KU LEUVEN

KU Leuven, Faculty of Economics and Business
Dept. of Management, Strategy and Innovation (MSI)

Acquiring R&D projects: who, when, and what?

Evidence from antidiabetic drug development

Jan Malek, Melissa Newham, Jo Seldeslachts and Reinhilde Veugelers

MSI Discussion Paper No. 2401

Acquiring R&D projects: who, when, and what? Evidence from antidiabetic drug development*

Jan Malek†, Melissa Newham‡, Jo Seldeslachts§, Reinhilde Veugelers¶

February 5, 2024

Abstract

This paper analyzes M&A patterns of R&D projects in the antidiabetics industry. For this purpose, we construct a database with all corporate individual antidiabetics R&D projects over the period 1997 - 2017, and add detailed information on firms' technology dimension using patent information, next to their position in product markets. This allows us to identify the identity of targets and acquirers (*who*), the timing of acquisitions along the R&D process (*when*), and which type of R&D projects changes hands in terms of technology novelty (*what*). The main results can be summarized as follows. First, most of the action in M&As is in early R&D stages, still far from product markets. Second, most of the early-stage projects that change hands are high-risk/high-gain novel projects. Third, the industry leaders in the product markets are rather inactive in acquiring those novel early-stage projects. The likely acquirers of such projects are small or pipeline firms. Our results put in perspective the narrative that large incumbents acquire small targets with low-risk projects close to product launch.

Keywords: M&As, innovation, R&D, pharmaceutics, technology, novelty, patents
JEL Codes: L41, L65, O31

*We thank seminar participants at KU Leuven, DIW Berlin, Bruegel, the EuHEA seminar series, the CISS 2019 summer school, the 2022 EARIE conference, the 2022 CMA Mergers and Innovation workshop, and the 2022 MaCCI Annual Conference for valuable comments and suggestions. We are grateful to Ekaterina Khmelnitskaya, Manuel Gigena, Dennis Verhoeven, and Sam Arts for sharing data with us.

†KU Leuven. Jan Malek acknowledges support from the FWO through project 3H190094. Email: jan.malek@kuleuven.be

‡ETH Zurich. Email: mnewham@ethz.ch

§KU Leuven, Universiteit van Amsterdam and DIW Berlin. Email: jo.seldeslachts@kuleuven.be

¶KU Leuven, Breugel and PIIE. Email: reinhilde.veugelers@kuleuven.be

1 Introduction

It has recently been documented that firms that exert market power grab larger market shares, i.e., the so-called "superstar firm" phenomenon (Autor et al., 2020). For example, for the US economy, De Loecker et al. (2020) show that there is a reallocation of economic activity from low to high markup firms. Related, that same study documents that during the period from 1985 to 2016, global M&A activity increased more than tenfold, while the aggregate markup increased with about thirty percentage points, suggesting that firms with initially high market power are becoming larger and more powerful through acquisitions. These trends, i.e., a higher concentration and rising markups, are especially perceived in industries where technological change and innovation are most intense (Diez et al., 2018).

Competition authorities have taken notice and are adapting their merger guidelines. Among others, two of the largest jurisdictions, the EU and US, have taken action. The competition arm of the European Commission, the Directorate of Competition (DGComp) updated its merger guidelines in 2021.[1] The Federal Trade Commission (FTC) and Department of Justice (DoJ) announced in January 2022 a joint initiative to conduct a comprehensive analysis of their merger guidelines.[2] Both jurisdictions include, as a key new element, the topic of mergers and "potential competitors" or "nascent competitors," and deals where large incumbents take over small promising targets to impact future competition. These deals may have stayed "under the radar" of competition authorities in the past as these targets are often too small in terms of revenues, or have no revenue at all when their products are still in the R&D stages.

Besides the digital industries, also the pharmaceutical industry is a key industry in terms of innovation and R&D spending (Grassano et al., 2021). It is also high on the agenda (competition-) policy wise. Recent merger cases in Europe show that

[1] http://competitionlawblog.kluwercompetitionlaw.com/2021/04/01/eu-commission-launches-major-merger-control-reform/

[2] https://www.reuters.com/legal/transactional/back-drawing-board-ftc-doj-rethink-merger-guidelines-2022-03-07/

enforcers are increasingly paying attention to the innovation dimension.[3] The FTC and DGComp, together with the UK Competition and Markets Authority and Canadian competition authorities, initiated in May 2021 a multilateral working group on pharmaceutical mergers on how to improve decision making in the area.[4]

Empirical academic research on M&As and innovation in pharma is following suit. Earlier research looking at the innovative performance of merged pharmaceutical companies typically discovered a negative relationship. Ornaghi (2009) found merged pharmaceutical companies to have on average a worse performance than the control group of non-merging pharma firms. Haucap et al. (2019) observed that not only average patenting and R&D of the merged entity declines in post-merger periods, but also of its rivals. More recently, the most interesting study is Cunningham et al. (2021). The paper finds that an incumbent – i.e., a company that has already launched a drug – when acquiring projects, has a higher likelihood to terminate potentially future competing projects in development. So-called "killer acquisitions" comprise about 5%-7% of the M&As in their sample.

However, while the above-cited research has a clear idea on how product markets can be defined and potential entrants therein can be identified, it is typically less focused on characterizing the technology dimensions. Indeed, most research on the topic of M&As and innovation in the pharmaceutical industry measures innovation at the firm-level using simple (citation-weighted) patent counts or R&D spending (e.g., Ornaghi, 2009; Haucap et al., 2019).

This paper aims to analyze M&A patterns along the innovation process in more detail. Therefore, it zooms in on the antidiabetics industry. We focus on the market for antidiabetic drugs because it is a large and growing market where we observe significant activity both in terms of innovation and M&As. By focusing on one therapeutic

[3]In May 2021, the European Commission (EC) fined the company Merck KGAA EUR 7.5m for a failure to disclose an R&D project while pursuing the acquisition of Sigma-Aldritch in 2015 (European Commission, 2021)

[4]https://competition-policy.ec.europa.eu/index/news/multilateral-working-group-pharmaceutical-mergers-launches-joint-public-consultation-2021-05-11_en

market, we are better able to connect numerous data-sources. However, as we further will illustrate, the framework we develop can be extended to other pharmaceutical markets. In particular, we construct a unique database with all corporate R&D activities for individual antidiabetics projects over the period 1997 - 2017. We add not only information on firms' position in product market segments, but also information on the technology characteristics of the projects and the involved firms, using detailed patent information. This allows us to identify the product and technology market position of targets and acquirers (*who*), the timing of acquisitions along the R&D process (*when*), and which type of R&D projects changes hands in terms of novelty (*what*). For the latter, we focus particularly on novel projects which have a high-risk profile, but if successful, can become the next disruptive breakthrough.

Our main results can be summarized as follows. First off, most of the action in M&As is in the early R&D stages, when projects are still far from being launched in a product market. Indeed, we find that 50% of acquisitions occur in the earliest (and most uncertain) stage of development., i.e., the preclinical stage (before the clinical trials have started). The majority of these deals take place between purely research-focused pipeline firms, which do not yet have any marketed antidiabetic drugs. Second, based on patent data, we can classify most of the early-stage projects that change hands as highly technologically novel, i.e., with a potential of high gains but with a high-risk profile. And third, the industry leaders in the antidiabetics product markets are rather inactive in acquiring projects. Our results show that the transaction landscape is rich and varied, going beyond the narrative that large incumbents acquire small targets with projects close to product launch. Rather technology uncertainty (early stages) and technological profile (high-risk/high-gain novelty) of projects are found to be key aspects of M&As in pharmaceutical markets. Although with high risk of failure and still far from the market, how these acquired high-risk/high-gain novel projects will evolve differently when acquired may nevertheless drastically impact future competition.

The rest of the paper is organized as follows. Section 2 gives a brief background on R&D in the pharmaceutical industry. Section 3 presents the data and construction of

variables. Section 4 describes the empirical implementation and section 5 presents the results. Lastly, section 6 concludes.

2 R&D in the pharmaceutical industry

The process of developing a new drug is uncertain, lengthy, and expensive. Drug development is funnel-shaped with many potential drug candidates entering the development process but only very few being successful. In our sample, only 4% of projects are launched. The development process is also lengthy. The average duration from the start of development to approval for our sample is almost 14 years.[5] The probability of success decreases and the costs of development increase as a project progresses. In the early stages, the probability of failure is very high, but the developmental costs are relatively low. There is a steep increase in costs when the drug is tested in large-scale later-stage clinical trials.

High costs of R&D and public disclosure of information through clinical trials create strong incentives to patent early. Patent filing mostly (but not exclusively) happens already before the drug enters into clinical trials. Companies file patents covering the drug's active ingredient, formulation and composition, as well as method of use. Patents are thus a relevant information base to assess the technology characteristics of projects.

3 Data

For each project, the innovation process centers around its therapeutic area - the disease the molecule or protein should target (e.g., Diabetes type II) - and its "Mechanism of action" (MoA), i.e., the biochemical process through which the drug produces the

[5]These figures are consistent with other studies. Pammolli et al. (2011) finds that only 6% of molecules pass the testing phases. Branstetter et al. (2014) report that the total development cycle takes, on average, nearly 12 years.

desired effect in the body.[6] Each project goes through a clearly defined set of development milestones to be eventually approved (pre-clinical, Phase I, Phase II and Phase III clinical trials) (FDA, 2018).[7] Public disclosure of these milestones and the corresponding data is required by the FDA.[8] This allows the progression of each project to be tracked from its inception until termination (if unsuccessful) or launch (if successful).

We use information at the project-level on patents and development milestones being reached, matched with (changes in) ownership data. Achieving this level of granularity involves substantial manual input. Our analysis, therefore, looks at one disease: diabetes. This approach can nonetheless be generalized across the pharma sector.

Diabetes is a widespread, chronic and rapidly rising condition (mostly type II diabetes, which is much more common than type I).[9] It has no cure and must be managed by life-long therapy. However, any therapy loses its pharmacological efficiency as the body gradually adapts to the treatment and becomes less responsive. Patients thus increasingly need new treatment options. The nature of the disease, together with its increasing market size, creates incentives for companies to be present in this market and develop new antidiabetics. The antidiabetic market has indeed been subject to substantial innovation activity over the period studied. The number of new drugs launched in the US market increased from 13 to 76 between 1997 and 2017. A few players dominate the market: Novo Nordisk is number one in terms of sales, and a few

[6]For example, a DPP-IV inhibitor is a particular MoA used to treat diabetes. These drugs inhibit the DPP-4 enzyme, thereby stimulating the secretion of insulin. This lowers the levels of glucose in the blood.

[7]Each drug begins with discovery and development in a laboratory. Once a drug candidate has been identified, researchers undertake pre-clinical studies using laboratory experiments and testing in animal subjects. If promising, the drug candidate progresses to three phases of testing in humans. In Phase I, the safety of the drug is tested with a small sample (ca. 20-100) of healthy individuals. In Phase II, the effect of the drug is tested in relatively small groups of people (ca. 50-300). Phase III trials test efficacy on large groups of subjects (ca. 300-3,000 or more).

[8]The information submission requirements for clinical trials is described in Section 801 of the Food and Drug Administration Amendments Act of 2007.

[9]The therapeutic market of antidiabetics is large and growing. In 2021, approximately 537 million adults were living with diabetes worldwide. This is projected to rise to 783 million by 2045 (International Diabetes Federation, 2021).

more companies have an average product market share of more than 10 % (Aventis, Eli Lily, Hoechst Marion Roussel, Merck & Co, Sanofi). However, a large share of the patents of the market leaders are expiring between 2018 – 2024.[10] With new biomedical technologies (like gene editing) fast developing, there is room for radically new projects from new players. Finally, M&A activity in the antidiabetics market is high. Taken together, this market provides an excellent setting to explore the M&A activities and how these link to innovation.

3.1 Data sources and data construction

The backbone of our dataset is the Pharmaprojects database from Citeline which provides a comprehensive list of global R&D activity in the pharmaceutical industry at the project level.[11] We identify all projects related to the treatment of diabetes, which results in a sample of 2711 projects for the period 1997-2017.

While providing information on drug names and sponsors, the Pharmaprojects database lacks information on the progression of projects. This information is available in the AACT database.[12] This database lists every study registered at ClinicalTrials.gov – a repository of clinical studies conducted around the world. We match these studies to the Pharmaprojects sample, using fuzzy string matching on sponsor and drug names. This information allows us to identify the start and end dates of the phases each project has passed.[13] In this way, we re-construct the development histories for 2378 projects (88% of projects).[14]

[10]The percentage of patents to expire by Novo Nordisk is 80%, of Eli Lilly 70%, of Sanofi 70%, and of Merck & Co. 60% (EvaluatePharma, 2018).

[11]Pharmaprojects is a commonly used database for studying the pharma sector. For example Adams and Brantner (2006); Kyle (2007); Blume-Kohout and Sood (2013); Branstetter et al. (2014) and Cunningham et al. (2021) all use Pharmaprojects data.

[12]Available at https://aact.ctti-clinicaltrials.org/

[13]In cases where complete histories could not be established, we impute the missing dates by estimating a log-normal distribution of duration per phase and randomly draw a project's phase duration from the estimated distribution. For each such imputation, we manually check that the sequence of development milestones is not violated.

[14]The remaining 12% are not matched due to a lack of information, as for some projects the Pharmaprojects database does not provide sufficient detail to make a link to the trials registry.

We identify changes of ownership for each project in our database by carefully unwinding the sequence of each project's consecutive owners.[15] To do so, we take the text information on project developers provided in the Pharmaprojects database as the starting point and apply text mining, algorithmic disambiguation, fuzzy string matching, and manual checks. To complement and verify that these changes indeed reflect ownership changes as opposed to name changes, we match relevant firms with the databases Zephyr and SDC Platinum. Manual checks and additional desktop searches for every company were performed to ensure correctness and completeness. The ownership changes considered in our analyses exclude large conglomerate transactions - situations where one big pharma company acquires another one. We exclude such cases since these deals are unlikely to relate specifically to diabetes R&D, and are thus not representative in explaining the transaction dynamics in antidiabetics.[16]

3.2 Construction of key variables

Leveraging the database described above, we construct variables that are central to our analysis. The construction of these key variables is described below.

3.2.1 Firm types

Our data comprises over 900 different pharmaceutical companies that have projects in the antidiabetics R&D pipeline during our sample period. These firms are very heterogeneous in terms of their size and previous experience (both in R&D and in marketing drugs). To understand which types of firms engage in M&A transactions, we cluster firms into bins ("firm types"). Our aim is to assign firms to bins such that firms in the same bin are similar in their capabilities, financial resources, and incentives to engage in M&As, and dissimilar across bins. We combine two dimensions, firm size and market incumbency.

[15] These ownership changes not only include mergers and acquisitions, but also deals involving sales of divisions, product lines, or individual assets.

[16] In unreported results, we check that the robustness of the results presented in this paper remains qualitatively the same when these transactions are included.

Size, our first bin-dimension, matters in the context of M&A decisions, as larger firms have deeper financial pockets and can benefit from economies of scale and scope when in clinical trials, obtaining regulatory approval, and also production and commercialization capacity and expertise (Arroyabe, 2021; Bena and Li, 2014; Danzon et al., 2007; Szücs, 2014). Big firms may find it disadvantageous to engage in an R&D race with small firms, but instead, gain access to innovation through acquisitions (Phillips and Zhdanov, 2013). From a small firm's perspective, a company can struggle in later phases of innovation, because of its lack of financial and commercialization capabilities. Therefore, selling out might be the most straightforward to get its products to market (Comanor and Scherer, 2013). Incumbency, our second bin-dimension, might impact acquisition motives as incumbents could have different incentives in comparison to non-incumbents, related to defending or expanding their current market position.

We define big firms as having a market share in the whole pharmaceutical industry of at least 1% on average during the entire sample period.[17] This group includes firms typically thought of as "big pharma" such as Johnson & Johnson and Pfizer. Using incumbency as a second bin dimension, we further split up this group into leaders, big incumbents and big non-incumbents. Incumbents, as opposed to non-incumbents, have at least one antidiabetic drug on the market. "Leaders" are incumbents with a significant market share in diabetes treatments. The antidabetics market has traditionally been dominated by just a handful of firms, with Novo Nordisk as largest, Sanofi, Merck & Co., Eli Lilly, Aventis, Hoechst Marion Roussel each having on average a market share in antidiabetics of at least 10% over the sample period (which is also much higher than the other firms with launched antidiabetics).

Small firms are defined as having at least one launched product in any pharma-

[17] We compute market shares using the 2003-2018 R&D Scoreboard data published by the European Commission (Hernández et al., 2014). For 1000 large firms (and 2500 firms in later years), the scoreboard data contains information on revenues and R&D spending categorized by sector. This allows us to approximate the market share of firms within the pharmaceutical sector for the large players. In the years before 2003 where data is not available, we search for this information in company reports. If the information could not be found, we set the values to those in the first 2003 Scoreboard report.

ceutical market, but having less than 1% market share on average during the entire sample period. We also split this group into small incumbents and non-incumbents. A few firms in the pharmaceutical industry, typically bio-pharma companies, have grown rapidly over the last 20 years (e.g., Gilead and Teva). We separate out these fast-growing companies and label these as "stars," where we define fast-growing as having a market share in pharma below 0.75% on entry into the sample and above 1% at sample end. None of these stars are yet incumbents in the antidiabetics market.

Finally, a large group consists of those firms that do not have any launched drugs (in any pharmaceutical market) and are purely engaged in R&D activities in the antidiabetics market, "pipeline firms." Since this is a large and heterogeneous group –and where we cannot make further cuts based on revenues– we use the filing date of a firm's first patent to split pipeline firms into two more homogeneous groups. In particular, "young pipelines" are defined as those firms with a first patent filing less than five year ago and "mature pipelines" are all firms with a first patent filing more than five years ago.[18] Table B.3 in Appendix B gives examples of firms belonging to every bin.

3.2.2 Technology novelty of projects

Projects differ in terms of their technological risks and their potential impact. Of particular interest are the projects that have the potential to generate the next breakthrough blockbuster drugs. Projects which have a high technology novelty, using new or previously unconnected pieces of knowledge, can bring new breakthrough solutions to the market, challenging incumbent market positions (Christensen, 2013; Fleming, 2001; Hall and Lerner, 2010). But they do so typically at a higher risk of failure, facing more uncertainty with their novel, unproven approach. They are thus "high-risk/high-gain." Foster et al. (2015) show that research introducing new combinations of chemicals is more likely to become highly cited, but also displays a higher variance

[18] After five years, a young pipeline firm switches to the mature pipeline bin in our data. Given that this switch is only based on the passage of time, such switching can be considered exogenous. We also adopted an alternative threshold of three years; results remain robust to this definition.

with regards to their citations. Krieger et al. (2018) show that riskier innovation makes drug candidates less likely to be approved by the FDA, but conditional on approval, these drugs are more valuable and earn higher revenues. Huvaj and Johnson (2019) show for firms in the medical device industry that large organizations are less likely to pursue more novel innovations.

Although all inventions have to be novel to be granted a patent, we use a more explicit measure for technology novelty, proxying for the high-risk/high-gain profile of projects. In particular, we use the Novelty in Technological Origins (NTO) indicator developed by Verhoeven et al. (2016) on the projects's patents in our sample.[19] This indicator measures the ex-ante technological novelty of a patent by assessing the extent to which a patent sources knowledge from previously unconnected fields.[20] Verhoeven et al. (2016) check in their data that NTO patents have indeed a higher dispersion of forward citations received, and are more likely to be among the least cited and the most cited patents, confirming their high-risk/high-gain profile. We consider a project to be high-risk/high-gain novel if at least one of the project's patents scores on the NTO indicator.[21]

3.2.3 Development stage

To analyze the timing dimension of M&As, we distinguish between preclinical, early (Phase I), and late (Phase II and Phase III) development stages. We group the two

[19] Alternative measures of drugs' novelty used are, for example, previous deployments of a drug's mechanism of action Dranove et al. (2020). Alternatively, Krieger et al. (2018) define a drug as novel if it is molecularly distinct from prior candidates. Verhoeven et al. (2016) verify that the NTO measure correlates with several existing constructs. Cunningham (2017) finds that a third of US medical start-ups produced novel products. In this work, a patent is novel if (i) it is the first instance of a (new) technology (USPTO subclass) on a patent, or (ii) it is the first instance of a particular pairwise combination of existing technologies on patents.

[20] In practice, a patent scores on NTO if it combines its own International Patent Classification code (IPC) and an IPC code from its referenced patents that has not yet occurred in the years previous to the application year of the patent.

[21] Since the probability to score on novelty mechanically increases with a rising number of assigned patents, we control for the number of project's assigned patents in the regression analysis. For the projects that don't change ownership, we use the patents assigned before entering clinical trials to define novelty; for the acquired projects, we use all patents up to an M&A event to assess their novelty.

smaller samples, Phase II and III. Both stages are aimed at testing the safety and efficacy profile of a drug and are frequently run in parallel with each other. We do not consider the launched phase, as it is very different from the R&D phases and not the focus of this study.

We trace each project per phase. A project is dropped from our database after the phase in which it was discontinued (if unsuccessful) or until launch (if successful). A project can thus contribute at most three observations (Pre-clinical, Early, and Late). The final database amounts to 2916 project-phase observations relating to 1860 projects. The sample contains 1787 preclinical projects. As projects progress, their number gradually reduces. Only 659 projects move to Phase I and 462 progress beyond phase I (with only 53 projects in phase III). Figure E.5 in Appendix E illustrates the funnel structure of the antidiabetics R&D.

3.3 Final database

For our final dataset of 1860 projects, of which 172 change ownership, we have the identity of targets and acquirers (*who*), the timing of acquisitions along the R&D process (*when*), and which type of R&D projects changes hands in terms of novelty (*what*). In Appendix D, we provide an example of a project in our sample with all dimensions of the data and with the variables defined in the previous sections.

Table 1 presents some first key summary statistics by various splits of the sample for the observations with and without ownership change ("M&A" vs "No M&A"). There are 186 ownership changes relating to 172 underlying projects. This represents 7% of observations and 10% of projects.[22] The first panel on timing (when) shows that many projects already change hands in the initial stages of the development process - 56% of projects are acquired while in the preclinical phase, and another 25% are acquired when in the early clinical stage, while they are lowest in the last clinical stage at 19% (although p-values do not pick up strong statistical differences between the percentage

[22]Note that the ownership changes include full M&As, affecting all projects of the target company but also partial ownership changes involving a portion of target's projects.

of acquired vs. non-acquired projects in each phase).

[Insert Table 1 about here]

Leaders are significantly less likely to be involved in acquiring projects (3% of projects experiencing M&As; see second panel). In contrast, stars are significantly more likely to be acquirers, as are small non-incumbent firms. The latter are the largest set of acquirers in our sample (30%). For targets (third panel), not surprisingly, small non-incumbents (32%), mature pipelines (38%), and young pipelines (25%) are the most common types. Only small non-incumbents and mature pipelines are statistically over-represented as targets.

Finally (last panel), NTO projects are no "outliers" in our sample, i.e., the overall share of projects which are novel is high (1071 out of 2730 observations). These rates are not unusual for the pharma sectors.[23] In any case, the share of NTO projects is significantly higher among the projects that changed hands (55%) compared to the share among projects that did not change hands (39%). In the following section, we further detail these descriptives and test them in an econometric framework.

4 Empirical implementation

Our main analysis on who acquires whom, and what when, is centered around simple linear probability models where we control for various company characteristics, project characteristics, and fixed effects with the aim to better isolate the key drivers of M&As in the antidiabetics market. The generic regression we estimate has the following form:

$$Prob(Dep.Var_{it} = 1) = \alpha + \beta X_{it} + \gamma FE + \epsilon_{it}. \quad (1)$$

The exact specification of dependent and independent variables depends on the question at hand. For example, when analyzing who is a likely acquirer, the dependent

[23]For example, Krieger et al. (2018) use novelty derived from the chemical dissimilarity between projects and find that between 45% and 55% of projects are novel, depending on the size of the originating firm.

variable is a binary indicator that equals one for a project (i) and phase (t) affected by an ownership change and zero otherwise, while the vector X contains then binary indicators for the acquirer types. The associated β coefficients indicate which acquirer types are positively or negatively associated with the likelihood of an ownership change.

The vector FE used across specifications includes fixed effects for the time point when the project is initiated ("cohort fixed effects"), the project's mechanism of action (MoA), the project's technological area, and the geographical location of the company. The first set of fixed effects, cohort fixed effects, group together projects initiated around the same time in order to control for time and/or technological trends. The sample is split into seven cohorts in three-year windows (1997-1999, 2000-2002,...).

The second set of fixed effects controls for the MoA of a project, determining how a drug produces its effect in the body. From a development perspective, significant heterogeneity exist between various MoAs in the underlying technology, success rate and the extent of R&D activity within an MoA. From a demand-side perspective, MoAs are closely linked to the side effects and suitability of treatment in different patient populations (Berger and Iyengar, 2011; Association et al., 2019; Chaudhury et al., 2017). For this reason, drugs with the same MoA are often considered to be more substitutable, and hence MoAs are often used to delineate antitrust product markets for launched drugs. To capture these differences, we introduce 19 MoA fixed effects into our regressions.[24]

The third set of 48 fixed effects captures differences between the technological areas, as indicated by the IPC subgroup codes of the patents assigned to projects (for example, A61P). Since a project typically relates to many technological areas, it will usually score on more than one of these fixed effects.[25]

[24] A total of 389 different MoAs are present in our sample. Since the number of MoAs is high and many MoAs contain only a very low number of projects, we introduce separate fixed effects only for the 17 largest MoAs where at least 30 projects were developed during our sample period. All other MoAs are aggregated into a single "rest" category, representing one MoA fixed effect. A last MoA fixed effect is then included for all projects with unknown MoA. Approximately 14% of the projects are combinatory, relying on two MoAs and 1% of projects rely on 3 MoAs. In these cases, projects can be be included in multiple categories.

[25] Similarly to MoAs, many technological areas contain only very few projects. We thus introduce

Finally, the last set of fixed effects refers to the geographical regions where targets and acquirers originate from to capture any different transaction propensities.[26]

We employ ordinary least squares (OLS) as the primary estimation method. Where our sample size or econometric specification allows, we also test the robustness of these estimates by estimating logistic regressions. Standard errors are clustered at the project level as the strong path-dependence of the pharmaceutical R&D process makes the independence of individual project-level error terms unlikely.[27]

5 Results

We examine which firms are the acquirers and which firms are the targets (*who*) in 5.1. Conditioning on a project changing ownership, we thereafter continue by investigating the matching between acquirers and targets, i.e., who acquires who (5.2). Section 5.3, in turn, looks at the role of the timing of the change of ownership (*when*). Section 5.4, finally, investigates the role of a project's NTO characteristics in the ownership changes (*what*).

5.1 Who acquires and who sells?

Table 2 (Column 1) investigates acquirer types in a regression framework, using our size-incumbency bins and with the acquiring status dummy as dependent variable. We take big incumbents as the base category, as these are in light of the current debate a natural benchmark (Cunningham et al., 2021; Argentesi et al., 2021; Gautier and Lamesch, 2021). In line with the descriptive analysis, the table shows that compared

separate fixed effects only for the 47 largest technological IPC classes where at least 100 projects were developed during our sample period. All the remaining IPC subgroups are aggregated into a single category. Overall, we thus have 48 IPC fixed effects.

[26] Carril-Caccia et al. (2022) show that the number and value of M&As between same-country firms are five times larger than between firms of different countries. We introduce fixed effects for regions depending on a company's headquarters. These regions include South-East Asia (Japan, China, India, Singapore and Taiwan), Europe, Northern America (Canada and the US), and the rest of the world.

[27] Our results are robust to the clustering of errors at the firm level.

to big incumbents, leaders are significantly less likely to acquire projects (at 1% significance level), whereas stars appear more likely acquirers (at 1% significance level). Furthermore, and also surprisingly confirming the descriptives, small companies and mature pipeline firms are as likely to acquire projects as big incumbents (their coefficients are not significant).

Also, the regression results with the target status dummy as dependent variable in Table 2 echo the insights from the summary statistics (Column 2). Using young pipelines as the base category, small companies and mature pipeline firms are significantly more likely to be targets. This result is non-trivial as it indicates that the acquirers in our sample prefer somewhat more seasoned targets, i.e., either more mature pipelines or small companies, already having a launched product in a pharma market. All other types of companies (grouped here into one bin), are less likely to be sellers of projects as compared to young pipelines.

[Insert Table 2 about here]

5.2 Who acquires whom?

In this section, we focus on the sample of the 186 ownership changes to analyze the pairing between acquirers and targets. The majority of deals (63%) take place among small non-incumbents and pipelines as targets *and* acquirers (see Figure A.1 in the Appendix), which is why we take those bins as separate dependent variable in Table 3.

Table 3 reports regressions estimating which type of targets are taken over by which types of acquirers (where we take big incumbents as benchmark). Column (1) investigates which type of firms acquire small non-incumbent targets. The analysis indicates that, relative to the base group (big incumbents), the big non-incumbents are the most likely acquirers of small non-incumbents. Pipeline firms, on the other hand, are *less* likely to acquire small non-incumbents. Column (2) looks at which firms are more likely acquirers of mature pipelines. Here small firms and pipeline firms are more likely acquirers than big incumbents. In Column (3) we focus on who

acquires the young pipeline firms. The results show that relative to big incumbents, big non-incumbents and small firms are significantly less likely to acquire young pipeline firms. Other pipelines, on the other hand, are not significantly less likely than big incumbents to acquire. Thus, mature pipeline firms are relatively more attractive targets for pipeline firms and small firms, while young pipeline firms are less attractive targets for big non-incumbents, who prefer small firms as targets.

[Insert Table 3 about here]

5.3 Who acquires when?

When looking at who acquires when, we see in column (1) of Table 4 that compared to big incumbents, stars, small firms, and mature pipelines are significantly more likely to acquire projects in the preclinical phase, where these preclinical projects are high risk with high probability, as it is uncertain they will reach final stages.[28] In contrast, in column (2), we see that the big non-incumbents are more likely to acquire late (as compared to big incumbents). These firms are most likely to enter into the antidiabetics by acquiring close-to-launch projects.

[Insert Table 4 about here]

5.4 What is transacted?

5.4.1 Who holds NTO projects?

On the question of *what* is transacted, we look at whether a project is novel, reflecting its high-risk/high-gain profile. To this end, we first examine who is likely to develop NTO projects by regressing our binary NTO indicator on the firm type bins in Table 5.[29] The results in column (1) indicate that projects with NTO patents are more likely to

[28] The descriptive statistics of Table A.2 confirm these results.

[29] As additional controls we include the number of patents assigned to a particular project (before a transaction occurs) and an indicator for projects with no assigned patents, as this affects the likelihood of scoring on NTO.

be owned by small non-incumbents and mature pipeline firms. This is very much in line with the lower incentives for incumbents, large or small, leader or not, to launch novel innovations, potentially cannibalising existing projects.

[Insert Table 5 about here]

5.4.2 When are NTO projects bought?

Table 6 tests whether and *when* NTO projects are bought. First, NTO projects are more likely to be taken over in general (column 1). Having an NTO profile is a strong predictor of transaction activity, as the NTO status is associated with a 3.6 percentage point increase in the probability of being acquired (50% increase in the probability to be taken over relative to baseline probability), as potential high-gain projects are interesting to acquire, despite their higher risk. When timing is interacted with the NTO status (column 2), we see that NTO projects are more likely to be acquired in the preclinical and early stages, and not in the late stage, further enforcing the risk status of these transactions. Non-NTO projects, when acquired, are significantly less likely to be acquired in preclinical phases (column 2). These are the low risk targets, both in terms of what and when.

[Insert Table 6 about here]

5.4.3 Who acquires and who sells NTO projects?

Given that small non-incumbent firms and especially mature pipeline firms are more likely to be the originators of NTO projects, we explore whether these projects are more likely to be involved in a transaction. Table 7 analyses which types of firms are likely to buy (column 1) or sell (column 2) NTO projects in transactions where they are involved.

As acquirers, only stars and mature pipelines are more likely to acquire NTO projects, relative to big incumbents (column 1). Big non-incumbents and leaders are

not more likely to acquire NTO projects. Looking at the identity of the targets (column 2), we see that mature pipeline firms as targets are significantly more likely to involve NTO projects, compared to young pipeline firms, whereas other firm types are not.

[Insert Table 7 about here]

So far, we found that mature pipelines are more likely to hold and sell NTO projects. Second, NTO projects are found to be more likely to change hands in the earlier phases. Third, mature pipelines and star firms are most likely to acquire NTO projects. Table 8 brings all these findings together. Column (1) looks into how the matches between target and acquirer differ for transactions involving NTO projects compared to others. The results show that deals of mature pipelines as target and acquirer are significantly more likely to involve NTO projects. Also when star firms acquire mature pipeline projects, they are more likely to involve NTO projects. Column 2 in Table 8 shows that these significant differences in matches involving NTO projects are primarily happening in the pre-clinical phase.

[Insert Table 8 about here]

To conclude, projects with an NTO profile are more likely candidates for ownership changes. These projects are likely to originate from mature pipeline firms and change ownership at the beginning of their development in the preclinical stage. NTO projects seem to play a distinctive role in the acquisition strategies of star firms which acquire targets owning such projects, and are especially central to transactions between mature pipeline firms, but are less on the radar of the large incumbents, including the leaders.

6 Conclusion

This paper takes a dive into the R&D activities of the companies working towards developing antidiabetic drugs and explores the changes in ownership of these R&D projects. Our dataset tracks ownership changes for all antidiabetic projects between

1997-2017, along with their progression through the R&D pipeline. We look at the market and technology profiles of the acquirers and their targets (who) and the timing of their acquisition decisions (when). We further examine in detail the technological characteristics of the projects changing hands (what). More specifically, we characterize the breakthrough nature of the R&D projects and their high-risk/high-gain profile as proxied by the technological novelty of the projects. To characterize projects in terms of their technology novelty and high-risk/high-gain profile, we use project-patent technology links.

Our results show that the M&A transaction landscape is more rich and varied than typically commented. Most of the changes in ownership happen already in the early preclinical stages when projects are still far from being launched into the product market and have high failure rates. Most of these deals occur between small and pipeline firms, which do not yet have any marketed antidiabetic drugs. Also stars, i.e., fast-growing biopharma companies not yet selling antidiabetic drugs, are active as acquirers in these early stages. Novel projects, which are more likely to originate in pipeline firms, are more likely to change ownership, and this early on while still in the preclinical stage, again with the involvement of pipeline companies and stars as most likely acquirers.

In contrast, the industry leaders in the antidiabetics product markets are rather inactive in acquiring projects, challenging the narrative that large incumbents acquire small targets with projects close to product launch to neutralize future competition. Instead, non-incumbent pharma companies are relatively more active as large acquirers, driven by a motive to enter the attractive antidiabetics market. They do so mostly in later stages and avoid acquiring the novel projects, i.e., their acquisitions typically seem to have a lower risk profile.

Taken together, our findings highlight the key roles that technology uncertainty (early stages) and technological profile (novelty) of projects play in M&As in pharmaceutical markets, affecting the risk/reward motives of ownership changes and thus their likely impact. These technology dimensions are largely neglected by academic re-

searchers and policymakers. Their current focus on acquisitions by big incumbents of projects close to the market seems thus too narrow. We argue that one should broaden the scope of investigations when studying the interaction of M&As and innovation, i.e., to include also the transactions between small and pipeline companies typically occurring in the early stages.

Future work should consider the (innovation) implications of the multiple early-stage acquisitions of novel projects by mature pipelines and stars. Although these projects, being high risk, may have a high probability of not making it to market, their potential breakthrough nature calls for a careful analysis of the impact of a change in ownership on their pathway to success. On the one hand, the potential for killer acquisitions, typically only associated with large incumbent acquirers, could also be present in transactions with small innovative firms in the early R&D stages. On the other hand, the uncertainty around early-stage novel projects might lead to many of these acquisitions ending in failure without a killer motive. Furthermore, some of these transactions might yield positive outcomes due to synergies, as small innovative acquirers are also more likely to have a novel project portfolio.

Finally, while the current paper focuses on antidiabetics, given the data requirements, this exercise can in principle be expanded to other pharmaceutical markets to broaden and generalize the scope of the analysis and our findings.

References

Adams, C. P. and V. V. Brantner (2006). Estimating the cost of new drug development: is it really $802 million? *Health affairs 25*(2), 420–428.

Argente, D., S. Baslandze, S. Moreira, and D. Hanley (2019). Patents to products: innovation, product creation, and firm growth. *Unpublished, Penn State*.

Argentesi, E., P. Buccirossi, E. Calvano, T. Duso, A. Marrazzo, and S. Nava (2021). Merger policy in digital markets: An ex post assessment. *Journal of Competition Law & Economics 17*(1), 95–140.

Arroyabe, M. F. (2021). The role of patent expiration in acquisition decision and target selection in the pharmaceutical industry. *R&D Management 51*(5), 521–537.

Arts, S., J. Hou, and J. C. Gomez (2021). Natural language processing to identify the creation and impact of new technologies in patent text: Code, data, and new measures. *Research Policy 50*(2), 104144.

Association, A. D. et al. (2019). 9. Pharmacologic approaches to glycemic treatment: Standards of Medical Care in Diabetes-2019. *Diabetes care 42*(Supplement 1), S90–S102.

Autor, D., D. Dorn, L. F. Katz, C. Patterson, and J. Van Reenen (2020). The fall of the labor share and the rise of superstar firms. *The Quarterly Journal of Economics 135*(2), 645–709.

Bena, J. and K. Li (2014). Corporate innovations and mergers and acquisitions. *The Journal of Finance 69*(5), 1923–1960.

Berger, S. I. and R. Iyengar (2011). Role of systems pharmacology in understanding drug adverse events. *Wiley interdisciplinary reviews: systems biology and medicine 3*(2), 129–135.

Blume-Kohout, M. E. and N. Sood (2013). Market size and innovation: Effects of Medicare Part D on pharmaceutical research and development. *Journal of public economics 97*, 327–336.

Branstetter, L., C. Chatterjee, and M. J. Higgins (2014). Starving (or fattening) the golden goose?: Generic entry and the incentives for early-stage pharmaceutical innovation. Technical report, National Bureau of Economic Research.

Carril-Caccia, F., A. Garmendia-Lazcano, and A. Minondo (2022). The border effect on mergers and acquisitions. *Empirical Economics 62*(3), 1267–1292.

Chaudhury, A., C. Duvoor, R. Dendi, V. Sena, S. Kraleti, A. Chada, R. Ravilla, A. Marco, N. S. Shekhawat, M. T. Montales, et al. (2017). Clinical review of antidiabetic drugs: implications for type 2 diabetes mellitus management. *Frontiers in endocrinology 8*, 6.

Christensen, C. M. (2013). *The innovator's dilemma: when new technologies cause great firms to fail*. Harvard Business Review Press.

Comanor, W. S. and F. M. Scherer (2013). Mergers and innovation in the pharmaceutical industry. *Journal of health economics 32*(1), 106–113.

Cunningham, C. (2017). When does novelty pay? In *Academy of Management Proceedings*, Volume 2017, pp. 16957. Academy of Management Briarcliff Manor, NY 10510.

Cunningham, C., F. Ederer, and S. Ma (2021). Killer acquisitions. *Journal of Political Economy 129*(3), 649–702.

Danzon, P. M., A. Epstein, and S. Nicholson (2007). Mergers and acquisitions in the pharmaceutical and biotech industries. *Managerial and Decision Economics 28*(4-5), 307–328.

De Loecker, J., J. Eeckhout, and G. Unger (2020). The rise of market power and the macroeconomic implications. *The Quarterly Journal of Economics 135*(2), 561–644.

Diez, M. F., M. D. Leigh, and S. Tambunlertchai (2018). *Global market power and its macroeconomic implications*. International Monetary Fund.

Dranove, D., C. Garthwaite, and M. I. Hermosilla (2020). Expected profits and the scientific novelty of innovation. Technical report, National Bureau of Economic Research.

European Commission (2021, 5). Commission fines Sigma-Aldrich EUR 7.5 million for providing misleading information during Merck takeover investigation. `https://ec.europa.eu/commission/presscorner/detail/en/IP_21_2181`.

EvaluatePharma (2018). EvaluatePharma World Preview 2018, Outlook to 2024. Technical report, EvaluatePharma.

FDA (2018, April). The drug development process.

Fleming, L. (2001). Recombinant uncertainty in technological search. *Management science 47*(1), 117–132.

Foster, J. G., A. Rzhetsky, and J. A. Evans (2015). Tradition and innovation in scientists' research strategies. *American Sociological Review 80*(5), 875–908.

Gautier, A. and J. Lamesch (2021). Mergers in the digital economy. *Information Economics and Policy 54*, 100890.

Grassano, N., H. Hernandez, P. Fako, A. Tuebke, S. Amoroso, A. Georgakaki, L. Napolitano, F. Pasimeni, F. Rentocchini, R. Compano, et al. (2021). The 2021 eu industrial r&d investment scoreboard. Technical report, Joint Research Centre (Seville site).

Hall, B. H. and J. Lerner (2010). The financing of R&D and innovation. In *Handbook of the Economics of Innovation*, Volume 1, pp. 609–639. Elsevier.

Haucap, J., A. Rasch, and J. Stiebale (2019). How mergers affect innovation: Theory and evidence. *International Journal of Industrial Organization 63*, 283–325.

Hernández, H., N. Grassano, A. Tuebke, L. Potters, P. Gkotsis, and A. Vezzani (2014). The 2018 EU industrial R&D investment scoreboard. *EU R&D Scoreboard Report*.

Huvaj, M. N. and W. C. Johnson (2019). Organizational complexity and innovation portfolio decisions: Evidence from a quasi-natural experiment. *Journal of Business Research 98*, 153–165.

International Diabetes Federation (2021). IDF diabetes atlas: Diabetes facts & figures. Available: https://idf.org/aboutdiabetes/what-is-diabetes/facts-figures.html (Accessed 26 January 2022).

Jensen, K. and F. Murray (2005). Intellectual property landscape of the human genome. *Science 310*(5746), 239–240.

Krieger, J. L., D. Li, and D. Papanikolaou (2018, May). Missing novelty in drug development. Working Paper 24595, National Bureau of Economic Research.

Kyle, M. K. (2007). Pharmaceutical price controls and entry strategies. *The Review of Economics and Statistics 89*(1), 88–99.

Ornaghi, C. (2009). Mergers and innovation in big pharma. *International journal of industrial organization 27*(1), 70–79.

Pammolli, F., L. Magazzini, and M. Riccaboni (2011). The productivity crisis in pharmaceutical r&d. *Nature reviews Drug discovery 10*(6), 428.

Phillips, G. M. and A. Zhdanov (2013). R&d and the incentives from merger and acquisition activity. *The Review of Financial Studies 26*(1), 34–78.

Sampat, B. and H. L. Williams (2019). How do patents affect follow-on innovation? evidence from the human genome. *American Economic Review 109*(1), 203–36.

Schmoch, U. (2008). Concept of a technology classification for country comparisons. *Final report to the world intellectual property organisation (wipo), WIPO*.

Szücs, F. (2014). M&A and R&D: Asymmetric effects on acquirers and targets? *Research Policy 43*(7), 1264–1273.

Verhoeven, D., J. Bakker, and R. Veugelers (2016). Measuring technological novelty with patent-based indicators. *Research Policy 45*(3), 707–723.

White, J. R. (2014). A brief history of the development of diabetes medications. *Diabetes Spectrum 27*(2), 82–86.

Figures and Tables

Tab. 1: Summary statistics

	No event Count	Mean	Event Count	Mean	Difference p-value
Timing:					
Preclinical	1687	0.62	104	0.56	0.12
Early	612	0.22	47	0.25	0.39
Late	431	0.16	35	0.19	0.31
Acquirer:					
Leader	325	0.12	6	0.03	0.00
Big inc	334	0.12	25	0.13	0.64
Big non-inc	220	0.08	22	0.12	0.12
Star	16	0.01	10	0.05	0.00
Small inc	21	0.01	1	0.01	0.68
Small non-inc	543	0.20	56	0.30	0.00
Mature pip	575	0.21	45	0.24	0.34
Young pip	696	0.25	21	0.11	0.00
Target:					
Leader	323	0.12	2	0.01	0.00
Big inc	334	0.12	1	0.01	0.00
Big non-inc	222	0.08	3	0.02	0.00
Star	16	0.01	1	0.01	0.93
Small inc	21	0.01	3	0.02	0.37
Small non-inc	543	0.20	59	0.32	0.00
Mature pip	575	0.21	70	0.38	0.00
Young pip	696	0.25	47	0.25	0.95
Project characteristics:					
NTO	1071	0.39	102	0.55	0.00
Observations	2730	2730	186	186	2916

Tab. 2: Who acquires and who sells?

	(1) Who acquires	(2) Who sells
Leader	-0.069***	
	(0.018)	
Big non-inc	0.017	
	(0.023)	
Star	0.292***	
	(0.104)	
Leader/Big/Star		-0.066***
		(0.011)
Small	0.012	0.043**
	(0.020)	(0.017)
Mature pip	-0.016	0.042***
	(0.020)	(0.015)
Young pip	-0.061***	
	(0.018)	
Cohort FE	Yes	Yes
MoA FE	Yes	Yes
Country FE	Yes	Yes
Obs	2916	2916
Adj. R2	0.042	0.046
Base	Big inc	Young pip

Note: This table presents the OLS estimates of the likelihood to acquire projects (column 1) and to sell projects (column 2). The dependent variable is a binary indicator equal to one for project i in phase t if it was acquired and zero otherwise. For acquirers, small incumbents and small non-incumbents are aggregated, due to a low number of observations. For targets, leaders, big incumbents, big-non incumbents, and stars are aggregated. Errors are clustered at the project level and displayed in parentheses. Table A.1 in Appendix A shows that these results are robust when employing logit estimation or clustering standard errors at the firm level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Tab. 3: Who acquires whom?

	(1) T is small non-inc	(2) T is mature pip	(3) T is young pip
Leader	-0.249	0.394*	-0.053
	(0.246)	(0.236)	(0.280)
Big non-inc	0.278**	0.106	-0.270**
	(0.140)	(0.113)	(0.122)
Star	-0.086	0.097	0.020
	(0.182)	(0.195)	(0.211)
Small	-0.056	0.455***	-0.275**
	(0.138)	(0.119)	(0.121)
Mature pip	-0.347***	0.615***	-0.174
	(0.112)	(0.115)	(0.125)
Young pip	-0.302**	0.402**	-0.021
	(0.128)	(0.165)	(0.158)
Cohort FE	Yes	Yes	Yes
MoA FE	Yes	Yes	Yes
Country FE A+T	Yes	Yes	Yes
Obs	186	186	186
Adj. R2	0.306	0.145	0.171
Base	Big inc	Big inc	Big inc

Note: This table presents the OLS estimates of the likelihood to acquire projects of a particular target type. For acquirers, small incumbents and small non-incumbents are aggregated, due to a low number of observations. In column (1), the dependent variable equals one if the target is a small non-incumbent and zero otherwise. In column (2), the dependent variable equals one if the target is a mature pipeline firm and zero otherwise. In column (3), the dependent variable equals one if the target is a young pipeline firm and zero otherwise. For acquirers, small incumbents and small non-incumbents are aggregated, due to a low number of observation. Errors are clustered at the project level and displayed in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Tab. 4: Who acquires when?

	(1) Preclinical	(2) Late
Leader	0.119	0.285
	(0.229)	(0.225)
Big non-inc	0.127	0.228*
	(0.170)	(0.134)
Star	0.464***	-0.042
	(0.154)	(0.108)
Small	0.329**	-0.114
	(0.133)	(0.096)
Mature pip	0.256*	-0.048
	(0.133)	(0.103)
Young pip	0.049	0.009
	(0.158)	(0.112)
Cohort FE	Yes	Yes
MoA FE	Yes	Yes
Country FE	Yes	Yes
Obs	186	186
Adj. R2	0.196	0.201
Base	Big inc	Big inc

Note: This table presents the OLS estimates of the likelihood to acquire projects in a particular development phase. In column (1), the dependent variable equals one if the project was taken over in preclinical phase and zero otherwise. In column (2), the dependent variable equals one if the project was taken over in the late phase and zero otherwise. Small incumbents and small non-incumbents are aggregated, due to a low number of observations. Errors are clustered at project level and displayed in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Tab. 5: Who holds which projects?

	(1) NTO
Leader	0.070
	(0.047)
Big non-inc	0.068
	(0.046)
Star	0.136
	(0.169)
Small inc	0.157
	(0.111)
Small non-inc	0.080*
	(0.042)
Mature pip	0.084**
	(0.041)
Young pip	0.036
	(0.039)
Cohort FE	Yes
MoA FE	Yes
Country FE	Yes
IPC	Yes
Patent nb.	Yes
No patents	Yes
Obs	2916
Adj. R2	0.347
Base	Big inc

Note: This table presents the OLS estimates of the likelihood to own NTO projects. The dependent variable is a binary indicator for the NTO indicator. Errors are clustered at project level and displayed in parentheses. The results are robust to clustering of errors at the firm level (unreported). Logit estimates produce similar results (unreported). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Tab. 6: What is acquired and when?

	(1) What	(2) What and when
NTO	0.036***	-0.016
	(0.012)	(0.028)
Preclinical		-0.042**
		(0.021)
Early		-0.031
		(0.023)
NTO × Preclinical		0.063**
		(0.029)
NTO × Early		0.055*
		(0.033)
Cohort FE	Yes	Yes
MoA FE	Yes	Yes
Country FE	Yes	Yes
IPC	Yes	Yes
Patent nb.	Yes	Yes
No patents	Yes	Yes
Obs	2916	2916
Adj. R2	0.045	0.046
Base		Late

Note: This table presents the OLS estimates of the likelihood of when NTO projects are acquired. The dependent variable is a binary indicator equal to one if project i was acquired in phase t and zero otherwise. Errors are clustered at the project level and displayed in parentheses. The results are robust to the clustering of errors at the firm level and logit specifications (unreported). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Tab. 7: Who acquires and who sells NTO projects?

	(1) Who acquires NTO	(2) Who sells NTO
Leader	0.030	
	(0.258)	
Big non-inc	0.073	
	(0.189)	
Star	0.451*	
	(0.262)	
Leader/Big/Star		0.358
		(0.225)
Small	0.138	0.207*
	(0.134)	(0.115)
Mature pip	0.270*	0.308***
	(0.138)	(0.099)
Young pip	0.019	
	(0.213)	
Cohort FE	Yes	Yes
MoA FE	Yes	Yes
Country FE	Yes	Yes
IPC	Yes	Yes
Patent nb.	Yes	Yes
No patents	Yes	Yes
Obs	186	186
Adj. R2	0.303	0.323
Base	Big inc	Young pip

Note: This table presents the OLS estimates of the likelihood to buy (column 1) or sell (column 2) NTO projects. The dependent variable equals one if the acquired (sold) project was an NTO project and zero otherwise. For acquirers, small incumbents and small non-incumbents are aggregated, due to a low number of observations. For targets, leaders, big incumbents, big-non incumbents, and stars are aggregated. Errors are clustered at the project level and displayed in parentheses. The results are robust to the clustering of errors at the firm level (unreported). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Tab. 8: Which matches are more likely for NTO projects?

	(1) NTO	(2) NTO in preclin
A mature pip + T mature pip	0.261**	0.262**
	(0.122)	(0.124)
A star + T mature pip	0.775***	0.793***
	(0.214)	(0.190)
A other + T mature pip	0.127	0.028
	(0.101)	(0.097)
Cohort FE	Yes	Yes
MoA FE	Yes	Yes
Country FE	Yes	Yes
IPC	Yes	Yes
Patent nb.	Yes	Yes
Obs	186	186
Adj. R2	0.333	0.326
Base	All other combinations	All other combinations

Note: This table presents the OLS estimates of how likely a particular match of acquirer and target is for NTO projects. In column (1), the dependent variable equals one if the acquired project was an NTO project and zero if the acquired project was not an NTO project. In column (2), the dependent variable equals one if the acquired project was an NTO project in the preclinical phase and zero otherwise. Errors are clustered at the project level and displayed in parentheses. The results are robust to the clustering of errors at the firm level (unreported). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Appendix

A Robustness checks and extensions

Fig. A.1: Who acquires whom?

Notes: The figure shows which acquirers buy projects of which targets. The numbers indicate the number of projects changing hands, colored depending on the frequency (186 in total).

Tab. A.1: Robustness: Who acquires and who sells?

	Logit		Cluster SE: Firm	
	(1)	(2)	(3)	(4)
	Acquirers	Targets	Acquirers	Targets
Leader	-1.608***		-0.069***	
	(0.507)		(0.015)	
Big non-inc	0.221		0.017	
	(0.328)		(0.029)	
Star	1.945***		0.292**	
	(0.604)		(0.129)	
Leader - Star		-2.316***		-0.066***
		(0.425)		(0.015)
Small	0.149	0.598**	0.012	0.043*
	(0.292)	(0.236)	(0.021)	(0.022)
Mature pip	-0.233	0.580***	-0.016	0.042**
	(0.312)	(0.208)	(0.020)	(0.020)
Young pip	-1.222***		-0.061***	
	(0.350)		(0.016)	
Cohort FE	Yes	Yes	Yes	Yes
MoA FE	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes
Obs	2916	2916	2916	2916
Pseudo R2	0.106	0.140		
Adj. R2			0.042	0.046
Base	Big-inc	Young pip	Big-inc	Young pip

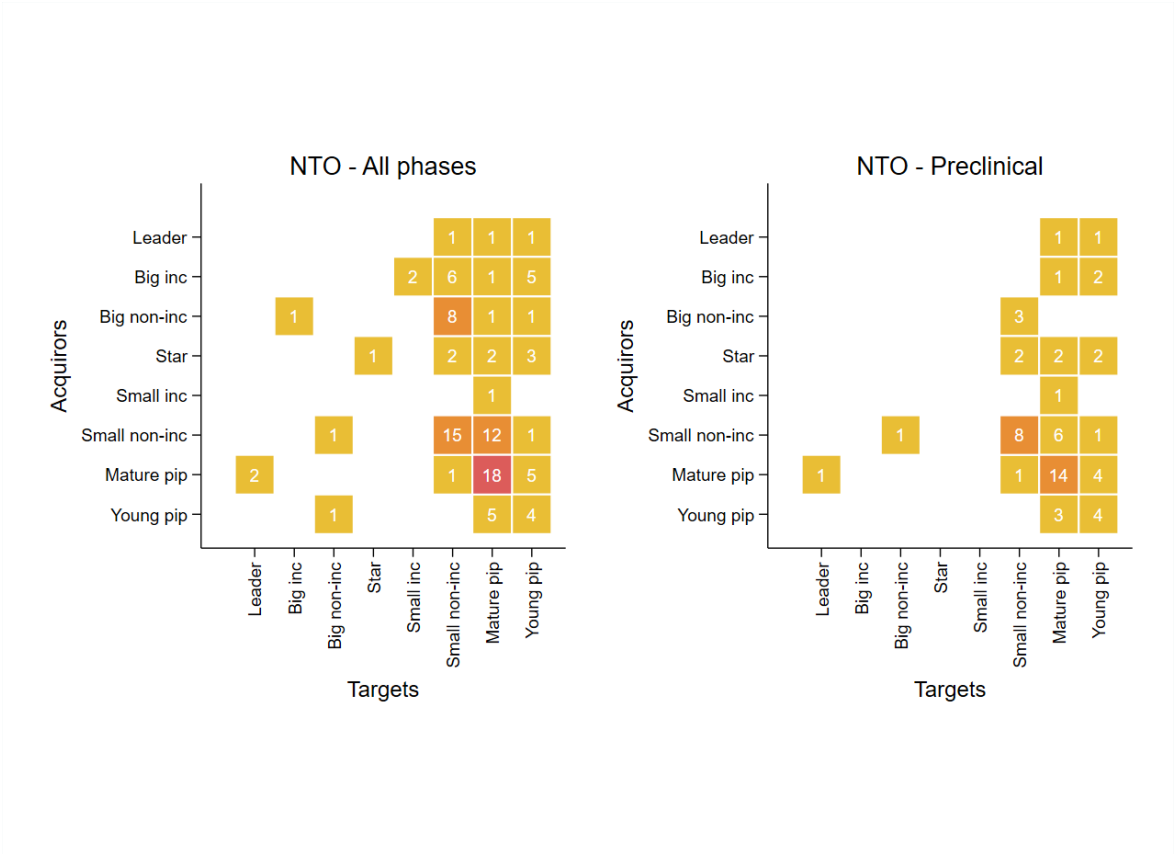
Note: This table presents robustness checks of the likelihood to acquire projects (odd columns) or sell projects (even columns). Columns (1) and (2) use a logistic regression instead of OLS. Columns (3) and (4) cluster standard errors at the firm level (instead of at the project level), allowing for a correlation within firms. The dependent variable in each column is a binary indicator equal to one if project i was acquired in phase t (treated) and zero otherwise. Aggregation of bins follows the baseline specification. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Tab. A.2: Who acquires when?

Acquirer	Preclinical	Early	Late	Total
Leader	4 (67%)	0 (0%)	2 (33%)	6 (100%)
	4%	*0%*	*6%*	*3%*
Big inc	10 (40%)	12 (48%)	3 (12%)	25 (100%)
	10%	*26%*	*9%*	*13%*
Big non-inc	9 (41%)	3 (14%)	10 (45%)	22 (100%)
	9%	*6%*	*29%*	*12%*
Star	8 (80%)	1 (10%)	1 (10%)	10 (100%)
	8%	*2%*	*3%*	*5%*
Small inc	1 (100%)	0 (0%)	0 (0%)	1 (100%)
	1%	*0%*	*0%*	*1%*
Small non-inc	31 (55%)	16 (29%)	9 (16%)	56 (100%)
	30%	*34%*	*26%*	*30%*
Mature pipeline	31 (69%)	8 (18%)	6 (13%)	45 (100%)
	30%	*17%*	*17%*	*24%*
Young pipeline	10 (48%)	7 (33%)	4 (19%)	21 (100%)
	10%	*15%*	*11%*	*11%*
Total	104 (56%)	47 (25%)	35 (19%)	186 (100%)
	100%	*100%*	*100%*	*100%*

Note: This table presents when each type of acquirer takes over projects. The even rows give the number of acquired projects in each phase together with the share of phases of each acquirer's type transactions in parentheses (row-wise percentages). The odd rows give the share of the different types of acquirers on the transaction within a given phase (column-wise percentages).

Fig. A.2: When are NTO projects bought?



Note: The figure shows which acquirers buy NTO projects of which targets. In the left panel, the numbers represent counts of ownership changes involving the NTO projects (102 in total). In the right panel, the numbers represent counts of ownership changes involving the NTO projects in the preclinical phase (58 in total). Colouring depends on observed frequency.

B Examples of companies in the bins

Tab. B.3: Examples of firms in the bins

Bin	Number of firms	Examples (up to 10 firms)
Leaders	**7**	Aventis, Eli lilly, Hoechst Marion Roussel, Merck & Co., Novo Nordisk, Sanofi, Sanofi-Aventis
Big incumbents	**11**	Astrazeneca, Bayer, Boehringer Ingelheim, Bristol-Myers Squibb, Glaxosmithkline, Johnson & Johnson, Merck kgaa, Pfizer, Roche, Takeda
Big non-incumbents	**19**	Abbott, American Home Products, Astellas Pharma, Daiichi Sankyo, Glaxo Wellcome, Glaxosmithkline, Johnson & Johnson, Sanofi-Synthelabo, Schering-Plough, Wyeth
Stars	**8**	Allergan, Biogen, Biogen Idec, Celgene, Gilead Sciences, Shire, Teva, Valeant Pharmaceuticals
Small incumbents	**10**	Ajinomoto, Amylin, Andrx, Depomed, Mannkind, Mitsubishi Tanabe Pharma, Nektar Therapeutics, Pharmacia, Veroscience, Zealand pharma
Small non-incumbents	**195**	Alexion, Baxter International, Chiron, Eisai, Kos Pharmaceuticals, Mochida, Solvay, Pliva, Warner-Lambert, Tanabe Seiyaku
Mature pipelines	**300**	Genmedica, Kemia, Develogen, Olipass, Catabasis pharmaceuticals, Genmedica, Alize Pharma, Biocure Pharma, Cloud Pharmaceuticals, Insmed
Young pipelines	**408**	Escoublac, Halsa, Hanall Biopharma, Limerick Biopharma, Energesis Pharmaceuticals, Mitochon Pharmaceuticals, Hansoh Pharmaceuticals, Crititech, Kinex Pharmaceuticals, Sprint Bioscience

Note: The table provides examples of companies in each bin. The total number of companies in the bin is indicated in column 2 and up to 10 examples are provided. Firms might appear in multiple bins if they switch bins over time. Some listed firms do not exist today as they have been incorporated into other firms through M&A deals between 1997 and 2017.

C Patent matching

This appendix is a guide to matching patents to antidiabetic projects. According to our knowledge, there is no publicly available database providing project-patent links beyond launched drugs. The private databases, on the other hand, do not provide sufficient coverage (e.g. Cortellis) are in-transparent as to how they assign patents to projects.

C.1 Patent databases

We focus on US patents exclusively. Three sources of patent data are used: the USPTO patent database[30], PATSTAT database, and the LENS patent database.[31] In these databases, we particularly utilize the information regarding patent application dates, patent grant dates, information on patent extensions, priority dates, IPC patent classes, patent families, patent assignees, and backward patent references.

Our matching starts by establishing so called 'candidate patent sets' for each of the 2387 projects in the sample. The candidate patent set contains all granted US patents of the firms that were involved in the development of a project (originator, the final owner, and all owners in the chain in between) and that were filed for between the initiation date and the discontinuation date of a project (retrieved from the progression through clinical trials).[32] To narrow down the patent universe to technologies which plausibly relate to pharmaceutical markets and drug development, we follow Schmoch (2008) and own analysis based on the FDA's Orange book and consider only patents with at least one IPC subclass in A61K, A61M, A61P, C07C, C07D, C07F, C07K, C07H, C08F, C08G, C12N or C12P. After this step, each project has a set of patents filed for during its lifetime and belonging to the relevant firms - so called candidate set - in which we look for the patents relevant for a specific project.

There are two major issues when searching for patents belonging to a specific project. First, there is significant heterogeneity between the projects in our sample, ranging from small molecule drugs to large molecule drugs. Second, a substantial number of projects in the Pharmaprojects database

[30] Available at https://patentsview.org/download/data-download-tables

[31] Available at https://lens.org

[32] To identify the patents belonging to each candidate set, we perform fuzzy string matching on the company names and patent assginee names. To improve precision of the matching routine, we first standardize the names. We remove the legal forms of companies, clean the names from non-alphanumerical characters, remove generic words and combinations of words that do only distinguish separate legal entities, but refer to the same underlying company (eg 'pharmaceutical products', 'intellectual properties', 'healthcare systems'). After this, we standardize the company names by taking the first word of the company's name, keeping the most numerous one and using it to substitute the other names referring to that company. Following the fuzzy string-matching routine, we manually checked the correctness on a random sub-sample of firms. The procedure yielded a minimum of false matches.

misses information along the relevant dimensions. Our matching procedure therefore consists of several complementary approaches that try to overcome these issues.

C.2 Chemistry matching

Patents for 'small molecule' or chemical drugs are matched based on their chemical properties. Using the Surechem[33] database and various identifiers (CAS numbers, SMILES Chemical structures, and UNII identifiers), we link project's underlying chemical compounds to PubChem[34]. PubChem contains information on the patents protecting specific molecules for some chemical entities. If we find an overlap between the 'candidate set' and the set of patents retrieved from PubChem, we consider these patents as assigned to a project[35]. This procedure results in 411 matches which are excluded from further matching.

C.3 Gene matching

Patents for 'large molecule' or biological drugs are matched based on gene sequences. Since the early 1990s, US patent applications claiming genes as intellectual property must disclose the exact DNA or protein sequences claimed in the text of the patent. The sequences are listed in USPTO patent applications in a standard format, labelled with the text 'SEQ ID NO'. Bio-informatics methods can be used to compare these sequences against the census of human genome to annotate each sequence with standard gene identifiers. In turn, these can be linked to outside databases, including the Pharmaprojects database.

We broadly follow the methodology suggested in Sampat and Williams (2019). First, we extract standard gene identifiers (known under Entrez gene ID) from the Pharmaprojects database. Using the GeneBank generated crosswalks[36], we connected each of the gene IDs to a list of mRNA, RNA and protein RefSeq accession/version numbers and extracted the corresponding nucleotide sequences from the GeneBank's webpage, including start and end positions of the chain in the sequence, if applicable. To capture the full universe of known sequences relating to a particular gene ID, we utilize both the sequences relating to the annotated human genome as well as the sequences maintained independently

[33]SureChEMBL provides free access to chemical data extracted from the patent literature. Available here: ftp://ftp.ebi.ac.uk/pub/databases/chembl/SureChEMBL/

[34]PubChem is a publicly available, open chemistry database at the National Institutes of Health (NIH). Available here: https://pubchemdocs.ncbi.nlm.nih.gov/downloads

[35]If match on company level is found but the patent lies outside of the development window of a project, we check whether other member of the patent's family lies in the development window of a project. If so, this is considered a match as well.

[36]Available here: ftp://ftp.ncbi.nih.gov/refseq/release/release-catalog/release97.accession2geneid.gz

of the annotated genome.

Following the methodology pioneered by Jensen and Murray (2005), we use the BLAST (Basic Local Alignment Search Tool) search engine to compare the above sequences to census of sequences disclosed in the US patents. To arrive at the true set of matches, we only consider blast matches with an E-value of less than 1e-50[37]. This yields a final mapping between projects and patents referring to the respective gene via the disclosed sequences. Again, we only consider patents as assigned to a project when they fall within the development window and belong to one of the relevant firms. This procedure results in 222 matches which are excluded from further matching.

C.4 MoA keyword matching

We complement the two above approached by text analysis using the mechanism of action (MoA). For projects with known MoA, we first perform cleaning to standardize MoA names,[38] obtaining a set of tokens ("keywords"). Using a combination of TF-IDF algorithm and a manual check, we also retrieve all relevant synonyms relating to a particular MoA[39] and add those to the relevant "keywords". To find the counterparts of these "keywords" in patents and establish matches, we utilize a database of Arts et al.. Arts et al. (2021) pre-process the text in the patents by concatenating the title and abstract and claims text, lowercasing the text, tokenizing all words, and eliminating stop words based on a manually compiled list, removing words with only one character, numbers, and words that appear only once across all patents. We pair a patent to a project if all MoA "keywords" or an abbreviation are found in a patent document. We then check that only patents are kept that belong to the relevant firms and which were applied for during the development window of a drug project. This procedure results in 487 matches which are excluded from further matching.

C.5 Remaining matching

If no match for a project has been found so far, we proceed by various plausible exclusion restrictions. For example, projects with a single patent in the candidate set are considered matched. Similarly, all patents are assigned to a project if the firm had only one project under development. We also matched

[37]Sampat and Williams (2019) use an E-value of exactly 0. However, we apply less strict threshold as it was confirmed by bio-informatics specialist that our threshold level is commonly applied in the field and a threshold of strict zero might be too restrictive.

[38]This includes tokenization, removal of special characters or words comprised of single letter only, and stemming using Porter's stemmer.

[39]For example, the MoA "glucagon-like peptide" is often only mentioned using its abbreviation glp1 or glp-1. The fibroblast growth factor 21 is known under fgf21 or fgf-21. The DPP-IV mentioned above is sometimes referred to as DPP-4, DPP4 or dipeptidyl peptidase 4 inhibitor.

projects to 0 patents if a firm had no US granted patent (we checked all these instances manually to verify the absence of the US patents). Lastly, the 93 launched projects were matched to patents based on information in the FDA's Orange Book. In total, this yielded an additional 660 matches.

The remaining set of 570 unmatched projects were checked manually. For each of these projects, patent text was compared to the above project's properties (where available). We hired a chemistry student to then read these text fields, compare those, and decide which patents from the set of candidate patents should be relevant for a particular project. This resulted in another 161 matched project. We drop the final unmatched projects from our sample. This leaves us with a final sample of 1941 matched projects and 437 unmatched projects, representing a match rate of 81%. A total of 4999 patents were assigned to the projects.

C.6 Summary statistics on matching

Although the Orange Book only considers drugs that were eventually launched on the market, and thus only captures the most selective subset of successful projects not representative of the entire pipeline, it is currently the only source of information on patent-project links. Below, we present basic descriptive statistics to put the results of the matching procedure into a perspective.

Tab. C.4: Comparison of patent statistics between the Orange Book and the matched project sample

Matched sample			
Patents		Projects	
Mean - projects per patent	1.8	Mean - patents per project	4.32
Max - projects per patent	25	Max - patents per project	113
Share of patents with single assigned project	67.3	Share of projects with single assigned patent	38.14
		Share of projects with no assigned patent	9.64
Orange Book			
Patents		Projects	
Mean - projects per patent	1.95	Mean - patents per project	10.91
Max - projects per patent	12	Max - patents per project	46
Share of patents with single assigned project	66.16	Share of projects with single assigned patent	9.76
		Share of projects with no assigned patent	0

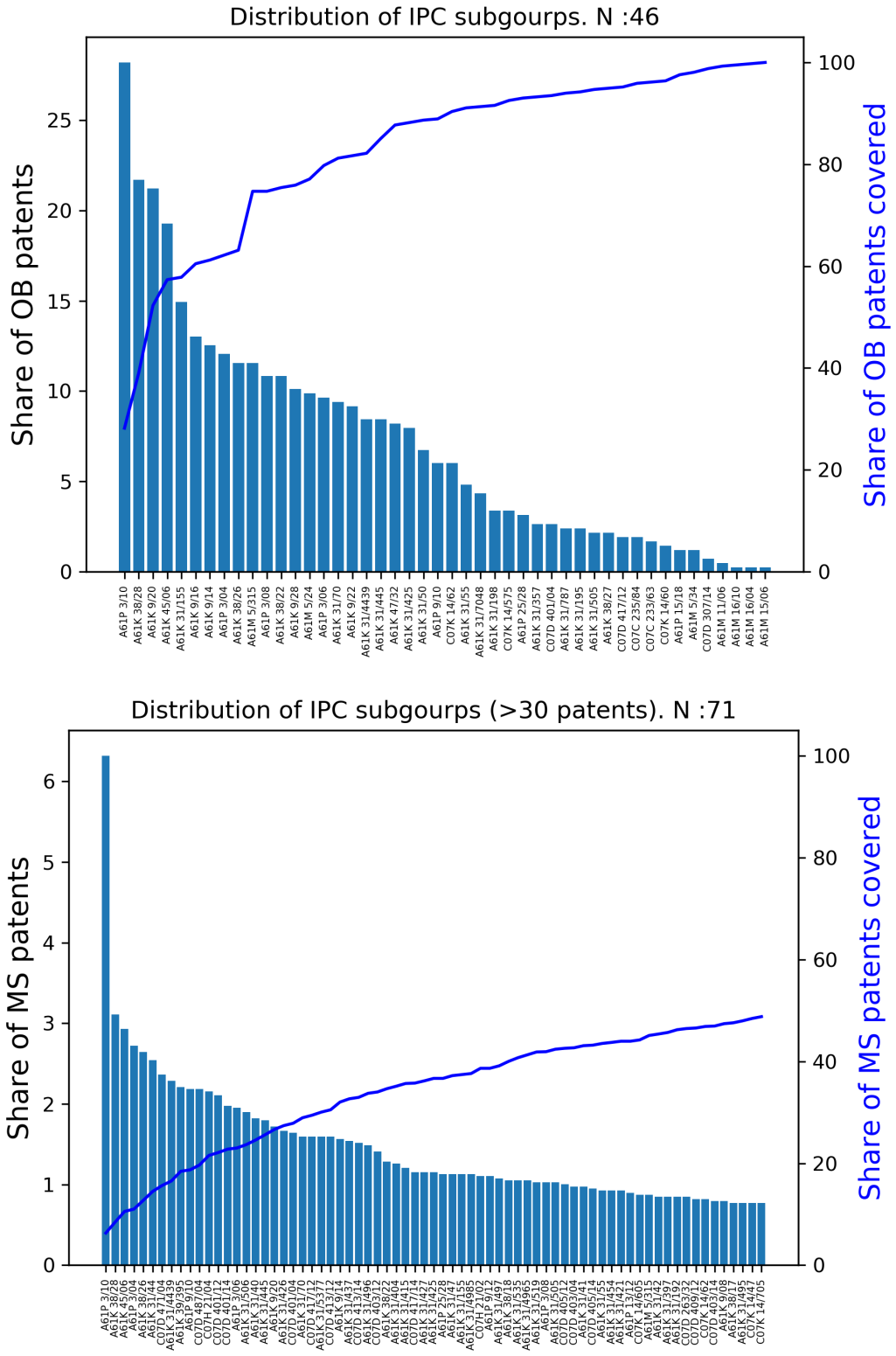
Since a project can have more than one patent assigned, we are presenting a project level and a patent perspective in Table C.4. Overall, the presented figures lend credibility to the outcome of the patents matching. Form the patent level perspective, the average number of projects assigned per patent amounts to 2 across both samples and in both cases with around 67% of patents assigned to a single

project. From the project level perspective, a project has on average 4.32 patents in the matched sample and almost 11 patent in the Orange book. Considering that the matched sample includes much broader set project in earlier development phases, a much higher fraction of the projects have only a single patent assigned compared to the Orange Book (38% in the matches sample vs. 10% in the Orange Book). This result should be expected since it indicates that successful projects that are launched to the market have substantial patent protection, with none of the launched drugs being unpatented. On average, in line with the findings of Argente et al. (2019), average number of patents per project in the matched sample is lower than in the Orange Book sample.

The Figure C.3 provides a comparison between the types of patents that were assigned to the projects (lower part) and how these compare to the Orange book (upper part). Using IPC classification subgroups level (the most granular classification available in the IPC classification), we plot the distributions of the occurrence of the IPC classes in the two sets. We should expect that whilst the matched sample will include many more IPC categories compared to the Orange Book due to the dispersion of the pipeline R&D activities, as least the top classes in both sets should be similar if properly assigned.

The distribution of the matched sample is indeed much broader compared to the Orange Book, and each IPC subgroup occurs less frequently. For example, the most frequent subgroup A61P 3/10 occurs in more than 25% of patents in the Orange Book and in slightly more than 6% of cases in the matched sample. However, comparing the ordering of the different IPC subgroups, we can find that among the most 5 frequent groups in each sample, 3 subgroups are shared and have high relevance to diabetes drugs: A61P 3/10 - drugs for hyperglycaemia, e.g. antidiabetics, A61K 38/28 - insulins, and A61K 38/28 - Mixtures of active ingredients without chemical characterization, e.g. antiphlogistics and cardiaca. In addition, third most frequent subgroups in the matched sample - A61K 38/26 - refers to Glucagons. This shows that the matched patents indeed relate closely to diabetes and at least the most frequent technological subgroups closely mirror the sample of drugs launched on the market.

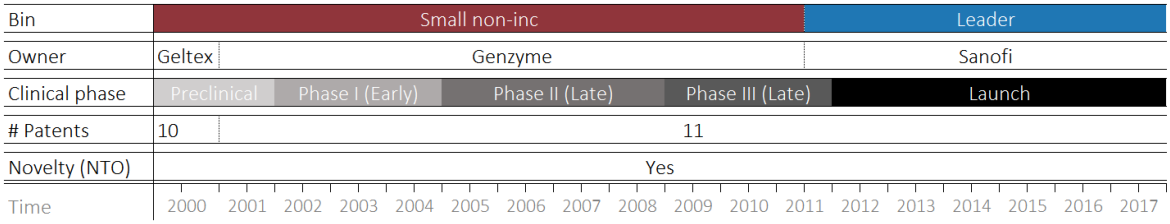
Fig. C.3: Distribution of IPC subgroups in Orange Book and matched samples



D Illustrative example

Figure D.4 uses the drug project colesevelam hydrochloride (brand name: Welchol) as an example to illustrate how all dimensions connect together. Colesevelam hydrochloride is an HMG-CoA reductase inhibitor that has a dual effect of (i) lowering cholesterol and (ii) reducing blood glucose levels (White, 2014). The development of colesevelam hydrochloride was initiated by Geltex, a small non-incumbent, in the year 2000. Based on the 10 US patent families protecting the compound at the beginning of its development (before entering the clinical trials), the project was identified as novel, as at least one of its patent got the NTO label. While still in the preclinical stage, the project was acquired by Genzyme, another small non-incumbent. Shortly before launch, i.e., in the "late phase," it was acquired by Sanofi, labeled a "leader" in our setting. Whenever a project changes hands, we code it as an event that takes on the value of 1.

Fig. D.4: Project example and relevant variables



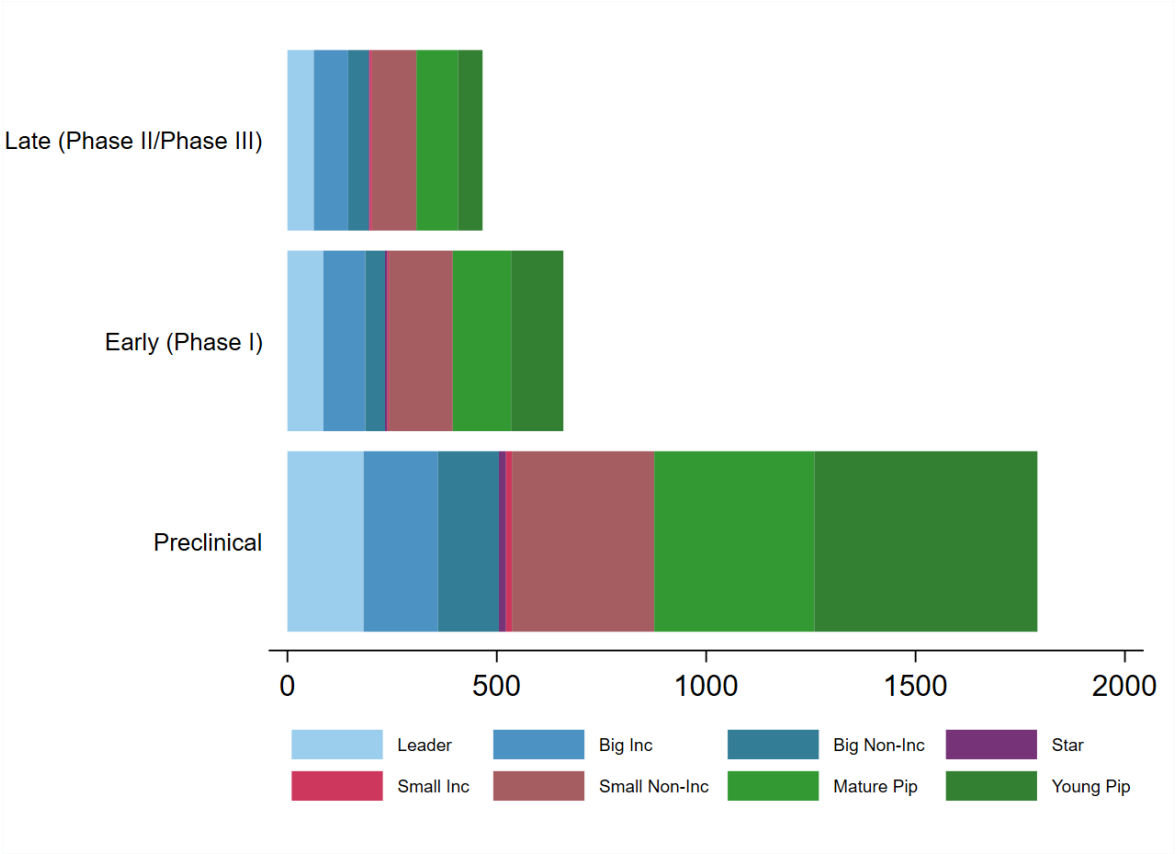
As illustrated in Table D.5, the drug Welchol adds three observations to the database. The first observation is made before the acquisition by Genzyme in the preclinical phase (2000h1). The second is made at the beginning of the early phase (Phase I) in 2002h1. The third and last is made before the acquisition of the drug by Sanofi (2010h2).

Tab. D.5: Database extract

Drug ID	Phase	Date	type	Event	Acquiror	A bin	Target	T bin	NTO	Patents
4827	Preclin	1/07/2000	A	1	Genzyme	Small non-inc	Geltex	Small non-inc	1	10
4827	Early	1/01/2002		0	Genzyme	Small non-inc	Genzyme	Small non-inc	1	11
4827	Late	1/01/2011	A	1	Sanofi	Leader	Genzyme	Small non-inc	1	11

E Funnel structure of R&D

Fig. E.5: R&D Funnel



Note: Each bar shows the number of projects developed in each phase by type of firm.

KU LEUVEN
Faculty of Economics and Business
Management, Strategy and Innovation (MSI)
Naamsestraat 69 bus 3535
3000 LEUVEN, Belgium
tel. + 32 16 32 67 00
msi@econ.kuleuven.be
https://feb.kuleuven.be/research/MSI/