

Predicting New Target Conditions for Drug Retesting Using Temporal Patterns in Clinical Trials: A Proof of Concept

Zhe He, PhD and Chunhua Weng, PhD

Department of Biomedical Informatics, Columbia University, New York, NY

Abstract

Drug discovery is costly and time-consuming. Efficient drug repurposing promises to accelerate drug discovery with reduced cost. However, most successful repurposing cases so far have been achieved by serendipity. There is a need for more efficient computational methods for predicting new indications for existing drugs. This paper conducts a retrospective analysis of the temporal patterns of drug intervention trials for every drug in a pair of different conditions in ClinicalTrials.gov, including 550 drugs used for 451 conditions between 2003 and 2013. We found that drugs are often targeted towards conditions that are related by similar or identical eligibility criteria. We demonstrated the preliminary feasibility of predicting new target conditions for drug retesting among conditions with similar aggregated clinical trial eligibility criteria and confirmed this hypothesis using evidence from the literature.

Introduction

Drug discovery is expensive. It is estimated that it takes up to 17 years and over \$800 millions to develop a new drug¹. Failures during development often cost a fortune for research sponsors. To accelerate drug discovery while reducing costs, methods have been sought for efficient discovery of novel indications for existing drugs on the market². This process, known as drug repurposing, repositioning, or re-profiling, promises to accelerate drug discovery due to known safety issues and reduced risk of failure^{3,4}. Some drugs have been successfully repurposed. Duloxetine was initially designed to treat *depression* but later successfully repurposed by Eli Lilly to treat *stress urinary incontinence* for women⁵. However, such discoveries have been primarily driven by insights or serendipitous observations⁶. It is not until recently that computational methods have been proposed to predict new indications for existing drugs using networks analysis of genetic, proteomic, and metabolic data⁷.

To date, ClinicalTrials.gov has archived more than 170,000 trials and is a valuable resource for studying clinical trial design patterns. There is a saying: “the best predictor of future behavior is past behavior.” Previously, the clinical evidence in ClinicalTrials.gov was used to verify drug repurposing targets predicted by a similarity-based computational framework⁸. In this work, we analyzed the drug retesting patterns in drug intervention trials from 2003 to 2013 with a focus on drugs that were used in every pair of different conditions over time. Trial summaries contain structured metadata such as start date, intervention(s), and free-text eligibility criteria for patient selection. This study explored the feasibility of leveraging these metadata in drug intervention trials to identify temporal patterns of drug retesting and to narrow the search for drug repurposing targets.

Methods

Step 1: Dataset Preparation

We identified 59,716 drug intervention trials between 2003 and 2013 covering 1,487 conditions in ClinicalTrials.gov. Then we leveraged a previously developed a database called COMPACT (Commonalities in Target Populations of Clinical Trials)⁹ to retrieve the information for these trials. For each trial, COMPACT contains structured trial descriptors and discrete common eligibility features (CEFs) (e.g., BMI, and HbA1c) associated with the condition that the trial investigated. The CEFs were present in the eligibility criteria section for at least 3% of all the trials that investigated the same condition¹⁰. We extracted the drug names from the structured “intervention” field in the XML format summary of each trial, which may use one or more drugs as the intervention. We included all the drugs that each was an intervention for at least five trials for the same condition in one year, within the time window being years 2003-2013. We empirically chose “five” as the threshold because most generic drugs were retained at this threshold after filtering out drug names that contained a mixture of brand names and dosage. We formulated each retesting case as a quintuple (*drug, initial condition, first year tested for initial condition, retested condition, first year tested for retested condition*), and stored the quintuples in COMPACT. The aforementioned Duloxetine example can be represented as (Duloxetine, depression, 1990s, stress urinary incontinence (SUI) for women, 2004). We excluded drug “Placebo” because it has no therapeutic effect and is often used as a control when testing new drugs¹¹.

Step 2: Analysis of drug retesting temporal patterns

For the ten-year time window, we constructed a 10 x 10 matrix, with row i and column j being each year during the time window and each cell containing two values, i.e., d_{ij} and c_{ij} . Variable d_{ij} represents the number of distinct drugs that were first studied for one condition in year i and later for a different condition in year j , while c_{ij} represents the number of distinct pairs of conditions in which a drug was tested for one condition in year i and later for a different condition in year j . With this matrix, we analyzed the temporal trends in drug retesting cases for each drug and each condition, respectively. The network of drug retesting patterns was visualized using Cytoscape¹². The heat map with the temporal patterns for the most retested drugs was visualized in MATLAB.

Step 3: Analysis of condition relatedness by their shared CEFs

We hypothesized that drugs were often retested among conditions whose trials employed similar eligibility criteria, i.e., using similar variables for patient selection. We assessed the similarities for each condition pair that involved drug retesting by analyzing their shared CEFs regardless of inclusion and exclusion status of these CEFs because there is no standardized way of writing free-text eligible criteria, e.g., “HbA1c > 7.0%” in the inclusion criteria is equivalent to “HbA1c <= 7.0%” in the exclusion criteria. There might be multiple retested drugs for each pair of conditions. We aggregated the retested drugs that were investigated with the same pair of conditions and analyzed the distribution of the number of condition pairs over counts of retested drugs.

Results

Among 59,716 drug intervention trials conducted between 2003 and 2013 that used one or more drug interventions, 40,167 drugs were used for 1,487 conditions. To analyze the drug retesting patterns, we considered only those drugs that were tested for a condition in at least five trials. Hence, we reduced the total number of drugs and conditions to 550 and 451, respectively. The total number of drug-condition pairs was 4,351. The number of drugs per condition varied, ranging from 93 drugs for *gastrointestinal diseases* and *digestive system diseases* to one drug (Fludarabine) for *skin neoplasms*. There were 118 other medical conditions with only one drug.

Analysis of drug retesting cases

Out of all 202,950 (451x450) plausible condition pairs, only 12,774 (6.3%) pairs included two different conditions, each testing the same drug in at least five trials in two different years between 2003 and 2013.

Figure 1 visualizes the drug retesting networks for two example conditions, i.e., *asthma* and *hypertension*. For example, *asthma* was the retested condition for four different drugs (i.e., GW685698X, Ciclesonide, Omalizumab, and Budesonide)

that were previously tested for seven other conditions. *Hypertension* was the retested condition for three drugs (i.e., Tadalafil, Sildenafil, and Amlodipine) that were previous tested for five other conditions (i.e., *mental disorders*, *vascular diseases*, *prostatic diseases*, *psychotic disorders*, and *erectile dysfunction*). A node indicates a condition, while an arrow represents a drug. The arrow ends and arrowheads are initial and retested conditions, respectively.

Table 1 shows the temporal matrix of drug retesting cases between 2003 and 2013. Year 1 is when a drug is initially tested for a condition and year 2 is when the same drug is later retested for a different condition. Because year 1 < year 2, the pairs of years in the lower half of the table are not applicable. The numbers of drugs are consistently smaller than the numbers of conditions pairs, indicating that a drug may have been used for more than one condition pairs. More retesting cases occurred between 2003 and 2004 than other pairs of years. Looking at one row at a time, we can see that as the time window widens, the counts of retested drugs and condition pairs decrease.

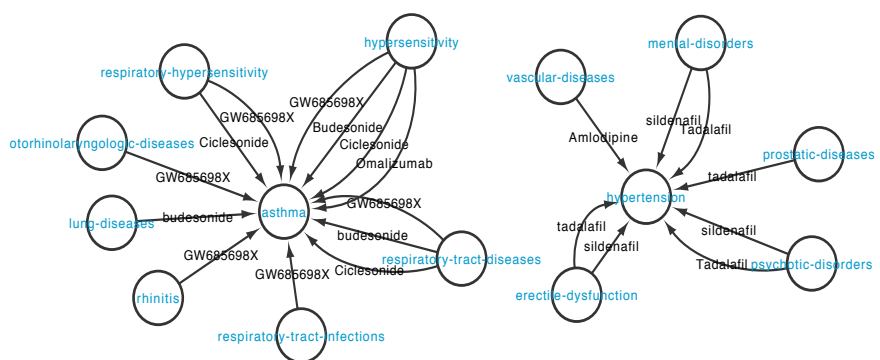


Figure 1. Network visualization of drug retesting patterns for *asthma* and *hypertension*; each arrow represents a transition from a prior drug indication to new drug indication and is labeled with the name of a retested drug.

Table 1. Pairwise temporal analysis of drug retesting cases (count of retested drugs / count of condition pairs).

Yr 2 \ Yr 1	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
2003	46/2982	34/2278	26/1212	18/1035	22/864	13/560	9/221	9/284	9/155	11/251
2004	--	39/1276	31/787	24/491	21/516	13/333	9/236	5/47	3/20	10/95
2005	--	--	31/821	30/554	18/180	15/471	9/231	8/95	3/11	8/67
2006	--	--	--	24/454	20/256	15/435	14/292	11/108	7/61	7/57
2007	--	--	--	--	19/333	17/218	14/179	10/129	4/82	7/28
2008	--	--	--	--	--	22/183	16/152	8/61	3/17	5/20
2009	--	--	--	--	--	--	13/385	13/91	4/24	5/33
2010	--	--	--	--	--	--	--	13/144	5/50	4/11
2011	--	--	--	--	--	--	--	--	13/143	6/86
2012	--	--	--	--	--	--	--	--	--	7/80

Figure 2 displays the count of different conditions that a drug was retested on each year for the top 20 drugs that were retested on most conditions between 2004 and 2013. Each color block represents the number of different conditions that the drug was retested on compared to the previous year(s). The most retested drug (i.e., Bevacizumab) resides at the bottom of the figure. Note that since our time window is from 2003 to 2013, the first year that a drug could be retested for another condition is 2004. Most retested drugs were used in chemotherapeutic activities. One reason could be that chemotherapy usually uses multiple drugs to kill or control tumor cells. Meanwhile, chemotherapy drugs are often used to treat different types of neoplasms and cancers.

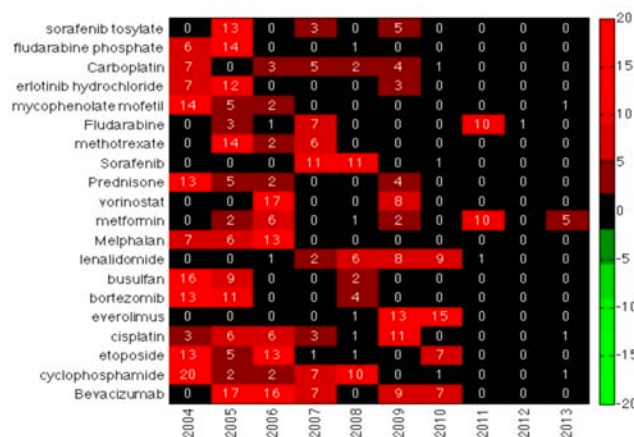


Figure 2. The numbers of different conditions that the top 20 most retested drugs were retested on.

Table 2 shows the most frequent initial conditions and retested conditions, respectively. The second column gives the number of condition pairs in which the initial condition is specified in the first column. The third column shows the number of drugs that were tested for the initial condition specified in the first column and later retested for a different condition. The fifth column gives the number of condition pairs in which the retested condition is specified in the fourth column. The sixth column shows the number of drugs that were previously tested for some other conditions and later retested for the condition specified in fourth column.

Table 2. Most frequent initial and retested conditions in the existing drug retesting cases

The top five frequent initial conditions	No. of condition pairs	No. of retested drugs	The top five frequent retested conditions	No. of condition pairs	No. of retested drugs
Respiratory tract diseases	173	35	Skin diseases	140	14
Carcinoma	167	46	Digestive system diseases	133	30
Vascular disease	167	30	Gastrointestinal diseases	133	30
Immunoproliferative disorders	164	39	Urologic diseases	124	10
Lymphoproliferative disorders	164	39	Neoplasm metastasis	117	19

Figure 3 illustrates the number of condition pairs and the average number of shared CEFs for pairs of conditions with the same number of retested drugs. The x-axis shows the number of drugs used for a condition pair. The left y-axis shows the number of condition pairs. The right y-axis shows the average number of shared CEFs between two conditions in a pair. On average, each condition has 172 CEFs. The average number of CEFs shared by any two conditions is 52, whereas the average number of CEFs shared by condition pairs involving drug retesting is 139. 64.6% of these condition pairs have 100-200 shared CEFs, while only 2.9% condition pairs have fewer than 50 shared CEFs, indicating that drug retesting often occurred between conditions with a large number of shared CEFs. Previously, Boland et al. used CEFs shared among diseases to identify disease relatedness¹⁰. The average number of

shared CEFs increases with the number of retested drugs, which indicates that conditions with more shared CEFs, implying the research on these two conditions tend to use similar criteria for patient recruitment, are more likely to use the same drug as an intervention on these conditions. For example, 15 drugs (e.g., Bendamustine, Bortezomib, brentuximab vedotin) that were tested for *lymphoproliferative disorders* were later retested for *leukemia*. *Lymphoproliferative disorders* and *leukemia* share 199 CEFs (e.g., electrocorticogram, alanine transaminase, creatinine clearance). However, some successful repurposed drugs also occurred in non-similar diseases. For example, metformin was initially tested for *diabetes mellitus* and later tested for treating *breast neoplasm*. These two conditions shared only 61 CEFs.

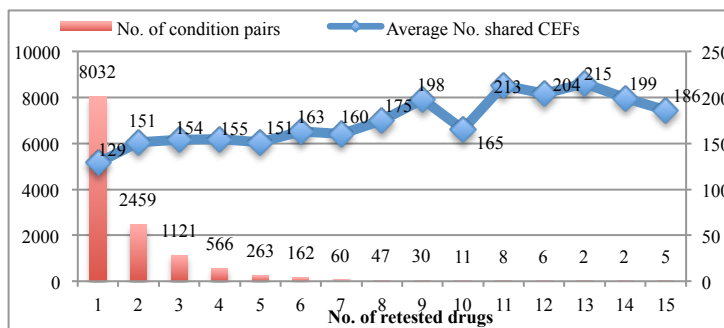


Figure 3. Number of condition pairs and average number of shared CEFs for pairs of conditions over counts of retested drugs.

The basis for predicting drugs for different conditions

Leveraging the observed drug retesting patterns, we designed a basis for predicting drugs for retesting on different conditions given a threshold value for the minimum number of shared CEFs between the initial condition and the possible different condition. Each prediction consists of a drug and a possible different condition. A prediction was made if (1) a drug has been tested for the initial condition but has never been tested for the possible different condition, (2) there exists another drug that has been tested for both conditions, and (3) the number of shared CEFs between two conditions is above a threshold. **Figure 4** shows the number of drug predicted and the number of different conditions for threshold values between 20 and 200. Higher thresholds yielded fewer predictions, which may also be more clinically relevant. The number of drugs is consistently greater than the number of different conditions, showing that a drug may be predicted for multiple conditions.

A few predictions have been confirmed by evidence from the literature. For example, “Ranolazine” was predicted as a drug to be retested for *myocardial infarction*, because (1) “Ranolazine” was tested for *ischemia* but has never been tested for *myocardial infarction*, (2) there exists another drug (i.e., Ticagrelor) that was tested for *ischemia* first and then retested for *myocardial infarction*, (3) *ischemia* and *myocardial infarction* had 112 shared CEFs. This prediction was confirmed by Hale et al.¹³ Similarly, a paper by Yoon et al.¹⁴ confirmed our prediction of “Everolimus” for treating *rheumatic disease*.

Discussion

Based on the observed drug retesting patterns, we provided a proof-of-concept of a method for predicting a set of drugs for retesting on different conditions, which can serve as a basis for developing more sophisticated methods. In this study, we analyzed the longitudinal drug retesting patterns of clinical trials between 2003 and 2013 in ClinicalTrials.gov. A trial may study comorbidities, which was taken into account so that a drug used in a trial studying multiple conditions was not considered as drug retesting for these conditions. Drug retesting often occurred among various types of cancer and neoplasms, many of which share a large number of CEFs. Some interesting drug retesting cases were found among quite different conditions. For example, Letrozole, which was a drug developed for *breast neoplasm*, was later used to treat *infertility*.

Our analysis has several major limitations. Since the drug indication predictions were made based on retrospective trials, this approach does not work for new conditions and drugs. We considered only drugs that were used in at least five trials in a year. Future research is warranted to test other threshold values after normalizing drug names that are mixtures of brand names, trade names, and dosage. We limited our analysis between 2003 and 2013. Thus, Sildenafil, which was a drug tested for treating *hypertension* before 2003 and later tested for treating *erectile*

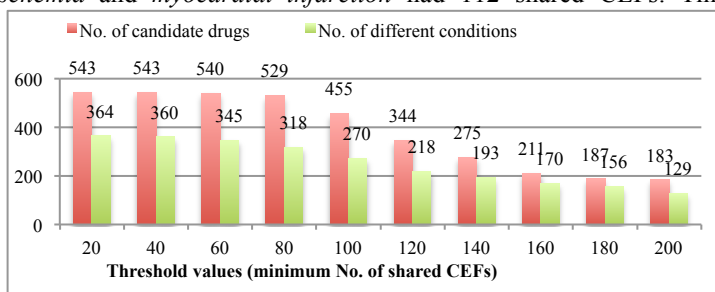


Figure 4. Number of drugs and number of retested conditions predicted for various thresholds.

dysfunction, was conversely deemed to be a drug originally tested for *erectile dysfunction* and later retested for *hypertension*. Even though it is unlikely that such cases prevail in our analysis, one should interpret “initial condition” in the context of our dataset. Another limitation is that our similarity analysis for conditions was at the concept-level using n-grams; ideally a more sophisticated similarity analysis should be done at the rule level so that we could use more complete meaning such as “myocardial infarction within the last five years” to represent a common eligibility feature. We will refine our analysis for inclusion criteria and exclusion criteria after eligibility features are enriched by contextual information, e.g., negation. A third limitation is the data quality issues in ClinicalTrials.gov. Moreover, the “intervention” field for every clinical trial does not specify which drug is primarily tested if multiple drugs are used in a trial. In this work, we removed the control “Placebo” from our analysis but all other drugs listed as intervention for a trial were included in our analysis. Automated techniques are desired to rank the importance of drugs within a trial to produce more precise analysis. The conditions assigned to each trial may not be normalized and hence may introduce condition-indexing errors.

In this study, we only analyzed the drug retesting patterns between pairs of conditions. In the future, it would be interesting to analyze drug-retesting path linking multiple conditions over time, which may reveal more interesting patterns. To enrich the method for predicting drugs for retesting on different conditions with domain knowledge, we can use ontologies such as SNOMED CT to quantify the similarities between the conditions in terms of distance in the topological structure. We can leverage the drug-target interaction data from DrugBank and adverse events data from openFDA to filter out clinically irrelevant predictions for further verification. With a knowledge-enriched method, we may identify more meaningful drugs for retesting on different conditions. Finally, future work should formally quantify the PPV, NPV, Sensitivity, and Specificity of using this method for drug repurposing.

Conclusions

Drug retesting has often occurred between conditions whose trials used similar eligibility criteria for participant selection; therefore, we could predict target conditions for drug retesting based on condition similarities in eligibility criteria. In contrast to existing approaches to drug repurposing based on compound or agents knowledge, our method leverage the design patterns in drug intervention trials to identify potential new conditions for drug retesting. This study only provides very preliminary proof of concept; more sophisticated models should be developed to further test this idea.

Acknowledgments

This study is sponsored by National Library of Medicine Grant **R01LM009886** (PI: Weng) and National Center for Advancing Translational Sciences (NCATS) **U01 TR000040** (PI: Ginsberg).

References

1. Reichert JM. Trends in development and approval times for new therapeutics in the United States. *Nat Rev Drug Discov.* Sep 2003;2(9):695-702.
2. Colonizing therapeutic space: the overlooked science of drug husbandry. *Nat Rev Drug Discov.* Feb 2004;3(2):101.
3. Bastos LF, Coelho MM. Drug repositioning: playing dirty to kill pain. *CNS drugs.* Jan 2014;28(1):45-61.
4. Corbett A, Williams G, Ballard C. Drug repositioning: an opportunity to develop novel treatments for Alzheimer's disease. *Pharmaceuticals.* 2013;6(10):1304-1321.
5. Li J, Yang L, Pu C, Tang Y, Yun H, Han P. The role of duloxetine in stress urinary incontinence: a systematic review and meta-analysis. *Int Urol Nephrol.* Jun 2013;45(3):679-686.
6. Ashburn TT, Thor KB. Drug repositioning: identifying and developing new uses for existing drugs. *Nature Rev Drug Discov.* Aug 2004;3(8):673-683.
7. Hopkins AL. Network pharmacology: the next paradigm in drug discovery. *Nat Chem Biol.* Nov 2008;4(11):682-690.
8. Zhang P, Wang F, Hu J. Towards Drug Repositioning: A Unified Computational Framework for Integrating Multiple Aspects of Drug Similarity and Disease Similarity. *AMIA Annu Symp Proc.* 2014;In press.
9. He Z, Carini S, Hao T, Sim I, Weng C. A Method for Analyzing Commonalities in Clinical Trial Target Populations. *AMIA Annu Symp Proc.* 2014;In press.
10. Boland MR, Miotto R, Weng C. A method for probing disease relatedness using common clinical eligibility criteria. *Stud Health Technol Inform.* 2013;192:481-485.
11. What Is the Placebo Effect? <http://www.webmd.com/pain-management/what-is-the-placebo-effect>. Accessed Sep 9, 2014.
12. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* Nov 2003;13(11):2498-2504.
13. Hale SL, Kloner RA. Ranolazine treatment for myocardial infarction? Effects on the development of necrosis, left ventricular function and arrhythmias in experimental models. *Cardiovasc Drugs Ther.* Oct 2014;28(5):469-475.
14. Yoon KH. Proliferation signal inhibitors for the treatment of refractory autoimmune rheumatic diseases: a new therapeutic option. *Ann N Y Acad Sci.* Sep 2009;1173:752-756.