The Knowledge-Incentive Trade-off: Understanding the Relationship between Organization Design and Innovation

APPENDICES

INTRODUCTION

This appendix document contains additional methodological and results details beyond those presented in the main paper. The additional methodological details pertain to how the key analysis variables are developed, additional information on the qualitative analysis included in the paper and the full sample of firms included in the study and those interviewed. The additional results details provide more information on the change in sample structural variables over time, qualitative interview outcomes, matching approaches used, robustness checks and analysis of knowledge and incentives mechanisms.

ADDITIONAL METHODOLOGICAL DETAILS

Dependent variables: invention (H1/2) - determining patent assignees

Two separate approaches are used to define the assignees of patents pertaining to firms in the study sample. First, the limited number of firms in the sample enables the manual matching of patent assignees (as defined by DOC_STD_NAME in the Patstat parent database) to sample firms. Using the Bureau van Dijk "Orbis" database, a list of subsidiaries for each sample firm is developed. Any Patstat assignee that contains a focal firm's subsidiary or parent name text string (and multiple variants of this text string) is captured. This subset of patent assignments per focal firm is then manually checked for each of the 49 firms in the sample to arrive at an intermediate set of Patstat assignees. As the Orbis database provides a snapshot of ownership at a specific point in time (2015), assignees that were subsidiaries of parent companies had to be checked to ensure whether they should be allocated to the parent company or whether when the patent was filed, the subsidiary was an independent company. Using the Zephyr database from Bureau Van Dijk, merger and acquisition (M&A) activity in the industry is controlled for by ensuring assignees represent the original corporate entity filing a patent rather than the parent owner in 2015 provided by the Orbis database. As a

result, prior to the specific M&A event, patents are retrospectively assigned to the acquired firm from the acquiring firm.

Second, following the process of Arora, Belenzon, and Rios (2014) patent assignees were matched against firm and subsidiary names obtained from Bureau Van Djik's "Icarus" database following cleaning of names using a standardized name-cleaning algorithm. This was an iterative process involving the adjustment of matching rules and manual checking. Again, using a similar process to that described above, the Zephyr database was used to control for M&A activity and retrospectively reassign patents to acquired firms from the acquiring firm prior to the M&A event.

Both approaches used to develop standardized names provided similar results with 99.7 % of assignees being the same for each sample patent. Those patents that did not have the same assignees from both methods were manually checked and reassigned appropriately.

Dependent variables: development (H3/4) - allocating drug candidates to parent firms

To ensure that a drug candidate in the clinical development process is allocated to the appropriate firm using Pharmaprojects data, two key steps are taken. First, transactions are examined using the Recap database to ensure that the firm assigned to a drug in the Pharmaprojects database is the actual firm managing the development of that drug candidate. These transactions include deals in which a selection of drug-candidates are sold from one company to another, a complete firm is acquired or merges with another and strategic alliances between firms in which an invention may be created through an alliance and then subsequently pursued through clinical trials by another firm. If a transaction is observed in Recap, the firm managing the development of that drug is adjusted accordingly.

Second, prior to 2012 Pharmaprojects retrospectively assigns a drug candidate to an acquiring firm following acquisition of another firm. As a result even prior to the acquisition year that drug candidate will be assigned to the acquiring firm rather than the acquired firm which was at that time an independent entity. Adjustment of these assignments requires a careful assessment of the "Overview" section of the Pharmaprojects record of a drug which indicates which firm was initially responsible for a drug-candidate

prior to the respective deal. For M&A activity post 2012, Pharmaprojects correctly allocates the firm responsible for the original development of a drug candidate. For M&A activity post 2012, drug-candidates were reassigned to the acquiring firm the year after the acquisition. Merger and acquisition data from Recap and the Zephyr database from Bureau Van Dijk were used to reassign drug candidates following post-2012 M&A activity.

In the absence of any transaction in Recap or additional information on a drug-candidate provided in Pharmaprojects "Overview" section the original firm assignment in the Pharmaprojects database is utilized. Further, it is noted whether a drug-candidate was developed internally, acquired via an M&A deal, acquired from another firm or was originally created through an alliance.

Independent variables: structural measures

Obtaining data on a commercial firms' internal organizational structures is highlighted as a significant challenge for the management scholar (e.g., Greenwood & Miller, 2010; Sathe, 1978; Walton, 1981). A review of the management literature highlights three methods by which internal organizational structures are inferred.

First, scholars use publicly available firm administrative records such as high level organizational charts and company annual reports to directly infer organizational structures (e.g., Albert, 2018; Girod & Whittington, 2015; Guadalupe, Li, & Wulf, 2014). Although this data is readily accessible it is limited in its coverage and can result in an incomplete picture of a firm's organizational structure being derived. For example, public administrative record information is often limited to the senior most levels of the organization and focuses on direct reports to the CEO – the executive team. As a result, it is challenging to accurately infer organizational structure lower down the organization.

Second, indirect proxies are used to determine structure using publicly available information. For example, Arora et al. (2014) use patent assignee data to define the level of centralization of a firm's R&D function based on whether patents are assigned to the parent company or a subsidiary. Other studies also focusing on the R&D function examine the number of employees in corporate and divisional laboratories

(Argyres & Silverman, 2004). Using the ratio of employees in both types of laboratory a degree of centralization percentage can be estimated. This approach requires careful consideration of construct validity as the indirect measure may not correlate perfectly with organizational structure.

Finally, the most common tool used in organizational structure research is survey analysis in which firms are questioned directly about their organizational structure in a standardized manner (e.g., Hill, Hitt, & Hoskisson, 1992; Markides & Williamson, 1996; Turner & Makhija, 2012). This enables scholars to tailor questions to better capture the information that they need, and helps them to observe organizational structure at a greater level of depth. Survey studies generally use multiple questions to measure a variety of specific organizational constructs. For example, Russell and Russell (1992) use surveys to measure structural components such as the degree of centralization, integration and breadth of control. Turner and Makhija (2012) measure whether firms are organic (more decentralized and less bureaucratic) or mechanistic (more centralized and process focused). The survey approach is limited by the usual factors associated with any form of survey research e.g. accessing the right survey respondents, and the extended period of time required to conduct survey.

In this study a combination of the first and third methods are used to develop three organizational structural measures: *R&D Decentralization, R&D Functional Differentiation* and *Corporate Decentralization*. First, company administrative records such as annual reports can be used to identify the executive level of management of each pharmaceutical firm. Each executive level management team role corresponds to a structural element (e.g. R&D, manufacturing) and these can be coded systematically to enable an estimate of the structural parameters described above. Second, survey-type interviews are conducted with sample firms to validate and expand upon the measures captured from archival sources.

Qualitative analysis: methodological description

To enrich the archival data analysis multiple managers within 28 firms from the sample of 49 firms (see Table 1 for further details of firms interviewed) and five industry experts were interviewed. In total 61 interviews were conducted. The managers interviewed were senior level R&D and strategy managers who had

a good understanding of the structure of both R&D and their organization as a whole. The interviews were conducted between 2015 and 2018. The interviews were conducted via teleconference and each interview typically lasted between 30 and 90 minutes with outline questions distributed to the respondent in advance to enable suitable preparation and follow-up clarification questions being conducted post-interview through email. Detailed notes were collected during each interview. Notes from all 61 interviews were reviewed to determine key issues pertaining to three areas.

First, interviews were used to validate the relevant decentralization measures that were developed through coding of firms' top management team structures using publicly available data sources (*R&D Decentralization, Corporate Decentralization* and *R&D Functional Differentiation*). Second, the mechanisms through which managers perceive firms' organizational design choices impact their innovation outcomes were also examined. The focus of these interview questions related to the incentives- and knowledge-based mechanisms through which decentralization could impact innovation. Third, a sequence of questions was asked relating to the product development decision-making process in these pharmaceutical companies and which parts of the organization are involved at different stages of the drug development process.

I supplement the data collected from these interviews through review of the more qualitative aspects of firms' annual reports and 10-K filings. A research assistant (RA) reviewed the letter to shareholders, business description, operational review and R&D overview of firms' annual financial filings for each year in the period 1995-2015. The text extraction was focused on descriptions of organizational design and references to incentives and knowledge-flows. First, any evidence pertaining to the R&D or the overall organizational structure of each firm was captured. This data was again used to validate the measures developed for each form of decentralization using the approach described in the main paper's methods section. Second, any evidence for managerial discussion pertaining to how organizational design choices could impact incentives and knowledge flows was also captured.

In the second phase of work these data extracts were further examined and common, major themes that are used to inform the qualitative commentary below were captured. These insights were complemented with relevant findings from the interviews with strategy and R&D managers. It should be emphasized that

this analysis is not intended to be a rigorous case-based form of qualitative analysis (e.g., Eisenhardt, 1989). It is simply designed to add greater insight into and confidence in the main quantitative archival analysis.

----------Insert Table 1 about here ----------

RESULTS

Descriptive statistics

Figures 1-3 illustrate the sample mean variation across firms of each of the three key structural variables in this study over the period 1995-2015. From Figure 1, it can be seen that *R&D Decentralization* increases from 1998 to 2001 as more firms decentralized their R&D units. Then it drops in 2002 and remains relatively flat to 2008. R&D decentralization increases from 2009 to 2012 and then remains flat. Figure 2 illustrates that R&D Functional Differentiation fluctuates over time peaking in 2004 and then dropping to a relatively steady value between 2005 and 2015. In contrast Corporate Decentralization has been relatively steady over the study period (Figure 3). These results illustrate that design choices can be cyclical influenced by events such as merger and acquisition activity or, potentially, firms attempting to replicate the structures of other firms. This viewpoint was referred to multiple times in managerial interviews. For example:

> *"Organizational design changes seem to go in waves across the industry, at one stage centralization is in, then it is all about being decentralized and nimble"*[1]

Table 2 illustrates the key descriptive statistics for the 28 firms interviewed. Although questions were asked about how the firms' structure changed over time, the data in this table pertains to their structures in the final year of the sample period (2015). It can be seen that, consistent with the overall sample, approximately 11 % of firms had decentralized R&D structures. Interestingly, the key way in which R&D was sub-divided was by functional area (68 % of firms interviewed had some form of functional sub-division in R&D). In a centralized R&D structure this will facilitate knowledge flows across therapeutic areas potentially facilitating invention and development outcomes. For those firms which had business units and were not functionally aligned, these business units were primarily organized along therapeutic area lines as opposed to

[1] Due to confidentiality associated with the study interviews, I cannot ascribe the comments to any specific firm or individual

geographies (83 % versus 17 %). Interestingly, these results highlight that R&D tends to be sub-divided by function (e.g. science area or stage of R&D) whereas the more commercial aspects tend to be more therapeutic area aligned.

----------Insert Table 2 and Figures 1 – 3 about here ----------

Key additional qualitative insights pertaining to knowledge and incentive-mechanisms

The majority of managers interviewed outlined in some form or other the importance of ensuring good cross-organizational knowledge flows to aid effective innovation (64 % as illustrated in Table 2). Greater organizational integration such as the creation of a more centralized R&D unit was one way of achieving this, but managers described other routes this could be achieved such as cross-organizational research forums and the use of various online knowledge management tools. Ensuring good knowledge flows was seen as especially important for ensuring the development of novel inventions and for facilitating their development into final products.

> *"Organizing to ensure greater integration across therapeutic areas is important as an idea in one area may be able to be translated into another therapeutic area. Quite often an indication may be unsuccessful in one therapeutic domain but have legs in another, however with the wrong structure scientists may not be able to take advantage of this"*
>
> *"It is important to get the viewpoint of multiple functions during clinical development and even earlier in the discovery phase"*

Managers frequently referred to the creation of organizational siloes with more decentralized structures that can result in poor knowledge flows and potential repetition of effort.

> *"Avoiding silos is an issue – we need to force people to collaborate with each other. Ultimately some technology will be replicated across the organization and this is ok if the cost of transporting a molecule is prohibitive, but the firm could improve in not replicating activities across labs in our more decentralized R&D organization"*
>
> *"It can always be difficult to get different teams collaborating as people fixate on the specific unit of the organization in which they are located"*

These poorer knowledge flows between business units could ultimately lead to inferior innovation outcomes.

> *"Drugs make great business units but business units do not make great drugs"*

Less attention was paid to incentives in firms' annual reports but interviews with R&D managers highlight that incentives could influence innovation outcomes and are related to a firm's organizational design attributes. 43 % of managers interviewed mentioned the importance of incentives and how these could shape R&D behavior (Table 2). The key theme that emerged was that R&D managers tended to be incentivized by the volume of inventions and ensuring that they progress through the innovation process rather than by the quality of the inventions being progressed. Greater integration was seen as being associated with lower powered incentives which some managers perceived could hinder innovation performance:

> *"The issue with incentives in a corporate (more centralized) setting is they are generally quite poor and under-reward good performance and over-reward poor performance i.e. people don't get fired"*

However, managers did highlight using higher powered incentives are not a panacea and could come at a cost:

> *"Ultimately there is a trade-off of getting ambitious performance and ensuring a good work environment and collaborative atmosphere"*

> *"Incentivizing people by counting compounds is not a way of incentivizing good science"*

Finally, many managers highlighted the organizational challenge firms' face in deciding the degree of organizational decentralization:

> *"You need to put in swing lanes to provide some discipline, the problem is that you make the swing lanes too narrow and people focus too narrowly and can be restricted in what they can do and may not collaborate effectively with individuals in other swing lanes"*

> *"Balance between being smaller more decentralized units and being agile like a biotech and being able to leverage scale of a larger organization"*

In summary, it appears that organizational design attributes can impact innovation outcomes through both knowledge flows and provision of incentives. Managers in pharmaceutical firms do discuss both mechanisms and how design choices can emphasize one over the other and some even highlight the trade-off firms' face when deciding to integrate more tightly or decentralize. However, no real mention was made as to the boundary conditions in which greater decentralization may be more appropriate.

Main analysis: propensity score matching - first stage regression & balance checks

Propensity score matching (PSM) models are used to generate matched samples of more decentralized and less decentralized firms across the three structural measures. In the first step, a logit regression is used to predict the likelihood that a firm will have the relevant differentiation or decentralization dimension based on a set of observable variables. Second, a standard regression of the pertinent innovation outcome against the appropriate structural variables using controls and fixed effects is undertaken for the matched sample identified using the first-stage logit regression. Matching is undertaken either using nearest neighbor (i.e. matching untreated and treated observations that have closest propensity scores) or caliper (i.e. setting a maximum propensity score difference between observations that are treated and untreated) methods (Caliendo & Kopeinig, 2008). Similar results are obtained using either approach. The focus of this analysis is to limit the possibility that firms' innovation outcomes result from inherent differences between firms which are more or less decentralized.

For H1 and H2, the relationship between *R&D Decentralization* and two different invention outcomes (originality of inventions, number of inventions) are examined. The first stage logit regression is highlighted in Table 3 and Table 4 presents the balance test across all the covariates in the first-stage regression.

For H3, the relationship between *R&D Decentralization* and progression of inventions through the early development process is investigated and the match is based on the dichotomous variable *R&D Decentralization*. For H4, the variable *corporate decentralization* is related to the progression of inventions through the later stages of the development process. As *corporate decentralization* is a continuous variable, this variable is dichotomized around the median and matching is undertaken using this variable. Several cut-points between 0.2 and 0.6 were used to dichotomize *corporate decentralization,* similar results were obtained for each cut-point. In the analysis of mechanisms, the variable *R&D functional differentiation* is related to the progression of inventions through the earlier stages of the development process and the match is based around this dichotomous variable. Table 5 illustrates the first stage logit regression results and Table 6 highlights the accompanying balance tests of the resulting matched samples which are used for subsequent analyses to test the relevant hypotheses. As can be seen from Table 4 and Table 6, the balance tests indicate that for the

majority of covariates the samples achieve balance. However, for H3, the decentralized sample is moderately smaller and has a smaller patent stock. For H4, the decentralized sample faces a moderately more competitive environment.

----------Insert Tables 3-6 about here ----------

Robustness tests

A series of additional robustness tests are also conducted and are outlined fully in Table 7 (H1/H2) and Table 8 (H3/4).

----------Insert Tables 7 and 8 about here ----------

Supplemental analyses: additional information

Knowledge flow mechanism

The full regression tables for these analyses are illustrated in Table 9 and Table 10. The variable *Tech. Diversity* measures the degree of diversity of firms' development portfolios and is operationalized through the breadth of therapeutic classes of a firm's current development portfolio. This variable is measured using a Herfindahl index (subtracted from 1 to ensure higher values represent more diverse portfolios) and is estimated using a similar approach to other studies in the pharmaceutical industry empirical context (Diestre & Rajagopalan, 2012). The key assumption in this analysis is that a more diverse development portfolio is associated with a broader array of knowledge within a focal firm. This assumption was validated through the managerial interviews that highlighted that the therapeutic classes associated with firms' portfolios provided a reflection of the internal knowledge base within a focal firm. For example:

> *"We dropped oncology from our portfolio and eventually lost our capability in the area meaning it will be difficult to pick up new candidates in this domain in the future"*

Consistent with my theoretical argumentation the interaction term *Tech. Diversity* x *R&D Decentralization* is statistically significant and negative for the pre-clinical to phase 1 transition examined in Hypothesis 3. Thus access to a broader array of technical knowledge enables firms with centralized R&D to solve a greater array of technical problems enabling such firms to progress even more drug candidates from

pre-clinical to phase 1 relative to firms with centralized R&D. Interestingly, for low levels of *Tech. Diversity* firms with decentralized R&D are able to progress more inventions than those with centralized R&D. This suggests that the benefits of rich intra-organizational knowledge flows, which are limited in the case of firms with a narrower array of technical knowledge, are outweighed by the stronger incentives associated with R&D decentralization.

Two alternative measures of the novelty of a firm's portfolio are *NCE* and *Novelty*. *NCE* represents the proportion of drug-candidates within a firm's portfolio that are new chemical entities. New chemical entities represent new drug candidates for which no component has been previously approved by the Federal Drug Administration. *Novelty* represents the mean novelty of firms' portfolios on a 0-2 scale (Klueter, 2013). Drug candidates whose mechanism of action and origin of material are new to the focal firm in a specific therapeutic class have a *Novelty* value of 2, drug-candidates where one of the mechanism of action or origin of material are new have a *Novelty* value of 1 and if neither the mechanism of action nor the origin of material within a specific therapeutic class are new to the focal firm then the *Novelty* value is 0. Drug candidates that represent a new mechanism in a specific therapeutic class for a firm entail greater challenges as scientists need to develop an understanding of both the mechanism and how to apply that mechanism in a drug candidate i.e. suitable pharmo-kinetics profile, appropriate delivery mechanism, understanding how the drug candidate impacts target receptors in the body. This increases the technical complexity in developing a drug candidate. Similar considerations apply if the origin of material is new to a specific therapeutic class e.g. if a firm has never used antibodies in oncology this provides a greater technical challenge. However, the firm may have experience of this mechanism or have used a material in the same class in in a different therapeutic class and be able to access this valuable information through cross-organizational knowledge flows.

R&D centralization should facilitate access to a firm's broader organizational knowledge thereby enabling more novel inventions to progress through the early stages of development. Empirically if this is the case the interaction terms *R&D Decentralization* x *Novelty* and *R&D Decentralization* x *NCE* should be negative and statistically significant. Table 10 illustrates support for this argumentation. Also, consistent with the hypothesis development in the main paper, the interaction term is only significant for early development i.e.

pre-clinical to phase 1 transition. This suggests that the importance of cross-organizational knowledge flows diminishes as an invention progresses through the development process.

----------Insert Tables 9-10 about here ----------

Incentives mechanism

The full regression tables for these analyses are illustrated in Tables 11-15. Table 11 illustrates the regression analysis examining how the mean patent grant-lag for a firm's set of patent families filed in a focal year is associated with *R&D Decentralization*. Consistent with an incentives based argumentation, R&D decentralization is associated with shorter lags between the filing and granting of patents. This is because greater efforts are undertaken by managers in firms with decentralized R&D to get patents granted.

Tables 12 and 13 illustrate a series of tests undertaken to evaluate whether greater overall decentralization and R&D functional differentiation are associated with the progression of inferior drug-candidates through development thereby explaining why more candidates progress for firms with these structures. To test whether this is the case, the likelihood of a drug that has been progressed from Phase 1 to 2 progressing into Phase 3 is examined for firms with and without functionally differentiated R&D (i.e. separate research and separate development units) and the likelihood of a drug candidate that has been progressed from Phase 2 to 3 progressing into Pre-Registration (PR) Status is evaluated for firms that are more or less decentralized at a corporate level. Two econometric approaches are used to examine the likelihood of a drug progressing from one phase to the next. First, using a maximum likelihood approach that accounts for the discrete nature of the time-element of the data set (i.e. clinical trial phase is only available per year), logit analyses are undertaken in which the unit of analysis is the drug-candidate-year and the dependent variable indicates whether the drug candidate moves from one phase to the next (Allison, 1982). A linear time function is used as one of the dependent variables, which is set to 1 when a firm enters the focal phase (phase 2 in the case of *R&D Functional Differentiation* and phase 3 in the case of *Corporate Decentralization*) and increases by 1 for each subsequent year. In the case of *R&D Functional Differentiation*, the focus is only on drugs in phase 2 and the progression focus is movement from phase 2 to 3. For *Corporate Decentralization*, the focus is

on drugs in phase 3 and the progression focus is from phase 3 to PR. Second, Cox proportional hazards model are used to examine the relative hazard of a drug moving from one phase to the next. The advantage of this approach is that that it is unconstrained in its underlying time function assumptions unlike the first approach. Again, the analysis is at the drug candidate-year level and the dependent variable and phase focus is the same as for the logit model used in the first approach.

In both approaches the same full set of controls and independent variables that are used to test Hypotheses 3-4 are utilized as well as including individual drug controls (if drug is NCE and if it is externally sourced), year, category and drug therapeutic class controls. If greater decentralization/ differentiation is associated with the progression of inventions that are less likely to progress through the later stages of the innovation process then the coefficient for *R&D functional Differentiation12* (i.e. R&D functional differentiation of the firm when a drug candidate moves from phase 1 to 2) and *Corporate Decentralization23* (i.e. the degree of corporate decentralization when a drug candidate moves from phase 2 to 3) should both be negative and statistically significant. Using this analysis no evidence is observed to suggest that functional differentiation of R&D and increased corporate decentralization are associated with the progression of inferior inventions that fail to progress through the later stages of the development process as the coefficients for *R&D functional Differentiation12* and *Corporate Decentralization23* are not statistically significant (see Table 12 and Table 13).

In Table 14 the average time taken for a drug candidate to move from Phase 2 to 3 across all firms in the sample is the focal dependent variable, with the unit of analysis being the firm. The negative coefficients for *Corporate Decentralization* indicate that greater corporate decentralization is associated with shorter times for progression from phase 2 to 3. This is consistent with an incentives-based argumentation as managers exert more effort to progress inventions when the firm is more decentralized.

In Table 15 the average compensation of R&D executives and executives overall (as defined as reporting to the CEO) is examined using Execucomp data. Consistent with the main incentives-based argumentation, I observe that firms with functionally differentiated R&D have higher total compensation on average (Table 15 Model 2), after controlling for a variety of firm-specific factors. This may partly explain why more drug-candidates progress from Phase 1 to 2 and consistent with the qualitative interview-based

evidence outlined earlier in this appendix.

I also observe that greater corporate decentralization is associated with lower salaries but unrelated to total compensation (Table 15 Models 3 and 4), after controlling for a variety of firm-specific factors. This implies that greater corporate decentralization is associated with a higher variable component of compensation. This is again consistent with the argumentation that decentralization is associated with the usage of higher-powered incentives. This analysis is limited in that I do not have access to compensation data for my complete sample of firm-years (e.g. lack of access of compensation data of Japanese-listed firms). However, this analysis is consistent with greater decentralization being associated with the use of higher powered incentives.

----------Insert Tables 11-15 about here ----------

APPENDIX FIGURES AND TABLES

Figure 1: Sample variation of R&D Decentralization over time. Each point is mean across firms in sample in that year

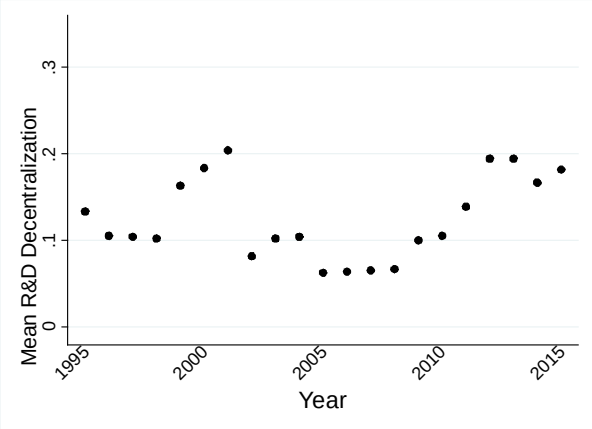

Figure 2: Sample variation of R&D Functional Differentiation over time. Each point is mean across firms in sample in that year

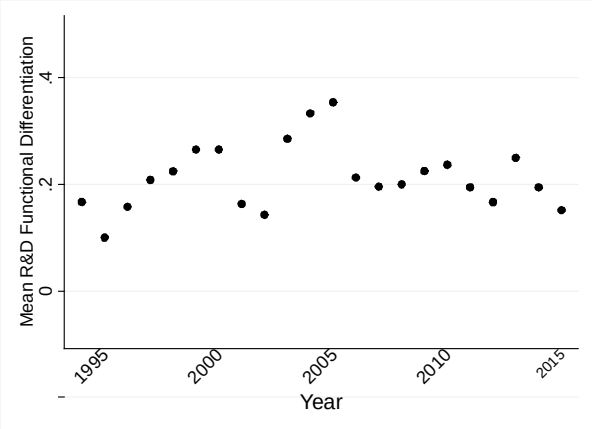

Figure 3: Sample variation of Corporate Decentralization over time. Each point is mean across firms in sample in that year

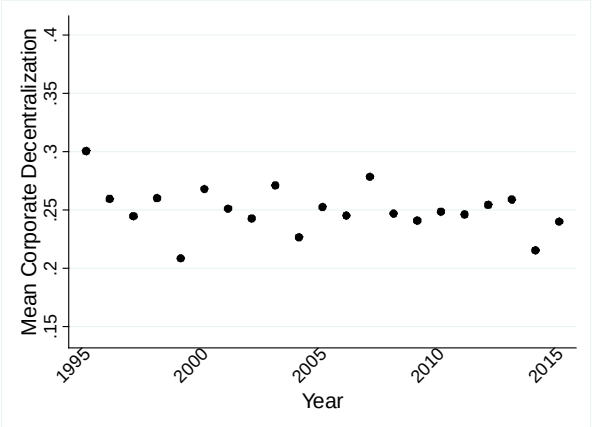

Table 1 Study sample firms and firms interviewed in this study

#	Firm	GVKEY	Interviewed
1	Abbott Laboratories	001078	Yes
2	Actavis	027845	Yes
3	Akzo Nobel	015334	
4	Allergan	015708	Yes
5	Altana	100004	
6	Amgen	001602	Yes
7	Ares-Serono	102045	
8	Astra Zeneca	028272	Yes
9	Aventis	013467	
10	Baxter International	002086	Yes
11	Bayer	100080	Yes
12	Biogen Idec	024468	Yes
13	Bristol-Myers Squibb	002403	Yes
14	Cephalon	023945	
15	Chugai Pharma.	100441	
16	CSL	223003	Yes
17	Daiichi Sankyo	100336	Yes
18	Eisai	100418	Yes
19	Eli Lilly	006730	Yes
20	Forest Labs	004843	
21	Genentech	005020	
22	Genzyme	012233	
23	Gilead Sciences	024856	Yes
24	GlaxoSmithKline	005180	Yes
25	Johnson & Johnson	006266	Yes
26	King	112033	
27	Kyowa Hakko Kirin	100516	
28	Lundbeck	232106	
29	Medlmmune	024008	
30	Merck & Co	007257	Yes
31	Merck KGaA	220301	
32	Mylan	007637	Yes
33	Novartis	101310	Yes
34	Novo Nordisk	008020	Yes
35	Pfizer	008530	Yes
36	Roche	025648	Yes
37	Sanofi	101204	Yes
38	Schering AG	101076	
39	Schering-Plough	009459	
40	Schwarz Pharma	108182	
41	Shire	212340	Yes
42	Solvay	101394	
43	STADA Arz.	214700	
44	Takeda	100718	Yes
45	Tanabe	100021	
46	Teva	014538	Yes
47	UCB	100751	Yes
48	Valeant Pharma. Int.	009340	Yes
49	Wyeth	001478	

Table 2: Key descriptive statistics for sample firms interviewed (n=28 firms and 61 interviews)

Interview Item	N	%
Decentralized R&D		
Centralized R&D	25	89
Decentralized R&D	3	11
Total	*28*	*100*
R&D Sub-division		
Functional	14	50
Mixed	5	18
Therapeutic	9	32
Total	*28*	*100*
Corporate Decentralization		
Divisional	18	64
Functional	10	36
Total	*28*	*100*
Business unit categories		
Therapeutic Area	15	83
Geography	3	17
Total	*18*	*100*
Respondents mentioning specific mechanism (unprompted)		
Knowledge Flows	18	64
Incentives	12	43

Table 3: H1-2 propensity score matching analyses. First stage logit regression

Dependent Variable	R&D Decentralization
R&D Functional Differentiation	-1.530**
	(0.410)
Corporate Decentralization	0.523
	(0.487)
performance	-1.435
	(1.711)
R&D Intensity	2.051**
	(0.773)
sga	-0.493+
	(0.287)
size	0.752**
	(0.291)
slack	0.0968
	(0.0847)
CEO	-0.307
	(0.378)
SBU	-0.164
	(0.108)
tech. diversity	-1.238
	(1.119)
patent stock	0.0710
	(0.133)
competition	-0.900
	(4.730)
Year grouping	Y
N	803
Pseudo-R^2	0.0652
Log Likelihood	-280.3

Standard errors in parentheses: $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$.
Standard Errors clustered at firm level

Table 4: H1-2 balance tests for propensity score matching model (Caliper=0.00035)

	Mean		p>|t|
	R&D Decentralization =1	R&D Decentralization =0	
R&D Functional Differentiation	0.093	0.093	1.000
Corporate Decentralization	0.296	0.309	0.735
performance	0.074	0.093	0.115
R&D Intensity	0.156	0.144	0.454
sga	7.865	7.990	0.578
size	8.782	8.904	0.575
slack	2.472	2.330	0.608
CEO	0.107	0.107	1.000
SBU	2.467	2.467	1.000
tech. diversity	0.771	0.763	0.633
patent stock	1.085	1.171	0.689
competition	0.961	0.957	0.455
Year grouping variable	2.693	2.560	0.448

Table 5: H3-4 Propensity score matching analyses. First stage logit regression

Hypothesis	**3**	**4**	**Mechanism**
Dependent Variable	**R&D Decentralization**	**Corporate Decentralization (dichotomized)**	**R&D Functional Differentiation**
Table/Model	**Table 6 Model 3**	**Table 7 Model 4/5**	**Table 6 Model 6**
Phase	**Phase 0 to 1**	**Phase 2 to 3**	**Phase 1 to 2**
R&D Decentralization		0.485^{+}	-1.215**
		(0.253)	(0.399)
R&D fnl. differentiation	-1.195**	-0.350^{+}	
	(0.391)	(0.200)	
Corporate Decentralizn.	0.574		-0.872*
	(0.499)		(0.437)
performance	-0.540	-1.109	0.705
	(1.711)	(1.299)	(1.376)
R&D Intensity	2.224**	-3.192**	1.377*
	(0.803)	(1.132)	(0.662)
sga	-0.394	0.191	-0.339
	(0.325)	(0.213)	(0.261)
size	0.695*	-0.378^{+}	0.257
	(0.323)	(0.215)	(0.257)
slack	0.0467	-0.0865	0.240**
	(0.0886)	(0.0682)	(0.0694)
CEO	-0.426	-0.238	-0.141
	(0.399)	(0.244)	(0.307)
SBU	-0.129	0.366**	-0.130
	(0.111)	(0.0752)	(0.0884)
patent stock	-0.0318	-0.0974	0.327**
	(0.135)	(0.111)	(0.124)
portfolio	0.00435	0.0112	-0.0527**
	(0.00642)	(0.0108)	(0.0158)
external	0.852	-0.375	-0.410
	(0.564)	(0.353)	(0.375)
NCE	-0.301	0.774^{+}	2.156**
	(0.674)	(0.443)	(0.562)
bio	-1.340^{+}	0.962^{+}	3.067**
	(0.766)	(0.520)	(0.652)
tech. diversity	0.0846	1.749**	2.006**
	(0.705)	(0.477)	(0.548)
competition	1.768	-6.932^{+}	0.298
	(5.755)	(4.205)	(5.700)
Year grouping variable	Y	Y	Y
Firm Fixed Effects	N	N	N
Category Fixed Effects	Y	Y	Y
N	787	762	764
Pseudo-R^2	0.071	0.119	0.137
Log Likelihood	-273.1	-464.7	-357.8

Standard errors in parentheses:$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$
Standard errors clustered at firm level.

Table 6: H3-4 balance tests for propensity score matching model

Hypothesis	H3: Means Post Balance			H4: Means Post Balance			Mechanism: Means Post Balance		
PSM Model	Table 6 Model 3 Nearest Neighbor (3)			Table 7 Models 4/5 Caliper (0.0002)			Table 6 Model 6 Caliper (0.008)		
Treatment Variable	R&D Decent.= 1	R&D Decent.= 0	p>|t|	Corporate Decent. =0	Corporate Decent.= 1	p>|t|	R&D FD=1	R&D FD=0	p>|t|
R&D Decentralization				0.113	0.081	0.547	0.052	0.065	0.733
R&D functional differentiation	0.085	0.099	0.738	0.242	0.210	0.671			
Corporate Decentralization	0.303	0.312	0.802				0.204	0.206	0.949
performance	0.089	0.073	0.224	0.092	0.088	0.783	0.081	0.077	0.612
size	9.177	8.746	0.042	9.228	9.175	0.805	8.756	8.544	0.153
sga	8.256	7.887	0.068	8.402	8.298	0.645	7.940	7.723	0.121
slack	2.223	2.538	0.188	2.098	1.907	0.260	2.532	2.822	0.105
R&D Intensity	0.151	0.199	0.134	0.164	0.140	0.189	0.183	0.195	0.622
patent stock	1.450	1.103	0.068	1.440	1.355	0.720	1.264	1.135	0.364
CEO	0.085	0.099	0.738	0.097	0.145	0.413	0.103	0.071	0.315
competition	0.952	0.957	0.230	0.950	0.959	0.053	0.960	0.965	0.104
SBU	2.404	2.323	0.629	2.403	2.645	0.225	2.368	2.348	0.894
portfolio	36.89	30.74	0.202	20.27	18.82	0.624	11.31	9.284	0.057
external	0.504	0.516	0.745	0.514	0.560	0.200	0.479	0.476	0.912
NCE	0.554	0.534	0.583	0.554	0.528	0.531	0.542	0.563	0.545
bio	0.215	0.201	0.640	0.244	0.268	0.485	0.305	0.247	0.060
tech diversity	0.726	0.678	0.186	0.762	0.724	0.159	0.618	0.606	0.648
Year grouping variable	2.585	2.447	0.412	2.565	2.452	0.545	2.477	2.452	0.825

Table 7: Summary of additional robustness tests – Hypotheses 1 and 2

	1. Coarsened Exact Matching (CEM)		2. Alternate Specification		5a. Lagged IV		5b. Rolling Average IV	
DV	**Originality**	**Quantity**	**Originality (OLS)**	**Log (Quantity)**	**Originality**	**Quantity**	**Originality**	**Quantity**
R&D Decentralization	**-0.190* (0.0753)**	**0.279** (0.0686)**	**-0.0372** (0.0132)**	**0.246+ (0.147)**	**-0.162* (0.0716)**	**0.186 (0.136)**	**-0.161* (0.0705)**	**0.224 (0.158)**
R&D Functional Differentiation	-0.0450 (0.0835)	0.132+ (0.0742)	0.00868 (0.0141)	0.164 (0.103)	0.0374 (0.0605)	0.0934 (0.0971)	0.0121 (0.0603)	0.125 (0.0975)
Corporate Decentralization	-0.118 (0.165)	0.0863 (0.133)	0.0271 (0.0270)	-0.0137 (0.189)	0.126 (0.121)	-0.0457 (0.153)	0.107 (0.123)	-0.0363 (0.180)
performance	-1.614* (0.743)	-0.737 (0.571)	-0.0895 (0.0769)	0.256 (0.668)	-0.460 (0.352)	0.268 (0.638)	-0.423 (0.346)	0.197 (0.620)
R&D Intensity	-0.138 (0.139)	0.0796 (0.161)	0.00248 (0.0293)	0.346 (0.288)	-0.0154 (0.185)	0.199 (0.309)	0.0169 (0.138)	0.175 (0.270)
size	0.0836 (0.0588)	0.348** (0.0572)	0.000748 (0.0102)	0.346** (0.0649)	0.000568 (0.0478)	0.287** (0.0634)	0.00144 (0.0452)	0.313** (0.0610)
slack	0.0657* (0.0316)	-0.0477* (0.0222)	0.00144 (0.00452)	-0.0118 (0.0332)	0.00639 (0.0224)	-0.00562 (0.0333)	0.00767 (0.0196)	-0.00391 (0.0299)
CEO	-0.120 (0.0863)	-0.0292 (0.0886)	-0.0149 (0.0125)	0.0162 (0.0708)	-0.0739 (0.0563)	0.0346 (0.0600)	-0.0744 (0.0543)	0.0306 (0.0595)
tech. diversity	-0.0876 (0.816)	1.571* (0.615)	0.202* (0.0769)	2.893** (0.561)	0.946** (0.351)	2.051** (0.646)	0.878** (0.341)	1.842** (0.646)
patent stock	-0.00867 (0.0562)	0.499** (0.0503)	-0.00230 (0.00908)	0.481** (0.0675)	-0.0131 (0.0418)	0.478** (0.0673)	-0.0116 (0.0398)	0.461** (0.0664)
competition	-1.140 (1.638)	-3.627+ (1.945)	-0.534* (0.251)	-3.210 (2.026)	-2.447+ (1.249)	-2.988+ (1.766)	-2.484* (1.162)	-3.238+ (1.696)
Year Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y
Firm Fixed Effects	N	N	N	N	N	N	N	N
Category Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y
N	329	329	803	803	773	773	803	803
R^2	0.0639	0.161	0.541	0.620	0.061	0.134	0.0631	0.132
Log Likelihood	-208.2	-1815.9	487.8	-758.6	-496.7	-4319.0	-513.0	-4480.0

Standard errors in parentheses: + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$ Standard errors clustered at firm level

Table 8: Summary of additional robustness tests – Hypotheses 3 and 4

Dependent variable	**Number of inventions progressing to next phase (prog)**					**Invention progress in focal year**		**Number of inventions progressing to next phase (prog)**					
Unit of analysis	**Firm-year**					**Invention-year**		**Firm-year**					
Robustness test	**1. CEM**			**2. Alternate Spec.**		**3. Individual Inv.**		**4. Novelty Measure**		**5a. Lagged IV**		**5b. Rolling Avg. IV**	
Hypothesis	**H3**	**H4**	**H4**	**H3**	**H4**	**H3**	**H4**	**H3**	**H4**	**H3**	**H4**	**H3**	**H4**
R&D Decentralization	**-0.216***	-0.0826	-0.214	**-0.150***	-0.201+	**-0.263****	-0.158	**-0.213***	-0.171	**-0.133**	-0.282**	**-0.237***	-0.283*
	(0.0919)	(0.145)	(0.174)	**(0.0707)**	(0.105)	**(0.0965)**	(0.117)	**(0.103)**	(0.104)	**(0.0984)**	(0.100)	**(0.114)**	(0.114)
R&D Functional Decentralization	0.304**	0.0394	-0.159	-0.125+	-0.0864	-0.0103	-0.179+	-0.0435	-0.0653	-0.112	-0.152+	-0.124	0.00547
	(0.109)	(0.124)	(0.156)	(0.0665)	(0.0907)	(0.0746)	(0.101)	(0.0792)	(0.0872)	(0.0770)	(0.0883)	(0.0888)	(0.0893)
Corporate Decentralization	0.209	**0.216***	**0.223+**	0.0790	**0.328***	0.144	**0.259+**	0.0886	**0.299+**	0.0617	**0.144**	0.124	**0.0732**
	(0.188)	**(0.0962)**	**(0.129)**	(0.133)	**(0.165)**	(0.155)	**(0.150)**	(0.120)	**(0.165)**	(0.130)	**(0.186)**	(0.130)	**(0.174)**
performance	-2.456**	-0.278	-0.0293	-0.375	0.445	-0.441	0.709	0.00416	0.527	0.209	0.637	0.0542	0.371
	(0.891)	(0.859)	(1.104)	(0.407)	(0.419)	(0.444)	(0.519)	(0.343)	(0.404)	(0.402)	(0.401)	(0.409)	(0.433)
R&D Intensity	-0.0991	0.829	1.366	-0.00748	0.273	-0.0577	0.816*	0.616**	0.314	0.768**	0.703**	0.613**	0.340
	(0.933)	(0.762)	(0.982)	(0.209)	(0.328)	(0.239)	(0.399)	(0.179)	(0.309)	(0.188)	(0.242)	(0.195)	(0.244)
size	0.338**	0.112	-0.0918	0.0208	0.0857	0.183*	-0.230	0.272**	0.0779	0.281**	0.153	0.283**	0.169**
	(0.0782)	(0.0883)	(0.185)	(0.0875)	(0.0916)	(0.0920)	(0.168)	(0.0516)	(0.0925)	(0.0503)	(0.0999)	(0.0485)	(0.0544)
slack	0.0433	0.0210	-0.0540	0.0145	-0.0554	0.0183	-0.0610	0.0353+	-0.0506	0.0388+	-0.0234	0.0429*	0.00599
	(0.0613)	(0.0561)	(0.0795)	(0.0288)	(0.0437)	(0.0243)	(0.0487)	(0.0204)	(0.0419)	(0.0226)	(0.0425)	(0.0200)	(0.0325)
CEO	0.0325	-0.0433	-0.0858	0.0136	-0.0162	0.0622	0.0242	-0.00819	-0.0144	0.0324	0.00785	-0.00889	-0.0248
	(0.112)	(0.135)	(0.147)	(0.0617)	(0.0810)	(0.0819)	(0.105)	(0.0716)	(0.0806)	(0.0753)	(0.0884)	(0.0675)	(0.0816)
patent stock	0.0284	0.0991	-0.107	0.129*	-0.00464	0.000	0.000	0.0935*	0.0146	0.0948*	-0.00870	0.101*	0.0816*
	(0.0523)	(0.0730)	(0.151)	(0.0553)	(0.0614)	(0.000)	(0.000)	(0.0386)	(0.0646)	(0.0403)	(0.0597)	(0.0393)	(0.0386)
portfolio	0.008**	0.036**	0.046**	0.006**	0.0260**	-0.006**	0.001	0.00641*	0.0245**	0.007**	0.023**	0.008**	0.0178**
	(0.003)	(0.009)	(0.011)	(0.002)	(0.004)	(0.002)	(0.006)	(0.003)	(0.004)	(0.002)	(0.004)	(0.002)	(0.004)
external	0.00197	0.821*	0.889*	-0.591**	0.0737	-0.113	0.228	-0.396*	0.193	-0.600**	0.0235	-0.647**	0.202
	(0.335)	(0.342)	(0.439)	(0.223)	(0.251)	(0.247)	(0.349)	(0.183)	(0.224)	(0.193)	(0.263)	(0.183)	(0.216)
NCE	-1.435**	-0.327	-0.484	-0.868**	-0.757*	-0.888**	-0.727*			-0.596**	-0.736*	-0.601**	-0.230
	(0.417)	(0.426)	(0.694)	(0.246)	(0.340)	(0.307)	(0.362)			(0.205)	(0.313)	(0.205)	(0.257)
bio	0.160	0.0968	-0.400	-0.126	-0.284	-1.196**	-1.038*			0.509+	-0.253	0.390	0.105
	(0.449)	(0.518)	(0.766)	(0.273)	(0.406)	(0.439)	(0.462)			(0.269)	(0.417)	(0.251)	(0.312)
tech. diversity	3.209**	1.157+	0.896	1.503**	1.145**	-1.082**	-0.557	1.269**	1.097**	2.097**	1.411**	1.911**	1.340**
	(0.650)	(0.634)	(1.019)	(0.337)	(0.389)	(0.389)	(0.655)	(0.253)	(0.341)	(0.304)	(0.478)	(0.248)	(0.363)
competition	-11.76**	7.156	4.769	-3.857+	-3.116	-4.023*	-0.971	-5.183*	-2.695	-6.421**	-3.896	-5.756**	-2.627
	(3.392)	(4.678)	(6.093)	(2.333)	(2.387)	(1.914)	(4.144)	(2.082)	(2.546)	(1.721)	(2.417)	(1.814)	(2.060)
novelty								-0.645**	-0.0978				
								(0.231)	(0.440)				
Year Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Firm Fixed Effects	N	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y
Category Fixed Effects	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y
Therapeutic Area Fixed Effects	N	N	N	N	N	Y	Y	N	N	N	N	N	N
N	256	392	392	785	762	21915	10616	787	762	736	713	787	762
Pseudo-R^2	0.313	0.170	0.215			0.0530	0.0367	0.235	0.218	0.241	0.196	0.240	0.194
Log Likelihood	-485.7	-515.0	-486.8	-1379.8	-871.5	-8552.1	-3034.6	-1573.4	-979.9	-1480.5	-953.8	-1562.2	-1009.8

Standard errors in parentheses: + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$. Standard errors clustered at the firm level.

Table 9: Negative binomial regression analyses examining how firms' diversity of technological knowledge (*Tech. Diversity*) moderates the association between *R&D Decentralization* and firm's development outcomes

Dependent Variable	**Number of inventions progressing to next phase**			
Phase Transition	**PC-1**	**1-2**	**2-3**	**3-PR**
R&D Decentralization	0.714*	-0.0239	0.831**	0.596+
	(0.306)	(0.614)	(0.310)	(0.350)
R&D Functional Differentiation	-0.0725	0.141+	0.0697	-0.0433
	(0.0775)	(0.0773)	(0.0778)	(0.0793)
Corporate Decentralization	0.168	0.113	0.0797	0.0126
	(0.129)	(0.179)	(0.136)	(0.138)
R&D Decentralization x Tech. Diversity	**-1.150****	**-0.157**	**-1.279****	**-0.800+**
	(0.393)	**(0.769)**	**(0.398)**	**(0.457)**
Tech. Diversity	2.032**	1.908**	1.655**	1.399**
	(0.268)	(0.256)	(0.343)	(0.225)
Performance	-0.00447	0.620+	0.302	0.673+
	(0.412)	(0.358)	(0.419)	(0.398)
R&D Intensity	0.586**	0.295	0.241	0.595**
	(0.202)	(0.202)	(0.266)	(0.186)
Size	0.287**	0.0892+	0.153**	0.105*
	(0.0498)	(0.0472)	(0.0508)	(0.0494)
Slack	0.0403*	-0.0112	0.00407	0.0480*
	(0.0205)	(0.0268)	(0.0319)	(0.0212)
CEO	-0.00382	0.00654	-0.00970	-0.183*
	(0.0696)	(0.0733)	(0.0801)	(0.0793)
Patent Stock	0.101**	0.0135	0.0819*	0.0132
	(0.0385)	(0.0363)	(0.0401)	(0.0316)
Portfolio	0.00723**	0.0286**	0.0178**	0.0545**
	(0.00250)	(0.00590)	(0.00345)	(0.00517)
External	-0.690**	0.114	0.217	0.0949
	(0.190)	(0.166)	(0.216)	(0.131)
NCE	-0.600**	0.299	-0.172	-0.652**
	(0.207)	(0.192)	(0.231)	(0.189)
Bio	0.410	0.887**	0.220	-0.166
	(0.252)	(0.266)	(0.265)	(0.201)
Competition	-6.109**	-1.593	-2.574	3.214*
	(1.871)	(2.292)	(2.109)	(1.360)
Year Fixed Effects	Y	Y	Y	Y
Category Fixed Effects	Y	Y	Y	Y
N	787	764	762	785
Pseudo-R^2	0.242	0.227	0.195	0.216
Log Likelihood	-1559.3	-1195.6	-1008.6	-1032.6

Standard errors in parentheses: $^{+}\ p < 0.1$, $^{*}\ p < 0.05$, $^{**}\ p < 0.01$
Standard errors clustered at firm level

Table 10: Negative binomial regression analyses examining how the novelty of firms' development portfolios (*NCE* and *Novelty*) moderates the association between R&D Decentralization and firm's development outcomes

Dependent Variable	**Number of inventions progressing to next phase**							
Phase Transition	**PC-1**	**1-2**	**2-3**	**3-PR**	**PC-1**	**1-2**	**2-3**	**3-PR**
R&D Decentralization	0.651**	-0.339	0.281	0.301	0.381	-0.392	-0.578	0.104
	(0.220)	(0.284)	(0.329)	(0.238)	(0.405)	(0.344)	(0.466)	(0.455)
R&D Functional Differentiation	-0.0728	0.148+	0.0570	-0.0480	-0.0460	0.187*	0.0753	-0.0674
	(0.0786)	(0.0772)	(0.0803)	(0.0784)	(0.0786)	(0.0781)	(0.0785)	(0.0765)
Corporate Decentralization	0.172	0.121	0.0695	0.0179	0.0830	0.0719	0.0531	-0.0345
	(0.131)	(0.182)	(0.135)	(0.138)	(0.120)	(0.179)	(0.139)	(0.130)
R&D Decentralization x NCE	**-1.474****	**0.336**	**-0.775**	**-0.605**				
	(0.363)	**(0.458)**	**(0.539)**	**(0.511)**				
NCE	-0.495*	0.250	-0.128	-0.586**				
	(0.199)	(0.207)	(0.235)	(0.191)				
R&D Decentralization x Novelty					**-0.669+**	**0.266**	**0.495**	**-0.140**
					(0.403)	**(0.415)**	**(0.525)**	**(0.527)**
Novelty					-0.619**	-0.349	-0.332	-0.428+
					(0.227)	(0.315)	(0.344)	(0.252)
Performance	-0.0257	0.616+	0.317	0.745+	0.0429	0.427	0.290	0.782*
	(0.410)	(0.360)	(0.420)	(0.393)	(0.341)	(0.346)	(0.394)	(0.368)
R&D Intensity	0.580**	0.301	0.241	0.652**	0.683**	0.255	0.244	0.669**
	(0.204)	(0.200)	(0.261)	(0.178)	(0.167)	(0.231)	(0.254)	(0.206)
Size	0.284**	0.0916+	0.159**	0.106*	0.274**	0.112+	0.154**	0.101+
	(0.0492)	(0.0473)	(0.0531)	(0.0503)	(0.0525)	(0.0653)	(0.0581)	(0.0516)
Slack	0.0435*	-0.0113	0.00643	0.0471*	0.0369+	-0.0116	0.00185	0.0399*
	(0.0205)	(0.0264)	(0.0329)	(0.0204)	(0.0204)	(0.0270)	(0.0332)	(0.0185)
CEO	0.0000474	0.00739	-0.0131	-0.183*	-0.00675	-0.000815	-0.0141	-0.166+
	(0.0689)	(0.0733)	(0.0793)	(0.0791)	(0.0717)	(0.0731)	(0.0787)	(0.0848)
Patent Stock	0.109**	0.00917	0.0845*	0.0150	0.0904*	0.0198	0.0771+	0.00151
	(0.0397)	(0.0365)	(0.0397)	(0.0296)	(0.0379)	(0.0380)	(0.0430)	(0.0297)
Portfolio	0.00717**	0.0286**	0.0182**	0.0542**	0.00639*	0.0308**	0.0191**	0.0557**
	(0.00254)	(0.00595)	(0.00338)	(0.00511)	(0.00252)	(0.00586)	(0.00345)	(0.00531)
External	-0.682**	0.107	0.209	0.0946	-0.401*	0.174	0.293	0.159
	(0.191)	(0.163)	(0.219)	(0.132)	(0.184)	(0.165)	(0.200)	(0.134)
Bio	0.402	0.850**	0.176	-0.169				
	(0.251)	(0.264)	(0.268)	(0.201)				
Tech. Diversity	1.961**	1.895**	1.423**	1.317**	1.285**	1.510**	1.121**	0.939**
	(0.249)	(0.251)	(0.330)	(0.227)	(0.256)	(0.319)	(0.291)	(0.202)
Competition	-6.000**	-1.616	-2.290	3.433*	-4.929*	0.114	-1.376	3.750*
	(1.887)	(2.313)	(2.034)	(1.438)	(2.051)	(2.227)	(2.085)	(1.470)
Year Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y
Category Fixed Effects	Y	Y	Y	Y	Y	Y	Y	Y
N	787	764	762	785	787	764	762	785
Pseudo-R^2	0.244	0.227	0.194	0.216	0.236	0.222	0.192	0.212
Log Likelihood	-1555.7	-1195.3	-1010.4	-1033.2	-1572.1	-1203.0	-1011.9	-1038.9

Standard errors in parentheses: $^{+}\ p < 0.1$, $^{*}\ p < 0.05$, $^{**}\ p < 0.01$
Standard errors clustered at firm level

Table 11: OLS regression examining how grant lag in days is associated with firms' structural measures

Dependent Variable	Grant Lag (Days)	Grant Lag (Days)
R&D Decentralization	**-81.65****	**-50.20***
	(28.72)	**(20.03)**
R&D functional differentiation	63.20+	15.69
	(33.63)	(29.00)
Corporate Decentralization	61.68	70.18
	(63.94)	(63.00)
originality	-200.3*	-134.8
	(96.97)	(91.80)
claims	-3.116	1.470
	(3.073)	(3.013)
Non-patent cites	5.528**	1.643
	(1.762)	(1.753)
performance	-334.1+	-214.7
	(169.0)	(139.4)
R&D Intensity	-91.11	4.625
	(68.25)	(45.67)
size	3.232	-36.07
	(22.44)	(24.66)
slack	0.478	-0.0646
	(8.024)	(9.731)
CEO	18.39	19.71
	(16.60)	(17.48)
tech. diversity	-170.3	-164.8
	(205.8)	(214.5)
patent stock	13.03	32.50
	(15.12)	(26.77)
competition	1585.9+	-157.4
	(795.5)	(530.2)
Year FE	Y	Y
Firm FE	N	Y
Category FE	Y	Y
N	782	782
R^2	0.618	0.644

Standard errors in parentheses
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$
Standard errors clustered at firm level.

Table 12: Logit and cox proportional hazard models examining likelihood or hazard of progression of refined inventions to next development stage. Drug candidates entering phase 2, likelihood and hazard of progressing into phase 3

Model	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Model Type	**Logit - linear time function**					**Cox Proportional Hazards model**		
R&D Functional Differentiation12	**0.0737**	**0.0828**	**0.0926**	**0.0949**	**0.0916**	**0.0586**	**0.0580**	**0.0868**
	(0.234)	**(0.238)**	**(0.249)**	**(0.251)**	**(0.255)**	**(0.166)**	**(0.169)**	**(0.175)**
Firm-year level controls	Y	Y	Y	Y	Y	Y	Y	Y
Category Fixed Effects	N	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effects	N	N	Y	Y	Y	Y	Y	Y
Drug-level controls	N	N	N	Y	Y	N	Y	Y
Therapeutic Area Fixed Effects (Drug level)	N	N	N	N	Y	N	N	Y
N	5216	5216	5216	5216	5168	4473	4473	4473
Pseudo R^2	0.0142	0.0181	0.0346	0.0347	0.0590	0.0203	0.0206	0.0360
Log Likelihood	-1354.4	-1349.0	-1326.3	-1326.2	-1289.2	-2096.8	-2096.2	-2063.2

Standard errors in parentheses $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$; Standard errors clustered at firm level.

Table 13: Logit and cox proportional hazard models examining likelihood or hazard of progression of refined inventions to next development stage. Drug candidates entering phase 3, likelihood and hazard of progressing into pre-registration

Model	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Model Type	**Logit - linear time function**					**Cox Proportional Hazards model**		
Corporate Decentralization23	**0.131**	**0.180**	**0.126**	**0.0551**	**0.107**	**-0.0178**	**-0.109**	**-0.0747**
	(0.330)	**(0.318)**	**(0.336)**	**(0.340)**	**(0.379)**	**(0.308)**	**(0.326)**	**(0.320)**
Firm-year level controls	Y	Y	Y	Y	Y	Y	Y	Y
Category Fixed Effects	N	Y	Y	Y	Y	Y	Y	Y
Year Fixed Effects	N	N	Y	Y	Y	Y	Y	Y
Drug-level controls	N	N	N	Y	Y	N	Y	Y
Therapeutic Area Fixed Effects (Drug level)	N	N	N	N	Y	N	N	Y
N	3578	3578	3578	3578	3578	2712	2712	2712
Pseudo R^2	0.00536	0.00898	0.0249	0.0281	0.0474	0.0156	0.0208	0.0287
Log Likelihood	-1393.6	-1388.5	-1366.3	-1361.7	-1334.8	-2083.3	-2072.3	-2055.5

Standard errors in parentheses $^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$; Standard Errors clustered at firm level.

Table 14: Evaluation of average time for drug-candidates to progress from Phase 2 to 3 clinical trials as a function of a variety of structural, firm-level and portfolio variables. The focus is on examining whether greater corporate decentralization is associated with the progression of inventions more rapidly through the development process consistent with an incentives-based argumentation. Unit of analysis is the firm.

OLS model using average values per firm over period 1995-2015		
DV= Average time progress P2-3 (Years)	**Model 2**	**Model 1**
R&D Decentralization		-0.495
		(0.744)
R&D Functional Differentiation		0.284
		(0.456)
Corporate Decentralization	**-1.898***	**-1.741+**
	(0.899)	**(0.937)**
Size	-0.0355	-0.00840
	(0.168)	(0.172)
R&D Intensity	0.257	0.346
	(0.540)	(0.642)
Slack	-0.161	-0.163
	(0.157)	(0.170)
External	0.200	0.366
	(0.968)	(1.042)
NCE	0.241	-0.321
	(1.298)	(1.575)
Bio	-0.0663	-0.661
	(1.247)	(1.625)
Tech. Diversity	1.621	1.841
	(1.430)	(1.436)
Performance	0.151	-0.116
	(2.615)	(2.896)
SBU	0.189+	0.163
	(0.108)	(0.115)
Patent Family count	-0.000452	-0.000461
	(0.000646)	(0.000686)
Number of Firms	47	47
R^2	0.277	0.294
Log-Likelihood	-40.46	-39.87

Standard errors in parentheses
+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

Table 15: OLS regressions with dependent variable being the log of various compensation measures regressed against 3 key organizational design variables and a variety of firm level controls (Main paper Table 5) with category and year fixed effects. Sample sizes are below those used for main analyses in main paper Table 5 because compensation data is only available for a sub-sample of US-listed firms.

Function	**R&D Executives**		**All Executives including CEO**	
	Model 1	Model 2	Model 3	Model 4
Dependent Variable	Log (Salary)	Log (Total Compensation)	Log (Salary)	Log (Total Compensation)
R&D Decentralization	-0.0293 (0.0496)	0.139 (0.286)	-0.00833 (0.0226)	-0.307 (0.187)
R&D Functional Differentiation	0.0388 (0.0341)	**0.579*** **(0.255)**	0.0527 (0.0398)	0.234 (0.163)
Corporate Decentralization	-0.0711 (0.112)	0.205 (0.654)	**-0.0847*** **(0.0390)**	-0.0871 (0.318)
Firm level controls (Main Paper Table 4)	Y	Y	Y	Y
Category Controls	Y	Y	Y	Y
Year Fixed Effects	Y	Y	Y	Y
N	326	279	555	390
R^2	0.782	0.373	0.738	0.484

Standard errors in parentheses
$^{+}$ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$

REFERENCES

Albert, D. 2018. Organizational module design and architectural inertia: Evidence from structural recombination in universal banking. *Organization Science*.

Allison, P. D. 1982. Discrete-time methods for the analysis of event histories. *Sociological methodology*, 13: 61-98.

Arora, A., Belenzon, S., & Rios, L. A. 2014. Make, buy, organize: The interplay between research, external knowledge, and firm structure. *Strategic Management Journal*, 35(3): 317-337.

Caliendo, M., & Kopeinig, S. 2008. Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys*, 22(1): 31-72.

Diestre, L., & Rajagopalan, N. 2012. Are all 'sharks' dangerous? new biotechnology ventures and partner selection in R&D alliances. *Strategic Management Journal*, 33(10): 1115-1134.

Eisenhardt, K. M. 1989. Building Theories from Case Study Research. *Academy of Management Review*, 14(4): 532-550.

Girod, S. J., & Whittington, R. 2015. Change escalation processes and complex adaptive systems: From incremental reconfigurations to discontinuous restructuring. *Organization Science*, 26(5): 1520-1535.

Greenwood, R., & Miller, D. 2010. Tackling design anew: Getting back to the heart of organizational theory. *The Academy of Management Perspectives*, 24(4): 78-88.

Guadalupe, M., Li, H. Y., & Wulf, J. 2014. Who Lives in the C-Suite? Organizational Structure and the Division of Labor in Top Management. *Management Science*, 60(4): 824-844.

Hill, C. W. L., Hitt, M. A., & Hoskisson, R. E. 1992. Cooperative versus Competitive Structures In Related And Unrelated Diversified Firms. *Organization Science*, 3(4): 501-521.

Klueter, T. 2013. *Searching for Needles in a Haystack: Three Essays on the Role of R&D Partnerships in the Bio-Pharmaceutical Industry*. University of Pennsylvania, University of Pennsylvania.

Markides, C. C., & Williamson, P. J. 1996. Corporate diversification and organizational structure: A resource-based view. *Academy of Management Journal*, 39(2): 340-367.

Russell, R. D., & Russell, C. J. 1992. An Examination of the Effects of Organizational Norms, Organizational-Structure, and Environmental Uncertainty on Entrepreneurial Strategy. *Journal of Management*, 18(4): 639-656.

Sathe, V. 1978. Institutional Versus Questionnaire Measures of Organizational-Structure. *Academy of Management Journal*, 21(2): 227-238.

Turner, K. L., & Makhija, M. V. 2012. The role of individuals in the information processing perspective. *Strategic Management Journal*, 33(6): 661-680.

Walton, E. J. 1981. The comparison of measures of organization structure. *Academy of Management Review*, 6(1): 155-160.