

# Discovery of Noncancer Drug Effects on Survival in Electronic Health Records of Patients With Cancer: A New Paradigm for Drug Repurposing

Yonghui Wu, PhD<sup>1,2</sup>; Jeremy L. Warner, MD, MS<sup>3</sup>; Liwei Wang, MD, PhD<sup>4</sup>; Min Jiang, PhD<sup>1</sup>; Jun Xu, PhD<sup>1</sup>; Qingxia Chen, PhD<sup>3</sup>; Hui Nian, PhD<sup>3</sup>; Qi Dai, MD, PhD<sup>3,5</sup>; Xianglin Du, PhD<sup>1</sup>; Ping Yang, MD, PhD<sup>4</sup>; Joshua C. Denny, MD, MS<sup>3</sup>; Hongfang Liu, PhD<sup>4</sup>; and Hua Xu, PhD<sup>1</sup>

**PURPOSE** Drug development is becoming increasingly expensive and time consuming. Drug repurposing is one potential solution to accelerate drug discovery. However, limited research exists on the use of electronic health record (EHR) data for drug repurposing, and most published studies have been conducted in a hypothesis-driven manner that requires a predefined hypothesis about drugs and new indications. Whether EHRs can be used to detect drug repurposing signals is not clear. We want to demonstrate the feasibility of mining large, longitudinal EHRs for drug repurposing by detecting candidate noncancer drugs that can potentially be used for the treatment of cancer.

**PATIENTS AND METHODS** By linking cancer registry data to EHRs, we identified 43,310 patients with cancer treated at Vanderbilt University Medical Center (VUMC) and 98,366 treated at the Mayo Clinic. We assessed the effect of 146 noncancer drugs on cancer survival using VUMC EHR data and sought to replicate significant associations (false discovery rate < .1) using the identical approach with Mayo Clinic EHR data. To evaluate replicated signals further, we reviewed the biomedical literature and clinical trials on cancers for corroborating evidence.

**RESULTS** We identified 22 drugs from six drug classes (statins, proton pump inhibitors, angiotensin-converting enzyme inhibitors,  $\beta$ -blockers, nonsteroidal anti-inflammatory drugs, and  $\alpha$ -1 blockers) associated with improved overall cancer survival (false discovery rate < .1) from VUMC; nine of the 22 drug associations were replicated at the Mayo Clinic. Literature and cancer clinical trial evaluations also showed very strong evidence to support the repurposing signals from EHRs.

**CONCLUSION** Mining of EHRs for drug exposure-mediated survival signals is feasible and identifies potential candidates for antineoplastic repurposing. This study sets up a new model of mining EHRs for drug repurposing signals.

**Clin Cancer Inform.** © 2019 by American Society of Clinical Oncology

Licensed under the Creative Commons Attribution 4.0 License 

## INTRODUCTION

Cancer drug development is increasingly expensive and time consuming. The development of a new drug is estimated to cost \$648 million<sup>1</sup> to \$2.5 billion<sup>2</sup> and takes an average of 9 to 12 years before market availability.<sup>3</sup> The drug development success rate is less than 8% because of lack of efficacy, excess toxicity, declining research and development, cost of commercialization, and payer influence.<sup>4</sup> Cancer drugs are now the top sellers among all Food and Drug Administration–approved therapies.<sup>5</sup> Although many new cancer therapeutics are in development, new methods to accelerate drug discovery are needed. Drug repurposing has received great attention<sup>6,7</sup> in recent years as one potential solution. A recent study reported that the discovery of new indications of existing drugs accounts for 20% of new drug products.<sup>8</sup>

Electronic health records (EHRs) could be an important source for drug repurposing discovery, but EHRs, which are now present in 96% of health care systems,<sup>9</sup> have not been extensively leveraged for drug repurposing studies. Recent studies have demonstrated that EHR data can be used as an efficient, low-cost resource to validate drug repurposing signals detected from other sources.<sup>10,11</sup> Currently, limited research exists on using EHR data for drug repurposing, and most published studies have been conducted in a manner that requires predefined hypotheses. For example, recent evidence has suggested that metformin improves cancer survival<sup>12,13</sup> and decreases cancer risk in patients with diabetes,<sup>14</sup> which suggests clinical promise as an antineoplastic agent. We previously found in a retrospective EHR-based study that metformin is associated with superior

## ASSOCIATED CONTENT

### Data Supplement

Author affiliations and support information (if applicable) appear at the end of this article.

Accepted on March 12, 2019 and published at [ascopubs.org/journal/cci](https://ascopubs.org/journal/cci) on XXXX, 2019; DOI <https://doi.org/10.1200/CCI.19.00001>

## CONTEXT SUMMARY

### Key Objective

In this study, we proposed a new drug repurposing approach that mines large electronic health records (EHRs) to detect candidate noncancer drugs that can potentially be used for the treatment of cancer.

### Knowledge Generated

Using the EHRs at Vanderbilt University Medical Center, we identified 22 drugs from six drug classes associated with improved overall cancer survival (false discovery rate  $< .1$ ), and nine of the 22 drug associations were replicated using the EHRs at the Mayo Clinic. Literature and cancer clinical trial evaluations also showed very strong evidence to support these repurposing signals from EHRs.

### Relevance

This study demonstrated that mining of EHRs for drug exposure–mediated survival signals is feasible and identifies potential candidates for antineoplastic repurposing. It sets up a new model of mining EHRs for drug repurposing signals.

cancer-specific survival.<sup>10</sup> This hypothesis-driven method highly depends on domain experts to generate hypotheses and select variables.

In the current study, we take a data-driven approach to detect potential drug repurposing signals using EHR data, with the specific goal of identifying new cancer treatment signals. We evaluated 146 drugs in the Vanderbilt University Medical Center (VUMC) EHR that typically are taken long term for noncancerous conditions and assessed their effects on survival in patients with cancer. We then evaluated signals detected at VUMC by replicating significant associations using the Mayo Clinic's EHR, searching the biomedical literature for corroborating evidence, and checking cancer clinical trials for support.

## PATIENTS AND METHODS

### Primary Data Source

We used the synthetic derivative (SD),<sup>15</sup> which is a deidentified copy of VUMC's EHR. The SD contains comprehensive clinical data for more than 2.3 million patients, including billing codes, laboratory values, pathology/radiology reports, medication orders, and clinical notes. In addition, the SD contains data from the Vanderbilt Cancer Registry, which is maintained by certified tumor registrars according to the standards set forth by the state of Tennessee and the Commission on Cancer.

### Patient With Cancer Definition

This study used patients with cancer identified by the Vanderbilt Cancer Registry, which operates under the mandate of the Tennessee Cancer Registry and the Commission on Cancer. Patients were identified through automated parsing of pathology reports and billing codes.

### Identification of Candidate Drugs for the Study

In the SD, medication information is extracted from both structured (eg, electronic physician orders) and unstructured (ie, clinical notes) data using MedEx.<sup>16</sup> MedEx has proven high performance on extracting medication

names and their signature information in clinical notes.<sup>16</sup> Here, we required that a drug name must be followed by at least a dosage instruction to account for a prescription to a patient. We have previously shown that the requirement that a drug name be followed by a dosage instruction led to a very high positive predictive value.<sup>10</sup> To generate a candidate list, we followed two steps. First, we selected normalized drugs used by more than 5,000 individuals, which resulted in 301 candidates, and second, two physicians (J.L.W., J.C.D.) manually reviewed these to exclude known antineoplastics, drugs used in the supportive care of cancer (eg, opiates), over-the-counter drugs, and drugs with short-term indications (eg, antibiotics). Subsequently, 146 candidate drugs remained (see the Data Supplement for the full list). With the assumption that patients were followed for 5 years and the median survival time of the control group was 5 years, with a total of 2.3 million patients and 5,000 who received the drug of interest, we have 89% power to detect a true hazard ratio (HR) of 1.1 (a reduction of 6 months in median survival time assuming exponential distribution) with 5% detection of approximately 150 drugs and a false discovery rate (FDR)–adjusted  $P = .1$ . For each of the 146 candidate drugs, we developed a multivariable Cox proportional hazards regression model to assess its effect on cancer survival. All other drugs (including non-candidate drugs) that were used by more than 5,000 patients were adjusted as covariates in the multivariable Cox model.

### Study Design, Covariates, and Statistical Analysis

For each drug, we conducted a retrospective cohort study with two comparison groups: an exposure group that comprised patients with one or more prescriptions of the drug in their EHR and a nonexposure group that comprised patients with no prescription of the drug in their EHR. Prescription of a medication was determined by combining both structured electronic physician orders and unstructured clinical notes. Cox proportional hazards regression modeling was used to assess the association of

drug exposure with overall survival (ie, time from cancer diagnosis to death) or last medical record date in the EHR (censored). Study covariates were patient demographics (age, biologic sex, race); tumor information (type, stage); diseases with International Classification of Diseases, Ninth Revision, Clinical Modification, codes grouped into phenome-wide association study<sup>17</sup> phenotypes; and normalized drugs. Because the dimensionality of covariates was high, we conducted variable screening using a univariable Cox model for each disease-related covariate and kept those with  $P < .3$ . Other variables were directly used without any filtering. We assessed mortality using a multivariable Cox proportional hazards regression model that adjusted for all the selected covariates and reported the  $P$  values, HRs, and 95% CIs. We used a cutoff FDR-adjusted<sup>18</sup>  $P < .1$  to select the top-ranked drugs associated with cancer survival; this cutoff was chosen to minimize the risk of excessive false negatives at this hypothesis-generating stage.<sup>19</sup> All analyses were conducted using R 3.1 with the survival, Hmisc, and rms packages (<http://www.r-project.org>).

### Evaluation

We undertook several experiments to validate the detected signals.

**Replication using another large site.** Using the EHR and cancer registry at the Mayo Clinic, we replicated the study by following the same design and statistical analysis plan used for the VUMC EHRs. Drugs with a survival signal detected in both institutions' EHRs also were examined.

### Search of biomedical literature for supporting evidence.

For additional examination, we identified English-language original publications from PubMed by searching for the drug name plus the term cancer survival. If there was no result or the number of publications was fewer than 10, we also included publications identified by searching the drug name with only the term cancer. We reviewed the abstracts of 100%, 20%, or 10% if the total number of publications was fewer than 20, 21 to 200, or more than 200, respectively. If necessary, the body of available publications also was reviewed. After review, each publication was labeled as one of three categories: evidence to support an antineoplastic effect wherein the drug, alone or in combination, has a cytotoxic effect on cancer cells in vitro or in vivo; evidence to support a carcinogenic effect wherein the drug, alone or in combination, has a proliferative effect on cancer cells in vitro or in vivo; and inconclusive wherein no conclusion can be made about the drug's cytotoxic or proliferative effect in vitro or in nonrandomized in vivo studies, or the drug failed to demonstrate statistical superiority in a randomized in vivo trial.

**Search of human interventional cancer trials for supporting evidence.** In a previous study, 25,530 cancer treatment trials were collected from ClinicalTrials.gov.<sup>20</sup> Among them, we identified 1,068 cancer trials associated with the 146

noncancer drugs used in this study. This subset was manually reviewed and categorized as follows: Category A, the intended primary outcome is survival or a surrogate of survival, including direct effects on a tumor (eg, changes in proliferation indices), solely from the candidate drug (primary effect); category B, the intended primary outcome is survival or surrogate of survival (as in category A) on the basis of synergy between the candidate drug and one or more known antineoplastics (additive effect, including radiotherapy given with the candidate drug); category C, the candidate drug is being used for supportive care purposes or to counter adverse effects of other interventions; and category D, false positives. Of the trials identified as category A or B, we also required that the study be in patients with a current or former diagnosis of cancer; chemoprevention trials were excluded. We also tested whether our signal detection method is significantly different from random selection of drug candidates by using permutation analysis. Additional details are available in the Data Supplement.

## RESULTS

### Drug Repurposing Signals Detected From the VUMC EHR

At VUMC, we identified 43,310 patients with cancer diagnosed at age 18 years or older between January 1, 1995, and December 31, 2010. Patients were a median age of 57 years at diagnosis, 57% were male, and 93% were white. The major cancer types were prostate (5,673; approximately 13%), breast (3,968; approximately 9%), lung (3,346; approximately 8%), and colorectal (2,537; approximately 6%). We collected 2,630 variables for each individual, including three patient demographics (age, biologic sex, race), two tumor information (type, stage), 1,279 diagnoses, and 1,346 medications. We assessed 146 noncancer drugs and detected 30 significantly associated with survival (FDR-adjusted  $P < .1$ ), of which 22 were significantly associated with improved cancer survival. **Table 1** lists these 22 drugs, which include statins (rosuvastatin, simvastatin, atorvastatin), proton pump inhibitors (omeprazole, esomeprazole, lansoprazole), angiotensin-converting enzyme inhibitors (ramipril, lisinopril),  $\beta$ -blockers (metoprolol, carvedilol), nonsteroidal anti-inflammatory drugs (NSAIDs; diclofenac, celecoxib),  $\alpha$ -1 blockers (tamulosin), and several others.

### Replication Using the Mayo Clinic EHR

We then sought to replicate the study using 98,366 individual patients diagnosed with cancer at the Mayo Clinic between January 1, 1995, and December 31, 2010. Patients were a median age of 64 years, 57% were male, and 88% were white. The major cancer types were prostate (19,951; approximately 20%), breast (10,415; approximately 10%), lung (9,948; approximately 10%), and colorectal (6,829; approximately 7%). We collected 5,725 variables for each individual, including 1,279 diagnoses and 4,441 medications. Using the same approach, we identified 16 drugs significantly associated with improved

**TABLE 1.** Noncancer Drugs Associated With Improved Cancer Survival From the VUMC and Mayo Clinic EHRs

Drug Name	VUMC		Mayo Clinic	
	HR (95% CI)	FDR-Adjusted <i>P</i>	HR (95% CI)	FDR-Adjusted <i>P</i>
Detected from both VUMC and Mayo Clinic EHRs*				
Rosuvastatin	0.81 (0.69 to 0.95)	.0691	0.68 (0.50 to 0.92)	.0846
Simvastatin	0.84 (0.79 to 0.90)	< .001	0.82 (0.76 to 0.87)	< .001
Amlodipine	0.84 (0.79 to 0.90)	< .001	0.85 (0.78 to 0.93)	.0054
Tamsulosin	0.87 (0.80 to 0.96)	.0435	0.71 (0.59 to 0.85)	.0061
Metformin	0.88 (0.80 to 0.97)	.0571	0.87 (0.80 to 0.95)	.0173
Omeprazole	0.89 (0.84 to 0.94)	.0006	0.90 (0.85 to 0.96)	.0120
Warfarin	0.90 (0.85 to 0.96)	.0084	0.90 (0.84 to 0.96)	.0174
Lisinopril	0.91 (0.86 to 0.97)	.0328	0.93 (0.89 to 0.97)	.0173
Metoprolol	0.92 (0.86 to 0.98)	.0519	0.69 (0.61 to 0.77)	< .001
Detected from the VUMC EHR				
Olmesartan	0.72 (0.59 to 0.89)	.0200	0.90 (0.56 to 1.4)	.8827
Sildenafil	0.73 (0.65 to 0.82)	< .001	1.0 (0.74 to 1.4)	.9683
Phenobarbital	0.77 (0.63 to 0.94)	.0603	0.83 (0.60 to 1.2)	.5775
Carvedilol	0.78 (0.68 to 0.90)	.0090	0.96 (0.83 to 1.1)	.8763
Diclofenac	0.81 (0.69 to 0.96)	.0945	0.73 (0.55 to 0.98)	.1794
Carbamazepine	0.84 (0.73 to 0.97)	.0981	1.0 (0.82 to 1.3)	.9415
Ramipril	0.85 (0.76 to 0.95)	.0451	0.99 (0.81 to 1.2)	.9785
Epoetin	0.85 (0.79 to 0.93)	.0023	1.7 (0.98 to 2.8)	.2429
Olanzapine	0.85 (0.75 to 0.97)	.0981	1.1 (0.83 to 1.6)	.7084
Atorvastatin	0.86 (0.80 to 0.93)	.0018	0.92 (0.83 to 1.0)	.3921
Esomeprazole	0.89 (0.84 to 0.95)	.0040	0.88 (0.69 to 1.1)	.5669
Celecoxib	0.91 (0.84 to 0.98)	.0944	0.68 (0.48 to 0.96)	.1441
Lansoprazole	0.92 (0.86 to 0.98)	.0771	0.91 (0.80 to 1.0)	.4002
Detected from the Mayo Clinic EHR				
Midazolam	1.0 (0.94 to 1.0)	.9964	0.43 (0.27 to 0.67)	.0053
Pravastatin	0.87 (0.76 to 1.0)	.2091	0.64 (0.54 to 0.75)	< .001
Venlafaxine	0.95 (0.84 to 1.1)	.6654	0.69 (0.54 to 0.87)	.0190
Oxybutynin	1.0 (0.70 to 0.91)	.8439	0.79 (0.68 to 0.91)	.0174
Lovastatin	0.92 (0.81 to 1.0)	.4504	0.81 (0.72 to 0.92)	.0173
Captopril	0.99 (0.86 to 1.1)	.9057	0.85 (0.75 to 0.96)	.0488
Hydrochlorothiazide	0.99 (0.93 to 1.1)	.8307	0.92 (0.87 to 0.96)	.0054

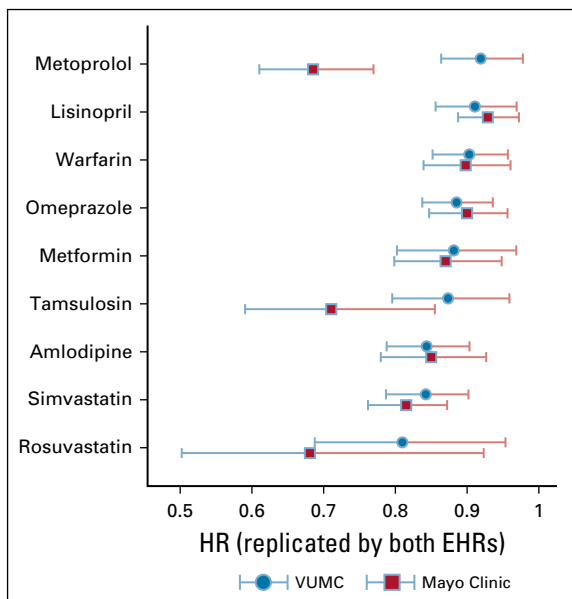
NOTE. Study covariates were patient demographics (age, biologic sex, race), tumor information (type, stage), diseases, and medications. Abbreviations: EHR, electronic health record; FDR, false discovery rate; HR, hazard ratio; VUMC, Vanderbilt University Medical Center.

survival (Table 1). Among the 22 initially detected drugs from the VUMC EHR, nine were replicated (Table 1). Figure 1 compares the HRs and 95% CIs for the nine replicated drugs. The Data Supplement shows the unadjusted Kaplan-Meier survival curves and associated 95% CIs for the nine drugs detected from both EHRs.

#### Validation Using Biomedical Literature

For each of the nine potential drugs detected from VUMC and found in the Mayo Clinic analysis, we searched

PubMed for corroborating evidence. A total of 1,348 relevant biomedical publications were found for all nine drugs. As listed in Table 2, all nine drugs have at least one publication that supported an antineoplastic effect, whereas five of them have at least one publication that reported a carcinogenic effect. For all nine drugs, there are more publications that supported their antineoplastic effect compared with their carcinogenic effect. Two drugs, simvastatin and metformin, have a substantial number of publications (20 and 57, respectively). Eighteen of 20



**FIG 1.** Comparison of the nine drugs detected from the Vanderbilt University Medical Center (VUMC) and replicated by the Mayo Clinic electronic health records (EHRs) by hazard ratio (HR). Study covariates were patient demographics (age, biologic sex, race), tumor information (type, stage), diseases, and medications.

publications supported simvastatin's antineoplastic effect. Similarly, 40 of 57 publications supported metformin's antineoplastic effect.

### Validation Using Clinical Trials

Manual review of 1,068 candidate trials identified 321 cancer efficacy trials, of which 105 (33%) explored the primary efficacy of the candidate drug (category A) and 216 (67%) explored additive efficacy to known antineoplastic drugs (category B). Of the 146 drugs in this study, 40 (27%) were tested in one or more trials, 28 (19%) were tested in two or more trials, and 17 (12%) were tested in five or more trials. Among the nine drugs with a survival signal replicated across the two clinical sites, four were identified as having completed or ongoing clinical trials (metformin, omeprazole, rosuvastatin, simvastatin). Two of these were among the most heavily studied (metformin and simvastatin, with 64 and 23 trials, respectively). In total, studies that involve the replicated drugs accounted for 30% (95 of 321) of the identified clinical trials.

We conducted permutation analysis<sup>21</sup> to compare the proposed method with random sampling. The 40 drugs that were tested by at least one cancer efficacy trial served as ground truth (see definition in Data Supplement). Among the 22 drug repurposing signals from VUMC, nine are in ground truth (precision, 41%; recall, 23%). Our method of detecting drug signals outperformed random sampling with  $P = .04$  on 100,000 permutations. Detailed results are provided in the Data Supplement.

## DISCUSSION

In this study, we mined large-scale EHR data to detect drug repurposing signals with potential cancer treatment implications. We found strong associations with improved overall cancer survival for statins, proton pump inhibitors, angiotensin-converting enzyme inhibitors,  $\beta$ -blockers, NSAIDs, and  $\alpha$ -1 blockers in two EHR systems. We also found evidence for these effects in the biomedical literature and clinical trials. Manual review of the biomedical literature and permutation analysis of cancer clinical trials also show that our proposed method generates potentially valid drug repurposing signals. These findings indicate that the use of EHRs is feasible as a new resource for drug repurposing signal detection. We believe that this study will set up a new model for drug repurposing signal detection using EHRs and thus complement existing methods for drug repurposing studies. For example, scientists have developed computational methods to detect new treatment signals for existing drugs, including structure-based screening,<sup>22,23</sup> adverse effect networks analysis,<sup>24,25</sup> genomic and gene expression analysis,<sup>26,27</sup> and biomedical literature mining.<sup>28,29</sup> Various data sources from genomics, drug chemical structure,<sup>30,31</sup> and phenotypic information<sup>24,25,32</sup> have been explored.

This study is different from previous EHR-based drug repurposing studies.<sup>10,11</sup> Most previous studies were conducted with a predefined hypothesis about drug and indication. These approaches highly depend on domain experts to define hypotheses and select relevant variables, which could be time consuming if we examine a large number of drugs. In this study, we have taken a data-driven approach that aimed to generate hypotheses. Instead of limited variables defined by domain experts, we included all available information (eg, patient demographics, diseases, drugs) as variables in the analysis. Of course, some important variables likely are not recorded in EHRs and thus not included in the analysis. For example, sociobehavioral determinants of health, including healthy behaviors, are rarely recorded in the current generation of EHRs.<sup>33</sup>

Some noncancer drugs identified in our study have strong evidence for cancer treatment from studies using other data sources. For example, many recent retrospective studies reported metformin associations with improved cancer survival,<sup>12,13</sup> and a chemoprevention trial in colorectal adenoma was positive.<sup>34</sup> We identified 64 ongoing or completed clinical trials studying metformin alone or in combination, whose anticancer effect could be related to mammalian target of rapamycin inhibition.<sup>35,36</sup> Ongoing cancer trials also are evaluating statins for cancer treatment (eg, a trial to assess the efficacy of simvastatin and capecitabine in locally advanced rectal cancer [ClinicalTrials.gov identifier: NCT02161822]). Recent studies

**TABLE 2.** Results of Literature Search for Corroborating Evidence for Drugs Associated With Improved Cancer Survival From Both VUMC and Mayo Clinic Cohorts

Drug Name	Reviewed Cancer-Relevant Studies	Evidence, No. (cell line, animal, human [trial type])		
		Antineoplastic Effect	Carcinogenic Effect	Inconclusive
Rosuvastatin	9	7 (5, 1, 1 [R, 1])	0	2 (1, 0, 1 [R, 1])
Simvastatin	20	18 (15, 1, 3 [R, 3])*	0	2 (0, 0, 2 [RCT, 2])
Amlodipine	6	5 (3, 2, 0)	1 (0, 0, 1 [R, 1])	0
Tamsulosin	5	0	0	5 (5, 0, 0)
Metformin	57	40 (15, 8, 17 [R, 15; NR, 1; RCT, 1])	2 (0, 0, 2 [R, 1; NR, 1])	13 (1, 0, 12 [R, 11; RCT, 1])
Omeprazole	10	8 (4, 0, 4 [NR, 3; RCT, 1])	2 (0, 1, 1 [R, 1])	0
Warfarin	17	5 (2, 1, 2 [R, 1; NR, 1; RCT, 1])	1 (0, 0, 1 [NR, 1])	11 (0, 1, 9 [R, 3; NR, 3, RCT, 3])
Lisinopril	7	2 (1, 1, 0)	1 (0, 0, 1 [NR, 1])	4 (0, 0, 4 [R, 1; NR, 2; RCT, 1])
Metoprolol	3	0	0	3 (2, 0, 1 [R, 1])

Abbreviations: NR, nonrandomized; R, retrospective; RCT, randomized controlled trial.

\*Total is > 100% because one trial reported both cell line and human results.

have reported that NSAIDs reduce the risk of a wide range of cancers (colon cancer,<sup>37</sup> oral cancer,<sup>38</sup> breast cancer,<sup>39</sup> melanoma<sup>40</sup>) through blocking cellular proliferation and by promoting apoptosis.<sup>37</sup> Of note, celecoxib (an NSAID) was identified as being the most heavily studied, with 92 ongoing or completed clinical trials,<sup>41</sup> but the signal for improved survival at VUMC did not replicate at the Mayo Clinic.

Repurposing signals have been found in population-based cohort studies, such as the signal for increased cancer survival in patients who take statins.<sup>42-44</sup> A smaller number of prospective repurposing trials have reported successes, such as a randomized trial of estradiol therapy of hormone receptor–positive, aromatase inhibitor-resistant advanced breast cancer<sup>45</sup>; a phase II study of pioglitazone in patients with stage IA to IIIA non–small-cell lung cancer ([ClinicalTrials.gov](https://clinicaltrials.gov/ct2/show/study/NCT01342770) identifier: NCT01342770); and an n-of-1 trial that combined metformin with trametinib in a patient with advanced ovarian cancer.<sup>46</sup> Although some repurposing trials, such as pravastatin added to standard chemotherapy for small-cell lung cancer, have been negative,<sup>47</sup> increasingly granular phenotyping efforts will lead to refined patient selection. In particular, the advent of routine germline sequencing, somatic tumor profiling, and immunophenotyping will allow for precise patient selection, as in the NCI-MATCH (National Cancer Institute Molecular Analysis for Therapy Choice) trial.<sup>48</sup> Currently, some drugs have no evidence, or sometimes conflicting findings, about their effects on treating cancer according to existing literature. For example, one study examined captopril and found no clear association between the use of antihypertensive drugs and prostate cancer.<sup>49</sup> However, another study that focused on users of captopril showed a lower risk of subsequent prostate cancer.<sup>50</sup> Our literature review was based on a sampling strategy and may have overlooked

human trials with strong evidence for antineoplastic effects. Of note, given that this is a repurposing study for candidate drugs not clearly known to have antineoplastic properties, much of the discovered literature was based on cell lines or was retrospective in nature. In addition, the well-known bias to selectively report positive results likely extends to a bias toward reporting antineoplastic results (eg, approximately 350,000 results were found using the medical subject headings term, antineoplastic agents, and only approximately 47,500 for the term carcinogens), which may have affected our findings. Five drugs, including amlodipine,<sup>51</sup> tamsulosin,<sup>52</sup> metformin,<sup>53</sup> warfarin,<sup>54</sup> and lisinopril<sup>55</sup> have published results that report an increased risk of cancer. These unsupported signals could be either false positive or novel findings. Additional research with more careful study designs or in-depth mechanism experiments is required to validate or reject these hypotheses.

This study has limitations. Similar to other epidemiologic studies using observational data, our study may suffer from incomplete information and/or unmeasured confounder effects. It is possible, although unlikely because of the time frame of the analysis, that certain clinicians were aware of the potential anticancer effects of some of the study drugs and were intentionally prescribing them for cancer treatment; temporal resolution of self-administered drug exposures is a difficult and as-yet unsolved problem in clinical data extraction.<sup>56,57</sup> To accommodate the large-scale analysis, our study design is relatively simple: Comparison groups were defined on the basis of mentions of the study drug only without considering the actual exposure details (timing of the drug exposure and drug doses administered) and other potential bias; overall survival, not cancer-specific survival, was used because there were no cancer-specific survival

data. However, because the goal of this study was to generate hypotheses, we expect that more carefully designed studies would evaluate such findings in functional models and/or randomized controlled trials. Furthermore,

the survival model only examined each drug without considering the combinations of variables. Therefore, our method cannot be used to identify the effect of combinations of drugs.

## AFFILIATIONS

<sup>1</sup>The University of Texas Health Science Center at Houston, Houston, TX

<sup>2</sup>University of Florida, Gainesville, FL

<sup>3</sup>Vanderbilt University Medical Center, Nashville, TN

<sup>4</sup>Mayo Clinic, Rochester, MN

<sup>5</sup>Department of Veterans Affairs, Tennessee Valley Healthcare System, Nashville, TN

## CORRESPONDING AUTHOR

Hua Xu, PhD, School of Biomedical Informatics, The University of Texas Health Science Center at Houston, 7000 Fannin St, Suite 800, Houston, TX 77030; e-mail: hua.xu@uth.tmc.edu.

## EQUAL CONTRIBUTION

Y.W. and J.L.W. contributed equally to this study.

## SUPPORT

Supported by grants from the Cancer Prevention & Research Institute of Texas (R1307), National Center for Advancing Translational Sciences (UL1TR000445), National Cancer Institute (P30CA068485, U24CA194215), National Library of Medicine (R01LM010681, R01LM010685), and National Institute of General Medical Sciences (1R01GM103859). The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

## AUTHOR CONTRIBUTIONS

**Conception and design:** Yonghui Wu, Jeremy L. Warner, Qingxia Chen, Ping Yang, Joshua C. Denny, Hongfang Liu, Hua Xu

**Financial support:** Joshua C. Denny, Hongfang Liu

**Administrative support:** Hongfang Liu

**Provision of study material or patients:** Joshua C. Denny, Hongfang Liu

**Collection and assembly of data:** Yonghui Wu, Jeremy L. Warner, Liwei Wang, Min Jiang, Hongfang Liu

**Data analysis and interpretation:** Yonghui Wu, Jeremy L. Warner, Liwei Wang, Jun Xu, Qingxia Chen, Hui Nian, Qi Dai, Xianglin Du, Ping Yang, Hongfang Liu, Hua Xu

## REFERENCES

- Prasad V, Mailankody S: Research and development spending to bring a single cancer drug to market and revenues after approval. *JAMA Intern Med* 177:1569-1575, 2017
- Avorn J: The \$2.6 billion pill—methodologic and policy considerations. *N Engl J Med* 372:1877-1879, 2015
- Dickson M, Gagnon JP: The cost of new drug discovery and development. *Discov Med* 4:172-179, 2004
- Gilbert J, Henske P, Singh A: Rebuilding big pharma's business model. *In Vivo* 21, 2003
- Romero D: To all involved - we have a problem. *Nat Rev Clin Oncol* 15:397, 2018
- Collins FS: Reengineering translational science: The time is right. *Sci Transl Med* 3:90cm17, 2011
- Weir SJ, DeGennaro LJ, Austin CP: Repurposing approved and abandoned drugs for the treatment and prevention of cancer through public-private partnership. *Cancer Res* 72:1055-1058, 2012
- Graul AI, Cruces E, Stringer M: The year's new drugs & biologics, 2013: Part I. *Drugs Today (Barc)* 50:51-100, 2014
- Henry J, Pylpchyk Y, Searcy T, et al: Adoption of electronic health record systems among U.S. non-federal acute care hospitals: 2008-2015. *ONC Data Brief* 35, 2016
- Xu H, Aldrich MC, Chen 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* 22:179-191, 2015

**Manuscript writing:** All authors

**Final approval of manuscript:** All authors

**Accountable for all aspects of the work:** All authors

## AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST Discovery of Noncancer Drug Effects on Survival in Electronic Health Records of Patients With Cancer: A New Paradigm for Drug Repurposing

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to [www.asco.org/rwc](http://www.asco.org/rwc) or [ascopubs.org/cci/author-center](http://ascopubs.org/cci/author-center).

**Jeremy L. Warner**

**Stock and Other Ownership Interests:** HemOnc.org

**Min Jiang**

**Employment:** Pieces Technologies, Eli Lilly

**Stock and Other Ownership Interests:** Eli Lilly

**Qingxia Chen**

**Employment:** HCA Healthcare (I), Cytel (I), J&J Medical Devices (I)

**Stock and Other Ownership Interests:** HCA Healthcare (I), Aimmune Therapeutics

**Consulting or Advisory Role:** Boehringer Ingelheim

**Joshua C. Denny**

**Patents, Royalties, Other Intellectual Property:** Royalties from research methods licensed by Vanderbilt to Nashville Biosciences, a company wholly owned by Vanderbilt (Inst)

**Hua Xu**

**Employment:** Melax Technologies

**Consulting or Advisory Role:** MORE Health, DCHealth Technologies, Hebta

**Patents, Royalties, Other Intellectual Property:** Receive royalties from software license from The University of Texas Health Science Center

No other potential conflicts of interest were reported.

11. Khatri P, Roedder S, Kimura N, et al: A common rejection module (CRM) for acute rejection across multiple organs identifies novel therapeutics for organ transplantation. *J Exp Med* 210:2205-2221, 2013
12. Landman GW, Kleefstra N, van Hateren KJ, et al: Metformin associated with lower cancer mortality in type 2 diabetes: ZODIAC-16. *Diabetes Care* 33:322-326, 2010
13. Currie CJ, Poole CD, Jenkins-Jones S, et al: Mortality after incident cancer in people with and without type 2 diabetes: Impact of metformin on survival. *Diabetes Care* 35:299-304, 2012
14. Evans JM, Donnelly LA, Emslie-Smith AM, et al: Metformin and reduced risk of cancer in diabetic patients. *BMJ* 330:1304-1305, 2005
15. Roden DM, Pulley JM, Basford MA, et al: Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clin Pharmacol Ther* 84:362-369, 2008
16. Xu H, Stenner SP, Doan S, et al: MedEx: A medication information extraction system for clinical narratives. *J Am Med Inform Assoc* 17:19-24, 2010
17. Denny JC, Ritchie MD, Basford MA, et al: PheWAS: Demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics* 26:1205-1210, 2010
18. Benjamini Y, Hochberg Y: Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc B* 57:289-300, 1995
19. McDonald JH: *Handbook of Biological Statistics*. Baltimore, MD, Sparky House, 2009
20. Xu J, Lee HJ, Zeng J, et al: Extracting genetic alteration information for personalized cancer therapy from ClinicalTrials.gov. *J Am Med Inform Assoc* 23:750-757, 2016
21. Ernst MD: Permutation methods: A basis for exact inference. *Stat Sci* 19:676-685, 2004
22. Keiser MJ, Setola V, Irwin JJ, et al: Predicting new molecular targets for known drugs. *Nature* 462:175-181, 2009
23. Ha S, Seo YJ, Kwon MS, et al: IDMap: Facilitating the detection of potential leads with therapeutic targets. *Bioinformatics* 24:1413-1415, 2008
24. Yang L, Agarwal P: Systematic drug repositioning based on clinical side-effects. *PLoS One* 6:e28025, 2011
25. Bisgin H, Liu Z, Fang H, et al: A phenome-guided drug repositioning through a latent variable model. *BMC Bioinformatics* 15:267, 2014
26. Lussier YA, Chen JL: The emergence of genome-based drug repositioning. *Sci Transl Med* 3:96ps35, 2011
27. Chen B, Ma L, Paik H, et al: Reversal of cancer gene expression correlates with drug efficacy and reveals therapeutic targets. *Nat Commun* 8:16022, 2017
28. Tari LB, Patel JH: Systematic drug repurposing through text mining. *Methods Mol Biol* 1159:253-267, 2014
29. Andronis C, Sharma A, Virvilis V, et al: Literature mining, ontologies and information visualization for drug repurposing. *Brief Bioinform* 12:357-368, 2011
30. Swamidass SJ: Mining small-molecule screens to repurpose drugs. *Brief Bioinform* 12:327-335, 2011
31. Pihan E, Colliandre L, Guichou JF, et al: e-Drug3D: 3D structure collections dedicated to drug repurposing and fragment-based drug design. *Bioinformatics* 28:1540-1541, 2012
32. Ye H, Liu Q, Wei J: Construction of drug network based on side effects and its application for drug repositioning. *PLoS One* 9:e87864, 2014
33. Adler NE, Stead WW: Patients in context—EHR capture of social and behavioral determinants of health. *N Engl J Med* 372:698-701, 2015
34. Higurashi T, Hosono K, Takahashi H, et al: Metformin for chemoprevention of metachronous colorectal adenoma or polyps in post-polypectomy patients without diabetes: A multicentre double-blind, placebo-controlled, randomised phase 3 trial. *Lancet Oncol* 17:475-483, 2016
35. Lamming DW, Ye L, Sabatini DM, et al: Rapalogs and mTOR inhibitors as anti-aging therapeutics. *J Clin Invest* 123:980-989, 2013
36. Sinnott-Smith J, Kisfalvi K, Kui R, et al: Metformin inhibition of mTORC1 activation, DNA synthesis and proliferation in pancreatic cancer cells: Dependence on glucose concentration and role of AMPK. *Biochem Biophys Res Commun* 430:352-357, 2013
37. Ghanghas P, Jain S, Rana C, et al: Chemopreventive action of non-steroidal anti-inflammatory drugs on the inflammatory pathways in colon cancer. *Biomed Pharmacother* 78:239-247, 2016
38. Will OM, Purcz N, Chalaris A, et al: Increased survival rate by local release of diclofenac in a murine model of recurrent oral carcinoma. *Int J Nanomedicine* 11:5311-5321, 2016
39. Forget P, Bentin C, Machiels JP, et al: Intraoperative use of ketorolac or diclofenac is associated with improved disease-free survival and overall survival in conservative breast cancer surgery. *Br J Anaesth* 113:i82-i87, 2014
40. Gowda R, Kardos G, Sharma A, et al: Nanoparticle-based celecoxib and plumbagin for the synergistic treatment of melanoma. *Mol Cancer Ther* 16:440-452, 2017
41. Falandry C, Debled M, Bachelot T, et al: Celecoxib and exemestane versus placebo and exemestane in postmenopausal metastatic breast cancer patients: A double-blind phase III GINECO study. *Breast Cancer Res Treat* 116:501-508, 2009
42. Cardwell CR, Hicks BM, Hughes C, et al: Statin use after colorectal cancer diagnosis and survival: A population-based cohort study. *J Clin Oncol* 32:3177-3183, 2014
43. Cardwell CR, Mc Menamin Ú, Hughes CM, et al: Statin use and survival from lung cancer: A population-based cohort study. *Cancer Epidemiol Biomarkers Prev* 24:833-841, 2015
44. Nielsen SF, Nordestgaard BG, Bojesen SE: Statin use and reduced cancer-related mortality. *N Engl J Med* 367:1792-1802, 2012
45. Ellis MJ, Gao F, Dehdashti F, et al: Lower-dose vs high-dose oral estradiol therapy of hormone receptor-positive, aromatase inhibitor-resistant advanced breast cancer: A phase 2 randomized study. *JAMA* 302:774-780, 2009
46. Castro MP, Whitcomb BP, Zajchowski DA, et al: Successful use of next generation genomic sequencing (NGS)-directed therapy of clear cell carcinoma of the ovary (CCCO) with trametinib and metformin in a patient with chemotherapy-refractory disease. *Gynecol Oncol Res Pract* 2:4, 2015
47. Seckl MJ, Ottensmeier CH, Cullen M, et al: Multicenter, phase III, randomized, double-blind, placebo-controlled trial of pravastatin added to first-line standard chemotherapy in small-cell lung cancer (LUNGSTAR). *J Clin Oncol* 35:1506-1514, 2017
48. Coyne GO, Takebe N, Chen AP: Defining precision: The precision medicine initiative trials NCI-MPACT and NCI-MATCH. *Curr Probl Cancer* 41:182-193, 2017
49. Ronquist G, Rodríguez LA, Ruigómez A, et al: Association between captopril, other antihypertensive drugs and risk of prostate cancer. *Prostate* 58:50-56, 2004
50. Ronquist G, Frithz G, Wang YH, et al: Captopril may reduce biochemical (prostate-specific antigen) failure following radical prostatectomy for clinically localized prostate cancer. *Scand J Urol Nephrol* 43:32-36, 2009
51. Chae YK, Dimou A, Pierce S, et al: The effect of calcium channel blockers on the outcome of acute myeloid leukemia. *Leuk Lymphoma* 55:2822-2829, 2014
52. Keledjian K, Kyprianou N: Anoikis induction by quinazoline based alpha 1-adrenoceptor antagonists in prostate cancer cells: Antagonistic effect of bcl-2. *J Urol* 169:1150-1156, 2003
53. Baghrioli MI, de Celis Ferrari AC, Pfiffer TE, et al: Phase II trial of metformin and paclitaxel for patients with gemcitabine-refractory advanced adenocarcinoma of the pancreas. *Ecancermedicallscience* 9:563, 2015



54. Meyer G, Marjanovic Z, Valcke J, et al: Comparison of low-molecular-weight heparin and warfarin for the secondary prevention of venous thromboembolism in patients with cancer: A randomized controlled study. *Arch Intern Med* 162:1729-1735, 2002
55. Friedman GD, Asgari MM, Warton EM, et al: Antihypertensive drugs and lip cancer in non-Hispanic whites. *Arch Intern Med* 172:1246-1251, 2012
56. Lin C, Dligach D, Miller TA, et al: Multilayered temporal modeling for the clinical domain. *J Am Med Inform Assoc* 23:387-395, 2016
57. Sun W, Rumshisky A, Uzuner O: Evaluating temporal relations in clinical text: 2012 i2b2 challenge. *J Am Med Inform Assoc* 20:806-813, 2013

