely this overestimation can also be driven by the fact that the full sample captures both the extensive and intensive margins of innovation. Specifically, market size not only influences the probability of success among initiated trials but also plays a critical role in firms' decisions to enter the innovation process in the first place. By focusing solely on marketed products, the analysis conditions on successful outcomes, thereby excluding failed or discontinued trials. As a result, part of the variation that reflects market size's role in encouraging the initiation of new trials is no longer captured. This restriction reduces the observed responsiveness of innovation to market size, leading to a smaller coefficient. Moreover, conditioning on successful trials may introduce a form of selection bias, as the sample no longer represents the full distribution of innovation attempts. Consequently, the coefficient estimated on the restricted sample is likely to understate the total effect of market size on the innovation pipeline.
The coefficient of the average age of firms and its square is in line with past observations (see, e.g., Huergo and Jaumandreu (2004) and Balasubramanian and Lee (2008) for specific studies on the topic). This effect evidences how the oldest firms tend to introduce less innovation than entrants in their early years. However, firms above intermediate ages *appear almost as active in process innovations as entering firms, and even more in product innovations* (Huergo and Jaumandreu (2004)).
Moreover, innovation decreases with the share of generics within a market. Thus, the effect theorised in Dubois et al. (2015) of decreasing margins of innovation proportionally to the entrance of generics is revealed to be correct (see also Lanjouw (2005)).
In line with Acemoglu and Linn (2004) and Rake (2017), technological advancements as measured by detrended papers are positively related to innovation. It is therefore reasonable to suppose that more trials emerge in markets where scientific research is prolific.

The divergence in the magnitude of coefficients observed between the main specification (Column (CF-IV)) and Columns (P) and (NB) in Table 4 can be attributed to two main factors. First, reverse causality between market size and innovation may bias the naive coefficient on market size upward. Innovation can itself expand observed market size, so that contemporaneous sales partly reflect past or unobserved innovative success. If market size is treated as exogenous, as in

[21]The total number of observations decreases when considering only trials resulting in marketed innovations, leading to an increase in ATCs summing to 0 over the panel.

Columns (P) and (NB), the regressor may therefore be positively correlated with the error term. In this case, the estimated coefficient on market size is upward biased, which is precisely why the market-size literature emphasizes the endogeneity of observed revenues (Acemoglu and Linn, 2004). By instrumenting market size with product recalls, our preferred specification isolates exogenous variation in sales and thereby recovers a causal estimate of the effect of market size on innovation.

Second, Columns (P) and (NB) do not adequately control for ATC-specific heterogeneity and common time variation. As a result, they mix between-ATC and within-ATC variation and may attribute to market size patterns that are in fact driven by omitted heterogeneity or aggregate trends. In our data, market size tends to increase over time, while innovation declines in the final years of the sample. Failing to account for these time patterns may therefore generate a spurious negative association between market size and innovation, attenuating the estimated coefficient in the naive specifications.

The net bias in Columns (P) and (NB) is therefore ambiguous ex ante. Reverse causality tends to bias the coefficient upward, whereas omitted heterogeneity and unmodelled time patterns may bias it downward. Since the coefficients in Columns (P) and (NB) are substantially smaller than in Column (CF-IV), our results suggest that the downward force dominates overall in the naive specifications. This explains why the preferred control-function IV Poisson estimate is larger than the corresponding coefficients obtained from the simple Poisson and Negative Binomial models.

5.2.1 Robustness checks

Robustness checks are conducted as shown in Table 4. Specifically, Columns (A-B)–(NR) of Table 4 assess the robustness of the relationship between market size and innovation. To enhance the preferred specification, three additional models are included. Column (A-B) replicates the analysis by Acemoglu and Linn (2004), incorporating controls for technological flows that may vary over time through the addition of lagged trials among the regressors. Given the non-linear nature of the estimating equation in Column (A-B), this approach includes the residuals from the first stage as part of the instrumentation strategy. In contrast, Column (A-B linear) follows the procedure of Column (A-B) but linearizes the dependent variable and omits the residuals, thereby disregarding the potential non-linearities and issues of endogeneity.

Incorporating lags of the dependent variable proves to be a substantive exercise. According to Acemoglu and Linn (2004), the predominant challenge to the identification strategy for innovation relates to changes in the innovation flow rate per dollar invested in research on a drug. It is important to note that permanent differences in innovation rates are controlled for via ATC fixed effects. Variations in these flow rates imply that achieving technological progress may be more challenging in certain domains than in others. The parameter representing the flow rate of innovation is an integral part of the theoretical model outlined by Acemoglu and Linn (2004). Following this perspective, if the innovation flow rate is variable over time, it is likely to exhibit serial correlation. Incorporating the lag of log innovation into the preferred specification provides a straightforward method to address these concerns. The lagged trials are instrumented using their respective lags through a system GMM one-step procedure. The Hansen test of overidentification results in Column (A-B linear) yield a p-value of 0.175, which lies within the acceptable tolerance levels of 0.1 to 0.25 as suggested by Roodman (2009). The results of the Arellano-Bond test for autocorrelation are presented in Table 5.

Table 5: Arellano-Bond test for autocorrelation of first differenced residuals of GMM

	z-score	p-value
Arellano-Bond test for AR(1) in first differences:	z = -10.47	Pr > z = 0.000
Arellano-Bond test for AR(2) in first differences:	z = 0.88	Pr > z = 0.377
Arellano-Bond test for AR(3) in first differences:	z = -1.46	Pr > z = 0.145
Arellano-Bond test for AR(4) in first differences:	z = 0.33	Pr > z = 0.740

When the idiosyncratic errors are independently and identically distributed (i.i.d.), the first-differenced errors are first-order serially correlated. Thus, as expected, Table 5 presents strong evidence against the null hypothesis of zero autocorrelation in the first-differenced errors at order 1.

As suggested in Roodman (2009), *"in the context of an Arellano-Bond GMM regression, which is run on first differences, AR(1) is to be expected, and therefore the Arellano-Bond AR(1) test result is usually ignored in that context"*. Moreover, this output presents no significant evidence of serial correlation in the first-differenced errors at orders 2, 3, and 4.
In Column (A-B linear), the assumption is that market size is exogenous, with fixed effects appropriately controlled for. The model specified in Column (A-B linear) is linear. For consistency across models, trials data have been converted to a logarithmic scale. Essentially, Column (A-B linear) serves as a critical control, as it accounts for fixed effects but neglects potential non-linearities and endogeneity.
Moreover, Column (NR) presents the model without additional controls, estimated according to the control-function approach of the preferred specification. The purpose of Column (NR) is to evaluate the impact of omitting controls for certain regressors on the robustness of the main specification estimates.

The findings from Columns (A-B)–(NR) of Table 4 corroborate the main specification's assertion of a positive impact of market size on innovation. All models in Columns (A-B)–(NR) of Table 4 indicate a positive relationship between market size and innovation.
Specifically, the analysis in Column (A-B) aligns with the outcomes presented in Acemoglu and Linn (2004). Consistent with that study, the inclusion of lagged trials does not materially alter the positive effect of market size on innovation.
Furthermore, in Column (A-B linear) of Table 4, the coefficient for lagged trials is positive and statistically significant at the 5% level. Market size continues to exhibit a strong and positive correlation with innovation, although the coefficient is smaller than in the preferred CF-IV specification. This suggests that part of the variation may be absorbed by the lagged dependent variable and that ignoring endogeneity and non-linearity may attenuate the estimated effect of market size.
Notice that the effect of market size in Column (A-B linear) of Table 4 bears similarities to Columns (P) and (NB) of the same table, which also do not account for endogeneity. By contrast, the impact of market size is more pronounced in models that incorporate corrections for endogeneity. It is therefore reasonable to suppose that the lower magnitude of the coefficient on market size in Columns (P), (NB), and (A-B linear) of Table 4 is primarily a consequence of treating market size as exogenous.
It is possible to provide further checks by controlling for possible overidentification of the instrumented lagged dependent variable. To do so, Table 6, Column (1), reports the two-step robust system GMM estimates of Column (A-B linear) of Table 4, which instead was estimated using a one-step system GMM.

Table 6: Coefficients of market size and lag dependent when a two-step GMM is employed. Column (2) includes suspended and withdrawn trials in the dependent

	(1) *log Trials*	**(2)** *log Trials*
$trials_{t-1}$	0.0592	0.380^{*}
	(0.0426)	(0.189)
Log sales	0.1208^{***}	-0.533^{**}
	(0.159)	(0.179)
Year Dummies	Yes	Yes
Obs.	1664	1664
Groups	208	208

Standard errors in parentheses
$^{*}\ p < 0.05, ^{**}\ p < 0.01, ^{***}\ p < 0.001$

Huber-White robust and clustered at ATC-3 level.(1) is the two-step GMM version of Column (A-B linear) of Table4. Only the critical coefficients are included. (2) is equal to (1) where also suspended and withdrawn trials are included in the dependent. Only the critical coefficients are included. Both equations are linearised to enable a simple comparison with Column (A-B linear) Table4.

The coefficient associated with market size in Table 6 is elevated compared to that observed in Column (A-B linear) of Table 4. Additionally, the lagged dependent variable does not reach statistical significance, aligning with findings from Acemoglu and Linn (2004). This result holds consistent when the count of trials is used as the dependent variable, as explored by Nutarelli (2021). Importantly, the direction of the market size estimates remains unchanged relative to the preferred model.
A final robustness check was conducted by incorporating all trials, including active, suspended, and withdrawn trials. While the number of classes remained constant, the total number of trials increased by 0.57%. These results are detailed in Table 6, Column (2). As demonstrated in our analysis, the results do not support the hypothesis proposed by Dubois et al. (2015); rather, we observe a coefficient that is both lower in magnitude and significance. This discrepancy may be attributed to

the inclusion of non-active trials, which are less sensitive to changes in market size and may skew the estimates towards randomness. For example, firms exclusively holding suspended trials are not impacted by governmental price regulations that reduce treatment costs. Consequently, expansions in market size may have a diminished effect on these firms, which have already minimized expenses due to their inactive trials. This suggests the presence of endogeneity in our model.
Additional robustness checks have been undertaken by modifying the proxy for market size to align with the methodology outlined in Acemoglu and Linn (2004). These modifications include transitioning to a different database for the collection of sales data, specifically using Evaluate sales, and utilizing all available recalls to instrument market size. Specifically, Table 7 presents the results from using Class II and Class I recalls as instruments for market size. Moreover, Table 8 estimates market size based on the number of patients within an ATC-3 category, consistent with the approach in Acemoglu and Linn (2004).
Because the number of patients is highly correlated with sales and it is employed as a natural alternative to sales, we have adopted recalls as an instrument for the number of patients.
The F-test amounts to 12 for the analysis with Evaluate and to 4 for the analysis with the number of patients.

Table 7: Column (First-stage all recalls) and Column (Second-stage all recalls) represent first and second stage results using all of the recalls at our disposal. Data are aggregated at the ATC-3 level. The impact of market size on innovation is tested using the Evaluate database in the (First-stage Evaluate) and (Second-stage Evaluate) columns. Furthermore, Column (First-stage Minor Recalls) shows the poor strength of minor recalls (i.e., recalls whose motivation pertains to labeling and packaging).

	(First-stage Minor Recalls) Log sales	**(First-stage all recalls)** Log sales	**(Second-stage all recalls)** Trials	**(First-stage Evaluate)** Log sales	**(Second-stage Evaluate)** Trials
$\tilde{recalls}$	0.00138	0.0009		−0.260***	
	(0.00302)	(0.00345)		(0.0059)	
$\tilde{recalls}_{t-1}$	−0.0017	−0.0223***		−0.0180	
	(0.00308)	(0.00678)		(0.0105)	
Log sales			0.636*		0.710**
			(0.255)		(0.275)
Residuals			−0.802***		−0.802***
			(0.216)		(0.275)
$\frac{K_{t+1}}{P_{-1}}$	0.187***	0.191**	−0.0926	−0.173	−0.147
	(0.0617)	(0.0621)	(0.0909)	(0.154)	(0.276)
average age firm	0.151	0.153	−0.133***	0.0470	0.224***
	(0.0916)	(0.0914)	(0.0377)	(0.0678)	(0.0045)
average age firm^2	−0.0019	−0.00195	0.00214***	−0.0004	−0.0814***
	(0.00127)	(0.00127)	(0.000488)	(0.0008)	(0.0316)
hhi	1.243***	1.239***	−0.320	1.721***	−0.60
	(0.259)	(0.259)	(0.290)	(0.419)	(0.495)
share generics in ATC	−0.155	−0.181	−0.317**	−0.372	0.00640
.	(0.341)	(0.339)	(0.112)	(0.328)	(0.157)
average age prod.	0.0533	0.0539	−0.0592**	−0.0330	−0.0732**
	(0.0587)	(0.0586)	(0.0190)	(0.0504)	(0.0225)
average age prod.^{2}	−0.003	−0.0037	0.00109	0.00138	0.0009
	(0.00216)	(0.00216)	(0.000966)	(0.00206)	(0.0009)
papers	−0.0264	−0.0267	0.156*	−0.139	0.265**
	(0.0510)	(0.0514)	((0.0750)	(0.0803)	(0.0917)
# firms	0.0077	0.00734	0.000825	−0.0106	0.0224***
	(0.00699)	(0.00701)	(0.00346)	(0.0095)	(0.0045)
Year Dummies	Yes	Yes	Yes	Yes	Yes
Obs.	1664	1664	1664	1136	1056
Groups	208	208	208	142	132

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Huber-White robust and clustered at ATC-3 level.

In the first column of Table 7, the main analysis is replicated, using minor rather than major recalls. In contrast to major recalls, minor recalls are recalls due to labeling or packaging issues. This exercise provides intuition for the differences in the motivations of the recalls, showing that minor recalls are, in fact, a weak instrument for market size (neither minor recalls in t nor lagged recalls are significant in the first stage). This weakness is also due to the fact that minor recalls are mainly voluntary, and hence can be easily anticipated by firms. This enforces the usage of major recalls as an instrument for market size (as explained in the main analysis).
Moreover, Table 7 reports the first and second stages of employing Class I and Class II recalls as an instrument for market size in the second and third columns. The remaining columns are devoted to the results obtained using the Evaluate database to collect market sales. The results of the primary analysis are confirmed in both exercises.
Employing all recalls decreases both the magnitude and the significance of the coefficients of sales. Moreover, only the lag of recalls is a good instrument at the market level. These two effects are expected because minor recalls may attenuate the drop in sales consequent to a recall. Indeed, within Class II, temporary recalls (e.g., recalls due to a labeling error) are also comprehended, which may not be unexpected to the firm (most of them are voluntary). For this reason, they might not

be taken into account by the company's management. Therefore, losses in terms of sales are well compensated. Furthermore, minor recalls are not publicised and cannot damage the image of the company or the market in which they happen.
Hence, adding minor recalls overtakes the strong and negative impact of current recalls, and consequently affects the estimates of the market size in the second stage. However, because Class II recalls often regard minor but persistent issues[22], a cumulative effect intervenes, and lagged recalls remain an excellent instrument.
The outcomes of the principal analysis remain robust when data on sales are collected from the Evaluate database.
Table8 reports the second stage results of the analysis with the number of patients as a measure of market size. The first stage results are investigated in the Appendix.

Table 8: Impact of market size on innovation using number of patients as a proxy of market size (only second-stage)

	(1) Trials
Log patients	3.274^{***}
	(0.648)
residuals	-3.291^{***}
	(0.647)
$\frac{K_{t+1}}{P_{-1}}$	1.476^{***}
	(0.377)
average age firm	0.589^{***}
	(0.154)
average age firm^2	-0.00478^{**}
	(0.0016)
hhi	0.145
	(0.260)
share generics in ATC	-0.648^{**}
	(0.244)
average age product	-0.278^{***}
	(0.0514)
average age product^2	0.0011^{***}
	(0.0022)
papers	0.546^{***}
	(0.147)
papers^2	0.193
	(0.114)
# firms	0.0895^{***}
	(0.0144)
Year Dummies	Yes
Obs.	1056
Groups	132

Standard errors in parentheses
$^{*}\ p < 0.05, ^{**}\ p < 0.01, ^{***}\ p < 0.001$

Huber-White robust and clustered at ATC-3 level. (1) employs MEPS database and matches the ATC-3 present in our database. Market size is measured through the number of patients within ATC-3.

Using the number of patients as a proxy for market size confirms the first and second-stage results compared to the main specification, and also highlights the effectiveness of recalls in instrumenting demand-side measures of market size. However, the outcomes are less significant than those from our main specification. Therefore, recalls do not seem to be a vital instrument for the number of patients. On the one hand, a larger number of patients in an ATC-3 class might increase the probability of an adverse event, thereby increasing the likelihood of a major recall. On the other hand, there is no reason to believe that an adverse event would happen in a more populated class, which might refer to commonly employed (and therefore well-tested) medicines.
Moreover, a recall in a class reduces the number of patients adopting pharmaceuticals in the questioned ATC-3 class, thereby offsetting the possible positive effect of a higher probability of adverse events. In the second stage, Table 8 supports the results of Acemoglu and Linn (2004), which found a coefficient for market size between 3 and 4. Therefore, the significance of residuals confirms the presence of endogeneity.

[22]Minor recalls often pertain to the manufacturing of the product accessories. Their cause ranges from label mix-up, particulate matter in specific lots, and packaging issues. Though minor recalls do not directly threaten the health of the patients, they are difficult to correct in the short-term by firms.

6 Conclusions

A long tradition of studies underscores the importance of market size in shaping innovation in the pharmaceutical industry. Following the critique by Cerda (2007), identifying exogenous variation in market size remains a central challenge, while recent contributions emphasize the need to properly account for competition and technological opportunities (Rake, 2017; Dubois et al., 2015).

This paper contributes to this literature by providing new causal evidence on the relationship between market size and pharmaceutical innovation. Exploiting major product recalls as plausibly exogenous shocks to expected market rewards, we show that larger markets induce significantly more innovation activity. Our preferred CF-IV estimate implies an elasticity of active trials with respect to market size of 0.636. Given that identification comes from recall-driven contractions in class-level sales, this coefficient is most directly interpreted as the response of innovation to negative shocks in expected market rewards. Under the maintained log-link specification, this elasticity corresponds to roughly a 6.3% change in trial activity for a 10% change in market size.

Our findings highlight the central role of expected profitability in shaping firms' innovation decisions. By linking recall-driven contractions in submarket-level revenues to subsequent changes in innovation activity, we provide direct support for demand-pull theories of innovation in a setting where market size is an endogenous equilibrium outcome. Importantly, the theoretical framework developed in this paper clarifies that what matters for firms' innovation choices is not contemporaneous sales per se, but the expected stream of market rewards, and that major recalls generate plausibly exogenous variation in profitability at the therapeutic-class level.

At the same time, the magnitude of this relationship depends critically on how innovation is measured. When focusing only on trials that culminate in marketed products, the estimated elasticity is substantially lower. This result is consistent with recent findings in (Dubois et al., 2015), which report lower elasticities when innovation is measured using realized outcomes. More generally, our results suggest that broader measures of innovation may overstate the responsiveness of commercially successful innovation to market size, and underscore the importance of distinguishing between innovation effort and realized outcomes.

From a methodological perspective, this paper introduces product recalls as a novel source of exogenous variation in market size. Unlike traditional approaches based on demographic shifts, which may suffer from endogeneity concerns (Cerda, 2007), recalls provide sharp and plausibly unanticipated shocks that affect expected market rewards at the therapeutic-class level. The empirical strategy, based on a control-function IV framework (Lin and Wooldridge, 2019), further allows us to account for both idiosyncratic and heterogeneity-driven endogeneity, providing a robust basis for causal inference.

Our analysis also contributes to a better understanding of how market-level shocks propagate in the pharmaceutical industry. While the existing literature has largely focused on firm-level effects of recalls, our results show that major recalls generate persistent contractions in ATC3-level revenues, thereby affecting innovation incentives at the market level. This highlights the importance of correctly defining submarkets, as captured in our analysis by the ATC-3 classification, in line with recent contributions that emphasize market structure and technological opportunities (Rake, 2017).

Our findings have important policy implications. If innovation is driven by expected market rewards, policies affecting pricing, reimbursement, and competition can substantially affect firms' incentives to invest in R&D. At the same time, our results suggest that such policies may have heterogeneous effects across stages of the innovation process, with stronger impacts on early-stage research than on commercially successful outcomes.

More broadly, our results contribute to the debate on innovation in small markets, such as orphan drugs (see, e.g., Denis et al., 2010; Grabowski, 2002; Dupont and Van Wilder, 2011) or in the case of personalized medicine (Garassino et al., 2025). These markets have limited demand but often require significant technological advances and high-risk investment. Establishing a clear causal link between market size and innovation suggests that targeted policies to expand expected market rewards may play a crucial role in fostering innovation in these contexts.

Finally, although our analysis focuses on the pharmaceutical industry, the framework developed in this paper may apply to other innovation-intensive sectors where market size and innovation are jointly determined. Future research could extend this approach by using more recent data, incorporating additional sources of variation – such as recalls from compounders and repackaging firms – or exploring the dynamic response of innovation to market-size shocks over longer periods.

Appendices

A Literature review table

The following table summarises the findings from the literature on the relationship between market size and innovation in the pharmaceutical industry[23]

Table Appx.1: The table reports relevant papers about the relation of innovation and market size after Acemoglu and Linn (2004). Details on the entries are reported as footnotes as ^{d_k}. Critiques related on the entries are reported as footnotes as ^{c_k}.

Paper	Data and sample	Unit of observation	Measurement of innovation	Estimation method	Report estimate of size	Proxy of market size	IV
Acemoglu and Linn (2004)	US; March CPS, 1965–2000; March CPS, 1965–2000; FDA; OECD	Broad ATC-2 classess	NME ^{c_2}	QML	>0 ^{d_1}	Demographic measures	No
Cerda (2007)	US; FOIA request; gov. funds on R&D $^{d_{2,i}}$; & U.S statistical abstract$^{d_{2,ii}}$; 1968-1997	15 drug categories $^{d_{2,iii}}$	NME	FE, GLS, IV, Tobit	>0 $^{d_{2,iv}}$	Demographic measures	No
Lichtenberg (2006)	US; GLOBOCAN; WHO;IARC & National Library of Medicine, others; 1952-2001	Mesh terms for cancer related diseases	Articles in scientific journals; number of distinct chemotherapy regimens for treating a cancer site	FE	0.53	None	No
Rake (2017)	US; R&D; FDA; OECD; ClinicalTrials.gov^{d_3}; 1974-2008	Disease classes	NDA; NME; Phase II and Phase III trials	QMLE (Poisson, 1995)	0.3444 (NME); 0.3521 (NDA)	Demographic measures	No
Dubois et al. (2015)	14 countries ^{d_4,c_1}; 1997-2007;IMS, WHO	Chemical entity; Dummies for ATC-1 and ATC-2	NCE (elasticity) ^{c_3}	OLS,2SLS,CF approach (Wooldr.,2002)	0.23 (average across ATC classes)	Deaths and GDP	Yes
Blume-Kohout and Sood (2013)	US; 1998-2010; Pharmaprojects ^{d_5}; MEPS; OECD; NIH	49 therapeutic classes	R&D $^{d_{5,ii}}$	Negative Bin.; Poisson	0.26; 0.41; 0.51	Demographic shifts	No
Agarwal and Gaule (2022)	US; 2015-2020; GBD; clinicaltrials.gov	75 Indication level	Clinical trials	Poisson	0.43	None	No

List of Abbreviations:

Abbreviation	*Definition*
ATC	Anatomical Therapeutic Chemical class
CF	Control Function
CPS	Current Population Survey
FDA	Food and Drug Administration
FE	Fixed Effects
GLS	Generalized Least Squares
IMS	Intercontinental Marketing Services
IV	Instrumental Variable
MEPS	Medical Expenditure Panel Survey
NME	New Molecular Entities
NCE	New Chemical Entities
NDA	New Drug Approval
QML	Quasi Maximum Likelihood
QMLE	Quasi Maximum Likelihood Estimate
2SLS	Two Stage Least Squares
WHO	World Health Organization

[23] In particular, the focus is on relevant works coming after Acemoglu and Linn (2004). The reason for this choice is that Acemoglu and Linn (2004) represents a milestone in investigating the relation between market size and innovation in pharmaceuticals. It overcomes issues emerging in previous studies (e.g., and above all, the one of endogeneity) and is taken as a reference point by authors willing to further delve into this stream of literature.
Furthermore, the literature review only reports the relationship between market size and innovation in the pharmaceutical industry. Indeed, different industries have different definitions of recalls.

Further details:

^{d_1} The estimates suggest that a 1 percent increase in the potential market size for a drug category leads to a 6 percent increase $^{d_{2,i}}$ Data on government funds used on the R&D process of the pharmaceutical sector; $^{d_{2,ii}}$ population data for market size; $^{d_{2,iii,iv}} > 0$ means that the exogenous increase in market size is initially associated with approximately 0.08 more drugs introduced in the market. These new drugs reduce the mortality rates of individuals aged 65 and older by 0.8 percent. This decrease in mortality rate leads to increases in market size (more demand), producing an additional increase of drugs equal to 0.096

^{d_3} Both Cerda and Rake consulted the 19^{th} edition of the Drug Information Handbook published by Lexi-Comp and the American Pharmaceutical Association (Lacy et al., 2010). This handbook is comparable to a pharmaceutical dictionary providing a list of drugs' active ingredients, the medical conditions the drug is used for, and further information such as adverse effects. The work takes into account only those medical conditions which can be found on the FDA-approved label. Hence, unlabeled and investigational uses are not present. For the period 1974 to 2008, FDA approved 599 unique NMEs and 1665 unique NDAs. These approvals refer to the 208 diseases or medical indications analyzed in this study. However, an NME or NDA may be used as therapy for several medical indications. In this case, an NME or NDA is counted as innovation for all the medical indications for which it is approved.

^{d_4} Data come from IMS and include all product sales in 14 countries (Australia, Brazil, Canada, China, France, Germany, Italy, Japan, Mexico, Korea, Spain, Turkey, United Kingdom, USA). Dubois et al. have data on the ATC-4 (they report 607 different classes), the main active ingredient of the drug (they report 6216 different active ingredients), the name of the firm producing the drug, whether it has been licensed, the patent start date, and the format of the drug (the work reports 471 different formats). Products in the same ATC-4 by definition have the same indication and mechanism of action. The authors do not consider OTC drugs. Quantities are given in standard units, one standard unit corresponding to the smallest typical dose of a product form, as defined by IMS Health.

$^{d_{5,i}}$ Pharmaprojects trend data "snapshot"; $^{d_{5,ii}}$ (focus on R&D): focus only on one instance of innovation as explained in Hall and Rosenberg (2010). Authors specify the adoption of clinical trials (from pre-clinical Phase to Phase III) not taken from ClinicalTrials.gov (see below).

Further related evidence:

For a drug class with average Medicare market share (41%, in 2004–2005), Duggan and Scott Morton's result translates to an 11% increase in revenues following Medicare Part D. Their Phase I estimates correspond, for a drug class with average Medicare market share, to a 26% increase for 2004–2005, a 33% increase post-implementation in 2006–2007, and a lagged 51% increase in 2008–2010. These estimates imply an elasticity of Phase I clinical trials of 2.4 to 4.7 compared to the market size, bracketing Acemoglu and Linn's estimated elasticity of 3.5 for approved new molecular entities (NMEs). However, when considering all clinical trials combined—including Phase III trials for supplemental indications the estimated elasticity of clinical trials with respect to market size is somewhat lower than Acemoglu and Linn's estimated elasticity of 6 for all new drug approvals, but certainly still more prominent than the Dubois et al. (2015) estimate of about 0.25. To summarize *"The results indicate that the increase in outpatient prescription drug coverage provided through Medicare Part D has had a significant impact on pharmaceutical R&D"*

Critiques: ^{c_1} Blume-Kohout and Sood (2013) states that several of the countries chosen regulate prescription drug prices, and regulations may change rapidly over time. Thus, given the lower expected profit per consumer and greater uncertainty about future profits and prices, firms' R&D decisions are likely to be less responsive to a unit change in expected revenues for all these countries combined versus the same unit change in the US market (Sood et al., 2009).

^{c_2} Blume-Kohout and Sood (2013) measured firms' innovative activities via clinical trials, whereas Dubois et al. (2015) and Acemoglu and Linn (2004) evaluate the responsiveness of approved and marketed drugs to changes in market size. Use of demographic shifts.

^{c_3} Dubois et al. (2015): the authors recognize to Blume-Kohout and Sood (2013) the fact of having exploited an innovative measure of Market Share (policy change in Medicare Part D). Use of demographic shifts.

List of controls:

Acemoglu and Linn (2004) Potential Supply-Side Determinants of Innovation (changes in scientific incentives); Proxies for pre-existing time trends across sectors; lag dependent var; life-years lost; public funding; pre-existing trends; major category trends; health insurance market size; (see page 1077-1080 for further details on variables).

Cerda (2007): Gov. expenditure (Medicare and social security); Gov. research efforts (grants on research); year dummies; some demographic information such as prevalence rates of disease *i* on males (fraction of males/white/married attending hospital due to *i*), blacks, whites, and married individuals as well as the average age of individuals affected by disease *i*.

Rake (2017) The empirical analysis draws upon the literature concerning the "demand-pull" versus "technology-push" debate and takes into account demand- and supply-side factors as the explanatory variables for pharmaceutical innovation. Regressors used comprise knowledge stock (consisting of the scientific publications (Pub_{it}) related to medical indication i and published in year t (BioPharmInsight database); Regulatory stringency (average time between the submission of a new drug approval to the FDA and its final approval); pre-sample mean of new pharmaceuticals; mortality rate per medical indication in 1983 to account for differences in the pre-sample prevalence of medical indication; pre-sample technological opportunities are constructed as the average annual growth rate of the knowledge stock from 1979 to 1983.

Blume-Kohout and Sood (2013) prescription drugs; funding grants for each disease class.

B Control-function identification and Mundlak representation

This appendix provides the formal justification for the control-function fixed-effects Poisson estimator used in the main text. The goal is to show why including the first-stage residual in the Poisson mean function corrects for idiosyncratic endogeneity of market size, while ATC-3 fixed effects account for time-invariant heterogeneity.

B.1 Baseline count model with endogenous market size

Let N_{it} denote the number of active trials in ATC-3 market i and year t. The structural conditional mean is

$$\mathbb{E}[N_{it} \mid M_{it}, z_{it}, c_i, \kappa_{it}] = c_i \exp\left(\beta_1 M_{it} + z_{it}'\gamma + \kappa_{it}\right), \tag{15}$$

where M_{it} is log market size, z_{it} is a vector of exogenous controls, c_i is time-invariant ATC-specific heterogeneity, and κ_{it} is an idiosyncratic error term. Endogeneity arises when M_{it} is correlated with κ_{it}, even after conditioning on c_i and z_{it}.

If κ_{it} were absent, or at least conditionally mean independent of M_{it}, then a standard fixed-effects Poisson quasi-maximum likelihood estimator would be consistent. The difficulty in our setting is that market size is an equilibrium outcome and may co-move with time-varying unobservables affecting innovation.

To address this issue, let the reduced form for market size be

$$M_{it} = z_{it}'\Pi + R_{it}'\delta + c_{i2} + u_{it2}, \tag{16}$$

where R_{it} is the vector of excluded instruments, here the contemporaneous and lagged normalized major recalls, c_{i2} is a market-specific effect, and u_{it2} is the reduced-form idiosyncratic component.

The central control-function assumption is that, conditional on the reduced-form disturbance u_{it2}, the remaining dependence between M_{it} and κ_{it} is removed. Following Lin and Wooldridge

(2019), this can be represented by the index restriction

$$\mathbb{E}[\exp(\kappa_{it}) \mid u_{it2}, z_i, c_i, c_{i2}] = \exp(\psi + \rho u_{it2}), \tag{17}$$

for some constants ψ and ρ. Under this formulation, u_{it2} summarizes the idiosyncratic endogeneity of M_{it}.

It is useful to distinguish two sources of endogeneity. First, M_{it} may be correlated with the time-invariant heterogeneity c_i. This is the standard correlated random-effects problem and is handled by fixed effects. Second, M_{it} may be correlated with the idiosyncratic component κ_{it}. Fixed effects do not eliminate this second source of endogeneity, because it varies over time within ATC-3 markets. This is why an additional control-function correction is needed.

In other words, fixed effects purge the between-market component of endogeneity, while the first-stage residual purges the within-market idiosyncratic component.

B.2 Mundlak representation

A convenient way to formalize the control-function argument is through a Mundlak decomposition. Write the reduced form for M_{it} as

$$M_{it} = \theta + z_{it}'\Pi + R_{it}'\delta + \bar{z}_i'\Phi_z + \bar{R}_i'\Phi_R + v_{it2}, \tag{18}$$

where $\bar{z}_i$ denotes the vector of within-market time averages of the exogenous regressors, and v_{it2} is the residual from the pooled projection on z_{it} and $\bar{z}_i$.

The Mundlak device projects the unobserved heterogeneity onto the time averages of the regressors. In particular, one may write

$$c_i = \eta + \bar{z}_i'\lambda + \bar{v}_{i2}\pi + a_i, \tag{19}$$

where a_i is orthogonal to $\bar{z}_i$ and $\bar{v}_{i2}$ by construction. This decomposition is useful because it separates the time-invariant part of the dependence structure from the within-market residual variation.

Combining (17) with (18) and (19), the conditional mean of N_{it} can be written as

$$\begin{aligned}\mathbb{E}[N_{it} \mid M_{it}, z_{it}, c_i, v_{it2}] &= c_i \exp\big(\beta_1 M_{it} + z_{it}'\gamma\big)\, \mathbb{E}[\exp(\kappa_{it}) \mid v_{it2}, z_i, c_i] \\ &= g_i \exp\big(\beta_1 M_{it} + z_{it}'\gamma + \rho v_{it2}\big), \end{aligned} \tag{20}$$

where g_i absorbs the time-invariant terms. Therefore, once v_{it2} is included in the Poisson mean, the remaining endogeneity in M_{it} is controlled for.

In practice, the residual included in the second-stage Poisson regression is obtained from the fixed-effects first stage. Let

$$\ddot{M}_{it} = \ddot{z}_{it}'\Pi + \ddot{R}_{it}'\delta + \ddot{u}_{it2}, \tag{21}$$

where double dots denote deviations from ATC-specific means. The estimated within residual is

$$\hat{\ddot{u}}_{it2} = \ddot{M}_{it} - \ddot{z}_{it}'\hat{\Pi} - \ddot{R}_{it}'\hat{\delta}. \tag{22}$$

Because the within transformation removes the time-invariant component, $\hat{\ddot{u}}_{it2}$ is the empirical counterpart of the within part of v_{it2}. Hence, under the assumptions above, replacing v_{it2} by $\hat{\ddot{u}}_{it2}$ yields a valid control function in the fixed-effects Poisson specification.

B.3 Second-stage estimating equation

The second-stage conditional mean therefore becomes

$$\mathbb{E}[N_{it} \mid M_{it}, z_{it}, c_i, \hat{\ddot{u}}_{it2}] = c_i \exp\big(\beta_1 M_{it} + z_{it}'\gamma + \rho \hat{\ddot{u}}_{it2}\big). \tag{23}$$

Equation (23) is estimated by fixed-effects Poisson QMLE. A Wald test of

$$H_0 : \rho = 0$$

provides a direct test of idiosyncratic endogeneity. Failure to reject H_0 would support the standard fixed-effects Poisson model without a control function; rejection implies that market size is endogenous through time-varying unobservables and that the control-function correction is required.

In our setting, M_{it} is log class-level sales and the excluded instruments R_{it} are contemporaneous and lagged normalized major recalls. The first stage isolates the component of market size shifted by major recalls, while the residual $\hat{u}_{it2}$ captures the remaining endogenous variation in revenues. Including this residual in the Poisson innovation equation corrects for the correlation between observed market size and unobserved profitability shocks that also affect trial activity.

Accordingly, the coefficient β_1 in (23) can be interpreted as the causal semi-elasticity of clinical-trial activity with respect to market size, under the maintained assumptions that recalls are relevant instruments and satisfy the exclusion restriction discussed in the main text.

C Time-to-event analysis

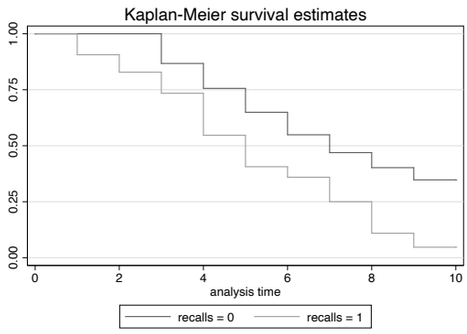


Figure Appx.1: Kaplan-Meyer time-to-event analysis. The x-axis represents the number of years until death. Recalled products (recalls =1) are already scaled down at 2 years of survival time with respect to not recalled products (recalls =0). The median survival time for recalled medicines is about 4 years, while the one for not-recalled medicines is about 7 years. The survival function of recalled products persists in its falling below the survival function of not recalled drugs. This means that recalls affect sales for a long period of time. In other words, within the market of the recalled product there will be a lack of potential sales left by the recalled products. Missed sales are hence not a temporary event. This demonstrates the length of the lack that should be covered to fill the gap provoked by the product's recall.

D Abnormal values (firm and product levels)

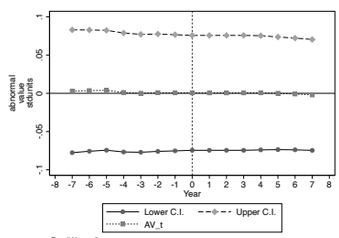
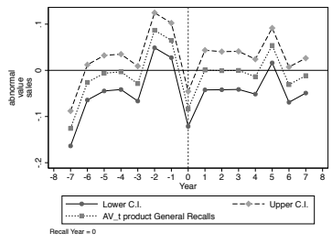
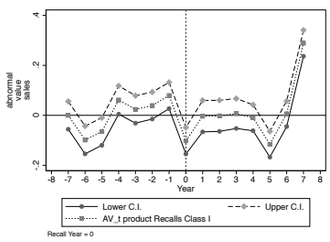


Figure Appx.2: Effect of recalls on market sales once firms having undergone a major recall are cancelled out. The absence of any effect (i.e., increases of sales due to recalls of competitors) at recall time for products other than the ones of the recalled firm witnesses the absence of compensations both at time 0 or soon after the recall.

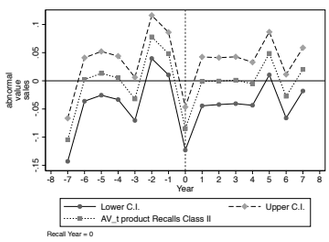
The following figures represent abnormal values at firm and product level for different typologies of recall (according to their gravity).



Product: general recalls sales

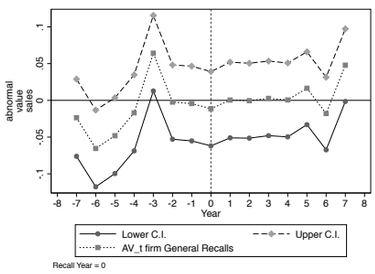


Product: Class I recalls sales

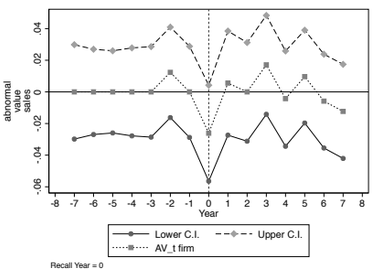


Product: Class II recalls sales

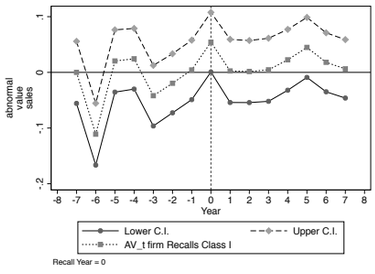
Figure Appx.3: Abnormal values for product aggregation. Years are normalised. Year 0 represents the year of recall. The three scenarios include the path of sales before and after the recall year, using three different definitions of recalls: Class I recalls, general recalls and Class II recalls. As shown in the pictures, sales at product level drop at recall year.



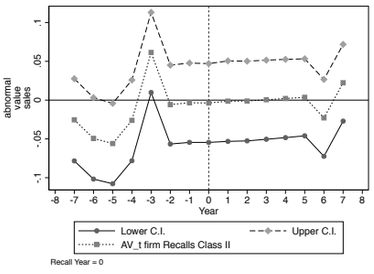
Firm: general recalls sales



Firm: Maj.recalls sales



Firm: Class I recalls sales



Firm: Class II recalls sales

Figure Appx.4: Abnormal values at firm aggregation. Years are normalized. Year 0 represents the year of the recall. The four scenarios include the sales path before and after the recall year, using four different definitions of recalls: major, Class I, general, and Class II. Apart from major recalls, catching firms unaware, the other types of recalls do not affect firms' sales. This might be due to compensation of sales within firms.

Below are reported summary statistics at the firm and product level. The results show how recalls are more common in old established firms where, however, the effect is null due to compensations.

Table Appx.3: Summary statistics at product and firm level. Databases at the product and firm level are unbalanced.

		Prod.			Firm		
Variable		Full Sample	Subs. recalls	Subs. no recalls	Full Sample	Subs. recalls	Subs. no recalls
Sales (log)	Overall mean	12.868	15.872	12.844	14.509	19.808	14.146
	Overall Std. Dev.	3.791	3.674	3.787	4.064	3.278	3.826
	Between Std. Dev	3.657	3.367	3.653	4.080	3.565	3.873
	Within Std. Dev	1.764	1.957	1.764	1.459	1.080	1.486
Age prod./ firm	Overall mean	11.910	13.588	11.909	17.880	31.022	16.850
	Overall Std. Dev.	10.692	11.718	10.687	14.849	15.805	14.024
	Between Std. Dev	9.977	10.821	9.974	13.901	16.723	13.110
	Within Std. Dev	2.748	3.042	2.746	3.022	3.385	2.998
New mol. market (%)	Overall mean	.085	.079	.0851	.	.	.
	Overall Std. Dev.	.279	.271	.279	.	.	.
	Between Std. Dev	.213	.109	.213	.	.	.
	Within Std. Dev	.256	.262	.256	.	.	.
Avg. age prod. by firm/ATC	Overall mean	.	.	.	11.002	11.581	10.921
	Overall Std. Dev.	.	.	.	8.974	6.039	9.130
	Between Std. Dev	.	.	.	8.602	5.858	8.685
	Within Std. Dev	.	.	.	2.985	2.656	3.018
Share generics by firm/ATC	Overall mean	.	.	.	.665	1.534	.623
	Overall Std. Dev.	.	.	.	3.366	3.963	3.361
	Between Std. Dev	.	.	.	3.407	4.486	3.386
	Within Std. Dev	.	.	.	1.668	1.637	1.685
Outflow rate	Overall mean	.	.	.	.086	.085	.086
	Overall Std. Dev.	.	.	.	.226	.148	.231
	Between Std. Dev	.	.	.	.234	.108	.238
	Within Std. Dev	.	.	.	.186	.124	.190

The impact of recalls is observable across all aggregations except at the firm level, where the effect is discernible but not readily apparent, indicating a potential area for future investigation and development. Detailed hypotheses regarding this phenomenon are expounded upon in the main text.

E First stage rob. checks

Significant coefficients from the first stage, using the number of patients as a measure of market size.

Table Appx.4: First stage of the robustness check using the number of patients as a measure of market size

	(1) *# patients*
$\tilde{recalls}$	0.0519
	(0.0333)
$\tilde{recalls}_{t-1}$	-0.262^{*}
	(0.149)
Year Dummies	Yes
Obs.	1056
Groups	132

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Huber-White robust and clustered at ATC-3 level standard errors are in parentheses. (1) First stage results when the MEPS database is employed, and the market size is measured by the number of patients. Second stage results are in Table 8.

References

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

A. Acquatella, K. M. Ericson, and A. Starc. Lagged-price reimbursement contracts: The impact of medicare part b on pharmaceutical price growth. *Journal of Public Economics*, 256:105595, 2026.

R. Agarwal and P. Gaule. What drives innovation? lessons from covid-19 r&d. *Journal of Health Economics*, 82:102591, 2022.

K. J. Arrow. The economic implications of learning by doing. *The review of economic studies*, 29 (3):155–173, 1962.

S. Ayerst. Distorted technology adoption. *The Economic Journal*, 135(668):1167–1190, 2025.

R. Bala, P. Bhardwaj, and P. K. Chintagunta. Pharmaceutical product recalls: Category effects and competitor response. *Marketing Science*, 36(6):931–943, 2017.

N. Balasubramanian and J. Lee. Firm age and innovation. *Industrial and Corporate Change*, 17(5): 1019–1047, 2008.

G. Ball, J. T. Macher, and A. D. Stern. Recalls, innovation, and competitor response: Evidence from medical device firms. *Harvard Business School Working paper*, (Working Paper 19-028), 2018a.

G. P. Ball, R. Shah, and K. D. Wowak. Product competition, managerial discretion, and manufacturing recalls in the us pharmaceutical industry. *Journal of Operations Management*, 58:59–72, 2018b.

M. F. Bellemare, T. Masaki, and T. B. Pepinsky. Lagged explanatory variables and the estimation of causal effect. *The Journal of Politics*, 79(3):949–963, 2017.

E. R. Berndt and J. P. Newhouse. Pricing and reimbursement in us pharmaceutical markets. *The Oxford Handbook of the Economics of the Biopharmaceutical Industry*, page 201, 2012.

J. M. Bertoni, J. P. Arlette, H. H. Fernandez, C. Fitzer-Attas, K. Frei, M. N. Hassan, S. H. Isaacson, M. F. Lew, E. Molho, W. G. Ondo, et al. Increased melanoma risk in parkinson disease: a prospective clinicopathological study. *Archives of Neurology*, 67(3):347–352, 2010.

M. E. Blume-Kohout and N. Sood. Market size and innovation: Effects of medicare part d on pharmaceutical research and development. *Journal of Public Economics*, 97:327–336, 2013.

A. Bonaime and Y. E. Wang. Mergers, product prices, and innovation: evidence from the pharmaceutical industry. *The Journal of Finance*, 79(3):2195–2236, 2024.

R. A. Cerda. Endogenous innovations in the pharmaceutical industry. *Journal of Evolutionary Economics*, 17(4):473–515, 2007.

J. Cheng. An antitrust analysis of product hopping in the pharmaceutical industry. *Columbia Law Review*, 108:1471, 2008.

A. Civan and M. T. Maloney. The effect of price on pharmaceutical r&d. *The BE Journal of Economic Analysis & Policy*, 9(1), 2009.

W. M. Cohen and D. A. Levinthal. Innovation and learning: the two faces of r & d. *The economic journal*, 99(397):569–596, 1989.

A. Denis, L. Mergaert, C. Fostier, I. Cleemput, and S. Simoens. Issues surrounding orphan disease and orphan drug policies in europe. *Applied health economics and health policy*, 8(5):343–350, 2010.

J. A. DiMasi, R. W. Hansen, and H. G. Grabowski. The price of innovation: new estimates of drug development costs. *Journal of Health Economics*, 22(2):151–185, 2003.

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

D. Dranove, C. Garthwaite, and M. Hermosilla. Pharmaceutical profits and the social value of innovation. Technical report, National Bureau of Economic Research, 2014.

E. A. Drotleff. The medicare part d prescription drug benefit: Who wins and who loses. *Marq. Elder's Advisor*, 8:127, 2006.

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

M. Duggan and F. Scott Morton. The effect of medicare part d on pharmaceutical prices and utilization. *American Economic Review*, 100(1):590–607, 2010.

M. G. Duggan and F. S. Morton. The medium-term impact of medicare part d on pharmaceutical prices. *American Economic Review*, 101(3):387–392, 2011.

A. G. Dupont and P. B. Van Wilder. Access to orphan drugs despite poor quality of clinical evidence. *British Journal of clinical pharmacology*, 71(4):488–496, 2011.

J. Feng, T. Hwang, Y. Liu, and L. Maini. Mergers that matter: The impact of m&a activity in prescription drug markets. *Journal of Political Economy: Microeconomics (forthcoming)*, 2024.

L. Gallelli, C. Palleria, A. De Vuono, L. Mumoli, P. Vasapollo, B. Piro, and E. Russo. Safety and efficacy of generic drugs with respect to brand formulation. *Journal of Pharmacology & Pharmacotherapeutics*, 4(Suppl1):S110, 2013.

M. C. Garassino, K. Odunsi, M. Siniscalchi, and P. Veronesi. On the economic infeasibility of personalized medicine, and a solution. Technical Report 33539, National Bureau of Economic Research, 2025.

P. A. Geroski and C. F. Walters. Innovative activity over the business cycle. *The Economic Journal*, 105(431):916–928, 1995.

C. Giaccotto, R. E. Santerre, and J. A. Vernon. Drug prices and research and development investment behavior in the pharmaceutical industry. *The Journal of Law and Economics*, 48(1):195–214, 2005.

S. Glied and P. C. Smith. *The Oxford handbook of health economics*. Oxford University Press, 2013.

H. Grabowski. Patents, innovation and access to new pharmaceuticals. *Journal of International Economic Law*, 5(4):849–860, 2002.

B. H. Hall and N. Rosenberg. *Handbook of the Economics of Innovation*, volume 1. Elsevier, 2010.

C. Hansen, J. Hausman, and W. Newey. Estimation with many instrumental variables. *Journal of Business & Economic Statistics*, 26(4):398–422, 2008.

I. Hashi and N. Stojčić. The impact of innovation activities on firm performance using a multi-stage model: Evidence from the community innovation survey 4. *Research Policy*, 42(2):353–366, 2013.

J. A. Hausman, B. H. Hall, and Z. Griliches. Econometric models for count data with an application to the patents-r&d relationship, 1984.

V. J. Hawk, B.E. and H. H.L. Recent developments in eu merger control. *Antitrust*, (15):24, 2000.

J. Heath. Big pharma's lobby: When public health and profit conflict. *Harvard Health Policy Review*, 14(1), 2013.

E. Huergo and J. Jaumandreu. How does probability of innovation change with firm age? *Small Business Economics*, 22(3-4):193–207, 2004.

T. J. Hwang, A. S. Kesselheim, and B. N. Rome. New reforms to prescription drug pricing in the us: opportunities and challenges. *JAMA*, 2022.

A. Janssen. Generic and branded pharmaceutical pricing: Competition under switching costs. *The Economic Journal*, 133(653):1937–1967, 2023.

B. Jovanovic. Product recalls and firm reputation. Technical report, National Bureau of Economic Research, 2020.

B. Jovanovic. Product recalls and firm reputation. *American Economic Journal: Microeconomics*, 13(3):404–442, 2021.

S. Keyhani, M. Diener-West, and N. Powe. Are development times for pharmaceuticals increasing or decreasing? *Health Affairs*, 25(2):461–468, 2006.

A. Kleinknecht and B. Verspagen. Demand and innovation: Schmookler re-examined. *Research Policy*, 19(4):387–394, 1990.

S. Klepper and F. Malerba. Demand, innovation and industrial dynamics: an introduction. *Industrial and Corporate Change*, 19(5):1515–1520, 2010.

S. Kolluru and P. Mukhopadhaya. Empirical studies on innovation performance in the manufacturing and service sectors since 1995: A systematic review. *Economic Papers: A Journal of Applied Economics and Policy*, 36(2):223–248, 2017.

J. L. Krieger, X. Li, and R. T. Thakor. Find and replace: R&d investment following the erosion of existing products. *Management Science*, 2022.

M. K. Kyle and A. M. McGahan. Investments in pharmaceuticals before and after trips. *Review of Economics and Statistics*, 94(4):1157–1172, 2012.

J. O. Lanjouw. Patents, price controls, and access to new drugs: how policy affects global market entry. Technical report, National Bureau of Economic Research, 2005.

F. R. Lichtenberg. Pharmaceutical innovation as a process of creative destruction. *Knowledge Accumulation and Industry Evolution: The Case of Pharma-Biotech*, page 61, 2006.

W. Lin and J. M. Wooldridge. Testing and correcting for endogeneity in nonlinear unobserved effects models. In *Panel Data Econometrics*, pages 21–43. Elsevier, 2019.

C.-j. Luan, C. Tien, and Y.-c. Chi. Downsizing to the wrong size? a study of the impact of downsizing on firm performance during an economic downturn. *The International Journal of Human Resource Management*, 24(7):1519–1535, 2013.

F. Malerba. Innovation and the evolution of industries. In *Innovation, Industrial Dynamics and Structural Transformation*, pages 7–27. Springer, 2007.

A. Markham. Lurbinectedin: first approval. *Drugs*, pages 1–9, 2020.

L. Martin, M. Hutchens, C. Hawkins, and A. Radnov. How much do clinical trials cost?, 2017.

K. Mellahi and A. Wilkinson. A study of the association between level of slack reduction following downsizing and innovation output. *Journal of Management Studies*, 47(3):483–508, 2010.

D. Mowery and N. Rosenberg. The influence of market demand upon innovation: A critical review of some recent empirical studies. *Research Policy*, 8(2):102–153, 1979.

D. C. Mowery and N. Rosenberg. *Paths of innovation: Technological change in 20th-century America*. Cambridge University Press, 1999.

F. Nutarelli. At the intersection between machine learning and econometrics: theory and applications ((unpublished doctoral dissertation). *IMT Lucca for Advanced Studies*, 2021.

I. J. Onakpoya, C. J. Heneghan, and J. K. Aronson. Worldwide withdrawal of medicinal products because of adverse drug reactions: a systematic review and analysis. *Critical Reviews in Toxicology*, 46(6):477–489, 2016.

A. Pakes and M. Schankerman. The rate of obsolescence of patents, research gestation lags, and the private rate of return to research resources. In *R&D, Patents, and Productivity*, pages 73–88. University of Chicago Press, 1984.

F. Pammolli, L. Magazzini, and M. Riccaboni. The productivity crisis in pharmaceutical r&d. *Nature Reviews Drug Discovery*, 10(6):428–438, 2011.

B. Rake. Determinants of pharmaceutical innovation: the role of technological opportunities revisited. *Journal of Evolutionary Economics*, 27(4):691–727, 2017.

D. Roodman. How to do xtabond2: An introduction to difference and system gmm in stata. *The Stata Journal*, 9(1):86–136, 2009.

F. M. Scherer. Demand-pull and technological invention: Schmookler revisted. *The Journal of Industrial Economics*, pages 225–237, 1982.

J. Schmookler. *Invention and Economic Growth*. Harvard University Press, 2013.

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

G. Siotis, C. Ornaghi, and M. Castanheira. *Market definition and competition policy enforcement in the pharmaceutical industry*. ECARES, 2020.

V. B. Siramshetty, J. Nickel, C. Omieczynski, B.-O. Gohlke, M. N. Drwal, and R. Preissner. Withdrawn—a resource for withdrawn and discontinued drugs. *Nucleic Acids Research*, 44(D1): D1080–D1086, 2016.

G. N. Stock, N. P. Greis, and W. A. Fischer. Firm size and dynamic technological innovation. *Technovation*, 22(9):537–549, 2002.

P. Stoneman. *Soft Innovation: Economics, Product Aesthetics, and the Creative Industries*. Oxford University Press, 2010.

J. Sutton. *Technology and market structure: theory and history*. MIT press, 2001.

G. Symeonidis. Innovation, firm size and market structure: Schumpeterian hypotheses and some new themes. 1996.

M. Thill-Tayara and M. Provost. Dominance in the pharmaceutical sector: An overview of eu and national case law. *e-Competitions Bulletin*, (Pharma & Dominance), 2020.

S. Thirumalai and K. Sinha. Product recalls in the medical device industry: An empirical exploration of the sources and financial consequences. *Management Science*, 57:376–392, 2011.

A. Vaishnav. Product market definition in pharmaceutical antitrust cases: Evaluating cross-price elasticity of demand. *Columbia Business Law Review*, page 586, 2011.

Ž. Vujović. A case study of the application of weka software to solve the problem of liver inflamation. 2021.

WHO. The anatomical therapeutic chemical classification system with defined daily doses-atc/ddd. 2009.

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