WP. No. : SAUFE-WP-2018-004

Rethinking Innovation: The Case of Diabetes Drug Discovery

Debdatta Saha
Assistant Professor
Faculty of Economics, South Asian University
Akbar Bhawan, Chanakyapuri
New Delhi 110021, INDIA
Email: debdatta@sau.ac.in

T.M. Vasuprada
Research Scholar
Faculty of Economics, South Asian University
Akbar Bhawan, Chanakyapuri
New Delhi 110021, INDIA
Email: tm.vasuprada@students.sau.ac.in

Working Paper Number: SAUFE-WP-2018-004

http://www.sau.int/fe-wp/wp004.pdf

FACULTY OF ECONOMICS
SOUTH ASIAN UNIVERSITY
NEW DELHI
May, 2018

Rethinking Innovation: The Case of Diabetes Drug Discovery

Debdatta Saha[*] and T.M.Vasuprada[†]

[*]Assistant Professor, Faculty of Economics, South Asian University

[†]Doctoral Student, Faculty of Economics, South Asian University

[†]Corresponding author: tm.vasuprada@students.sau.ac.in

May 1, 2018

Abstract

This paper provides a combinatorial perspective on innovation in pharmaceuticals by modeling clinical trials as a research network which resembles a hypergraph with supra-dyadic connections. Third phase interventionist clinical trials are interpreted as hyperedges and molecules tested as a part of these trials as nodes in the network. Our proxy for radical innovation uses the concept of structural holes which is measured using variance of eigenvector centrality of molecules, thereby extending the standard measure of structural holes in dyadic connections to hypergraphs. Using eigenvector centrality of trials and this measure of structural holes, a four-way classification of clinical trials is obtained: Star Performers, Incremental Innovation, Black Sheep and Low Hanging Fruits. Depending on the relative abundance of these categories, twin characteristics of "strength" and "resilience" of the network are defined to discuss quality of innovation. This methodology is demonstrated for clinical trials in diabetes using data over a five year period from one of the largest worldwide registries for clinical trials (ClinicalTrials.gov). Despite the fact that most clinical trials are privately funded, there is no evidence of success bias in trial molecule composition nor is there a bias against novelty. The network exhibits innovative "strength" but not "resilience". While a disease-by-disease approach is necessary to implement this methodology, we feel differential incentives in drug research at the clinical trial stage for different diseases validates our approach towards understanding novelty and incentives in innovation for pharma.

Keywords: innovation in pharma; structural holes; hypergraphs; clinical trial network; eigenvector centrality

JEL Classification Codes: D85, O31, L65

1 Introduction

Existing literature on innovation has predominantly used patents and R&D expenditure as the yardstick for measuring innovation (Papageorgiadis and Sharma (2016), Allred and Park (2007), Tyagi et al. (2014)), Mansfield et al. (1981), Mansfield (1986)). While the former is a leading output-based measure, R&D expenditures are a proxy for input based measures. Other measures such as total innovation expenditures, sales of innovative products, innovation counts, patent citations (Narin and Olivastro (1988)), patent claims (Tong and Frame (1994)), patent families (Dernis and Khan (2004)) and renewal fees (Pakes et al. (1989)) have recently been used as contenders to patents and R&D expenditure as they improve upon some dimensions of the earlier measures as far as innovation is concerned (Kleinknecht (1996), Archibugi and Planta (1996)).

Measures based on patents additionally suffer from an inherent survival bias because only success in the first stage of research and development (particularly the pre-clinical stage in pharmaceuticals) leads to a patent[1](Grabowski (2002)). Many failures which never resulted in a patent application, crucial to the process of learning cannot possibly be accounted for within this system. Scotchmer (1991) points out that looking at *patents in isolation* misses the larger picture. Pharmaceutical patents have to account for differential value of innovation, ranging from a breakthrough discovery to marginal contributions to existing knowledge in drug research (Harhoff et al. (1999), Gambardella et al. (2017)). Since patent counts do not differentiate between patents in terms of their inherent value, indices based on patent citations are a potential improvement in evaluating the value of patents (Trajtenberg (1990)). One issue with patent counts is that it captures quantity (just as other measures such as number of drugs marketed) rather than quality of innovation (Cohen (2005)). More substantially, as a "measure of knowledge flows", patent citations are deeply flawed. For instance, (Alccer and Gittelman (2006)) finds that approximately 63 per cent of the final citations accompanying an average patent (across a pool of 442,839 granted patents between January 2001 and August 2003 with the USPTO) are externally added by patent examiners and are not generated by the innovator filing for patent.

In general, the process of research and development in pharmaceuticals can be decomposed in two stages (Henderson and Cockburn (1996)). The first comprises the process of drug discovery wherein a new molecule is found in the laboratory and the second comprises of clinically testing the new molecule for both safety and efficacy. In order to locate the space of learning through which innovation happens in pharmaceuticals, we adopt a "combinatorial novelty in science" approach (Wang et al. (2017), Arthur (2009), Nelson and Sidney (1982)) that is similar to the literature on patent citations. However, we work directly with data on molecule combinations in clinical trials, rather than indirect citations. Our perspective is that the manner in which different molecules are combined in clinical trials are revelatory of both the learning process of sponsors as

[1]https://www.biology.iupui.edu/biocourses/Biol540/4pipeline09Full.html

well as the differential quality in terms of novelty of research in the process of drug discovery.

Current standards in drug launch ensure that a new drug discovery cannot happen without the second stage. The results of the clinical trials are verified by regulatory authorities, with successful trials leading to the marketing of new drugs or more efficient methods of drug delivery. Failed trials result often in research papers, that feed into the first stage. Additionally, this industry requires considerable committed sunk expenditure primarily in clinical trials, prior to the actual production of drug formulations (Cohen (2005), Cockburn and Henderson (2001), Mestre-Ferrandiz et al. (2012)). Over time, biotechnology collaborating with pharmaceutical research (Stuart et al. (2007), Danzon et al. (2005), Nicholson et al. (2002)) the frontiers of innovation have expanded to include many new drug formulations, such as biologics in cancer research, where live culture is combined with chemical compounds to increase the efficiency of the intervention. The importance of clinical trial testing to understand drug efficacy on various disease conditions. For instance, Cohen (2005) shows that between 1976 and 2002, the expenditure on non-clinical/ pre-clinical research has sharply declined from around 48 per cent of total annual R&D expenditure to approximately 28 per cent using data from annual PhRMA surveys in the United States.

The combinatorial perspective on novelty of innovation is mostly limited to citation and bibliometric analysis in the first stage of pre-clinical trial R&D (Klavans and Boyack (2013), Henderson and Cockburn (1996), Powell et al. (1996)). Note our non-standard usage of research network: we fix a disease and explore the research network as represented by third stage clinical trials using various molecule combinations by sponsors. Our focus is not on using the person-specific exploration of research networks, as in Burt (2004) or in the pre-clinical trial stage research network (Powell et al. (1996)), or in papers exploring patent citations (Alccer and Gittelman (2006), Trajtenberg (1990)). These papers explore the pattern of interactions between pairs of agents in the research network, either through citations or actual collaborative contacts. As citations themselves do not reveal the exact learning process of the innovator (additions to knowledge from ubiquitous sources such as informative blogs mostly do not find a space of citations and bibliography and cannot be captured), we infer the learning process of the pharmaceutical companies and public organizations that sponsor clinical trials from their choice of combination of molecules in clinical trials. To the best of our knowledge, there is no other attempt similar to ours in the literature on innovation in pharmaceuticals.

In order to model the clinical trial process as a research network, we limit ourselves to third stage trials where two or more molecules are combined and tested on a particular population of participants[2]. The clinical trials themselves are interpreted as hyperedges and molecules are the nodes in the network.

We discuss the issue of novelty in research (to differentiate between path-breaking discoveries (Schumpeter (1939)) as opposed to incremental innovation) using the concept of "structural holes". Lichtenberg

[2]A number of trials explore some aspects of a particular disease or investigate the side-effects of one single molecule itself. Our methodology does not apply to these trials.

(1998) provides an exploration of only Schumpetarian discoveries in pharma (used interchangeably with pharmaceuticals). The concept of structural holes has emanated from the sociology literature on the potential for dissimilar connections to exchange information that they did not possess apriori leading to innovation in networks of human agents (Burt (2004)). Economic analysis of innovation has recently used this concept (Raider (1998), Ahuja (2000), Foster et al. (2015), Chen et al. (2009), Rzhetsky et al. (2015), Shi et al. (2015)). As "structural holes" represent missing or brokerage connections, they are ideal candidates for modeling path-breaking research ideas in a clinical trial research network. The presence of both dyadic and supra-dyadic combinations (as trials include two or more than two molecules) necessitate an application of hypergraphs. These are generalized graphs in which there is no restriction on the size of the edges. Our paper presents an opportunity to demonstrate the workings of hypergraphs, which are present in many social contexts but have not been explored much in the literature[3].

As we use hypergraphs, the measure using constraints to capture structural holes (Burt (2004)) has to be modified. In the research network of clinical trials, we work with the concept of eigenvector centrality (evc henceforth) of the trials and molecules, using the two-mode data construct of Bonacich et al. (2004) in the context of hypergraphs. This measure of centrality is appropriate in our context, as a molecule becomes "central" in the network by its association with more "central" connections (clinical trials). The more "central" the molecule, researchers are more commonly aware of its properties and lower is the possibility that it can contribute to path-breaking discoveries in pharma research, unless combined with some "less central" molecules in some other clinical trials. Neither degree nor betweenness nor other common measures of centrality serve the purpose of highlighting the quality of information that a molecule conveys in the overall context of pharma research through its location in the hyperedges. Additionally, as a measure of innovation it has some desirable properties: first, it captures learning without a survival bias towards successes; second, it does not interact with the incentives for innovation; third, this measure is replicable over time; and fourth, it reflects the differential learning costs which exist across diseases. Though most of the literature on drug research using patent data clubs across all diseases, it attracts criticism that incentives for research differ across diseases. For instance, the incentives for conducting clinical trials for relatively common medical conditions such as diabetes type II is very unlikely compared to those for a rare medical condition like Rett's Syndrome.

Variance of evc of molecules (nodes) distinguish between structural holes (or uncommon molecule combinations) and pedestrian combinations in clinical trials (hyperedges): the larger the variance of evc for molecules, the higher is the potential for spectacular discoveries through clinical trial research. Hence, more uncommon the combination (a notion which finds resonance in Fleming (2001) where "familiar combinations" of technology subclasses in patent classification are marked as less novel leading to higher average

[3]One exception is Bonacich et al. (2004) in the context of historical networks.

patent citations with lower variance in citations compared to less novel ones), as reflected by a high variance of evc of molecule eigenvector centrality, the higher is its potential to be a structural hole in the network. Further, the location of new information is unique to the clinical trial or the hyperedge and is not present in the entire network. Every new and costly clinical trial contains within it the possibility of path-breaking research and we posit that for supra-dyadic connections which use clinical trial data, the location of structural holes should be studied within each non-trivial drug combination, i.e. the hyperedge of the clinical trial hypergraph, which is not discussed in the research network literature. For instance, Powell et al. (1996) contends that the entire research network is a locus for innovation, rather than specific sub-parts of the network.

In our paper, trials which use obvious combinations become more "common" with high trial evc, whereas unique trials with less common combinations have a lower trial evc. The median of both the variance of molecule and trial evc are used as a cut-off to classify "high" and "low" centrality score, leading to a four-way classification of trials, which we creatively label "Star Performers" (high trial as well as variance of molecule eigenvector centralities), "Black Sheep" (low trial but high variance of evc for molecules), "incremental innovation" (low trial and variance of molecule evc) and "low-hanging-fruit" (high trial but low variance of evc for molecules). The quality of popular trials in terms of novelty of drug combinations is read off from the "strength" of the network, whereas "resilience" applies to novelty in uncommon trials. This classification has the advantage of treating structural holes not as a homogeneous entity, as strength and resilience differentiate between the brokerage connections across popular and uncommon trials. As novelty is more likely in uncommon trials leading to resilient networks, we ascribe strength to networks where popular trials are ubiquitously marked by novel structural hole molecule combinations. To this end, our research adds to the literature questioning the extent to which structural holes yield better outcomes for agents located at these brokerage connections (Ahuja (2000)). At the same time, we do not ignore incremental research which are a part of small and sometimes failed clinical trials, which avoids the survival bias afflicting patent measures. Our treatment of clinical trials as a research network for learning and for potential drug discoveries allows us to study both the trivial as well as the novel in a unified framework, which papers like Cohen (2005) argue in favor of to understand the innovative process in pharma.

We demonstrate the application of the hypergraph model and research network classification to third stage interventionist clinical trial data for diabetes type II, using data from the largest clinical trial registry for all diseases (ClinicalTrials.gov) for a time period of five years. At present, diabetes has an alarming incidence worldwide. International Diabetes Federation reports that in 2017, 60 per cent of the adult population between 20-79 years are diabetic from ten countries including China, India and the US, accounting for 69 per cent of the global healthcare expenditures on diabetes[4]. Unlike the citation-based combinatorial exploration

[4]Refer to http://diabetesatlas.org/resources/2017-atlas.html

result on delays in recognition for novel research (Wang et al. (2017)), we find no evidence of differential rates of repetitions of novel structural hole combinations of molecules compared to common non-structural hole combinations within or across years. It is mostly the same sponsor who repeats such combinations, be it for structural holes or non-structural holes. As these are big pharmaceutical majors, our intuition is that there is a race in research for novelty in modern clinical trial research for diabetes type II so that sponsoring firms do not discriminate against novel (and therefore risky) combinations. In corroboration with this result, we also find no evidence of obvious biases in our data in favor of low hanging fruits in clinical trial research for diabetes type II (through obvious combinations which are more likely to yield successful trials and therefore, marketable drugs). In terms of quality of the research network of clinical trials for this disease, while the criterion of strength is fulfilled, resilience is absent in the network. This lends further support to our intuition that big pharma majors, engaged in stiff competition to discover the next big success, do not shy away from undertaking the risk of combining unusual molecules in sponsoring large popular trials. Low resilience of the network in fact indicates that the less popular trials that fall short of adequate numbers of structural holes, indicating less risk-taking and lower potential for breakthrough innovation.

The structure of this paper is as follows: section 2 presents some issues with the definition of innovation in pharma, the difficulties in linking novelty with quality of innovation and a brief description of interventionist clinical trials, section 3 models clinical trials as a research network using hypergraphs and examines measures based on this network which are valid markers for innovation in pharma, section 4 presents our empirical analysis with an application of our methodology for diabetes clinical trials from 2010 to 2014. Section 5 concludes.

2 Clinical Trials and Pharmaceutical Innovation

2.1 Definition of Innovation in Pharma

We start with the onerous task of defining innovation relevant for the pharmaceutical industry. We define innovation in pharma broadly *to include all processes that have the potential for drug discovery.* In the particular, our usage of this definition is applied to clinical trials which are a part of the process of drug discovery. The closest to our definition is that of Wardell and Diraddo (1980) which acknowledges the difficulty inherent in defining pharmaceutical innovation, they nonetheless define it as "any development that is intended to produce a therapeutic advance". Narrower definitions such as Bouet (2015), Morgan and Lopert (2008) and Hollis (2004) bind innovation to the space of pharmaceutical products which retains an inherent survival bias because successful innovations leading to product development. We include both successful and failed clinical trials to emphasize the point that any incremental step towards drug discovery has to be seen in the larger perspective wherein failed attempts only lead to further innovation in an attempt

to obtain a successful outcome in the future.

2.2 Novelty and Quality

Wang et al. (2017) mentions that citation-based measures of novelty suffer from potential unobserved and uncontrolled heterogeneity in published paper quality. As we work with the institution of clinical trials, we rely on objective parameters such as trial size and other technical requirements which have been standardized worldwide, which apply uniformly to trials with novel molecule combinations and those with common ones. There is no variation in terms of quality standards for clinical trials, though their sizes, funding and molecule combinations may vary. This is discussed in detail below.

2.3 Brief Exposition on Clinical Trials

Pharmaceutical research is no longer conducted in a adhoc manner and has clear pre-defined objectives, which are essentially classified in two steps (Henderson and Cockburn (1996)). The first step involves investigation into molecules and the second step is the drug development process itself, which includes clinical trials. These trials are unique to the pharma industry and is one of the largest sources of risks in drug discovery. For one, success in the trial is not guaranteed (though tampering with the probability of success using common molecule combinations is not ruled out: we test for this in the case of diabetes). Second, clinical trials decrease the patent term granted to molecules since a substantial period of time is invested in the conduct of clinical trials and the required marketing approvals (Budish et al. (2015)).

The USFDA defines clinical trials as a way of testing "potential treatments on human volunteers" to establish whether these treatments can be used by the population at large. Treatments here are defined as a drug (technically, molecules), medical devices or a biologics. However, potential treatments must be first tested for safety on animals before testing on human volunteers. Once safety on animals is proven than only they are taken to clinical trials. The USFDA also takes care that participants in clinical trials are not subject to unethical treatment by researchers by ensuring the participants are fully informed about the risks involved. The standard of "informed consent" goes beyond written consent so that potential participants can clarify their doubts about the process.

Clinical trials are conducted in different phases at hospitals and research centers with each phase being more stringent than the previous in terms of rigour. Pocock (1983) classifies clinical trials into four phases which corresponds to the USFDA's classification. To summarise, phase 1 trials rule out any toxicity, phase 2 trials are a preliminary clinical investigation for treatment effects, phase 3 trials are a full scale evaluation of treatment and phase 4 trials involve post marketing surveillance. By the stage 3, enrollment of subjects is completed and the trial is set for yielding results. Description of the clinical trial mentions the molecules

which are a part of the clinical trial. Most commonly, the sponsor of the clinical trial (if it happens to be a large pharmaceutical firm) conducts the trial on its own, hiring agents as per the directives of the clinical trial. The sponsor typically puts up one or more of its own patented formulation for testing in the interventionist trial. Nonetheless, there are more complex ways in which the link between patented drugs and clinical trials are formed. The sponsor also includes patented formulations of other firms as part of the clinical trial as, for instance, comparators. This is most evident for diabetes, which affects multiple organs and a cocktail of drugs is prescribed simultaneously to address the simultaneous and multiple affectations. This fact also necessitates a network structure exploration to understand the locus of innovation, rather than a partial by-firm profit maximizing approach to understand the link between patents and innovation. The trials themselves have multiple objectives: checking for reaction to the injectable form of the drug, evaluating differential responses to a combination of drugs for a particular disease condition etc.

Since the literature on clinical trials tends to be biased towards successful and promising trials (Friedman et al. (1998)), a move towards registering all clinical trials was made in the United States in 1997 [5]. An advantage of registering trials is that unnecessary expense in duplicating unsuccessful trials can be avoided. The International Committee of Medical Journal Editors (ICMJE) mandates that it is "an ethical obligation to share data generated by interventional clinical trials" because participants have undertaken some risk in the conduct of the trial (Taichman et al. (2016)) giving an impetus to the reporting of both successful and failed clinical trials.

3 Methodology: Modeling Clinical Trials as a Research Network

We model the research network as a finite collection of clinical trials for a particular disease in a given year k using hypergraphs, where such a structure defined as:

$$H = \{\mathbb{V}, \mathbb{E} = (e_i)_{i \in I}\} \tag{1}$$

where $\mathbb{V}$ = collection of nodes or vertices

$\mathbb{E}$ = set of non-empty collection of edges, e_i of V

I = the finite set of indices

The simple graph is nothing but a degenerate hypergraph which only considers two vertices at a time. A simple graph is defined as:

$$G = (V, E) \tag{2}$$

[5] India, according to the International Diabetes Federation Report (2015) is only second to China in terms of the number of diabetes patients, has launched its own clinical trial registry in 2007. Registration has become compulsory from 2009 leading to 6000 trials till July, 2015.

where V is a set of non-empty vertices or nodes and E is the set of edges, where each edge contains at most two nodes. Hypergraphs, on the other hand, can account for multiple nodes which are elements of the same hyperedge.

While a number of economic interactions involving networks involve supra-dyadic connections, the only application is in Bonacich et al. (2004). We define H to be a research network where $\mathbb{V}$ is a finite collection of molecules which are elements of interventionist clinical trials $\mathbb{E}$ for a particular disease in a given year. While $\mathbb{E}$ might include a single, two or more molecules in various combinations, we consider only those trials which contain at least two or more molecules, as we are interested in the information embedded in these combinations. This is depicted in Figure 1[6], which represents the hypergraph for diabetes type II clinical trials in phase 3 for the year 2014. The numbered nodes are molecules which were included in a total of 24 trials. The exact molecule names for the number codes shown in Figure 1 is detailed in Appendix A (Table 7). Note clinical trials as hyperedges in the research network, sometimes include more than two molecules (for instance, nodes 50, 14 and 21 belong to a single clinical trial). We assume that the sponsor of the

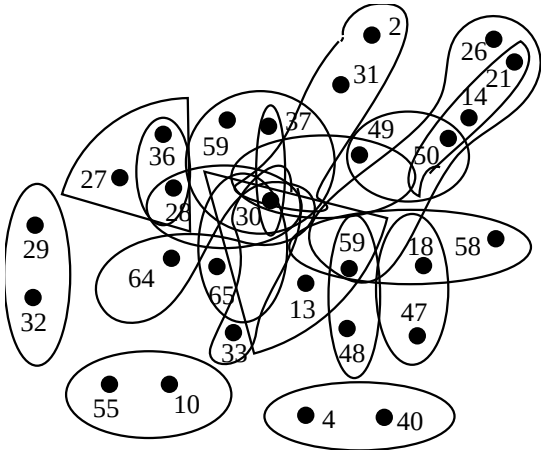

Figure 1: Diabetes Clinical Trial-Molecule Network for 2014

clinical trial exercises choice about novelty or lack thereof through combinations of molecules in these trials. Measurement of novelty uses a metric based on evc (eigenvector centrality, as explained in the introduction). Evc is based on the notion that a node becomes central in the network if it is part of a more central hyperedge and an hyperedge is more central is it contains more central nodes. Clinical trials acquire centrality by testing common (popular) molecules and common (popular) molecules are tested by central clinical trials. However, more central is the node (molecule), more information exists about its properties and therefore, if a trial with common/central nodes is considered one which is testing obvious combinations, with low potential for

[6]For the sake of clarity, two clinical trials were omitted from Figure 1 and two trials with the same composition (26, 21, 14, 50, 59, and 30, 37, 59) sponsored by Merck Sharp together with Pfizer and Sanofi are repeated in 2014.

novelty. For instance, molecule metformin (coded 59 in figure 1) has either the highest or the second highest evc molecule centrality in our network. It was discovered as early as 1922 and has been a part of diabetes medication since 1957. Trials which only include molecules like metformin experiment with information that has been part of medical literature for a sufficiently long period of time. Evc centrality for these trials are therefore relatively high compared to trials experimenting with less common molecules.

Working directly with hypergraphs is cumbersome (Newman (2010)). We represent the hypergraph as a bipartite/two-mode network (Bonacich et al. (2004)). The two modes of the data for our network are trials and molecules included in the trials. The incidence matrix W_{ij} reveals whether molecule i is part of trial j. We appeal to the Perron-Frobenius Theorem which states that a symmetric $n \times n$ matrix A which has only positive entries is guaranteed to have a unique maximum eigenvalue to ensure the existence of a unique eigenvalue of the incidence matrix. The evc scores for nodes and trials are read off from the normalized eigenvector corresponding to the maximum eigenvalue, where we use the column total of the eigenvector to ensure that evc scores are less than one. Note that rows of the incidence matrix represent trials and columns represent molecules. The manner of calculation requires the simultaneous solution of the following equations (Busseniers (2014), Bonacich et al. (2004)):

$$c_1 v_i = \sum_j W_{ij} f_j \tag{3}$$

$$c_2 f_j = \sum_i W_{ij} v_i \tag{4}$$

Using the median value of evc for trials and molecules as the cut-off, we classify them into "high" (evc scores higher than or equal to the median) and "low" (evc scores lower than the median).

Hypothesis 1. *If there exists a trial with evc f_j with value higher than the median evc for all trials in that year, then there is at least one molecule that is part of that tial which has evc v_i that is also higher than the median evc for all molecules. Similarly, if a molecule has "high" v_i, then it is part of a trial with "high" f_j.*

In order to measure novelty using structural holes, we modify the methodology of Burt (2004) to apply to hypergraphs. We use the variance of evc for molecules included in a clinical trial for identifying these points of "gaps in information flows", as relatively rare combinations of molecules contain information that is not very commonly held in the research community already. Using the median variance of evc of molecules as the cut-off, all trials with variance of evc of molecules higher than or equal to that the median are labeled "high" and those below the median as "low". High variance of evc of molecules imply that molecules with both high and low centrality scores have been combined in the trial, which is indicative of experiments with common/popular molecules as well as less common ones. These combinations indicate the presence of a 'structural hole' in the clinical trial research network, as our second hypothesis posits.

Definition 1. *Structural Hole:*

For any trial t_j, if the variance in eigenvector molecule centrality is classified to be 'high', then there exists a structural hole connection and hence, innovative potential. Hence, the presence of a structural hole connect in a trial signifies novelty in research objectives sought to be fulfilled by the trial.

We can now use a two-way classification of the research network, along the dimensions of evc of trials and variance of evc of molecules in the following way:

- **Star Performers:** Trials with high eigenvector centrality and high variance of eigenvector molecule centrality
- **Black Sheep:** Trials with low eigenvector centrality and high variance of eigenvector molecule centrality
- **Incremental Innovation:** Trials with low eigenvector centrality and low variance of eigenvector molecule centrality
- **Low Hanging Fruits:** Trials with high eigenvector centrality and low variance of eigenvector molecule centrality

Using this classification, we can characterize the clinical trial research network along the dimensions of "strength" and "resilience" as discussed in the following hypothesis.

Hypothesis 2.

- *A clinical trial network is **strong** if "Star Perfomers" are higher in number than "Low Hanging Fruit" trials consistently for at least five years.*
- *A clinical trial network is **resilient** if "Black Sheep" trials is greater than "Incremental Innovation" trials consistently for at least five years.*

To maximize innovative potential, a research network should ideally satisfy both characteristics of "strength" as well as "resilience" for a sufficiently long period of time. Given the dynamics of new research, we have fixed this term as five years for our definition. The network is strong in its research innovativeness if a significant number of popular trials are characterized by innovative potential (measured by structural holes). This is of great importance as one would expect less popular trials to experiment with uncommon molecule combinations (and therefore have more structural holes). As long as the trials central to the network have the potential for innovation, we believe the criterion of strength is met. On the other hand, the less popular trials, if a majority of them have innovative potential/structural holes, then the research network is marked by potential for Schumpetarian innovation on the margin, leading to high resilience of the network. These properties are desirable because all kinds of trials whether they include commonly known formulations

or whether they experiment with less-known prescriptions, have an element of new information (potential structural hole) and hence, a possibility that results of the trial will not be a marginal extension of existing information.

4 Application to Third Phase Interventionist Clinical Trials in Diabetes Type II

4.1 Data Description

We have collected the data on clinical trials from ClinicalTrials.gov[7], which is maintained by the U.S. National Library of Medicine and is the world's largest clinical trials registry[8]. This is a public registry of public as well as privately funded clinical trials from 2008 onward. It comprehensively includes trials conducted at multiple locations worldwide. Studies like B.R. et al. (2013), Inrig et al. (2014) and Pasquali et al. (2012) have used the database to explore results from oncology, nephrology and pediatric clinical trials. A shortcoming of the this database is that the methodological approach used in analysing the results of clinical trials is not consistent across all clinical trials (Califf et al. (2012)). Further, the reliability of the database depends upon the truthful revelation of the results by researchers conducting the clinical trials (Zarin et al. (2011)) and timely updation of the database. However, Jeong et al. (2017) claim that this database is the most accessible and the largest registry of clinical trials worldwide and use this database to study the characteristics of clinical trials so as to explain the pattern of globalization of clinical trials from 2011 to 2013. For a randomly selected sample of clinical trials, there exists an additional problem of bias in publication of results (Ross et al. (2009), Saito and Gill (2014)). This fact underlines the importance of this database because even if there is no reporting of results in a publication, there is a record of the trial having been conducted in the registry.

These clinical trials conducted are categorized as interventionist and observational with interventionist trials being conducted to assess the effect of a particular intervention assigned to participants in the trial. While observational studies encompass an account of various factors that may affect the result without administering the intervention to participants. We use data on diabetes type II interventionist clinical trials, which have completed the third phase of testing. A completion of a trial is described as the trial that is no longer treating participants and hence the results can be evaluated. Since for the years 2015 and 2016 there were only 11 and 4 completed diabetes type II clinical trials respectively, we have restricted the data for the five year period from 2010 to 2014. This time period is sufficient to capture the changing landscape of clinical

[7]https://clinicaltrials.gov/ct2/home

[8]Note that http://www.circare.org/registries.htm, which provides a comprehensive list of all public registries declares this database to be the most comprehensive one.

Table 1 Description of Diabetes Type II Clinical Trials with Third Phase Completed

Year	Number of Trials	Percentage of Dyadic Trials	Percentage of Supra-dyadic Trials	Enrollment size of the largest trial (Sponsor)	Enrollment size of the smallest trial (Sponsor)	Average size of enrollment
2010	40	15.00	10.00	1549(Boehringer Ingelheim and Eli Lilly)	16(Radboud University)	555
2011	26	11.25	5.00	2705(Boehringer Ingelheim and Eli Lilly)	26(Astrazeneca)	678
2012	38	14.37	9.37	1413(Boehringer Ingelheim and Eli Lilly)	32(AstraZeneca and Bristol-Myers Squibb)	467
2013	32	14.37	5.64	7637(Novo Nordisk A/S)	7(Hoffmann-La Roche)	730
2014	24	8.14	6.87	1291(Merck Sharp and Dohme Corp. and Pfizer)	10(Medical University of Vienna)	525

Source: Authors' own calculations based on data from ClinicalTrials.gov

trials. Trials in phase 1 and 2 often get terminated due to lack of participants and financing issues and our focus is solely on the innovative potential of these trials, we restrict our dataset to phase 3 trials. The total number of completed trials in our dataset are 160. Since research and development budgets are allocated annually, we first create year wise cohorts containing studies with strictly greater than one molecule. Table 1[9] gives a year-wise description of the phase three trials conducted between 2010-14. Though the composition of molecules for some of the trials may be same, the trials themselves are different. We discuss an example from our dataset in a later section.

A regularity check, following Bonacich et al. (1998), that was conducted on all year-wise cohorts was to check whether the size of the trial is correlated with trial evc scores. We find no evidence of consistent pattern in these correlations, as they range from a low (0.10) to high (0.84) in our data. Since, the correlation coefficients are not consistently high, we have not taken size into consideration when calculating our eigenvector centrality scores. In Table 2, we draw a parallel between the eigenvector trial centrality scores

[9]All tables in this section are authors' own calculations from ClinicalTrials.gov data for respective years.

Table 2 Year-wise Lowest and Highest Eigenvector Centrality Trials with Number of Diabetes Drugs Marketed by Sponsor

Year	ClinicalTrials.gov Identifier- NCT Number	Evc	Sponsor (type; diabetes drugs portfolio)	Trial Size by Enrollment
2010	NCT01000688	0.0016	Radboud University(Public; 0)	16
	NCT01106690	0.0629	Janssen Research LLC(Private; 6))	344
2011	NCT01301833	0.0014	Mitsubishi Tanabe(Private; 1)	462
	NCT01388361	0.1215	Novo Nordisk A/S(Private; 4)	413
2012	NCT01513590	0.0007	Novo Nordisk A/S(Private; 4)	394
	NCT01680341	0.0007	Novo Nordisk A/S(Private; 4)	272
	NCT01590771	0.0580	Merck Sharp(Private; 3)	498
2013	NCT01734785	0.0000	Boehringer and Eli Lilly(Private; 2)	607
	NCT01755156	0.0976	Merck Sharp(Private; 3)	402
2014	NCT02068443	0.0000	Takeda(Private; 1)	374
	NCT02131272	0.0000	Novo Nordisk A/S(Private; 4)	42
	NCT02220907	0.0000	Mitsubishi Tanabe Pharma(Private; 1)	153
	NCT02099110	0.1000	Merck Sharp and Pfizer(Private; 3)	1291
	NCT02036515	0.1000	Merck Sharp and Pfizer(Private; 3)	464

Source: Authors' own calculations based on data from ClinicalTrials.gov

with the number of drugs marketed by the sponsor before the start of the trial in consideration. We have taken the trials with the largest and smallest centrality score along with some intermediate trials. There is no obvious correlation between the number of drugs discovered and the centrality scores, reducing the possibility of a survival bias in our clinical trial data.

We find that **Hypothesis 1** holds for our data[10]. Table 8 in Appendix A, shows the results for year 2014 (a research network of 24 trials testing with 29 unique molecules in various combinations) where median values for trial and molecule evc are 0.042 and 0.014 respectively. Trials and molecules with evc equal to or greater than these values are labeled "high" (or "low" otherwise). Consider trial 1, it has 'high' trial evc and consists of molecules 30, 37 and 59, all of which are classified 'high' evc. On the other hand, trial 8 has 'low' trial evc and it comprises of molecules 27, 28 and 36 of which 27 and 36 have 'low' evc whereas 28 has 'high' evc.

In Table 4, we identify structural holes in the data for 2014 using the median value of 0.004 for variance

[10]We use MATLAB software for all our calculations of centrality scores.

of evc for molecules. All trials with higher or equal values for variance of evc of molecules are classified as structural holes. As mentioned earlier, this methodology of identifying a structural hole within the hyperedge extends the methodology of Burt (2004), which uses a network constraint measure to quantify these bridge connections in the network. Note that the latter methodology only applies for dyadic connections, whereas our measure works for supra-dyadic links[11].

4.2 Testing for Success Bias in Trials

As mentioned earlier, we do not expect to see quality biases in our data. However, a bias towards successful trials (by choosing combinations of molecules) cannot be ruled out. For clinical trials to serve as a valid platform to measure learning and innovation in research, there must be an element of risk-taking through unusual molecule combinations, so that patterns of combinations biased in favor of success of the trial can be ruled out. Given a clinical trial research network in year t, a success-biased sponsor will include combinations in for trial j in the trial network $H = \{\mathbb{V}, \mathbb{E}\}$ in order to maximize the probability of success:

$$\text{Max Probability}_j(\text{success}|j \in \mathbb{E} \text{ and } H = \{\mathbb{V}, \mathbb{E}\}) \text{ such that } \sum_j C_j(t) \leq \ B(t) \tag{5}$$

where $\sum_j C_j(t)$ is total expenditure on clinical trials driven by trial size (Mestre-Ferrandiz et al. (2012)) and $B(t)$ is the total annual R&D budget on clinical trials. This kind of choice formulation for the sponsor would imply strategic composition of elements in the trial, with an objective to pick winners.

We assume that the appropriate counterfactual to the strategic model is one of naive behavior that maximizes learning potential offered by combination of molecules in trial j in the trial research network.

$$\text{Max } \mathrm{L}_j(t; H = \{\mathbb{V}, \mathbb{E}\}) \text{ such that } \sum_j C_j(t) \leq \ B(t) \tag{6}$$

where the learning function L is represented by structural holes; i.e. $\mathrm{L}_j(t; H = \{\mathbb{V}, \mathbb{E}\}) = \ \mathrm{var}_j(\mathrm{evc}(\mathbb{V}); j \in H, t)$. As structural holes provide the potential for gaps in current knowledge to be filled in, they proxy for the learning potential in the network. Note that 97.5 per cent of our data from 2010-14 on third phase clinical trials in diabetes II are privately funded by pharma majors and likely to suffer from a larger success bias than publicly funded trials. Rejection of the success bias in these trials should be sufficient proof that in general clinical trial networks are appropriate platforms for inferring learning and innovation in pharmaceuticals.

We model the probability of success in the third phase clinical trial through a multinomial logit estimation. The reason for this specification is that clinical trials indicate a gradation/natural categorization of success rather than a binary success or failure kind of outcome. Hence, instead of directly modeling the probability of success, we first create three distinct categories of success for any clinical trial. The first category contains

[11] Newman (2010) notes a significant loss of information about the entire hypergraph if we forcefully restrict it to unimodal data format where the usual network constraint method for identifying structural holes is applicable.

Table 3 Year-wise Description of Success Categories

Success Category	2010	2011	2012	2013	2014	Total (Per cent out of 160)
1	19	9	13	11	13	65 (41)
2	16	8	19	10	5	58 (36)
3	5	9	6	11	6	37 (23)
Total	40	26	38	32	24	160

Source: Authors' own calculations

trials that have led to drug discovery indicating a very high probability of success. The second category contains successful trials which cannot be mapped uniquely to a new drug discovery program. Trials included in this category typically are associated with an already discovered drug and the test is conducted on a new population or against a new comparator. The last category comprises of both failed as well as inconclusive results. Table 3 shows that success categories 1 and 2 are almost similar in number over the five year period, but success category 3 is lower. While there might a potential bias against reporting failed results through public registries, we hope to correct for it by including trials with inconclusive results in this category.

With categories of success as the dependent variable, our independent variables are trial size (positively correlated with funding Mestre-Ferrandiz et al. (2012)) and controls for experience of the sponsor in conducting diabetes research prior to the start date of the particular clinical trial and importance of the trial in the research network. Experience is measured by the number of diabetes drugs that the sponsor has marketed prior to the trial. Evc scores for trials is our research-network based measure to capture trial importance. Additional controls for year (dummy variable) and trial type (dummy variable for whether a trial is a part of a large investigational series[12]) are included.

Using the third success category as the baseline, we find that trial size (measured by enrollment) is significant at the 5 percent level, for both success categories and the dummy representing series which is significant at the 1 percent level for the first category, all other variables are insignificant, discussed in Table 9 in Appendix B. Thr results do not change with other specifications such as logarithm of enrollment as the independent variable.

We test for the IIA (Independence of Irrelevant Alternatives) assumption using the Hausman test and find that it is satisfied since none of the categories of our dependent variable are significant (see Appendix B Table 10 for details). Therefore, the odds are independent of other alternatives. To check whether an independent variable has any affect on the dependent variable, we perform the likelihood ratio test and find

[12]Sponsors in their quest for drug discovery have often conducted trials which are part of a large investigational series exploring various aspects of the new drug/molecule in consideration before filing a new drug application. For diabetes some examples include the BEGIN series of trials conducted by Novo Nordisk, AWARD series by Eli Lilly.

that experience and evc of trials do not significantly effect the success of a clinical trial (see Appendix B Table 11 for details). Our estimation results summarized in Appendix B lend empirical support in favour of the naive counterfactual. We now comment on the quality of innovation and learning that we find in our data.

4.3 Delays in Replicating Novelty or Race to Stay Ahead?

Science policy based on bibliometric techniques to discover novelty in research is likely to suffer from problems of biases against novel research (Wang et al. (2017), Hicks et al. (2015), Butler (2003)). This arises from delays in recognition of research with novel combinations of citations in the empirical literature investigating these combinations using the Web of Science database. However, our clinical trials data on diabetes II over 2010-14 reveals the opposite: a probable race to stay ahead of the innovation curve by trial sponsors. Out of a total of 160 clinical trials from 2010-14, there exist 98 structural holes in our dataset. Within a given year or across years, we find the same pattern of sponsorship for structural hole as well as non-structural hole connections with the same molecule composition in Table 4 panels A and B respectively. For structural hole trials that repeat (in terms of molecule composition) within the same year, we find that they are funded by the same sponsor with the exception of one repeated trial in the year 2013 which has same composition (molecule codes 16 and 50) but different sponsors (Astrazeneca and Intarcia Therapeutics). Evidence of delay for novel research would have shown up in our data as significantly lower repeats of structural hole trials within and across years relative to non-structural holes. There is no significant difference in within and across year transmission of information about molecule composition between these two classes of trials. That large trials (proxied by size of enrollment) have a potential for novelty (structural holes) is corroborated by comparing the average enrollment size in trials with structural holes repeating in the same year against their non-structural hole counterpart or the all-trial average(see panels A and B of Table 4). With the exception of 2012, structural hole trials repeating within a year have larger enrollment. As mentioned earlier, trial size determines trial cost. Hence, we can infer that large pharma companies as sponsors of clinical trials invest in novelty through molecule combinations to presumably stay ahead of competitors.

Delays in recognition of novelty would imply lower repeats both within and across years for structural holes compared to other trials and more self-sponsorship. The remarkable absence of this points to a potential race between sponsors to stay ahead of the research curve, which we test through our second hypothesis in the following section.

Table 4 Year-wise Description of Structural Holes

Panel A: Within and Across Year Repeats for Structural Hole Trials			
Year	No. of Structural Holes with Average Enrollment	Repeating in the same year (Enrollment)	Repeating across years (Year of repetition, Enrollment)
(1)	(2)	(3)	(4)
2010	21 (Average enrollment for column (3) trials: 790; All trial average enrollment: 555)	[AstraZeneca and Bristol-Myers Squibb (1179), AstraZeneca and Bristol-Myers Squibb (400)]	Janssen Research(2010,469)→ Janssen Research (2011, 678); Astellas Pharma Inc(2010,168)→ Astellas Pharma Inc (2011, 171); Merck Sharp(2010, 884)→ Merck Sharp (2012, 337); Boehringer Ingelheim and Eli Lilly (2010, 574)→ Boehringer Ingelheim and Eli Lilly (2012, 876)
2011	21 (Average enrollment for column (3) trials: 924; All trial average enrollment: 678)	[Novo Nordisk A/S (1663), Novo Nordisk A/S (413)]; [Sanofi (807), Sanofi (811)]	AstraZeneca (2011,26)→AstraZeneca and Bristol-Myers Squibb (2012, 32); Eli Lilly and Boehringer Ingelheim(2011, 1516)→Boehringer Ingelheim and Eli Lilly (2012,1413);Eli Lilly and Boehringer Ingelheim(2011,1516)→Eli Lilly (2013,392); Eli Lilly and Boehringer Ingelheim(2011,1516)→ Eli Lilly(2014,68);Eli Lilly and Boehringer Ingelheim(2011,759)→Eli Lilly and Boehringer Ingelheim (2014, 489)
2012	20 (Average enrollment for column (3) trials: 353; All trial average enrollment: 467)	[Sanofi(538), Sanofi (167)]	Merck Sharp(2012, 751)→ Merck Sharp(2013, 402); Hoffmann-La Roche(2012, 200)→ Hoffmann-La Roche(2013,7); Merck Sharp(2012, 307)→ Merck Sharp(2013, 642); AstraZeneca and Bristol-Myers Squibb(2012,32)→ AstraZeneca (2014, 1136)
2013	23 (Average enrollment for column (3) trials: 2343; All trial average enrollment: 730)	[AstraZeneca(365), Intarcia Therapeutics(535)]; [Novo Nordisk A/S(833), Novo Nordisk A/S (7637)]	Novo Nordisk A/S(2013, 7637)→ Novo Nordisk A/S (2014, 721)
2014	13 (Average enrollment for column (3) trials: 953;; All trial average enrollment: 525)	[Sanofi(736), Sanofi(1170)]	-
Panel B: Within and Across Year Repeats for Non-Structural Hole Trials			
Year	No. of Non- Structural Holes with Average Enrollment	Repeating in the same year (Enrollment)	Repeating across years (Year of repetition, Enrollment)
(1)	(2)	(3)	(4)
2010	19 (Average enrollment for column (3) trials: 423; All trial average enrollment: 555)	[Novo Nordisk A/S(465), Novo Nordisk A/S(530)]; [Novo Nordisk A/S(460), Novo Nordisk A/S(435), Novo Nordisk A/S(460), Novo Nordisk A/S(467), Novo Nordisk A/S(143)]	Novo Nordisk A/S (460)→Novo Nordisk A/S (2012, 145)
2011	5 (Average enrollment for column (3) trials: NA; All trial average enrollment: 678)	No repeats	No repeats
2012	18(Average enrollment for column (3) trials: 362; All trial average enrollment: 467)	[Novo Nordisk A/S (394), Novo Nordisk A/S(272)]; [Novo Nordisk A/S (435), Novo Nordisk A/S(346)]	Novo Nordisk (2012, 272)→Novo Nordisk A/S (2013, 274); Novo Nordisk (2012, 272)→Novo Nordisk A/S (2013, 40), Novo Nordisk (2012, 435)→Novo Nordisk A/S (2014, 420)
2013	9 (Average enrollment for column (3) trials: NA; All trial average enrollment: 730)	No repeats	No repeats
2014	9 (Average enrollment for column (3) trials: 878; All trial average enrollment: 525)	[Merck Sharp and Dohme Corp. and Pfizer (464), Merck Sharp and Dohme Corp. and Pfizer (1291)]	-

Source: Authors' own calculations based on trial data from ClinicalTrials.gov

4.4 Strength and Resilience of the Clinical Trial Research Network for Diabetes Type II

We characterize our research network of clinical trials for diabetes type II using the concepts of "strength" and "resilience" that we introduce in **Hypothesis 2**. Table 5 shows an interesting bimodality. Most of our trials are either "Star Performers" or "Incremental Innovation", as we have large positive entries only on the diagonals of Panel A. Therefore, trials either display high-high evc for both trials and molecule variance or they are low-low evc for both trial and molecule variance. Popular trials in our data have an element of new learning possibilities, by combining various molecules with different information content. While this is a good outcome as far as biasing of incentives are concerned, the less popular trials, which are the potential sources of new information have very few structural hole possibilities.

Our data meets the criterion of "strength", as for each year in the data, the number of Star Performers outweigh the number of Low Hanging Fruit trials. However, we find that the number of Incremental Innovation trials outnumber the number of Black Sheep trials consistently over the five year period, leading to absence of "resilience" in the network. The presence of strength and absence of resilience is likely to indicate a race on the part of sponsors to stay ahead of the research curve, as trials sponsored commonly in our data have innovation potential. On the other hand, the presence of Incremental Innovation brings up the possibility that more unique trials are not likely to be innovative.

Based on revenue data collected from statista.com[13], we have summarized trial category-wise breakup for the top ten pharmaceutical companies in terms of their future revenues in 2015-16 in Table 6. The characteristic feature of network "strength" that we observe from 2010-14 continues to mark the industry top players (Star Performers outnumber Low Hanging Fruits). While the "strength" of the network is a uniform feature for these firms[14], the pattern is not uniform for "resilience" of the network. Two of the top ten performers (AstraZeneca and Sanofi) have sponsored an equal number or more of Black Sheep than Incremental Innovation trials. However, since majority of these firms still lack "resilience" in the clinical trial research network, we can place some confidence in our result that the diabetes drug discovery process is incentivized either to take risks in popular trials (race to innovate) or to innovate incrementally through less popular trials.

[13]Gullen and Plungis (2013) review this source as a reliable data source for 21 sectors with a focus on industry information and consumer interests)

[14]These firms combined control 93.20 per cent of the market for diabetes drugs worldwide in 2016 (https://www.statista.com/statistics/309730/top-anti-diabetic-pharmaceutical-companies-by-market-share-worldwide/).

Table 5 Category-wise Summary of Clinical Trials

Panel A: Percentage of Clinical Trials from 2010 (first entry) to 2014 (fifth entry)		
	Variance of Molecule evc	
Evc trial	High	Low
High	47.5, 50.0, 52.6, 50.0, 45.8 (Star Performers)	2.5, 0.0, 0.0, 0.0, 12.5 (Low Hanging Fruits)
Low	5.0, 30.8, 0.0, 21.9, 8.3 (Black Sheep)	45.0, 19.2, 47.4, 28.1, 33.3 (Incremental Innovation)

Panel B: Percentage of Clinical Trials as Star Performers and Incremental Innovation			
Year	Star Performers	Incremental Innovation	Total
2010	47.5	45.0	92.5
2011	50.0	19.2	69.2
2012	52.6	47.4	100.0
2013	50.0	28.1	78.1
2014	37.5	25.0	62.5

Source: Authors' own calculations

5 Discussion

An advantage of using the clinical trial research network as a framework for understanding the quality of innovation in pharmaceuticals is that we can infer measures about innovation using revealed preference of sponsors in trials. This avoids the indirect inference problem that is present in all citometric/bibliometric studies in the innovation literature. Potential biases against novel innovation that Wang et al. (2017) warns about in bibliometric analysis is not present in our investigation, at least for the diabetes clinical trials network. The drawback is that we cannot draw general inferences across diseases using the clinical trial data. One has to proceed disease by disease. Given differential incentives for research in different diseases, we think our approach is appropriate. However, our future intention is to test that claim and to extend the results for other diseases as well.

Table 6 Categorization of Single Sponsor Trials

S.No.	Sponsor name (Revenue in USD million)	Star Performers	Black Sheep	Incremental Innovation	Low Hanging Fruit	Percent of Total Single Sponsor Trials
1	Novo Nordisk (17,786)	10	2	24	0	27.48
2	Eli Lilly (7,963)	9	1	7	0	12.98
3	Sanofi (7,753)	9	2	2	0	9.92
4	Merck and Co. (6,446)	14	1	3	0	13.74
5	Boehringer Ingelheim (5,516)	0	0	0	0	0.00
6	AstraZeneca (3,687)	3	3	2	0	6.11
7	Johnson and Johnson (2,073)	6	0	0	1	5.34
8	Novartis (1,207)	1	0	0	0	0.76
9	Takeda (995)	2	0	2	0	3.05
10	Bayer (635)	0	0	0	0	0.00
Total		54	9	40	1	104

Source: Authors' own calculations together with revenue data taken from www.statista.com

6 Appendix A

Table 7 Codes Assigned to Molecules in Research Network

Molecule	Code	Molecule	Code
Acarbose	1	Ipragliflozin(ASP1941)	34
Albiglutide	2	Linagliptin	35
Aleglitazar	3	Liraglutide	36
Alogliptin	4	Lixisenatide (AVE0010)	37
Anagliptin	5	LMF237 (Vildagliptin and Metformin)	38
Atorvastatin	6	Lobeglitazone (CKD-501)	39
Bexagliflozin	7	Metformin hydrochloride	40
Biguanide	8	MP513/Teneligliptin	41
Biphasic insulin aspart	9	Nateglinide	42
Canagliflozin/TA-7284	10	Omarigliptin	43
CJ-30001/CJ-30002(Metformin/voglibose)	11	Pioglitazone	44
Dapagliflozin	12	Prandial insulin	45
Dulaglutide(LY2189265)	13	Ranolazine	46
Ertugliflozin	14	Rosuvastatin	47
Evogliptin	15	Saxagliptin	48
Exenatide	16	Semaglutide	49
Fasiglifam	17	Sitagliptin	50
Gemigliptin(LS15-0444)	18	Sitagliptin phosphate	51
Glibenclamide	19	SPIL1033	52
Gliclazide	20	Sulphonylruea	53
Glimepiride	21	Taspoglutide	54
Glinide	22	Teneligliptin	55
Glipizide	23	Thiazolidinedione/TZD	56
Glucagon	24	Trelagliptin (SYR-472)	57
Gosogliptin	25	Vildagliptin	58
Insulin	26	Metformin	59
Insulin aspart	27	Empagliflozin	60
Insulin degludec	28	Simvastatin	61
Insulin detemir	29	Insulin glulisine	62
Insulin glargine(Lantus-HOE901-U100)	30	HOE901-U300-Insulin glargine new (Sanofi)	63
Insulin Lispro	31	Mylan's Insulin glargine	64
Insulin NPH	32	Insulin glargine (Eli Lilly and Boehringer)	65
Insulin Peglispro/LY2605541	33	MK-1293- Insulin Glargine(Merck)	66

Source: Authors' code assignment based on molecule data from ClinicalTrials.gov

Table 8 Evc Scores and Classification of Structural Holes for 2014

Trial No	Trial evc	Trial	Molecule evc	Molecule	Variance	Structural Hole	Trial Category
1	0.086	high	0.015	high	0.004	**Yes**	Star Performer
2	0.042	high	0.000	low	0.017	**Yes**	Star Performer
3	0.059	high	0.000	low	0.008	**Yes**	Star Performer
4	0.086	high	0.008	low	0.004	**Yes**	Star Performer
5	0.000	low	0.025	high	0.000	**No**	Incremental Innovation
6	0.100	high	0.082	high	0.001	**No**	Low Hanging Fruit
7	0.100	high	0.013	high	0.001	**No**	Low Hanging Fruit
8	0.004	low	0.082	high	0.000	**No**	Incremental Innovation
9	0.077	high	0.084	high	0.009	**Yes**	Star Performer
10	0.042	high	0.001	low	0.018	**Yes**	Star Performer
11	0.003	low	0.016	high	0.000	**No**	Incremental Innovation
12	0.037	low	0.000	low	0.011	**Yes**	Black Sheep
13	0.000	low	0.201	high	0.000	**No**	Incremental Innovation
14	0.000	low	0.015	high	0.000	**No**	Incremental Innovation
15	0.042	high	0.000	low	0.018	**Yes**	Star Performer
16	0.043	high	0.014	high	0.016	**Yes**	Star Performer
17	0.024	low	0.002	low	0.002	**No**	Incremental Innovation
18	0.050	high	0.074	high	0.000	**No**	Low Hanging Fruit
19	0.023	low	0.000	low	0.003	**No**	Incremental Innovation
20	0.053	high	0.001	low	0.008	**Yes**	Star Performer
21	0.045	high	0.020	high	0.012	**Yes**	Star Performer
22	0.004	low	0.022	high	0.000	**No**	Incremental Innovation
23	0.042	high	0.097	high	0.018	**Yes**	Star Performer
24	0.038	low	0.000	low	0.008	**Yes**	Black Sheep
-	-	-	0.012	low	-	-	-
-	-	-	0.171	high	-	-	-
-	-	-	0.014	high	-	-	-
-	-	-	0.014	high	-	-	-
-	-	-	0.019	high	-	-	-

Source: Authors' own calculations based on data from ClinicalTrials.gov

7 Appendix B

Table 9 Multinomial Logistic Regression with Success Categories as Dependent Variable

Variables	Coefficient	P-value
Success Category 1: Trials Leading to Drug Discovery		
Trial evc	5.961	0.485
Experience	-0.0264	0.639
Enrollment	0.531	0.025
Series dummy	1.941	0.000
Year dummy		
2	- 1.557	0.054
3	- 0.188	0.809
4	- 1.444	0.048
5	- 0.596	0.460
Success Category 2: Trials Not Linked to Drug Discovery		
Trial evc	- 4.099	0.644
Experience	- 0.045	0.424
Enrollment	0.405	0.042
Series dummy	- 0.756	0.188
Year dummy		
2	-1.340	0.067
3	-0.007	0.992
4	-1.175	0.093
5	-1.077	0.197
Log likelihood = -137.198		
Number of observations = 160		
LR $\chi^2(16) = 68.77$		
Prob $> \chi^2 = 0.000$		
Pseudo $R^2 = 0.2004$		

Source: Authors' own calculations with Success Category 3 as the base category.

Table 10 Suest-based Hausman tests of IIA assumption (N=160)

	χ^2	df	P> χ^2
1	4.277	6	0.639
2	3.538	6	0.739
3	2.584	6	0.859
Ho: Odds are independent of other alternatives			

Source: Authors' own calculations

Table 11 Likelihood Ratio Tests for Independent Variables (N=160)

	χ^2	df	P > χ^2
Trial evc	1.396	2	0.497
Experience	0.645	2	0.724
Enrollment	6.735	2	0.034
Series Dummy	38.898	2	0.000
Year Dummy			
2	4.868	2	0.088
3	0.103	2	0.950
4	4.636	2	0.098
5	1.699	2	0.428

Source: Authors' own calculations

Acknowledgements and Declaration

A version of this paper was presented at The Winter School Conference held at Delhi School of Economics (December 2017) and the Research Scholar Workshop (Centenary Celebrations), (February 2018), Department of Economics, Calcutta University. We are grateful for the comments and suggestions made by Prof. Sunil Kanwar (Faculty of Economics, Delhi School of Economics) and Dr. Gagari Chakraborti (Faculty of Economics, Presidency University). Further, we would like to acknowledge comments and guidance from Dr. Deepa Sinha (Faculty of Mathematics and Computer Sciences, South Asian University), Dr. Rinkoo Devi Gupta (Faculty of Life Sciences and Biotechnology, South Asian University) and Prof. Syed Ahsan (Faculty of Economics, South Asian University). We declare that we have not received any funding from external organization for our research.

References

Ahuja, G. (2000): "Collaboration Networks, Structural Holes, and Innovation: A Longitudinal Study," *Administrative Science Quarterly*, 45, 425–455.

Alcer, J. and M. Gittelman (2006): "Patent Citations as a Measure of Knowledge Flows: The Influence of Examiner Citations," *The Review of Economics and Statistics*, 88, 774–779.

Allred, B. B. and W. G. Park (2007): "Patent Rights and Innovative Activity: Evidence from National and Firm-level Data," *Journal of International Business Studies*, 38, 878–900.

Archibugi, D. and M. Planta (1996): "Measuring Technological Change through Patents and Innovation Surveys," *Technovation*, 16, 451 – 519.

Arthur, W. B. (2009): *The Nature of Technology: What it is and How it Evolves*, Simon and Schuster.

Bonacich, P., A. C. Holdren, and M. Johnston (2004): "Hyper-edges and Multidimensional Centrality," *Social Networks*, 26, 189–203.

Bonacich, P., A. Oliver, and T. A. Snijders (1998): "Controlling for Size in Centrality Scores," *Social Networks*, 20, 135–141.

Bouet, D. (2015): "A Study of Intellectual Property Protection Policies and Innovation in the Indian Pharmaceutical Industry and Beyond," *Technovation*, 38, 31–41.

B.R., H., C. R.M., C. S.K., and et al (2013): "Characteristics of Oncology Clinical Trials: Insights from a Systematic Analysis of Clinicaltrials.gov," *JAMA Internal Medicine*, 173, 972–979.

Budish, E., B. N. Roin, and H. Williams (2015): "Do Firms Underinvest in Long-term Research? Evidence from Cancer Clinical Trials," *American Economic Review*, 105, 2044–85.

Burt, R. S. (2004): "Structural Holes and Good Ideas," *American Journal of Sociology*, 110, 349–399.

Busseniers, E. (2014): "General Centrality in a Hypergraph," *CoRR:Computing Research Repository*, abs/1403.5162.

Butler, L. (2003): "Explaining Australias Increased Share of ISI PublicationsThe Effects of a Funding Formula based on Publication Counts," *Research Policy*, 32, 143 – 155.

Califf, R. M., D. A. Zarin, J. M. Kramer, R. E. Sherman, L. H. Aberle, and A. Tasneem (2012): "Characteristics of Clinical Trials Registered in ClinicalTrials.gov, 2007-2010," *Jama*, 307, 1838–1847.

Chen, C., Y. Chen, M. Horowitz, H. Hou, Z. Liu, and D. Pellegrino (2009): "Towards an Explanatory and Computational Theory of Scientific Discovery," *Journal of Informetrics*, 3, 191–209.

Cockburn, I. M. and R. M. Henderson (2001): "Scale and Scope in Drug Development: Unpacking the Advantages of Size in Pharmaceutical Research," *Journal of Health Economics*, 20, 1033–1057.

Cohen, F. J. (2005): "Macro Trends in Pharmaceutical Innovation," *Nature Reviews Drug Discovery*, 4, 78–84.

Danzon, P. M., S. Nicholson, and N. S. Pereira (2005): "Productivity in Pharmaceutical- Biotechnology R & D- The Role of Experience and Alliances," *Journal of Health Economics*, 24, 317 – 339.

Dernis, H. and M. Khan (2004): "Triadic Patent Families Methodology," Working Paper 2004/02.

Fleming, L. (2001): "Recombinant Uncertainty in Technological Search," *Management Science*, 47, 117–132.

Foster, J. G., A. Rzhetsky, and J. A. Evans (2015): "Tradition and Innovation in Scientists Research Strategies," *American Sociological Review*, 80, 875–908.

Friedman, L. M., C. Furberg, D. L. DeMets, D. M. Reboussin, and C. B. Granger (1998): *Fundamentals of Clinical Trials*, Springer, vol. 3.

Gambardella, A., D. Harhoff, and B. Verspagen (2017): "The Economic Value of Patent Portfolios," *Journal of Economics & Management Strategy*, 26, 735–756.

Grabowski, H. (2002): "Patents, Innovation and Access to New Pharmaceuticals," *Journal of International Economic Law*, 5, 849–860.

Gullen, A. and J. Plungis (2013): "Statista," *The Charleston Advisor*, 15, 43–47.

Harhoff, D., F. Narin, F. M. Scherer, and K. Vopel (1999): "Citation Frequency and the Value of Patented Inventions," *The Review of Economics and Statistics*, 81, 511–515.

Henderson, R. and I. Cockburn (1996): "Scale, Scope, and Spillovers: The Determinants of Research Productivity in Drug Discovery," *The RAND Journal of Economics*, 27, 32–59.

Hicks, D., P. Wouters, L. Waltman, S. De Rijcke, and I. Rafols (2015): "The Leiden Manifesto for Research Metrics," *Nature*, 520, 429.

Hollis, A. (2004): "An Efficient Reward System for Pharmaceutical Innovation," *Unpublished Manuscript, Department of Economics, University of Calgary.*

Inrig, J. K., R. M. Califf, A. Tasneem, R. K. Vegunta, C. Molina, J. W. Stanifer, K. Chiswell, and U. D. Patel (2014): "The Landscape of Clinical Trials in Nephrology: A Systematic Review of Clinicaltrials. gov," *American Journal of Kidney Diseases*, 63, 771–780.

Jeong, S., M. Sohn, J. H. Kim, M. Ko, H.-w. Seo, Y.-K. Song, B. Choi, N. Han, H.-S. Na, J. G. Lee, et al. (2017): "Current Globalization of Drug Interventional Clinical Trials: Characteristics and Associated Factors, 2011–2013," *Trials*, 18, 288.

Klavans, R. and K. Boyack (2013): "Towards the Development of an Article-level Indicator of Conformity, Innovation and Deviation," in *18th International Conference on Science and Technology Indicators*.

Kleinknecht, A. (1996): *New Indicators and Determinants of Innovation: An Introduction*, Palgrave Macmillan, 1–11.

Lichtenberg, F. R. (1998): "Pharmaceutical Innovation, Mortality Reduction, and Economic Growth," Working Paper 6569, National Bureau of Economic Research.

Mansfield, E. (1986): "Patents and Innovation: An Empirical Study," *Management Science*, 32, 173–181.

Mansfield, E., M. Schwartz, and S. Wagner (1981): "Imitation Costs and Patents: An Empirical Study," *Economic Journal*, 91, 907–18.

Mestre-Ferrandiz, J., J. Sussex, and A. Towse (2012): *The R & D Cost of a New Medicine*, Office of Health Economics, London.

Morgan, S. and D. Lopert, R.and Greyson (2008): "Toward a Definition of Pharmaceutical Innovation," *Open Medicine*, 2, 4–7.

Narin, F. and D. Olivastro (1988): "Chapter 15 - Technology Indicators based on Patents and Patent Citations," in *Handbook of Quantitative Studies of Science and Technology*, ed. by A. V. Raan, Amsterdam: Elsevier, 465 – 507.

Nelson, R. R. W. and G. Sidney (1982): *An Evolutionary Theory of Economic Change*, Harvard University. Cambridge. US.

Newman, M. (2010): *Networks: An Introduction*, Oxford University Press, Oxford,UK.

Nicholson, S., P. M. Danzon, and J. S. McCullough (2002): "Biotech-Pharmaceutical Alliances as a Signal of Asset and Firm Quality," Working Paper 9007, National Bureau of Economic Research.

Pakes, A., M. Simpson, K. Judd, and E. Mansfield (1989): "Patent Renewal Data," *Brookings Papers on Economic Activity. Microeconomics*, 1989, 331–410.

Papageorgiadis, N. and A. Sharma (2016): "Intellectual Property Rights and Innovation: A Panel Analysis," *Economics Letters*, 141, 70 – 72.

Pasquali, S. K., W. K. Lam, K. Chiswell, A. R. Kemper, and J. S. Li (2012): "Status of the Pediatric Clinical Trials Enterprise: An Analysis of the US ClinicalTrials.gov Registry," *Pediatrics*, 130, e1269–e1277.

Pocock, S. J. (1983): *Introduction: The Rationale of Clinical Trials*, John Wiley & Sons Ltd, 1–13.

Powell, W. W., K. W. Koput, and L. Smith-Doerr (1996): "Interorganizational Collaboration and the Locus of Innovation: Networks of Learning in Biotechnology," *Administrative Science Quarterly*, 41, 116–145.

Raider, H. J. (1998): "Market Structure and Innovation," *Social Science Research*, 27, 1–21.

Ross, J. S., G. K. Mulvey, E. M. Hines, S. E. Nissen, and H. M. Krumholz (2009): "Trial Publication after Registration in ClinicalTrials.Gov: A Cross-Sectional Analysis," *PLOS Medicine*, 6, 1–9.

Rzhetsky, A., J. G. Foster, I. T. Foster, and J. A. Evans (2015): "Choosing Experiments to Accelerate Collective Discovery," *Proceedings of the National Academy of Sciences*, 112, 14569–14574.

Saito, H. and C. J. Gill (2014): "How Frequently Do the Results from Completed US Clinical Trials Enter the Public Domain? - A Statistical Analysis of the ClinicalTrials.gov Database," *PLOS ONE*, 9, 1–9.

Schumpeter, J. A. (1939): *Business Cycles: A Theoretical, Historical, and Statistical Analysis of the Capitalist Process*, McGraw-Hill.

Scotchmer, S. (1991): "Standing on the Shoulders of Giants: Cumulative Research and the Patent Law," *Journal of Economic Perspectives*, 5, 29–41.

Shi, F., J. G. Foster, and J. A. Evans (2015): "Weaving the Fabric of Science: Dynamic Network Models of Science's Unfolding Structure," *Social Networks*, 43, 73–85.

Stuart, T. E., S. Z. Ozdemir, and W. W. Ding (2007): "Vertical Alliance Networks: The Case of University-Biotechnology-Pharmaceutical Alliance Chains," *Research Policy*, 36, 477–498.

Taichman, D. B., J. Backus, C. Baethge, H. Bauchner, P. W. de Leeuw, J. M. Drazen, J. Fletcher, F. A. Frizelle, T. Groves, A. Haileamlak, A. James, C. Laine, L. Peiperl, A. Pinborg, P. Sahni, and S. Wu (2016): "Sharing Clinical Trial Data A Proposal from the International Committee of Medical Journal Editors," *New England Journal of Medicine*, 374, 384–386.

Tong, X. and J. Frame (1994): "Measuring National Technological Performance with Patent Claims Data," *Research Policy*, 23, 133 – 141.

Trajtenberg, M. (1990): "A Penny for Your Quotes: Patent Citations and the Value of Innovations," *The RAND Journal of Economics*, 21, 172–187.

Tyagi, S., V. Mahajan, and D. K. Nauriyal (2014): "Innovations in Indian Drug and Pharmaceutical Industry: Have they Impacted Exports?" *Journal of Intellectual Property Rights*, 19, 243–252.

Wang, J., R. Veugelers, and P. Stephan (2017): "Bias Against Novelty in Science: A Cautionary Tale for Users of Bibliometric Indicators," *Research Policy*, 46, 1416–1436.

Wardell, W. M. and J. Diraddo (1980): "The Measurement of Pharmaceutical Innovation," *The Journal of Clinical Pharmacology*, 20, 1–9.

Zarin, D. A., T. Tse, R. J. Williams, R. M. Califf, and N. C. Ide (2011): "The ClinicalTrials.gov Results Database - Update and Key Issues," *New England Journal of Medicine*, 364, 852–860.