

Product recalls, market size and innovation in the pharmaceutical industry

Federico Nutarelli*¹ and Massimo Riccaboni^{† ‡ 2}

^{1,2}AXES Group, IMT School for Advanced Studies

¹Invernizzi Center for Research and Innovation, Bocconi University

Abstract

This study revisits the well-documented premise that research investment is motivated by market incentives, a cornerstone of the literature on innovation economics (Schmookler, 1966; Acemoglu & Linn, 2004; Bryan & Williams, 2021). While it's recognized that demographic shifts, indicative of market size changes, correlate with an uptick in pharmaceutical innovations (Acemoglu & Linn, 2004; Dubois et al., 2015), discussions around reverse causality persist (Cerdeira et al., 2007). Our investigation contributes to this dialogue by examining how variations in market size influence innovation activity, specifically through the lens of active clinical trials within the U.S. market. In this paper, we analyze the impact of market size on innovation as measured by active clinical trials. We exploit product recalls as an innovative instrument that is found to be sharp, strong, and unexpected. This paper analyzes the relationship between U.S. market size and ATC-3-level innovation using an original dataset and the two-stage IV method proposed by Wooldridge et al. (2019). The results show a robust and significant positive response of the number of active trials to market size. Beyond reaffirming the nexus between market size and innovation, our findings offer fresh insights into the discourse on strategic investment in niche markets. By demonstrating the pivotal role of market dimensions in driving innovation, especially in fields poised for significant breakthroughs, this paper underscores the critical need for nuanced investment strategies. Such strategies not only cater to broad markets but also prioritize innovation in specialized areas, potentially catalyzing groundbreaking advancements in underserved sectors.

Keywords: Product Recalls; Innovation; Industrial Organisation; Pharmaceutical industry.
JEL Codes: O31, J10, J20, I11, L11

1 Introduction

The relationship between market rewards and innovation has been extensively studied in innovation economics (Scherer (1982), Schmookler (2013), Klepper and Malerba (2010)). This has opened up the possibility of redesigning public demand incentives to encourage innovation, such as in the case of orphan drugs. Schmookler's "demand-pull" hypothesis, which states that innovation is a function of market demand, has been challenged over the years. As early as the 1990s, Kleinknecht and Verspagen (1990) noted that the direction of causality between market size and innovation seems far from obvious. In particular, the authors suspected that there was a simultaneous relationship between demand and innovation, but could not control for it. More recently, Stoneman (2010) and Ball et al. (2018b) have developed more rigorous methods to detect this type of endogeneity.

*Corresponding author: federico.nutarelli@unibocconi.it

[†]massimo.riccaboni@imtlucca.it

[‡]The authors are grateful to Crisis Lab. 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 would also like to thank the FDA for providing data on recalls and clearing any doubt about recall timing and procedures. We are grateful to Evaluate data for allowing very up-to-date and detailed checks on clinical trial data. Furthermore, we are thankful to Springer AdInsight for enabling a 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, Dr. Armando Rungi, Dr. Francesco Serti, Dr. Laura Magazzini from Sant'Anna School for Advanced Studies (Pisa), the ASSA 2022, DRUID 2022, Unicatt 2023 and Research Policy Online Conference for Early Career Researchers participants for the valuable comments on the work.

Acemoglu and Linn (2004) developed a strategy to overcome the endogeneity bias at the market level. Specifically, they exploited changes in the market size for different drug categories driven by US demographic trends (Acemoglu and Linn (2004)). After the contribution of Acemoglu and Linn (2004), the focus moved from ascertaining the presence of the reverse causality of market size and innovation to detecting the best instrument for market size. Indeed, the instrument adopted in Acemoglu and Linn (2004) was later criticized by Cerda (2007) as itself endogenous. As detailed in Cerda (2007), although pharmaceutical innovation is increasing the age of patients, the fact that the average age of patients is increasing means that more patients need innovative products. This scenario also raises the problem of reverse causality: demographic trends affect market size, which affects innovation, and the latter in turn affects demographic trends. To the best of our knowledge, this gap has not yet been closed in the literature.

Inspired by Acemoglu and Linn (2004), most authors explored the relationship between market size and innovation by focusing on the pharmaceutical industry. The latter, indeed provides an ideal setting. In this sector, consumer needs are diverse and almost constant over time, so the market can be divided into independent submarkets based on patient needs (Bertoni et al., 2010; Sutton, 2001). In addition, investment in innovation is essential to the industry's existence. As one of the key outputs of the pharmaceutical industry, innovation is also relatively easy to measure. In the pharmaceutical industry, relevant submarkets are usually identified on the basis of the Anatomical Therapeutic Chemical (ATC) classification system, i.e., a drug classification system classifying *"the active ingredients of drugs according to the organ or system on which they act and their therapeutic, pharmacological, and chemical properties"* (WHO, 2009).

This paper analyzes the relationships between market size and innovation at the ATC-3 level by instrumenting U.S. market size with product recalls of drugs by the Food and Drug Administration (FDA). This work contributes to the literature in several ways.

First, and critically, we employ product recalls as an instrument to address the issue of endogeneity in measuring market size, an aspect often sidestepped in previous studies. Unlike the traditional Poisson approach, this study introduces a Control Function (CF) IV strategy, specifically harnessing the unpredictability and sharp nature of product recalls. The core of this improvement lies in the innovative utilization of recalls to instrument market size, thus directly tackling the endogeneity challenge. This approach is premised on the exogeneity of recalls, rooted in the fact that firms have no foresight over competitors' recalls at the market level. The sudden and unanticipated nature of significant recalls, which are not necessarily more prevalent in "riskier" ATC markets, offers a unique vantage point. Detailed in Section 4, our arguments underscore the validity of considering major recalls as sharp and unanticipated, thereby enhancing the robustness of our methodology. The significance of leveraging recalls in this manner fills a notable void in the existing literature, as elaborated in Section 2, marking a methodological leap forward.

Secondly, we introduce a novel metric for measuring innovation, specifically, the total number of (active) trials at the ATC-3 level, diverging from traditional indicators like cumulative R&D expenditures or the count of New Molecular Entities (NMEs). This essential modification surpasses the constraints inherent to the previously predominant metrics. R&D expenditures are associated with both a firm's strategic decisions for long-term profitability (Cohen, 2010) and, more critically, its size. Stoneman (2010) and others have posited that smaller, newer market entrants may be more disposed towards R&D investment compared to 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, gauged by the aggregate sales of the participating firms. The discourse on NMEs, as also explored in Acemoglu and Linn (2004), presents its own complexities. NMEs are innovative products that contain active moieties (i.e., significant parts of the molecule). These molecule parts were not previously approved by the FDA. Therefore, these may be innovative products that have not yet been utilized in clinical practice. Alternatively, they could also be connected to products that have already received approval. Despite being comprehensive, we believe that the definition of NME does not entirely encompass the market firms' intention to innovate. This is because the point at which NMEs are approved by the FDA, relative to the innovation metric used in this study, is critical. To expand on this matter, we must revisit the drug approval process. Pharmaceutical drug approval is a lengthy process. Initially, firms undergo a pre-clinical phase, a period of research that occurs before clinical trials (human testing)

can commence. During this pre-clinical phase, public agencies gather essential information regarding the trial’s feasibility, conduct iterative testing, and assess the safety of the drug under investigation. The clinical drug development stage consists of three phases. Phase 1 involves the company performing clinical trials on healthy volunteers to establish the drug’s fundamental characteristics and its 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. This phase essentially involves conducting the trial with a larger group of individuals compared to the earlier phase. Subsequently, Phase 3 entails comparing the new drug against a standard-of-care drug, thereby establishing its efficacy and safety. Thus, NMEs are entities approved by the FDA after successfully passing pre-clinical trials. The metric used in this study also accounts for the pre-clinical phase, meaning it includes both successful and potentially unsuccessful trials—those that do not advance to the clinical phase. This approach recognizes even the unsuccessful trials as indicative of a firm’s commitment to innovation ¹.

A further contribution consists in utilizing a more sophisticated segmentation of market categories, specifically through ATC-3 classifications, as opposed to the broader classifications commonly referenced in existing literature. The data categorized at the ATC-3 level precisely reflects the structure of sub-markets, which are often synthetically constructed or overlooked in the scholarly discourse. Traditionally, literature has predominantly focused on the ATC-1 and ATC-2 levels for analysis, with the ATC-3 level mainly serving as a basis for comparison with these broader classifications. Importantly, the adoption of the ATC-3 level in our analysis is not only for its finer granularity but also because it aligns with the categorizations used by antitrust agencies. This alignment underscores the relevance of the ATC-3 level for deriving policy implications, offering a framework that mirrors the operational parameters considered in regulatory assessments. For more detailed information, refer to Section 3.

Finally and importantly, we specifically address concerns in recent literature about the overestimation of market size effects due to a focus on trials that do not necessarily result in innovative outcomes. As anticipated, previous studies typically employed As anticipated, previous studies often used NMEs or NCEs as dependent variables to measure innovation, but these metrics may not fully capture the timing and scope of market-driven innovation incentives due to their focus on long-term, successful outcomes only—ignoring earlier-stage efforts and failures, which are also shaped by market size. In one of our preferred specifications we address potential biases by including an examination of only those trials that resulted in marketed innovations, revealing an 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 (Nelson (1962), Arrow (1962), Cohen and Levinthal (1989), Mowery and Rosenberg (1999), and Romer (2000)). The latter advocate for an enhanced focus on empirical studies that not only trace the lineage of innovation from conception to commercialization but also measure the economic impact of those innovations that culminate in successfully marketed products, thereby shaping industry standards and driving economic growth.

The utilization of recalls as an instrument represents a significant aspect. Product recalls are a particular form of supply disruption resulting from regulatory oversight.

The FDA refers to a recall as *"the most effective way to protect the public from a defective or potentially harmful product. A recall is a voluntary action taken by a company to remove a*

¹The length of the approval process is also an argument in favor of the exogeneity of recalls. In this analysis, we conduct our work at the ATC level, and we argue that a major recall at the ATC level occurs randomly. By this, we mean 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 there were prior suspicions by the regulator, 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. Indeed, 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 particular sector. In essence, they may delay innovation due to the recall. The underlying hypothesis is that recalls at $t - 1$ could influence firm behavior, particularly in terms of their investment in innovation. One possible resolution to this issue lies in the understanding 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 compel a firm to halt clinical trials that are already in progress, especially considering the substantial sunk costs associated with such trials. Subsequent analyses validate this argument.

defective drug product from the market. Drug recalls are conducted either on a company's initiative or by FDA request" (FDA U.S. Food & Drug (2019)). In a recall, the FDA's role is to oversee a company's strategy, assess the recall's adequacy, and classify the recall. According to their severity, the FDA classifies recalls into Class I (more severe), and Class II and Class III (least severe). Medicines may be recalled for several reasons, ranging from health hazards to potential contamination, adverse reaction, mislabelling, and poor manufacturing. However, recalls should not be confused with withdrawals. Contrary to the FDA's definition, scholarly literature frequently describes withdrawals as post-marketing recalls initiated by the FDA against companies because of significant severity and risk to human health. Thus, recalls might be anticipated and voluntarily conducted by companies in the case of minor issues, or unexpectedly and mandatorily enforced by the FDA in situations of extreme severity. Essentially, the literature often interprets withdrawals as post-marketing recalls, typically following a severe Class I recall. As highlighted by Onakpoya et al. (2016), despite differences across countries, approximately 72% of medications subject to recall procedures due to adverse effects ultimately lead to a withdrawal². For consistency with our recall data, we observed that the majority of major Class I recalls, specifically those citing "death" as part of their root cause, resulted in a product withdrawal.

Based on the definition of drug recalls, it is anticipated that a recall will lead to a decline in sales within the market. To illustrate this effect, one can examine the notable case of Merck's VIOXX recall in 2004. VIOXX was pulled from the market due to a heightened risk of severe cardiovascular incidents. This recall took both the market and the company by surprise. Following the announcement of VIOXX's recall in September 2004, there was a notable decrease in both Merck's stock value and its sales figures. This decline was widely covered by mainstream media (see, for example, Terence N. (2004), Bowe C. (2005)) and is also well-documented in academic literature (refer to, for instance, Tong et al. (2009) among others)

In this study, we focus on evaluating sharp and unexpected recalls, hereinafter referred to as "major recalls." The criteria for defining major recalls were established by analyzing the causes of Class I recalls. We selected recalls based on the significance of the cause, its potential threat to human life, and the actions taken by the FDA. Specifically, our definition includes Class I recalls and withdrawals that are identified by critical keywords in their causes, such as "contamination," "death/s," "overdose," "symptoms," "particulate matters," and "adverse reaction." We chose not to include Class II recalls in our primary analysis because we believe they represent a less significant indicator compared to major recalls. However, to ensure the robustness of our findings, we have included an analysis encompassing all types of recalls in Table 7. The initial column of Table 7 in the robustness checks demonstrates that using "minor" recalls (primarily Class II recalls related to packaging or labeling issues) as a proxy for market size constitutes an ineffective instrumentation strategy.

Notice that major recalls exert an impact on both the supply and demand sides. By reducing supply, they inherently augment the demand for a drug among patients³ (Azghandi et al. (2018), Hosseini-Motlagh et al. (2021)). While this heightened demand theoretically should lead to increased market rewards, this is not invariably the case within the Pharmaceutical Industry. After a recall, patients still require additional supplies of the recalled medicine, yet market markups might not increase at a general level due either to the presence of generics (Simoens et al. (2005)) or to the fact that only firms that both prescribe and dispense drugs – hence not wholesalers taken into account in the present paper – could take advantage of the mentioned markups (Iizuka (2009)). In other words, a supply-side instrument like recalls can effectively capture demand-side aspects of market size, such as heightened markups or greater patient medicine needs. This is well established in Appendix C and in Section.5.1.2 where it is shown how firms do not internally compensate for supply-side shocks provoked by recalls entailing that patients' needs remain unfulfilled. Thus recalls reflect a demand side shock through this channel.

The existing literature typically estimates the elasticity of market size to innovation at about 0.5. However, when focusing specifically on trials that led to marketed innovations, we discovered a significant overestimation of this elasticity. This overestimation primarily arises from an inadequate handling of the endogeneity issue, which involves the impact of market

²This estimate was computed based on the list provided by Onakpoya et al. (2016).

³Vice versa similar arguments can be made for the demand side instruments commonly adopted by literature for having supply-side consequences.

size on innovation. In light of these findings, our preferred model yields a more conservative estimate of 0.27, taking into account only trials that culminated in a marketed innovation (active trials)⁴.

The organization of this paper is outlined as follows: Section 2 offers a comprehensive review of the existing literature that attempts to quantify the correlation between market size and product innovation, underscoring the limitations inherent in previous methodologies. Section 3 delineates the characteristics of the dataset employed in this study and presents an array of descriptive statistics. In Section 4, we elucidate the econometric techniques employed and introduce the innovative instrument of major product recalls. The findings of this study are delineated in Section 5, which is bifurcated into two segments. The initial segment accentuates the repercussions of major recalls on market size, while the subsequent segment addresses the endogeneity inherent in market size and assesses its association with innovation, supported by robustness checks. In the robustness check subsection we estimated several models that are commonly used in the literature to test –among others– for either delayed effects of trials or the presence of bias if market size is assumed to be exogenous or if fixed effects are omitted. The paper culminates in Section 6, where we draw our conclusions.

2 Literature Review

The importance of market size in explaining the rate of innovation has been acknowledged in the literature for many years. In 1942, Schumpeter indicated that larger firms are more innovative than smaller ones. In the early 1960s, the focus shifted more broadly on the possible effects of demand on market size (see, e.g., Scherer (1982)). It was not yet clear whether the reverse causality of demand and innovation played a relevant role. Scherer (1982), for instance, argued that causality ran primarily from sales to innovation. However, this study has been criticised in several aspects. The definition of demand was still too broad and was not conclusive about the unique sign 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 the pharmaceutical sector nor did it consider the aggregate market level (see, e.g., Pakes and Schankerman (1984)). More recently, Kleinknecht and Verspagen (1990) denounced a clear reverse causality of demand and innovation, thus invalidating the prior studies. Soon after, Geroski and Walters (1995) empirically verified these conclusions and showed how innovations increase demand 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 Mukhopadhyaya (2017), among others). Specifically, the discussion has produced conflicting results, largely due to challenges in completely eliminating unobservable sources of endogeneity that vary over time. These unobservable factors may arise from strategic decisions within firms, which could be influenced by their 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 focus, 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. To the best of our knowledge, however, among the innovation measures, no work has exploited

⁴Refer, for example, to Aghion et al. (2022), which underscored how patented innovations often require a time lag to develop following a stimulus, thus potentially resulting in a diminished response to market size, a factor not accounted for in previous studies.

Investigational New Drugs (INDs) and early-stage clinical trials (i.e., pre-clinical and Phase I) together with Phase II and Phase III trials.

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 scenarios of market imperfection, current revenues (indicative of 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, authors 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 how although pharmaceutical firms made substantial investments in R&D, they did not produce innovation⁵. Other measures of innovation include clinical trials (see Kyle and McGahan (2012) among others) and changes in Medicare part D. Medicare part D is an optional US programme to help beneficiaries of the Medicare national health insurance to pay for self-administered prescription drugs. 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 suggested 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."* (see Blume-Kohout and Sood (2013) for further details⁶). Thus, Medicare Part-D may have boosted not only current sales but also, crucially, future sales through R&D expenditures. Researchers have observed that innovation positively reacts to fluctuations in market size. However, a persistent issue has been the simultaneous response of innovation to both the immediate and anticipated cash flows resulting from changes in market size. Consequently, it became challenging to differentiate between the immediate and long-term impacts within the estimated coefficient for market size. Furthermore, as highlighted in Section 1, it is crucial to note that changes in Medicare Part-D could be subject to endogeneity, stemming from lobbying activities by pharmaceutical firms.

In the present work, we ensured that our measure of innovation did not influence both current and future sales by including both successful innovation-generating projects and unsuccessful ones across all clinical and pre-clinical phases. This methodological enhancement significantly 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 been also quantified by the number of relevant medical journal articles (Licht-

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

⁶Blume-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 critiqued 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.

enberg (2006)). However, these measures do not consider innovations that occur exclusively within industries and focus mainly on academic and scientific discoveries, thus 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 NMEs, New Chemical Entities (NCEs), or new medicines approved by the FDA. The reasons for choosing active successful and unsuccessful clinical trials over NMEs have been discussed in Section 1. 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 research and development. 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, as observed in several studies within the industrial organization literature (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 how the introduction of new drugs could fundamentally alter market dynamics through their effects on 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.

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). To our knowledge,

(Dubois et al., 2015) and our study are the only ones that have utilized an Instrumental Variables (IV) approach in this area of research, providing a unique perspective that captures the complexities and nuances often overlooked by traditional methodologies. Our analysis, additionally, distinctly quantifies, for the first time, 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. This level of analysis is particularly relevant for antitrust authorities, and to the best of our knowledge, no other studies have specifically targeted this more granular level, which we believe offers a richer and more actionable insight into market dynamics. 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 (NCT), New Molecular Entities (NME), or New Drug Approvals (NDA), focusing excessively on the demand side might overlook critical supply-side dynamics. These dynamics are pivotal as they drive firms to pursue NDAs, NCTs, or NMEs. 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 an NME, NDA, or NCT. 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, 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.

The two more recent estimates of the relationship between market size and innovation have been provided in Rake (2017) and Dubois et al. (2015). The latter 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.

Finally, Rake (2017) adopted several measures of innovation from NCE to clinical trials in Phase II and Phase III. Rake (2017) found 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" (Rake (2017)). 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). Table Appx.1 in the Appendix provides a schematic literature review of previous estimates of the relationship between innovation and market size⁷

3 Data

The sales data we used are sourced from the Evaluate dataset. The control variables were derived from Evaluate, the Pharmaceutical Industry Database (PHID), and the FDA. Specifically, some of the regressors are based on an analysis of variables present in the PHID database. Sales data for the U.S. pharmaceutical market span from 2004 to 2015. Initially, these data were available at the product and molecule levels and were later aggregated at the ATC-3 level. In the ATC classification system, drugs are categorized into five levels: ATC-1 through ATC-5, with higher levels indicating more detailed classification. Acemoglu and Linn (2004) utilized the ATC-1 and ATC-2 categories to define market size. Specifically, they calculated market size by aggregating the average expenditure share of drugs within each ATC-1 (or ATC-2) category across all such categories.

The data available to us capture the varied layers of products within broader categories (ATC-1 and ATC-2), addressing both demand and supply dynamics. Medications classified at the ATC-1 or ATC-2 level cater to patients with vastly different needs, as they are formulated to treat a variety of diseases. A company investing within a specific ATC-2 category may extend its investments across multiple ATC-3 categories. Thus, utilizing ATC-1 or ATC-2 levels alone –as done in Acemoglu and Linn (2004)– might lead to the creation of less informative variables for innovation and market size, as these do not account for a firm’s specialization in specific sub-sectors within the same ATC-2 or ATC-1 class.

In our previous research, we explored various other levels of analysis, such as firm-specific, product-specific, and ATC-firm aggregations (see Nutarelli (2021)). However, for this study, we have chosen to focus on the ATC-3 level, recognizing its significance to antitrust authorities⁸. To illustrate this point, we refer to cases analyzed by scholars such as Provost et al. (2019), Markham (2020), Vaishnav (2011), Hawk et al. (2000), and Cheng (2008), among others, which predominantly relate to mergers and acquisitions, including the case of M.8889 - TEVA / PGT OTC ASSETS from 2018..

We have chosen not to use the ATC-4 level because, at this level of granularity, products within a specific ATC-4 class may not significantly differ from those in another ATC-4 class. This similarity could lead to dependencies between groups (e.g., innovations in one ATC-4 class could influence a closely related ATC-4 class), which might render any statistical inferences invalid. Additionally, at the ATC-4 level, compensation mechanisms between groups could interfere, potentially weakening the effectiveness of the instrumental variables used. Our dataset also includes the launch dates and ATC codes of products. In this study, we have concentrated on analyzing the global sales data of U.S. companies.

Data on NCTs for 2004 to 2015 at the product level come from the ClinicalTrials.gov website, while data on commercial Investigational New Drugs (IND) at product level derive from a Pharmaceutical Industry Database⁹.

Clinical trials are research studies conducted on people to assess medical, surgical, or behavioral interventions. An IND application in clinical trials allows a pharmaceutical company to obtain permission to begin human clinical trials and to transport an experimental drug across state lines before the drug’s marketing application has been approved.

Clinical trials progress through Phase I to Phase IV. Figure 1 presents the annual number of

⁷The 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.

⁸Appendix C provides some results of Nutarelli (2021).

⁹Maintained at IMT Lucca

trials and commercial INDs as reported by the sources mentioned.

The figure also illustrates the expected positive trend in sales within the pharmaceutical industry over time.

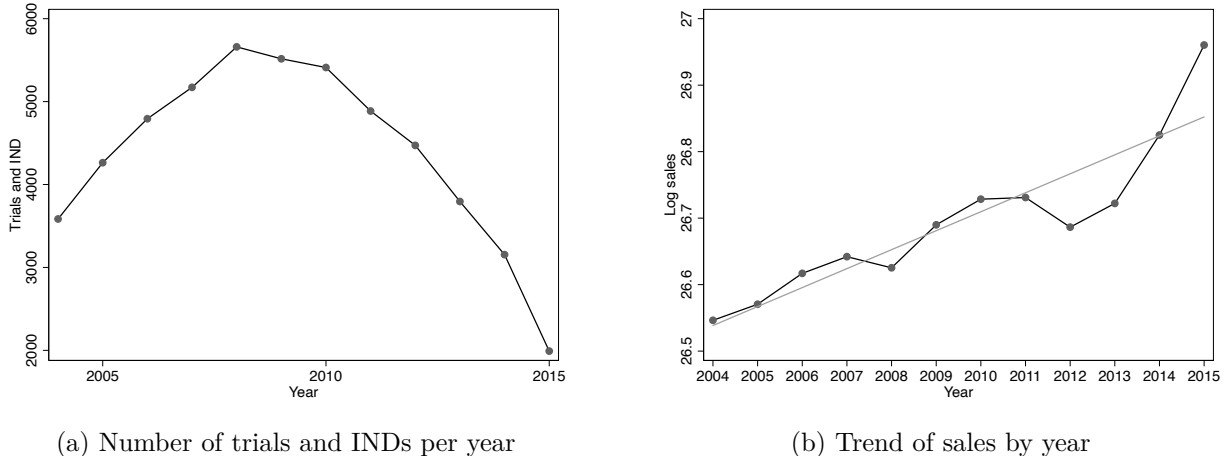


Figure 1: Overview of the trends for sales and trials

A considerable drop in trials and IND occurred after 2013, as is evident from Figure 1 (a). The reason for this drop is that, in general, clinical trials innovate drugs, approaches, and interventions. However, approaches and interventions are excluded from the count of trials to focus strictly on innovation coming from industrial sources which, as noted in Pammolli et al. (2011) have recently undergone a productivity crisis.

Recall data have been manually collected from different sources, including the FDA website, openFDA, various articles, and web sources (e.g., Onakpoya et al. (2016); WHOCC website, PubMed, Siramshetty et al. (2016) and others).

In total, 7.19% of the firm sample (i.e., 697 firms in total) have issued a Class II recall. Among the firms that have issued a recall, 51 firms underwent a recall of Class I, 27 of which issued a single recall of Class I, and just three firms issued more than nine Class I recalls.

The estimates provided must be interpreted in the context of the database’s potential limitations concerning the presence or absence of firms and products within it.

Additionally, the inclusion of recalls from pure compounders was only partial¹⁰, and when included, these recalls were attributed to the sole manufacturer or distributor listed in the database. Furthermore, recalls from repackaging firms were not exclusively attributed to the repackager (e.g., Aidapak) but to the labeler as specified in the National Drug Code (NDC). The necessity to exclude recalls from repackagers and compounders makes it challenging to identify a clear and consistent pattern of recalls over the years. This difficulty is exacerbated by the varying methodologies used by different sources to count recalls. For instance, a comparison of FDA Enforcement Statistics (2015) and Laguna Hospital (2019) highlights this inconsistency. FDA Enforcement Statistics (2015) suggest that the number of recalled products remained relatively stable, except for a decline of about 35% in 2010 and 2013. In contrast, Laguna Hospital (2019) reports a surge in recalls in 2013, with nearly 60 recalls, and in 2017, with 71 recalls, marking almost the highest number since 2009. Only the years 2011 and 2009 had more recalls, with 74 and 75, respectively.

Table 1 provides a list of the primary sources and the average number of recalls across them, as part of an effort to reconcile these discrepancies in data sources.

¹⁰Due to data unavailability. We confirmed that the representativeness of the recalls is maintained Nutarelli (2021).

Year	CNN	Regulatory Focus	Hall et al. (2016)	FDA Enforcement Reports	Laguna Hospital (2019)	AVERAGE
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

Table 1: Consulted sources with reported number of recalls

To address the discrepancies in the data sources of Table 1, we used the average number of recalls as a benchmark for comparison against the collected recall data. Figure 2 provides a detailed comparison between the benchmark recalls, represented by the average recalls from all sources, and the collected recalls. The recalls from Aidapak in 2011 are considered outliers and, therefore, are excluded from the collected recalls at this stage:

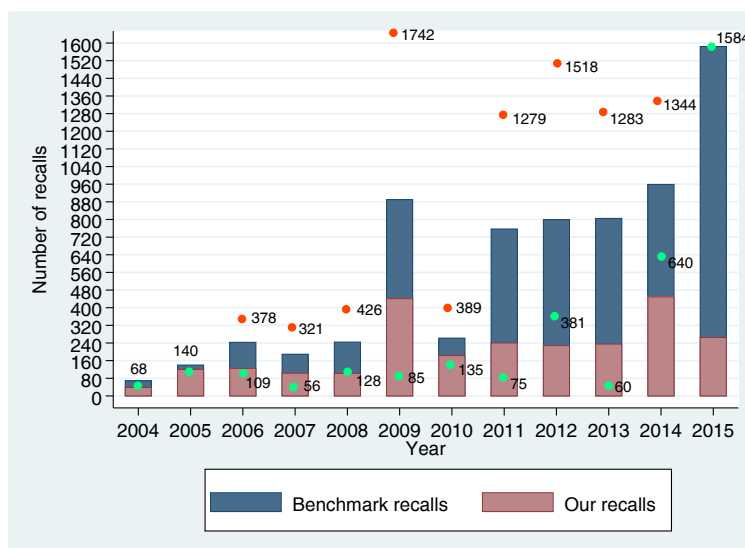


Figure 2: This histogram compares our number of recalls against the number of recalls used as a benchmark (average of sources). We include the minimum and the maximum number of recalls retrieved by the different sources. Mint colored points represent the minimum amount of recalls retrieved among all of the sources at our disposal. Red points represent the maximum number of recalls among all of the sources. A single mint point has been put whenever a single source was present for a year (i.e., 2004, 2005, and 2015).

Figure 2 underlines a disproportion in terms of the number of recalls starting from 2009, with respect to the benchmark. This deficiency pertains to the counting methodology and the structure of the database (see earlier).

Although the global trend is approximately reproduced, 2011, 2013, and 2015 represent problematic years. The dissimilarity of 2011 concerning the benchmark can easily be explained. Indeed, when excluding Aidapak’s recalls from the count of the collected recalls, the latter dropped. Furthermore, 2013 and 2015 have far fewer recalls than expected because more than 60% of the recalls in 2013 and nearly 75% of the recalls in 2015 were represented by compounding firms.

The recalls trend of the benchmark seems to be well reproduced. However, when recalls of pure compounders are excluded from the benchmark number and the sample of collected recalls, Figure 3 shows an accordance in trends. Aidapak’s recalls are included here. It is important to note here that Aidapak is a repackager and not a compounder. In addition, we want to show that 2011 does not constitute a problematic year once Aidapak’s recalls are considered.

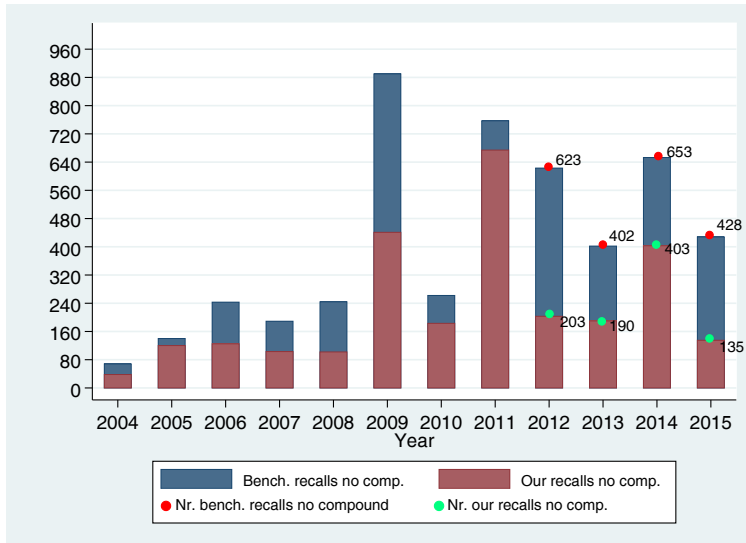


Figure 3: This histogram compares our number of recalls against the number of recalls used as benchmark without compounding recalls. We included the minimum and the maximum number of recalls retrieved by the different sources. Mint colored points represent the minimum amount of recalls retrieved among all of the sources at our disposal. Red points represent the maximum number of recalls among all of the sources. The situation is almost unchanged with respect to Figure 2 until 2011. From 2011 on, the recalls that are collected in our dataset follow the benchmark if the compounder’s recalls are excluded more precisely.

4 Methodology

The main theoretical framework we employ aligns with that described in Acemoglu and Linn (2004). Specifically, Acemoglu and Linn (2004) models innovation, which is the dependent variable in our current study, as being proportional to market size. For more detailed information, we direct the reader to Acemoglu and Linn (2004).

In our analysis, the measure of innovation is the number of clinical trials in all phases for the ATC-3 category i . The measure of market size is represented by the total sales of products within the i^{th} market. We also incorporated additional potential determinants, time effects, and category effects into our analysis. The theoretical model applied in our study is the well-established Poisson estimation model:

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

where E is the expectations operator; M_{it} represents endogenous market size; X_{it} captures age (e.g., the average age of products in category i weighted by the product’s size), diversification and innovation patterns (e.g., the scientific production)¹¹; μ_i denotes the ATC fixed effects and ζ_t denotes the time fixed effects.

Throughout, the problem of endogeneity in market size has been exposed as being intrinsic to market size. Hence, instrumentation of the endogenous M_{it} is required. Market size is instrumented through normalised recalls. The normalisation is on the number of products present in the market i at time t . Calling m the major recalls, the normalised recalls are denoted as follows:

$$\tilde{m} = \frac{m}{\#prod.} \cdot 100.$$

Normalisation is necessary to avoid endogeneity. Indeed, ATC markets with more products are more likely to undergo a recall by definition. Omitting this control would partly invalidate the estimates.

The belief is that markets experiencing major (normalized) recalls are likely to face a sudden negative shock in sales. The relevance of the instrument is tested in Section 5.

4.1 Identification

The estimation of (1) could lead to biased estimates for two main reasons: firstly, the non-linearity of (1) prevents the consistent estimation of fixed effects; secondly, the market size variable is endogenous.

¹¹A detailed list of covariates can be provided upon request.

In order to provide causal estimates, we rely on a control function (CF) instrumental variables (IV) approach, as outlined in Lin and Wooldridge (2019). This method marks an advancement over previous approaches by allowing us to simultaneously address (through testing and estimation) two potential sources of endogeneity: the first arises from the correlation of covariates with time-constant, unobserved heterogeneity; the second from their correlation with time-varying idiosyncratic errors. Additionally, this method is readily adaptable to non-linear models that include fixed effects. Specifically, if we denote by κ_{it} the idiosyncratic shock and by c_i the individual heterogeneity, the unobserved effects model that accommodates both idiosyncratic and heterogeneity endogeneity is represented as follows:

$$E[N_{it}|M_{it}, z_{it}, c_i, \kappa_{it}] = c_i \exp(x_{it}\beta_1 + \kappa_{it}) \quad (2)$$

where $x_{it} = (M_{it}, z_{it})$. z_{it} would typically include a full set of time effects, and M_{it} is the endogenous variable. All of the exogenous variables, which include the vector z_{it} , can be correlated with the heterogeneity (i.e., no random effects). There is also a set of excluded exogenous R_{it2} that serves as an instrument for the potentially endogenous variable. In the present work, R_{it2} includes \tilde{m} . Lin and Wooldridge (2019) pointed out that in the absence of idiosyncratic endogeneity, a fixed-effects Poisson estimator would be an attractive choice. This estimator, when viewed as a Quasi-Maximum Likelihood Estimator (QMLE), would require only the assumption of strict exogeneity for the idiosyncratic shocks to ensure consistency, namely

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

which holds as long as κ_{it} is not present and does not correlate with M_{it} . This assumption is utilized as a null hypothesis to test for idiosyncratic endogeneity, with the alternative hypothesis suggesting a complete dependency of the error term on the model specification for M_{it} and κ_{it} . An alternative approach involves leveraging the reduced form equation for the endogenous variable.

$$M_{it} = z_{it}\Pi + c_{i2} + u_{it2} \quad \forall t = 1, \dots, T \quad (3)$$

where because the z_{it} is strictly exogenous, the correlation between κ_{it} and functions of u_{it2} is tested. Lin and Wooldridge (2019) developed a straightforward procedure that enables testing for idiosyncratic endogeneity and producing consistent estimates in the presence of non-linearity, fixed effects, and both types of endogeneity. The procedure comprises the following steps:

1. Estimate the reduced form for the endogenous through fixed effects and obtain the fixed effects residuals $\check{u}_{it2} = \check{M}_{it} - \check{z}_{it}\hat{\Pi}$
2. Use fixed effects Poisson on the mean function

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

use the robust Wald test of $H_0 : \rho = 0$

The described procedure enables the consistent estimation of fixed effects in the presence of non-linearity¹². The intuition of the reason why this procedure provides consistent estimators of the coefficient of interest is presented below and passes through the introduction of Mundlak residuals.

Consistency, in particular, follows from the fact that \check{u}_{it2} can be substituted by the Mundlak residuals \hat{v}_{it2} that are the OLS residuals from estimating $M_{it} = \theta + z_{it}\Pi + \bar{z}_i\Phi + v_{it2}$, where $\hat{v}_{it2} = \hat{u}_{it2} + \hat{\kappa}_{i2}$ ¹³. As we will see, the Mundlak formulation explicitly decomposes the fixed effects and the first-stage residual into within-unit and between-unit components.

¹²The fixed effects Poisson model facilitates the removal of ATC-level fixed effects, thereby allowing for a conditional Maximum Likelihood (ML) consistent estimation. For more detailed information, we refer the reader to Cameron and Trivedi (2013)

¹³Remember that $\check{u}_{it2} = \check{M}_{it} - \check{z}_{it}\hat{\Pi}$. Notice also that Eq.(3) can be easily recovered from $M_{it} = \theta + z_{it}\Pi + \bar{z}_i\Phi + v_{it2}$. Indeed remember that $v_{it2} = M_{it} - \theta - z_{it}\Pi - \bar{z}_i\Phi$. By adding and subtracting $\Pi\bar{z}_i$ we obtain that $M_{it} = \theta + \Pi \underbrace{(z_{it} - \bar{z}_i)}_{=: \check{z}_{it}} + \underbrace{(\bar{z}_i\Pi - \bar{z}_i\Phi)}_{=: g_{i2}} + \underbrace{M_{it} - \bar{M}_i}_{=: \check{M}_{it}} + \underbrace{(\bar{z}_i\Pi - \bar{z}_i\Phi - \bar{M}_i)}_{=: c_{i2}}$.

Therefore, substituting the expression of $M_{it} - \theta$ in the expression of v_{it2} and adding and subtracting $\bar{z}_i\Pi$ we obtain $v_{it2} = \underbrace{c_{i2} + \Pi\check{z}_{it}}_{=: \check{M}_{it}} - \underbrace{\Pi(z_{it} - \bar{z}_i)}_{=: \check{z}_{it}} - \underbrace{\bar{z}_i(\Phi + \Pi) + \bar{M}_i}_{=: \kappa_{i2}}$.

By projecting the fixed effects onto the time averages of the instruments and the first-stage residuals, the Mundlak device separates out a unit-specific residual (a_1) that is, by construction, orthogonal to the regressors and the instruments. This clean separation matters because it justifies the assumption that the within-unit residual v_{it2} is conditionally exogenous once included in the model. Specifically, the Mundlak projection helps demonstrate that v_{it2} is uncorrelated with the exogenous covariates z_{it} and is the only channel through which the endogenous regressor M_{it} influences the idiosyncratic error κ_{it} . In other words, the Mundlak argument isn't strictly necessary for applying the two-step procedure, but it rigorously formalizes why the inclusion of the residual \hat{u}_{it2} actually solves the endogeneity problem. Namely, by removing the FE from \hat{v}_{it2} we can therefore obtain \hat{u}_{it2} which means that we can obtain sufficient conditions to correct for idiosyncratic endogeneity via the above described procedure by assuming that $(\kappa_{it}, u_{it2}) \perp (c_i, c_{i2}, z_i)$ (Lin and Wooldridge, 2019). The latter assumption means in turn that the Mundlak equation expresses a conditional expectation since when $(\kappa_{it}, u_{it2}) \perp (c_i, c_{i2}, z_i)$, $v_{it2} \perp z_{it}$ hence a functional form assumption can be made as follows following Lin and Wooldridge (2019) given that $(\kappa_{it}$ is independent on fixed effects and instruments the only potential pattern of dependencies is u :

$$E[\exp(\kappa_{it})|u_{i2}] = \exp(\gamma + u_{i2}\rho) = \exp[\gamma + (v_{it2} - a_1)\rho]$$

where a_1 derives from applying the Mundlak device to c_i . Crucially a_1 is shown to be uncorrelated both from z_i and with M_i ¹⁴. Substituting the latter in Eq.(1), we obtain

$$E[N_{it}|M_{it}, z_i, c_i, c_{i2}, v_{i2}] = c_i \exp(x_{it}\beta_1 + \theta + (v_{it2} - a_1)\rho) = g_i \exp(x_{it}\beta_1 + v_{it2}\rho)$$

where $g_i := c_i \exp(-a_1\rho)$. Using the above procedure β_1 and ρ can be consistently estimated using Poisson FE. This is because now v_{i2} –which are obtained via the Mundlak projection and therefore are the residuals of a regression that can be consistently estimated via OLS and thus controlled for– appear in the restated Eq.(1). The only requirement is that the Poisson distribution is correctly specified (Lin and Wooldridge, 2019).

Notice that a key requirement of Poisson fixed-effects (FE) models is that the dependent variable must be nonzero for at least one time period. The performance of the model is enhanced when there is a lower proportion of zeroes in the dependent variable. To satisfy this condition, we have excluded ATC categories that do not meet this criterion, which account for approximately 10% of the ATC-3 categories in our sample.

4.2 Exogeneity of the instrument and exclusion restriction

The instrument, recalls, is neither directly related to the dependent variable (i) nor endogenous with respect to market size (ii). The latter concern (ii) revolves around the possibility that pharmaceutical recalls are triggered by class actions from patients suffering from drug side effects. Since patients can be seen as a proxy for market size, any reverse causality involving the instrument would compromise the validity of the estimates, as market size could then be considered a determinant of recalls, and vice versa. However, the literature generally concurs that class actions, when initiated, typically occur after the recall process¹⁵. These actions are primarily aimed at gathering further evidence. Moreover, patients often find themselves in a disadvantaged position, as it is not initially clear whether the observed adverse events are directly attributable to the drug, are "*opportunistic consequences of pre-existing conditions*" (Cohn and Swick, 2010), or are even acknowledged as possible side effects by the FDA (Lecci et al., 2022). Hence, experts do often recommend the prosecution of a litigation once FDA has concluded the process of recall and has collected scientific data witnessing the failures of the drug (Cohn and Swick, 2010).

The primary hypothesis suggesting that normalized recalls may directly influence trials (as mentioned in i above) posits that the void created by recalls is often filled by innovations. Thus, sectors more susceptible to recalls should also be the most innovative. However, existing

¹⁴Specifically c_i is projected onto $(\bar{z}_i, \bar{v}_{i2})$ to obtain $c_i = \eta + \bar{z}_i\lambda + \bar{v}_{i2}\pi + a_1$. the fact that a_1 is independent from z_i and from M_i follows by the fact that applying a Mundlak device on c_i is akin to estimate an OLS model of c_i on $(\bar{z}_i, \bar{v}_{i2})$, i.e. by construction a_1 is independent both from \bar{z}_i and \bar{v}_{i2} . Since M_{it} depends on v_{it2} and \bar{z}_i a_1 is also independent from M_{it} .

¹⁵This is illustrated by the case of OxyContin's recall by Purdue Pharma, where legal actions were pursued after Purdue Pharma acknowledged the potential for OxyContin to lead to substance-related disorders, resulting in its removal from the market (Herder and Juurlink, 2018).

literature presents mixed evidence on this matter. While the hypothesis implies a positive relationship between recalls and innovation, recent developments seem to challenge this view. Notably, despite the increase in the number of recalls from 2004 to 2015 (as depicted in Figure 2), the innovation crisis within the pharmaceutical industry is well-documented and recognized in scholarly discussions (see, e.g., Pammolli et al. (2011), Price and Nicholson (2014) among others). It could be argued that various factors contributing to the decline in innovation have overshadowed any stimulative effect recalls might have on innovation, thus contributing to the overall downward trend in pharmaceutical innovation. Consequently, while the positive impact of recalls on innovation may still exist, it could be obscured by other overriding effects. Empirical research to further investigate the relationship between innovation and recalls, or product withdrawals in general, remains limited. The most significant recalls often receive extensive media attention, enabling researchers to gather data on the market's response to these adverse events (see, e.g., Pérez-Rodríguez and Valcarcel (2012)). Studies in this area have found that the impact of recalls and withdrawals on market innovation is highly variable: some recalls have had substantial effects, while others have had minimal or no impact. There appears to be no consistent method for identifying which recalls significantly influence innovation among the major recalls.

One possible channel linking recalls to innovation lies in the market's reaction to these events. When market reactions to recalls are particularly severe—such as significant stock price declines or intense media scrutiny—firms may become more risk-averse, potentially curbing future innovation. The market reaction to recalls is mainly triggered by the timing of the recalls' communication. Also, if firms perceive that poor timing or delayed recalls trigger harsher market punishments, they might adjust their innovation strategies, either by avoiding risky products or by prioritizing safer, higher-quality innovations to mitigate future recall risks. The role of market timing and the speed of recall communication, especially from regulators like the FDA, can therefore influence how recalls shape firms' innovation behavior. As anticipated, the literature has shown that market reactions are largely dependent on the timing of the recall and any potential delays in the FDA's communication about the recall. Generally speaking, the market does not consistently overreact to these shocks, which suggests there is no systematic link between recalls and innovation. This finding challenges the notion that recalls directly foster innovation by highlighting the variability and complexity of market and regulatory dynamics.

In summary, the direct connections between innovation and recalls are primarily influenced by fixed and time effects. The FDA's delays are challenging to control directly. However, during the period considered in this study, the FDA implemented precise guidelines and protocols for recall communication and announcements, thereby minimizing the impact of delays. Under regulations from the FDA, delays constitute a less significant issue.

Furthermore, through the Freedom of Information Act (FOIA) agreement, as well as access to openFDA and FDA enforcement reports, we have been able to verify the occurrence of delays. These sources provided insight into the time gaps between the initiation, classification, and termination of recalls. Recall communication forms part of the initiation process and, particularly in the case of severe recalls, must be prompt. On average, the time between initiation and termination for Class I and Class II recalls is about 23 months. There may be delays in the communication during the initial phase, which typically lasted about four months for any Class I and Class II recall. In our sample of major recalls, the average duration of the initiation phase was approximately 2 to 3 months, aligning with the criteria for prompt communication. This data supports the conclusion that delays have a limited impact on our analysis.

By dropping out unobserved heterogeneity and including time dummies in the primary specification, we can control for possible direct connections between recalls and innovation. Thus, the mentioned operations ensure that there is only an indirect effect of recalls through sales.

Further arguments in favor of the indirect effect of recalls on innovation follow.

In particular, the focus of our study is on severe recalls of marketed products. The time gap between various trial phases and the eventual marketing of a drug typically ranges from 8 to 14 years. This substantial time gap is crucial for predicting and understanding possible reactions from competitors to a drug recall within the same sector where a firm operates. We propose that a competitor undergoing a recall in a sector shared by multiple firms does not significantly alter the risk of innovation in the short run. Indeed, marketed products typically

undergo major recalls long after their commercialization.

Moreover, the market requires a considerable period to fully recover from the sales void created by the recall of a drug. Consequently, there is no immediate need to invest in clinical trials to capitalize on such a shortage in the short term. To further test this hypothesis, we conducted a time-to-event analysis presented in FigureAppx.1. This analysis includes all types of recalls and clearly demonstrates that although experiencing a recall decreases the survival probability of a drug, the probability that a recalled drug remains on the market two years post-recall is still substantial. The median survival time of a recalled drug is five years. Thus, after a recall, the drop in sales is likely to remain unfilled for years. Indeed, had firms found innovative replacements for recalled drug d , which allowed them to recover the shortages left by the recall of d , then there would be no reason to keep selling drug d for years. Therefore, recalled products leave a long-term lack in terms of sales within the ATC-3 market to which they belong.

A further argument against the logic of filling the gaps left by drug recalls through innovative products is that these shortages might instead be met by drugs already present in the market. These drugs could have started their trials before, soon after, or concurrently with the trials that led to the recalled drug. This scenario is plausible because our sample excludes suspended or terminated studies, meaning that ongoing clinical trials sponsored by competing firms are likely to bring to market products within the same therapeutic class.

Competition among wholesalers within an ATC category can become apparent in the early stages once it is evident that a firm will develop an innovative treatment. The race to develop alternative drugs is spurred from the onset of trial phases, where there are opportunities to be the first to market. Drugs substituting for recalled ones within the same ATC might be developed quickly after the recalled drug, motivated by a "first to arrive" competition rather than a strategy to "fill the gaps of recalls."

This dynamic could also be influenced by the demand for patented medicines of the type that was recalled, which was likely substantial when the trials for the recalled drug commenced. If demand persists at the time of the recall, existing generics or new ones may step in to fill the void. Notably, the development of generics is less time-consuming as they primarily need to demonstrate bio-comparability.

It is important to acknowledge that any potential positive relationship between recalls and innovation, which exploits gaps in the market, indirectly passes through market size. Indeed, the initiation of new trials within a market following a recall depends on the demand that the recalled product previously generated. If a recalled product had minimal underlying demand, it is reasonable to anticipate that no company would embark on an expensive trial solely to fill the void left by that product. Thus, the response of innovation appears to hinge not directly on the recall itself but on the underlying magnitude of demand for the recalled product (i.e., market size).

Another possible critique that could undermine the validity of the instrument is the potential for a recall of product i to provoke the recall of trials concerning similar products, creating a domino effect based on the reasons for the recall. If the recall is specific to the withdrawn product, then implications for other companies' products are less likely. For example, following the recall of the COX-2 inhibitor Vioxx due to cardiovascular side effects, it is conceivable that firms with ongoing trials targeting COX-2 inhibitors might have suspended or withdrawn these trials. To our knowledge, no extensive research has been conducted to explore this scenario within the drug market. The closest examination of this issue is found in Ball et al. (2018a), although this work focuses on the medical device industry, which operates under different recall regulations than the drug market. Typically, a device recall involves a temporary removal for repair or update before being reintroduced to the market.

Our approach to managing this circumstance was threefold. First, we focused solely on active trials, thereby excluding any that were suspended or withdrawn (including those halted due to the recall of other drugs). Second, we excluded trials from companies that were undergoing a recall. Lastly – whenever it was possible to link the reason for the severe recalls – if the reason for a severe recall was linked to adverse events caused by an active principle also used in ongoing trials, we excluded those trials using a similar active principle. This methodological decision aimed to minimize the impact of recall-related biases in our analysis.

To further validate our conclusions of an indirect relationship between innovation and recalls (see point (i) above), we performed further analyses. Firstly, we constructed a box plot to display the average number of trials (and their dispersion) across ATC markets that have

experienced a recall and those that have not, broken down by year. This analysis aids in understanding that major recalls do not necessarily occur in more innovative markets. Indeed, Figure 4 illustrates that the yearly number of trials in ATC-3 markets undergoing major recalls is nearly identical to the average number of trials across all other markets.

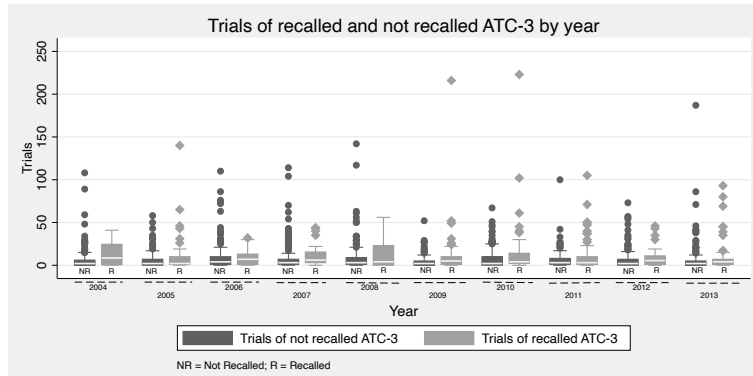


Figure 4: The box plots show how, on average, recalls do not necessarily happen in more innovative markets. The average number of trials amounts to 123 for ATC-3 markets having undergone a recall and to 118 for ATC-3 markets not having undergone a recall.

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 5 provides additional evidence supporting the presence of only an indirect relationship between innovation and recalls. Specifically, Figure 5 clearly demonstrates that the percentage change (either increase or decrease) in the number of activated trials¹⁶ following the year of a recall is minimal and consistent with the average change observed in the two years preceding the recall. It is important to note that in this context, the term "activated trials" refers to the difference between the number of active trials in year t and the number of active trials in the previous year, which by definition, determines the number of trials activated in year t .

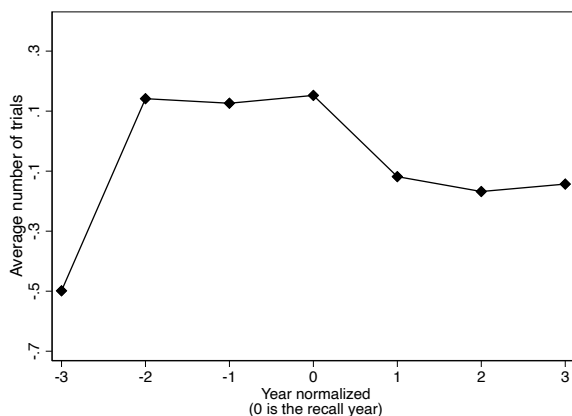


Figure 5: Change in the number of trials as measured by $\ln(Trials_t) - \ln(Trials_{t-1})$.

To conclude, a Monte Carlo experiment was conducted to assess the random occurrence of major recalls within ATC-3 markets. The purpose of this exercise is to test whether certain markets experience significantly more major recalls than would be expected by chance. If the observed number of recalls per ATC-3 market falls well outside what would typically occur in a random distribution, this could suggest that specific market characteristics or structural factors drive recall concentration. Conversely, if the observed data aligns with what random sampling would predict, this would indicate that the distribution of major recalls may not follow any systematic pattern. The simulation helps provide a benchmark for evaluating whether the observed recall frequencies are statistically meaningful or just the product of random variability. Specifically, we randomly sampled our data 1 million times. In each iteration, we calculated the average number of recalls for each ATC-3 in the sampled data and stored the results in a separate file. From this data, we constructed standard errors and confidence intervals for our estimates. We then checked whether the actual number of major

¹⁶Calculated as $\ln(Trials_t) - \ln(Trials_{t-1})$.

recalls in our dataset fell within these confidence intervals. It turns out that the average number of major recalls in the actual data is within the confidence intervals, ranging from [1.49, 2.14], with an exact average of 1.53 major recalls per ATC-3. In other words, there's no evidence that certain ATC-3 markets systematically experience more (or fewer) recalls than random chance would suggest.

As an additional exercise, we artificially allocated $B = 1000$ instances of N^{17} major recalls randomly within the ATC-3 markets and repeated the Monte Carlo experiment described above. In this scenario, the estimated average number of actual recalls also lies within the constructed confidence intervals [1.51, 1.87].

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 sub-samples—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. Database at ATC-3 level is balanced.

Variable		ATC-3			Description
		Full Sample	Subs. recalls	Subs. no recalls	
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	

¹⁷Where N represents the total number of recalls in our dataset

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 recall¹⁸, 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.

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.

¹⁸To 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 *recalls* in the subsequent sections of the analysis.

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

	(ATC-3-Firm Aggregation)	(ATC-3 Aggregation)
	Log sales	Log sales
$\tilde{recalls}$	-0.0053 (0.0033)	-0.0283*** (0.0056)
$\tilde{recalls}_{t-1}$	-0.0226** (0.0083)	-0.0267*** (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. *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 Table 7). 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 recalls¹⁹. 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 products are uniquely innovative and without substitutes.

¹⁹The t-stat is obtained after 30000 repetitions and amounts to 2.438.

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):

$$AV_{it} = y_{it} - E(y_{it}), \quad (4)$$

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

$$y_{it} = \alpha + \beta y_{st} + \gamma X_{it} + \mu_i + \lambda_t + u_{it}, \quad (5)$$

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 $t \in \{-(T-1), \dots, 0, \dots, (T-1)\}$, as follows:

$$\overline{AV}_t = \sum_{i=1}^{N_t} \widehat{AV}_{it}, \quad (6)$$

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:

$$Var(\overline{AV}_t) = \frac{\sum_{i=1}^{N_t} Var(\widehat{AV}_{it})}{N_t^2}, \quad (7)$$

where $Var(\widehat{AV}_{it})$ is the variance of the forecast error derived from estimation of Equation 5. 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 5; that is, the value at the whole market level.

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

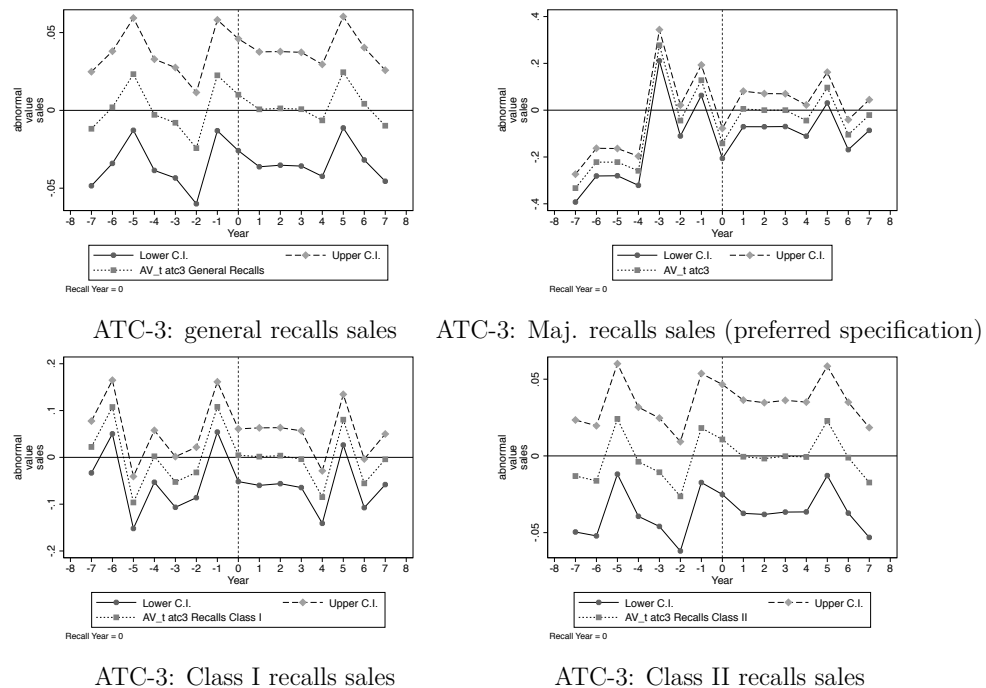


Figure 6: Abnormal values are analyzed at the ATC-3 level of aggregation, with years normalized such that year 0 corresponds to the year of the recall. The analysis encompasses four scenarios, delineating the sales trajectory before and after the recall year, according to four distinct recall classifications: 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 6 presents the abnormal values at the ATC-3 level, with confidence intervals constructed at the 95% level. As illustrated in Figure 6, following the initial decline in the year of the recall, sales typically rebound within one or two years post-recall, particularly evident in the figure associated with major recalls. This observation suggests that the instrument used in our study has a short-run effect. This short-term nature of the impact contrasts with the long-run effects of demographic shocks discussed in the work of Acemoglu and Linn (2004). 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.(1).

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 terms²⁰. 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 ($recalls_{it}$, and $recalls_{it-1}$)²¹. In the first stage of the analysis, we compute the residuals, \hat{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,

²⁰For 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

²¹We included two instruments following Hansen et al. (2008), which suggests that using more valid instruments leads to more accurate estimates.

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 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) <i>log Trials</i>	(NR) Trials
$trials_{t-1}$					-0.00741 (0.0005)	0.0732* (0.0335)	
Log sales	0.1378*** (0.0060)	0.122*** (0.0224)	0.6362** (0.2149)	0.2713*** (0.0756)	0.802** (0.266)	0.1176*** (0.0153)	0.8229** (0.3174)
residuals			-0.8018*** (0.2157)	-.3657*** (0.0745)	-0.862** (0.269)		-0.9711** (0.3177)
$\frac{K_{t+1}}{P_{t-1}}$	-0.5378*** (0.0847)	-0.423* (0.182)	-0.0926 (0.0909)	-.2635** (0.0983)	-0.484*** (0.147)	-0.0504 (0.0914)	
<i>average age firm</i>	0.2890*** (0.0139)	0.178*** (0.0333)	-0.1332*** (0.0377)	-0.0608*** (0.0116)	-0.106* (0.0398)	0.0634** (0.0214)	
<i>average age firm</i> ²	-0.0038*** (0.0002)	-0.00234*** (0.0004)	0.0021*** (0.0005)	0.0006*** (0.00015)	0.00178*** (0.0005)	-0.0008** (0.0003)	
$hh_i^{(a)}$	0.2245*** (0.0446)	0.548** (0.200)	-0.3199 (0.2903)	-0.457* (0.2403)	-0.145 (0.360)	0.1106 (0.1153)	
<i>share generics in ATC</i>	-0.5571*** (0.0404)	-0.526** (0.180)	-0.3168** (0.1124)	-0.0782* (0.0241)	-0.898*** (0.143)	-0.2036 (0.1068)	
<i>average age product</i>	-0.0658*** (0.0061)	-0.0512* (0.0240)	-0.0592** (0.0190)	-0.1023*** (0.0130)	-0.0928** (0.0323)	-0.0564*** (0.0143)	
<i>average age product</i> ²	0.0010*** (0.0002)	0.0005 (0.0007)	0.0011 (0.0010)	-0.0034*** (0.00053)	0.0100 (1.64)	0.0010* (0.0004)	
<i>papers</i>	0.5608*** (0.0728)	0.175 (0.237)	0.1558* (0.0750)	0.0492*** (.00482)	0.101 (0.0013)	-0.0443 (0.1477)	
<i>papers</i> ²	-0.6672*** (0.1056)	-0.316* (0.131)	-0.0067 (0.0838)	-0.0078 (0.0853)	-0.0929 (0.083)	-0.0083 (0.0883)	
<i># firms</i>	0.0090*** (0.0007)	0.0111*** (0.0028)	0.0008 (0.0035)	0.0120*** (0.0026)	-0.0032 (0.0792)	0.0048** (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 model²² 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.

²²This model assumes exogenous market size and does not control for fixed effects.

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.

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)²³). 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

²³The 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.

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 several factors. Firstly, reverse causality of market size and innovation. Innovation can itself expand market size, creating a feedback loop that violates the assumption of exogeneity in a naive model. If we treat market size as exogenous – as it is done in (P) and (NB) – we ignore the fact that part of the observed market size is a consequence of past innovation. This generates a positive correlation between market size and the error term, as market size now partially reflects unobserved factors driving innovation. The naive model assumes zero covariance between market size and the error term, but in reality, this covariance is positive due to reverse causality. As a result, the estimated coefficient on market size is biased downward, underestimating the true causal effect—a phenomenon known as attenuation bias. By employing an IV approach, such as using product recalls as an instrument, we can isolate the exogenous variation in market size that is unrelated to innovation. This allows us to recover a larger and unbiased estimate of the true impact of market size on innovation. Secondly, in Columns (P) and (NB) of Table 4, the temporal correlation of units is not accounted for. As a result, it is presumed that units are independent across both the cross-sectional and temporal dimensions—an assumption that imposes significant constraints in a longitudinal analysis. Specifically, this assumption treats an individual (or market) observed at two distinct times, t_0 and t_1 , as if they were completely independent entities. Hence, individual (or market) i at time t_0 is regarded as distinct from the same individual (or market) i at time t_1 . The primary implication of this assumption is that time-independent heterogeneities, which are unobserved, do not affect other individuals. Nevertheless, it is important to recognize that the same individual observed at two different times is treated as "two distinct individuals" in this framework. Consequently, in the model represented by Columns (P) and (NB), it is implicitly assumed that unobserved shocks affecting an individual (or market) i at time t do not influence the same individual (or market) i at time $t+k$. This approach conflates between and within individual effects: between effects arise when the time component is averaged out from the variables. Such between-effect settings capitalize on differences between units, which, in our analysis, are assumed to be independent by definition (refer to ATC-3 markets discussed in previous sections), thereby ignoring temporal variations. As a result, the variance in market size (time-demeaned) will be greater in a between-effect setting, as it only considers the average market size differences across independent ATC-3 markets.

Moreover, given the divergent time trends of trials and market size observed in Figure 1, the between effects of market size on innovation are likely to be understated in a between-effect setting. Specifically, while the trend in innovation decreases from a certain point in time, the trend in market size increases. However, because temporal variations are not accounted for in a between-effect framework, an inverse relationship between market size and innovation is apparent. Consequently, mixing between and within individual effects results in overall lower coefficients for market size in Columns (P) and (NB) compared to those in Column (CF-IV). In other words, failing to account for time trends may generate a spurious negative correlation between market size and innovation, simply because market size tends to increase over time while innovation rates decline. This inverse relationship is not necessarily causal but rather reflects divergent temporal patterns. To accurately estimate the causal effect of market size on innovation, it is essential to control for these underlying time trends, as they may confound the observed association.

Ultimately, the downward bias in the coefficients observed in Columns (P) and (NB) indicates that the unobserved heterogeneity is negatively correlated with trials.

To illustrate this point, consider a scenario in which an ATC experienced a significant positive shock (increased trials) in 2010, which is neither modeled nor measured. Assuming all other factors remain constant, the fixed effect for that ATC over the period 2004 to 2013 would appear to be anomalously high. However, according to existing literature, notably Bresnahan and Reiss (1991) among others, the margins on each product will be lower when a greater number of products are available to treat a particular clinical condition. Consequently, the unobserved positive shock for the i^{th} ATC would reduce the margins of all competing products in the same market, thereby depressing sales within that market. This negative correlation between the market size regressor and the error term leads to an underestimation of the

market size effect.

Conversely, in Column (CF-IV), where time dependency is controlled, this deflationary effect is mitigated. Therefore, the coefficient representing market size is notably higher in Column (CF-IV) than in Columns (P) and (NB), which do not account for these time dependencies. Columns (P) and (NB) do not to control for the reverse causality from market size to innovation. Ignoring this reverse causality contributes to an upward bias in the coefficient of market size, as discussed in the literature, notably by Acemoglu and Linn (2004). Thus, in Columns (P) and (NB), there are two conflicting effects: an upward bias due to endogeneity from reverse causality and a downward bias resulting from endogeneity due to unobserved heterogeneity.

These two biases do not offset each other, and the analysis shows that the negative bias from unobserved heterogeneity predominates over the bias from reverse causality endogeneity. This predominance leads to an overall deflation in the estimated impact of market size in Columns (P) and (NB) compared to what is observed when such endogeneities are appropriately controlled, as in Column (CF-IV).

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 notably 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.

	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:	$z = 0.33$	$\Pr > z = 0.740$

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

When the idiosyncratic errors are independently and identically distributed (i.i.d.), the first-differenced errors are first-order serially correlated. Thus, as expected, Table5 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 (misspecification) and endogeneity, thereby raising the possibility of serial correlation.

Moreover, Column (NR) demonstrates 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 uniformly indicate this positive relationship.

Specifically, the analysis in Column (A-B) aligns with the outcomes presented in Acemoglu and Linn (2004), showing no evidence of serial autocorrelation. Notably, the coefficient associated with lagged trials is negative and statistically insignificant, consistent with the findings reported by Acemoglu and Linn (2004). As the potential explanations for these observations have been comprehensively addressed in Acemoglu and Linn (2004), they are not reiterated in this study.

Furthermore, in Column (A-B linear) of Table 4, the coefficient for lagged trials is positively significant at the 5% level, indicating a robustness check in alignment with the non-instrumented scenarios suggested by Acemoglu and Linn (2004). In this scenario, the coefficient of the lagged dependent variable was positive but not statistically significant, a finding that is consistent with those reported in Acemoglu and Linn (2004). Furthermore, market size continues to exhibit a strong and positive correlation with innovation, although the coefficient is the smallest among the specifications analyzed thus far. This observation suggests that some variability may be captured by the lagged dependent variable. Additionally, there may be a misspecification bias resulting from the failure to correct for nonlinearity.

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 do not account for endogeneity. Additionally, the impact of market size is more pronounced in models that incorporate corrections for endogeneity. This suggests that the omission of controls for temporal dependence may not significantly impact estimation results, corroborated by the weak autocorrelation coefficient observed. It is, hence reasonable to suppose that the lower magnitude of the coefficient of size in Columns (P), (NB) and (A-B linear) of Table 4 is primarily a consequence of considering 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, performed 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) <i>log Trials</i>	(2) <i>log Trials</i>
<i>trials_{t-1}</i>	0.0592 (0.0426)	0.380* (0.189)
Log sales	0.1208*** (0.159)	-0.533** (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 Table 4. 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) Table 4.

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 Evaluate database in Columns (First-stage Evaluate) and (Second-stage Evaluate) Furthermore, Column (First-stage Minor Recalls) shows the poor strength of minor recalls (i.e., recalls whose motivation pertains labeling and packaging).

	(First-stage Minor Recalls)	(First-stage all recalls)	(Second-stage all recalls)	(First-stage Evaluate)	(Second-stage Evaluate)
	Log sales	Log sales	Trials	Log sales	Trials
$\tilde{recalls}$	0.00138 (0.00302)	0.0009 (0.00345)		-0.260*** (0.0059)	
$\tilde{recalls}_{t-1}$	-0.0017 (0.00308)	-0.0223*** (0.00678)		-0.0180 (0.0105)	
Log sales			0.636* (0.255)		0.710** (0.275)
Residuals			-0.802*** (0.216)		-0.802*** (0.275)
$\frac{K_{t+1}}{P_t}$	0.187*** (0.0617)	0.191** (0.0621)	-0.0926 (0.0909)	-0.173 (0.154)	-0.147 (0.276)
average age firm	0.151 (0.0916)	0.153 (0.0914)	-0.133*** (0.0377)	0.0470 (0.0678)	0.224*** (0.0045)
average age firm ²	-0.0019 (0.00127)	-0.00195 (0.00127)	0.00214*** (0.000488)	-0.0004 (0.0008)	-0.0814*** (0.0316)
hhi	1.243*** (0.259)	1.239*** (0.259)	-0.320 (0.290)	1.721*** (0.419)	-0.60 (0.495)
share generics in ATC	-0.155 (0.341)	-0.181 (0.339)	-0.317** (0.112)	-0.372 (0.328)	0.00640 (0.157)
average age prod.	0.0533 (0.0587)	0.0539 (0.0586)	-0.0592** (0.0190)	-0.0330 (0.0504)	-0.0732** (0.0225)
average age prod. ²	-0.003 (0.00216)	-0.0037 (0.00216)	0.00109 (0.000966)	0.00138 (0.00206)	0.0009 (0.0009)
papers	-0.0264 (0.0510)	-0.0267 (0.0514)	0.156* (0.0750)	-0.139 (0.0803)	0.265** (0.0917)
# firms	0.0077 (0.00699)	0.00734 (0.00701)	0.000825 (0.00346)	-0.0106 (0.0095)	0.0224*** (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, Table 7 replicates the main analysis while adopting minor recalls rather than major recalls. In contrast to major recalls, minor recalls are represented by the available recalls that are due to labeling or packaging issues. This exercise provides an intuition about the differences in the motivations of the recalls, showing how minor recalls are, in fact, a weak instrument of the 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 issues²⁴, 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)
<i>average age firm</i>	0.589*** (0.154)
<i>average age firm</i> ²	-0.00478** (0.0016)
hhi	0.145 (0.260)
<i>share generics in ATC</i>	-0.648** (0.244)
<i>average age product</i>	-0.278*** (0.0514)
<i>average age product</i> ²	0.0011*** (0.0022)
<i>papers</i>	0.546*** (0.147)
<i>papers</i> ²	0.193 (0.114)
<i># firms</i>	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.

Adopting the number of patients as a proxy for market size confirms the first and second-stage results compared to the principal specification highlighting also the power of recalls in instrumenting demand-side measures of market size. However, the outcomes are weaker in terms of significance than our main specification. Therefore, recalls do not seem a vital instrument for the number of patients. On the one hand, a more significant number of patients within an ATC-3 class might increase the probability of an adverse event than in a scarcely populated ATC-3 class, which would increase the probability 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 causes a decrease in the number of patients adopting pharmaceuticals in the questioned ATC-3 class, thus compensating for the possible positive effect implied by a higher probability of adverse events. In the second stage, Table8 enforces

²⁴Minor 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.

the results found in Acemoglu and Linn (2004) where a coefficient of market size between 3 and 4 was found. Therefore, the significance of residuals confirms the presence of endogeneity.

6 Conclusions

Recent research underscores the significance of market size in influencing the rate of innovation within the pharmaceutical industry. Following the critique by Cerda (2007), using demographic shifts as instruments remains a tenuous, albeit valid, approach. Additionally, newer studies emphasize the necessity of accurately modeling both competition and technological opportunities (refer to Rake (2017), Dubois et al. (2015)). For instance, numerous scholars highlight the pivotal role of advancements in molecular biology and related fields in enhancing the industry’s capacity for innovation and technological development (Rake (2017)). However, the existing literature is deficient in aggregation analyses that would facilitate straightforward derivation of policy implications. Consequently, this study employs the ATC-3 level of aggregation, which is utilized by antitrust authorities.

The empirical analyses are conducted using a unique database augmented with additional sources. This diversity of sources has enabled the precise collection and categorization of data on clinical trials and drug recalls within ATC-3 categories. The methodology adopted is both innovative and robust (Lin and Wooldridge (2019)), distinguishing itself from previous approaches by allowing for the control of both idiosyncratic and heterogeneity endogeneity. This method involves a two-stage process. Notably, a simple Wald test applied to the residuals’ coefficient in the second stage serves to confirm the presence of idiosyncratic endogeneity.

Recalls serve as an innovative instrumental variable introduced for the first time in this study. Data on recalls were meticulously gathered from multiple sources, including FDA Enforcement reports, openFDA, and a database established through FOIA agreements with the FDA. The recalls selected for this analysis are representative; major recalls were specifically chosen based on their sharpness, indirect effect on the dependent variable (innovation), and exogeneity. The initial stage of analysis revealed a significant negative impact of recalls on market size, thereby validating the effectiveness of the instrument. To the best of our knowledge, this research contributes a novel empirical perspective to the existing literature, which predominantly addresses the optimal management of recalls and provides theoretical discussions on their negative impacts at the firm level. However, there has been limited focus on the implications of drug recalls at the market level.

Data on clinical trials, spanning from the pre-clinical phase to Phase IV, were sourced from the ClinicalTrials.gov website and integrated with information on Investigational New Drugs (INDs). To address potential biases arising from the more pronounced response of market size to the entirety of trials compared to those trials most likely to lead to innovation, only activated trials were included in the analysis. This selection also provides a substantive response to concerns that the recall of a drug might necessitate the suspension of trials within the same drug family. Indeed, both suspended and withdrawn trials were excluded from our primary analysis. For robustness, additional estimates were calculated including these trials, which confirmed the presence of idiosyncratic endogeneity and a positive sign in the estimates, albeit with reduced magnitude and significance.

Our analysis aligns with previous findings, showing a 6.3% increase in innovation following a 10% increase in market size. Critically, when we specifically examine trials that culminated in marketed innovations, the coefficient notably diminishes to 27%. This substantial reduction highlights a potential overestimation bias pervasive in the literature, which can be attributed to the insufficient instrumentation of market size, irrespective of the measurement method employed. This observation calls for a more nuanced approach in assessing the true impact of market size on innovation.

Recent studies, such as those by Dubois et al. (2015), report a notably lower coefficient of 0.23%. The authors acknowledge the challenges of comparing their results with other studies that utilize different measures of innovation. They also suggest that their use of global data, rather than US-specific data, might contribute to a diminished responsiveness in their findings.

Our results demonstrate robustness across several specifications. The coefficients of independent variables align with expectations, including the minimal effect of lagged trials, previously examined in Acemoglu and Linn (2004). Additional analyses reveal a positive and

significant effect of market size on innovation, persisting even in models where fixed effects are uncontrolled, and market size is assumed exogenous. This finding supports, in terms of the sign and significance, the recent observations of Rake (2017), who found no evidence of reverse causality. However, the magnitude of the coefficient significantly decreases compared to our preferred specification, indicating a notable bias.

Importantly, our estimates retain their robustness even in the absence of control variables in the analysis.

This work elucidates significant policy implications concerning the stimuli for innovation and highlights the impact of recalls at the market level. Notably, governments should recognize that innovation is predominantly an economic phenomenon, driven by firms' pursuit of financial returns. The positive correlation between market size and innovation suggests that authorities and policymakers should avoid overly penalizing economic actors. To ensure the future welfare of their citizens, they should prudently promote research and smartly invest in new technologies, while managing competition from generics.

The present study adds to the ongoing debate on the insufficient investment in innovation concerning orphan drugs (see, e.g., Denis et al. (2010), Grabowski (2002), Dupont and Van Wilder (2011)). Orphan drugs, designed to treat rare "niche" diseases with small market sizes, demand breakthrough innovations and are typically developed by small biotech firms, which are the most risk-taking entities in the market but often lack the financial resources of more established firms. Establishing a clear relationship between market size and innovation could foster investments in these overlooked "niche" markets.

Furthermore, we highlight that recalls impact not only at the firm level but also at the market level. They induce adverse market shocks, thereby affecting economic stability and welfare. Authorities should thus enforce stricter regulations to prevent severe recalls. However, an increase in Class II and Class III recalls might be a natural outcome of increased market competition.

Future research should utilize more up-to-date data and consider including recalls from compounders and repackaging firms as an instrumental variable

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 industry²⁵

²⁵In 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.

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 classes	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 $d_{2,i}$; 1968-1997	15 drug categories $d_{2,i}$	NME	FE, GLS, IV, Tobit	>0 $d_{2,i}$	Demographic measures	No
Lichtenberg (2007)	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 $d_{2,i}$; 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 $d_{2,i}$; MEPS; OECD; NIH	49 therapeutic classes	R&D $d_{2,i}$	Negative Bin.; Poisson	0.26; 0.41; 0.51	Demographic shifts	No
Agarwal and Gaule (2022)	US; 2015-2020; GBD; clinicaltrials.gov	49 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

Further details:

^{d1} The estimates suggest that a 1 percent increase in the potential market size for a drug category leads to a 6 percent increase ^{d2,i} Data on government funds used on the R&D process of the pharmaceutical sector; ^{d2,ii} population data for market size; ^{d2,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

^{d3} Both Cerda and Rake consulted the 19th 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.

^{d4} 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.

^{d5,i} Pharmaprojects trend data "snapshot"; ^{d5,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: ^{c1} 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).

^{c2} 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.

^{c3} 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 Time-to-event analysis

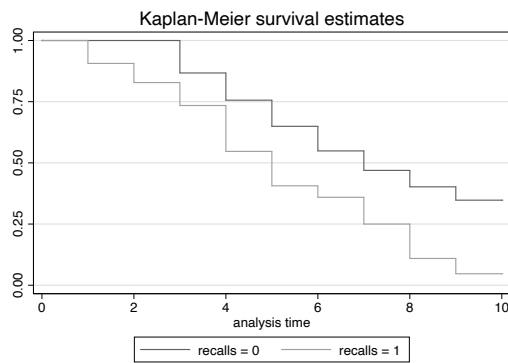


Figure Appx.1: Kaplan-Meyer time-to-event analysis. The x-axis represents the number of years until death. Recalled products (recalls =1) are already scaled down at 2 years of survival time with respect to not recalled products (recalls =0). The median survival time for recalled medicines is about 4 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.

C Abnormal values (firm and product levels)

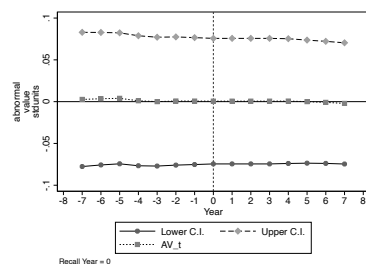
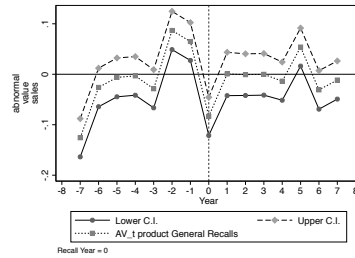
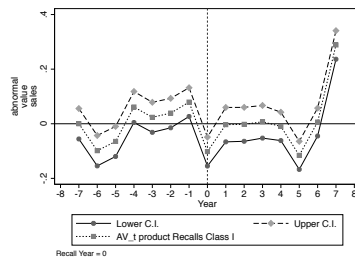


Figure Appx.2: Effect of recalls on market sales once firms having undergone a major recall are cancelled out. The absence of any effect (i.e., increases of sales due to recalls of competitors) at recall time for products other than the ones of the recalled firm witnesses the absence of compensations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

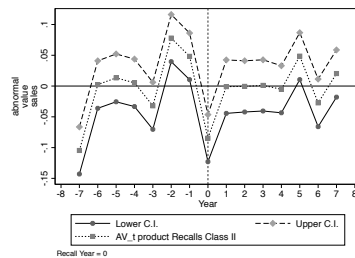
The following figures represent abnormal values at firm and product level for different typologies of recall (according to their gravity).



Product: general recalls sales

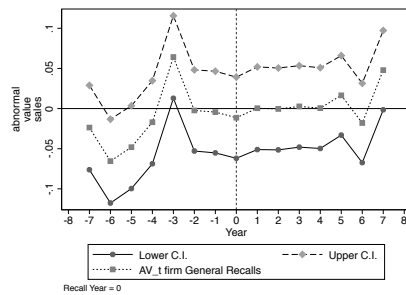


Product: Class I recalls sales

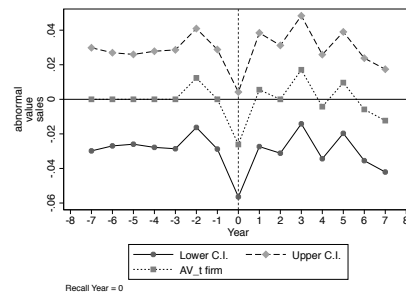


Product: Class II recalls sales

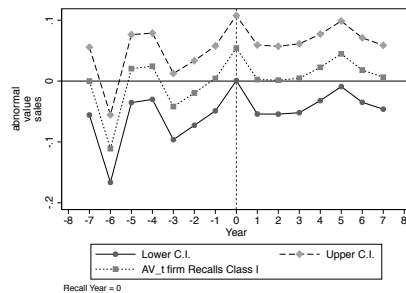
Figure Appx.3: Abnormal values for product aggregation. Years are normalised. Year 0 represents the year of recall. The three scenarios include the path of sales before and after the recall year, using three different definitions of recalls: Class I recalls, general recalls and Class II recalls. As shown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



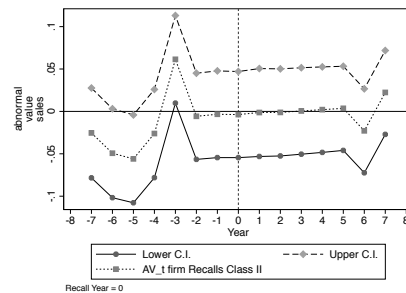
Firm: general recalls sales



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.

Variable	Prod.			Firm			
	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 mean665	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 mean086	.085	.086
	Overall Std. Dev.226	.148	.231
	Between Std. Dev234	.108	.238
	Within Std. Dev186	.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.

D First stage rob. checks

Significant coefficients of the first stage employing the number of patients as a measure for market size.

Table Appx.4: First stage of the robustness check using the number of patients as 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 with 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.
- R. Agarwal and P. Gaule. What drives innovation? lessons from covid-19 r&d. *Journal of Health Economics*, 82:102591, 2022.
- P. Aghion, A. Bergeaud, M. Lequien, and M. J. Melitz. The heterogeneous impact of market size on innovation: evidence from french firm-level exports. *Review of Economics and Statistics*, pages 1–56, 2022.
- K. J. Arrow. The economic implications of learning by doing. *The review of economic studies*, 29(3):155–173, 1962.
- R. Azghandi, J. Griffin, and M. S. Jalali. Minimization of drug shortages in pharmaceutical supply chains: A simulation-based analysis of drug recall patterns and inventory policies. *Complexity*, 2018:1–14, 2018.
- 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. 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.
- J. M. Bertoni, J. P. Arlette, H. H. Fernandez, C. Fitzner-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.
- Bowe C. Merck quarterly profits hit by viox recall, 2005. URL <https://www.ft.com/content/b507ea92-b25b-11d9-bcc6-00000e2511c8>. [Online; accessed 15-April-2021].
- T. F. Bresnahan and P. C. Reiss. Entry and competition in concentrated markets. *Journal of Political Economy*, 99(5):977–1009, 1991.
- A. C. Cameron and P. K. Trivedi. *Regression Analysis of Count Data*, volume 53. Cambridge university press, 2013.
- 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.
- S. R. Cohn and E. M. Swick. The sitting ducks of securities class action litigation: Biopharmas and the need for improved evaluation of scientific data. *Del. J. Corp. L.*, 35:911, 2010.
- 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.
- 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.
- 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.
- M. et al. Pharmaceutical antitrust law in european union. Dechert LLP, 2019.
- FDA U.S. Food & Drug. Cfr - code of federal regulations title 21 (subpart "a-general provisions", sec. 7.3 "definitions") no. 21cfr7.3, 2019. <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=7.3>.
- 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.
- 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.
- 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.

- K. Hall, T. Stewart, J. Chang, and M. K. Freeman. Characteristics of fda drug recalls: A 30-month analysis. *American Journal of Health-System Pharmacy*, 73(4):235–240, 2016.
- 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.
- 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.
- M. Herder and D. Juurlink. High-strength opioid formulations: the case for a ministerial recall. *CMAJ*, 190(48):E1404–E1405, 2018.
- S.-M. Hosseini-Motlagh, M. Nematollahi, and N. Nami. Drug recall management and channel coordination under stochastic product defect severity: A game-theoretic analytical study. *International Journal of Production Research*, 59(6):1649–1675, 2021.
- 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.
- T. Iizuka. Generic entry in a regulated pharmaceutical market. *The Japanese Economic Review*, 60:63–81, 2009.
- B. Jovanovic. Product recalls and firm reputation. Technical report, National Bureau of Economic Research, 2020.
- 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.
- L. Lecci, G. R. Page, J. R. Keith, S. Neal, and A. Ritter. The effects of health anxiety and litigation potential on symptom endorsement, cognitive performance, and physiological functioning in the context of a food and drug administration drug recall announcement. *Frontiers in Psychology*, 13, 2022.
- 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.

- 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.
- J. V. Pérez-Rodríguez and B. G. Valcarcel. Do product innovation and news about the r&d process produce large price changes and overreaction? the case of pharmaceutical stock prices. *Applied Economics*, 44(17):2217–2229, 2012.
- W. Price and I. Nicholson. Making do in making drugs: Innovation policy and pharmaceutical manufacturing. *BCL Rev.*, 55:491, 2014.
- 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 revisited. *The Journal of Industrial Economics*, pages 225–237, 1982.
- J. Schmookler. *Invention and Economic Growth*. Harvard University Press, 2013.
- S. Simoens, K. De Bruyn, M. Bogaert, and G. Laekeman. Pharmaceutical policy regarding generic drugs in belgium. *Pharmacoeconomics*, 23:755–766, 2005.
- 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.
- Terence N. Merck pulls vioxx painkiller from market, and stock plunges, 2004. URL <https://www.nytimes.com/2004/09/30/business/merck-pulls-vioxx-painkiller-from-market-and-stock-plunges.html>. [Online; accessed 15-April-2021].
- 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.
- C. H. Tong, L.-I. Tong, and J. E. Tong. The vioxx recall case and comments. *Competitiveness Review: An International Business Journal*, 2009.
- 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 inflammation. 2021.
- WHO. The anatomical therapeutic chemical classification system with defined daily doses-atc/DDD. 2009.