Translational and Clinical Pharmacology
TCP

Review

Check for updates

The use of real-world data in drug repurposing

Kyungsoo Park (iD) *

Department of Pharmacology, Yonsei University College of Medicine, Seoul 03722, Korea

*Correspondence to
Kyungsoo Park
Department of Pharmacology, Yonsei
University College of Medicine, 50-1 Yonsei-ro,
Seodaemun-gu, Seoul 03722, Korea.
E-mail: kspark@yuhs.ac

ORCID iDs
Kyungsoo Park (iD)
https://orcid.org/0000-0002-6972-1143

ABSTRACT

Drug repurposing, or repositioning, is to identify new uses for existing drugs. Significantly reducing the costs and time-to-market of a medication, drug repurposing has been an alternative tool to accelerate drug development process. On the other hand, 'real world data (RWD)' has been also increasingly used to support drug development process owing to its better representing actual pattern of drug treatment and outcome in real world. In the healthcare domain, RWD refers to data collected from sources other than traditional clinical trials; for example, in electronic health records or claims and billing data. With the enactment of the 21st Century Cures Act, which encourages the use of RWD in drug development and repurposing as well, such increasing trend in RWD use will be expedited. In this context, this review provides an overview of recent progresses in the area of drug repurposing where RWD was used, by firstly introducing the increasing trend and regulatory change in the use of RWD in drug development, secondly reviewing published works using RWD in drug repurposing, classifying them in the repurposing strategy, and lastly addressing limitations and advantages of RWDs.

Keywords: Drug Repurposing; Real World Data; Electronic Health Record; 21st Century Cures Act

INTRODUCTION

Drug repurposing

Drug repurposing, or repositioning, is to identify new uses for existing drugs [1]. Significantly reducing the costs and time-to-market of a medication as compared to a de-novo drug development, it has been an alternative tool to accelerate drug development process [2].

Repurposing approaches can be divided into experimental screening and in silico approaches, where in silico approaches are also called computational approaches.

Experimental screening uses in-depth high-throughput screening skill to screen known molecules either approved or failed with some knowledge about safety or the mode of action [3].

in silico approaches are based on the knowledge of drug activity and disease pathophysiology. It can be divided into knowledge-based, signature-based, and phenotype-based repurposing, where knowledge-based repurposing includes target-based, pathway-based, and targeted mechanism-based repurposing. These repurposing approaches were extensively addressed in the previous publication [4].

While in-silico methods do not require experimental work and are therefore cost-effective, their analytics are still within the molecular domain, limited in accurately predicting clinical outcomes.

Advent of RWD in drug development

In the healthcare domain, the term 'real world data (RWD)' refers to data collected from sources other than traditional clinical trials, including electronic health records (EHRs), claims and billing data, and registries among others [5-7].

RWD contains detailed patient information such as disease status, treatment, treatment outcomes, and comorbidities that are tracked longitudinally. The information generated from RWD provides important real-world evidence (RWE) to inform patient care, safety surveillance, therapeutic development, outcomes research, and comparative effectiveness studies [8].

While randomized controlled trials (RCTs) are gold-standards in drug development, besides the high cost and long development time, there have been more fundamental limitations as follows. The first limitation is the generalizability. Due to strict selection criteria, patients with conflicting comorbidities and/or co-medications are excluded, ending up with a very low representation of a specific subpopulation. Second, RCTs are highly controlled and patients should visit a clinic at a fixed time specified in the protocol, which in reality patients can hardly abide by. Therefore, RCTs do not accurately predict actual patterns of drug use in clinical practice.

In contrast, RWD does not suffer from the issue of cost and time, and is not constrained by the above limitations also. RWD studies based on EHRs guide clinical researches at a very little cost and does not have strict selection criteria, so broader populations and/or subpopulations of patients can be included. It provides information that represents the way most of the population receives the care. Clinical studies performed in the routine care environment help understand better how medicines behave when people have multiple diseases and use multiple medications. Accordingly, there has been an increasing trend toward using RWD instead of clinical trial data or in conjunction with it to inform medical decisions.

The key difference is that RWE, which is derived from analysis of RWD, informs effectiveness and safety in larger populations with greater power, allowing real life behaviour to be possible, with patients of co-morbidities and co-medications included.

Noticing such importance of RWD in drug development, the 21st Century Cures Act was enacted into the US law in December 2016, which aims to accelerate the FDA drug and medical device approval processes by replacing some of the data requirements from clinical trials with observational data or RWD settings [9]. It also placed additional focus in the area of drug repurposing, encouraging the use of RWD in getting the approval of new indications or label expansions for approved drugs.

These regulatory changes in the USA have become a basis to increase opportunities to use RWD in drug development, leading to FDA guidance on the use of EHR data [5] as well as guidance on incorporating RWD into regulatory submissions [10].

With this background, this paper will review the works that used RWD for drug repurposing.

METHODS

Literature search on the works that used RWD for drug repurposing revealed that drug repurposing was performed using different strategies, in terms of the modality of database used; either single modal database (EHR or another RWD or genomics), multimodal database (i.e., the combination of different modalities of data or multi-omics data), or multimodal database including animal data for validation. In this context, this section reviews the previous works, classifying them in the database modality used as follows.

Repurposing using single modal database

Recent evidence showed that, in patients treated with metformin, cancer survival increases [11,12] while cancer risk decreases [13], which suggests a repurposing hypothesis that metformin could be used as an antineoplastic agent.

Xu et al. [14] conducted a retrospective study to validate the above hypothesis. In their work, automated informatics methods including natural language processing (NLP) were applied to EHR data to identify patient cohorts and medication information, and then it was assessed whether metformin can be repurposed to cancer treatment. They found that metformin decreased mortality after cancer diagnosis compared with diabetic and nondiabetic cancer patients not on metformin.

In the work of Visanji et al. [15], using ML methods the authors have performed a computational analysis of published literature to rank several existing antihypertensive drugs that are predicted to reduce alpha synuclein oligomerization. Then, to provide evidence of a possible disease modifying effect in Parkinson's disease (PD), they analyzed RWD consisting of a cohort of individuals with incident hypertension, which was constructed using IBM MarketScanâ Research Databases containing healthcare claims information, and identified angiotensin receptor blockers in combination with dihydropyridine calcium channel blockers as a combination of potential disease-modifying effect in PD.

In another clinical drug repurposing study using EHR data, Kuang et al. [16] developed a ML-based drug repurposing approach, called baseline regularization, to predict the effects of drugs on different physical measurements such as fasting blood glucose to identify potential repurposing. They used the continuous self-controlled case series problem to solve for the pathway solution [17].

Wu et al. [18] proposed detecting drug repurposing signal by screening the effect of noncancer drugs on the survival of cancer patients using two large EHRs at Vanderbilt University Medical Center (VUMC) and Mayo Clinic. Based on EHR data at VUMC, they showed that, among 146 noncancer drugs analyzed, 22 drugs of 6 drug classes (statins, β-blockers, α-1 blockers, angiotensin-converting enzyme inhibitors, proton pump inhibitors, nonsteroidal anti-inflammatory drugs) improved overall cancer survival. When their results were replicated using EHR data at Mayo Clinic, 9 of the 22 drugs were validated.

Ozery-Flato et al. [19] and Laifenfeld et al. [20] presented a framework that systematically analyzes real-world longitudinal data for a large cohort of patients. Using causal inference methodology, the framework emulates a maximal number of RCTs based on observed healthcare data, while adjusting for selection and confounding biases. They applied the proposed framework in drug repurposing for PD to identify candidates for disease-modifying effects on PD progression. Constructing cohorts of PD patients sampled from medical databases, Explorys SuperMart (N = 88,867) and IBM MarketScan Research Databases (N = 106,395), they conducted an observational study and applied causal inference methods to estimate the effectiveness of 218 drugs on delaying dementia onset as a marker for slowing PD progression. As a result, they found that rasagiline, prescribed for PD motor symptoms, and zolpidem, a psycholeptic, are effective for delaying PD progression in both datasets.

Repurposing using multimodal database

Brilliant et al. [21] combined EHR and insurance claim data to support the protective potential of L-DOPA (Levodopa) against age-related macular degeneration (AMD), which was found in their previous work illustrating that L-DOPA activates GPR143 expressed in the retinal pigment epithelium, such that GPR143 signaling may protect from AMD [22,23].

The authors demonstrated that AMD was significantly delayed in patients receiving L-DOPA prescription compared with those not treated and found that the odds ratio for AMD development was significantly negatively correlated with L-DOPA use.

The work by Goldstein et al. [24] investigated associations between EHR phenotypes and genetic variants to identify drugs that could prevent or treat gestational diabetes mellitus (GDM). Identifying 129 active drugs and 196 genes associated, which are considered safe in pregnancy, they extracted 37,380 patients' data that include DNA samples and analyses from Vanderbilt University Medical Center's EHR, with patients de-identified using the Synthetic Derivative. Using the Illumina Infinium Human Exome Bead Chip that represents 306 SNPs in 130 genes among 196 genes of interest, they tested for associations between GDM and/or type 2 diabetes (DM2). A routine 50-gram glucose tolerance test (GTT) was also performed to test for the association with glucose tolerance during pregnancy. They found 11 drug classes had an association between their target genes and GDM/DM2. For changes in GTT, they found 6 drug classes were associated. Two drug classes, L-type calcium channel blocking antihypertensives (CCBs) and Serotonin receptor type 3 (5HT-3) antagonist antinausea medications, were identified in both analyses, where the former produced a decrease and the latter an increase in glucose level during GTT. In conclusion, CCBs were identified as a drug class considered safe in pregnancy and effective in preventing or treating GDM while 5HT-3 antagonists may worsen glucose tolerance.

In the work of Zhou et al. [25], an integrated drug repurposing strategy was presented for opioid use disorders (OUD) that integrates computational prediction, clinical corroboration using EHRs and mechanisms of action analysis. First, building a drug side effect-gene (DSEG) computational drug prediction system, the top 20 drug candidates to treat OUD were predicted. Second, using patient EHR data, for each of the top 20 candidate drugs, a retrospective case-control study was performed to evaluate the odds ratio for remission comparing the exposure group versus the comparison group in which both groups suffered OUD. Here, for EHR data, de-identified population-level data collected by the IBM Watson Health from 360 hospitals and 317,000 providers were used, which represented 20% of the US population. Five drugs of tramadol, olanzapine, mirtazapine, bupropion, and atomoxetine were selected as they were

associated with increased odds of OUD remission. Third, for the 5 repurposed drugs selected, genetic and pathway enrichment analysis showed that OUD-associated target genes include BDNF, CYP2D6, OPRD1, OPRK1, OPRM1, HTR1B, POMC, and SLC6A4, and target pathway includes opioid signaling, G-protein activation, serotonin receptors, and GPCR signaling.

Similarly combining drug–target interaction prediction and clinical corroboration, the authors applied another integrated drug repurposing strategy to identifying novel repositioned candidate drugs for Alzheimer's disease [26].

Repurposing using multimodal database including animal data

Nagashima et al. [27] conducted FAERS (FDA adverse event reporting system) analysis to search for a coexisting drug that can reduce the hyper-glycaemia risk of atypical antipsychotics. They found that a vitamin D analogue can significantly decrease quetiapine–induced adverse events relating hyper-glycaemia. Through signaling pathway and gene expression analyses, they showed quetiapine-induced downregulation of *Pik3r1*. They validated their results using a mouse model. These results suggest that, when co-administered, vitamin D can prevent antipsychotic-induced hyperglycaemia by reducing insulin resistance by PI3K upregulation.

Based on the assumption that similar drugs can treat similar diseases, Paik et al. [28] generated disease and drug pair similarity scores in genomics and EHR-extracted lab test data, independently. As a result, terbutaline sulfate, a β2-adrenoceptor agonist widely used for the treatment of asthma, was identified as a candidate for treatment of amyotrophic lateral sclerosis (ALS), on the one hand based on similarity between terbutaline sulfate and ursodeoxycholic acid, but on the other hand based on similarity between Kawasaki syndrome and ALS. Then, to validate the potential therapeutic benefit of terbutaline sulfate for ALS, using a zebrafish ALS model, prevention of defects in axons and neuromuscular junction degeneration was demonstrated.

DISCUSSION

As seen in the Methods section, the previous works using RWD in repurposing illustrates various repurposing strategies with different modalities of database used, which might be taken into account as a guide in designing a repurposing study at a given scope of data. It is noticeable that, when single modal RWD was used, another RWD (of the same modality) was also used for the validation purpose [14,18,19]. While most of the works tried to validate their repurposing results with another modality of data (e.g., results obtained from EMR were validated using genomic or multi-omic data or vice versa), it is hardly found that validation was made in human or in clinical trials. This is also true for the work validated with animal data [28].

One essential limitation with RWD studies is that many RWD sources have the data quality issue, associated with data inconsistency such as selection bias and missing data as in RWD collection across different data sources is usually heterogeneous and entails the lack of standardization and harmonization [29].

Nevertheless, on top of basic advantages addressed in the Introduction section, there are several advantages with RWD studies, some of which are described in the following:

First, if RWD incorporated, clinical trials can be simulated more realistically. Traditionally, clinical trial simulation (CTS) uses virtual populations to test various trial designs before conducting the actual clinical trial [30]. CTS incorporating RWD can simulate virtual populations more realistically.

Furthermore, the recent development of emulating trials with RWD [19, 20] enables the unbiased estimation of casual relationships [31]. Thus, if the traditional CTS approach is combined with the concept of modern trial emulation, different assumptions of a clinical trial can be systematically tested, which can be used to inform future trial design and produce RWD based causal results [32].

Another emerging trend of RWD approach to facilitate the drug development process is linking EHRs with other modality of data such as biobank data to better understand drug-phenotype and drug-gene relations [24,25,28].

Finally, the establishment of large observational research network would facilitate the sharing of RWD. One such example is found in Observational Health Data Sciences and Informatics (OHDSI) consortium [33].

REFERENCES

1. Langedijk J, Mantel-Teeuwisse AK, Slijkerman DS, Schutjens MH. Drug repositioning and repurposing: terminology and definitions in literature. Drug Discov Today 2015;20:1027-1034.
   PUBMED | **CROSSREF**

2. Ashburn TT, Thor KB. Drug repositioning: identifying and developing new uses for existing drugs. Nat Rev Drug Discov 2004;3:673-683.
   PUBMED | **CROSSREF**

3. Cha Y, Erez T, Reynolds IJ, Kumar D, Ross J, Koytiger G, et al. Drug repurposing from the perspective of pharmaceutical companies. Br J Pharmacol 2018;175:168-180.
   PUBMED | **CROSSREF**

4. Park K. A review of computational drug repurposing. Transl Clin Pharmacol 2019;27:59-63.
   PUBMED | **CROSSREF**

5. Center for Devices and Radiological Health. Use of real-world evidence to support regulatory decision-making for medical devices (August 2017) [Internet]. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/use-real-world-evidence-support-regulatory-decision-making-medical-devices. Accessed September 27, 2021.

6. FDA. Promoting effective drug development programs: opportunities and priorities for FDA's office of new drugs - November 7, 2019 (March 31, 2020). https://www.fda.gov/news-events/fda-meetings-conferences-and-workshops/promoting-effective-drug-development-programs-opportunities-and-priorities-fdas-office-new-drugs. Accessed September 27, 2021.

7. FDA. Real-World Evidence (retrieved March 6, 2020). https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence. Accessed September 27, 2021.

8. Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, et al. Real-world evidence—What is it and what can it tell us? N Engl J Med 2016;375:2293-2297.
   PUBMED | **CROSSREF**

9. Congress.Gov. 21st Century Cures Act, Pub. L. No. 114-255 (2016) [Internet]. https://www.congress.gov/bill/114th-congress/house-bill/34/text. Accessed September 27, 2021.

10. FDA. Submitting documents using real-world data and real-world evidence to FDA for drugs and biologics guidance for industry (April 29, 2020). [Internet]. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/submitting-documents-using-real-world-data-and-real-world-evidence-fda-drugs-and-biologics-guidance. Accessed September 27, 2021.

11. Landman GW, Kleefstra N, van Hateren KJ, Groenier KH, Gans RO, Bilo HJ. Metformin associated with lower cancer mortality in type 2 diabetes: ZODIAC-16. Diabetes Care 2010;33:322-326.
    PUBMED | **CROSSREF**

12. Currie CJ, Poole CD, Jenkins-Jones S, Gale EA, Johnson JA, Morgan CL. Mortality after incident cancer in people with and without type 2 diabetes: impact of metformin on survival. Diabetes Care 2012;35:299-304.
PUBMED | CROSSREF

13. Evans JM, Donnelly LA, Emslie-Smith AM, Alessi DR, Morris AD. Metformin and reduced risk of cancer in diabetic patients. BMJ 2005;330:1304-1305.
PUBMED | CROSSREF

14. Xu H, Aldrich MC, Chen Q, Liu H, Peterson NB, Dai Q, et al. Validating drug repurposing signals using electronic health records: a case study of metformin associated with reduced cancer mortality. J Am Med Inform Assoc 2015;22:179-191.
PUBMED | CROSSREF

15. Visanji NP, Madan P, Lacoste AM, Buleje I, Han Y, Spangler S, et al. Using artificial intelligence to identify anti-hypertensives as possible disease modifying agents in Parkinson's disease. Pharmacoepidemiol Drug Saf 2021;30:201-209.
PUBMED | CROSSREF

16. Kuang Z, Bao Y, Thomson J, Caldwell M, Peissig P, Stewart R, et al. A machine-learning-based drug repurposing approach using baseline regularization. Methods Mol Biol 2019;1903:255-267.
PUBMED | CROSSREF

17. Suchard MA, Zorych I, Simpson SE, Schuemie MJ, Ryan PB, Madigan D. Empirical performance of the self-controlled case series design: lessons for developing a risk identification and analysis system. Drug Saf 2013;36 Suppl 1:S83-S93.
PUBMED | CROSSREF

18. Wu Y, Warner JL, Wang L, Jiang M, Xu J, Chen Q, et al. Discovery of noncancer drug effects on survival in electronic health records of patients with cancer: a new paradigm for drug repurposing. JCO Clin Cancer Inform 2019;3:1-9.
PUBMED | CROSSREF

19. Ozery-Flato M, Goldschmidt Y, Shaham O, Ravid S, Yanover C. Framework for identifying drug repurposing candidates from observational healthcare data. JAMIA Open 2020;3:536-544.
PUBMED | CROSSREF

20. Laifenfeld D, Yanover C, Ozery-Flato M, Shaham O, Rosen-Zvi M, Lev N, et al. Emulated clinical trials from longitudinal real-world data efficiently identify candidates for neurological disease modification: examples from Parkinson's disease. Front Pharmacol 2021;12:631584.
PUBMED | CROSSREF

21. Brilliant MH, Vaziri K, Connor TB Jr, Schwartz SG, Carroll JJ, McCarty CA, et al. Mining retrospective data for virtual prospective drug repurposing: L-DOPA and age-related macular degeneration. Am J Med 2016;129:292-298.
PUBMED | CROSSREF

22. Lopez VM, Decatur CL, Stamer WD, Lynch RM, McKay BS. L-DOPA is an endogenous ligand for OA1. PLoS Biol 2008;6:e236.
PUBMED | CROSSREF

23. Falk T, Congrove NR, Zhang S, McCourt AD, Sherman SJ, McKay BS. PEDF and VEGF-A output from human retinal pigment epithelial cells grown on novel microcarriers. J Biomed Biotechnol 2012;2012:278932.
PUBMED | CROSSREF

24. Goldstein JA, Bastarache LA, Denny JC, Roden DM, Pulley JM, Aronoff DM. Calcium channel blockers as drug repurposing candidates for gestational diabetes: Mining large scale genomic and electronic health records data to repurpose medications. Pharmacol Res 2018;130:44-51.
PUBMED | CROSSREF

25. Zhou M, Wang Q, Zheng C, John Rush A, Volkow ND, Xu R. Drug repurposing for opioid use disorders: integration of computational prediction, clinical corroboration, and mechanism of action analyses. Mol Psychiatry. 2021 [In Press].
PUBMED

26. Zhou M, Zheng C, Xu R. Combining phenome-driven drug-target interaction prediction with patients' electronic health records-based clinical corroboration toward drug discovery. Bioinformatics 2020;36 Suppl_1:i436-i444.
PUBMED | CROSSREF

27. Nagashima T, Shirakawa H, Nakagawa T, Kaneko S. Prevention of antipsychotic-induced hyperglycaemia by vitamin D: a data mining prediction followed by experimental exploration of the molecular mechanism. Sci Rep 2016;6:26375.
PUBMED | CROSSREF

28. Paik H, Chung AY, Park HC, Park RW, Suk K, Kim J, et al. Repurpose terbutaline sulfate for amyotrophic lateral sclerosis using electronic medical records. Sci Rep 2015;5:8580.
    PUBMED | CROSSREF

29. Boland MR, Hripcsak G, Shen Y, Chung WK, Weng C. Defining a comprehensive verotype using electronic health records for personalized medicine. J Am Med Inform Assoc 2013;20:e232-e238.
    PUBMED | CROSSREF

30. Holford N, Ma SC, Ploeger BA. Clinical trial simulation: a review. Clin Pharmacol Ther 2010;88:166-182.
    PUBMED | CROSSREF

31. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. Am J Epidemiol 2016;183:758-764.
    PUBMED | CROSSREF

32. Chen Z, Liu X, Hogan W, Shenkman E, Bian J. Applications of artificial intelligence in drug development using real-world data. Drug Discov Today 2021:26:1256-126.
    PUBMED | CROSSREF

33. Hripcsak G, Duke JD, Shah NH, Reich CG, Huser V, Schuemie MJ, et al. Observational health data sciences and informatics (OHDSI): opportunities for observational researchers. Stud Health Technol Inform 2015;216:574-578.
    PUBMED