Original Investigation

Clinical Evidence Supporting Pharmacogenomic Biomarker Testing Provided in US Food and Drug Administration Drug Labels

Bo Wang, PharmD; William J. Canestaro, MSc; Niteesh K. Choudhry, MD, PhD

← Invited Commentary

+ Supplemental content at jamainternalmedicine.com

IMPORTANCE Genetic biomarkers that predict a drug's efficacy or likelihood of toxicity are assuming increasingly important roles in the personalization of pharmacotherapy, but concern exists that evidence that links use of some biomarkers to clinical benefit is insufficient. Nevertheless, information about the use of biomarkers appears in the labels of many prescription drugs, which may add confusion to the clinical decision-making process.

OBJECTIVE To evaluate the evidence that supports pharmacogenomic biomarker testing in drug labels and how frequently testing is recommended.

DATA SOURCES Publicly available US Food and Drug Administration databases.

MAIN OUTCOMES AND MEASURES We identified drug labels that described the use of a biomarker and evaluated whether the label contained or referenced convincing evidence of its clinical validity (ie, the ability to predict phenotype) and clinical utility (ie, the ability to improve clinical outcomes) using guidelines published by the Evaluation of Genomic Applications in Practice and Prevention Working Group. We graded the completeness of the citation of supporting studies and determined whether the label recommended incorporation of biomarker test results in therapeutic decision making.

RESULTS Of the 119 drug-biomarker combinations, only 43 (36.1%) had labels that provided convincing clinical validity evidence, whereas 18 (15.1%) provided convincing evidence of clinical utility. Sixty-one labels (51.3%) made recommendations about how clinical decisions should be based on the results of a biomarker test; 36 (30.3%) of these contained convincing clinical utility data. A full description of supporting studies was included in 13 labels (10.9%).

CONCLUSIONS AND RELEVANCE Fewer than one-sixth of drug labels contained or referenced convincing evidence of clinical utility of biomarker testing, whereas more than half made recommendations based on biomarker test results. It may be premature to include biomarker testing recommendations in drug labels when convincing data that link testing to patient outcomes do not exist.

Author Affiliations: Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts (Wang, Canestaro, Choudhry); Department of Pharmacy, University of Washington, Seattle (Canestaro).

Corresponding Author: Niteesh K. Choudhry, MD, PhD, Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, 1620 Tremont St, Ste 3030, Boston, MA 02120 (nchoudhry@partners .org).

The ability to target and tailor drug therapies based on genetic information has created much hope that personalization of pharmacotherapy will revolutionize health care. This enthusiasm is supported by a vast literature that evaluates a variety of biomarkers with polymorphisms that have predictive significance and can potentially improve a drug's efficacy or safety profile. On the basis of this literature, the US Food and Drug Administration (FDA) has included pharmacogenomic information in the physician prescribing information (drug labels) of more than 100 drugs. For example, the label for the oral anticoagulant warfarin suggests dosage adjustments based on a patient's genotype for *CYP2C9* and *VKORC1*, whereas the label for abacavir sulfate, an antiretroviral, recommends avoiding its use in patients who screen positive for the *HLA-B*5701* allele.

Prescription drug labels are an important source of information about drug therapies for many health care professionals,[1] and the information contained in them also appears in other frequently consulted references, such as UpToDate,[2] Micromedex,[3] and the *Physicians' Desk Reference*.[4] Because only 1 in 10 US physicians reports being adequately informed about the appropriate use of pharmacogenomic biomarkers,[5] the information and recommendations included in labels should be not only evidence based but also directly relevant to clinical decision making. However, despite their inclusion in drug labels, the use of many biomarkers does not appear to be clearly associated with health benefits. For example, although the label for warfarin contains a recommended dosing algorithm based on *CYP2C9* and *VKORC1* polymorphisms, the clinical utility of genotype-based dosing for this medication (ie, the ability to improve clinical outcomes compared with management without genetic testing[6]) remains unclear.[7-10] Furthermore, inclusion of potentially tenuous recommendations or associations within a drug's label may encourage health care professionals to order tests or lead them to change therapies based on limited evidence, even if the label does not recommend explicit action based on biomarker status. This outcome is particularly relevant given the increasing cost associated with genetic tests, which is projected to increase at more than twice the rate of overall health care spending and reach $10 billion in the United States by 2015.[11,12] We sought to determine the level of evidence that supports the use of pharmacogenomic biomarker testing in drug labels and how frequently their testing is recommended, as well as to examine how completely these supporting studies are cited in the labels.

Methods

Data Sources

Because all data came from public sources, this study was not human subject research and did not require IRB review or informed consent. We identified medications that contained biomarker information in their labels from the FDA Table of Pharmacogenomic Biomarkers in Drug Labeling, a publicly accessible database.[13] This database contains all FDA-approved drugs with pharmacogenomic information in their labels and is updated on a regular basis by the agency. For each

medication, we gathered the earliest available drug label that contained mention of its associated biomarker(s) from Drugs@FDA, another publicly accessible FDA database available through the agency's website that lists regulatory actions, including initial approvals and drug labeling changes.[14] We focused on this particular label to evaluate the quality of the cited evidence supporting testing recommendations and accessibility of these supporting studies at the time when physicians could have first become aware of this pharmacogenomic information through the drug label. We could not gather the drug labels from this FDA database for 4 of the drugs in our study (nefazodone, propafenone, protriptyline, and thioridazine) and instead gathered their labels from other databases.[15,16] Because information about a single biomarker may have been included in the label of several different drugs and because several drug labels contained information about more than one biomarker, the drug-biomarker combination was our unit of analysis.

Data Extraction and Evaluation

Two of the authors (B.W. and W.J.C.) independently evaluated the label(s) for the quality of cited clinical evidence, the completeness of the citation of supporting studies, and the presence of recommendation for use of biomarker testing, with disagreements resolved by consensus (eMethods in the Supplement).

Strength of Evidence for Clinical Validity and Utility

Using guidelines from the Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group,[17] we categorized the robustness of evidence for each biomarker to support its clinical validity (ie, the ability to predict phenotype) and clinical utility (ie, the improvement of clinical outcomes as a result of the use of the biomarker compared with management without genetic testing) as convincing, adequate, or incomplete (based on the criteria that the EGAPP Working Group uses to define data as inadequate) (**Table 1**). The EGAPP Working Group was formed by the Centers for Disease Control and Prevention and is charged with establishing a systematic, evidence-based approach to assessing genetic tests and other applications of genomic technology. To mirror how practicing health care professionals are likely to interpret the evidence cited within a label, we graded the quality of evidence that supports a biomarker's clinical validity and utility based on the information presented within the drug labels and evaluated other relevant studies only when a citation was included in the label, allowing for their location on the PubMed/MEDLINE database.[18] We did not gather or evaluate other sources of information physicians may use to inform their practice.

If the drug label contained evidence that addressed the clinical validity of a drug-biomarker association, including pharmacokinetic studies, but provided an insufficient description and citation to identify its design, we rated the evidence of clinical validity as incomplete. For example, we considered a label for drug X that listed clinical studies as having demonstrated a decreased hemoglobin level in patients with *G6PD* deficiency but that did not provide or cite information to iden-

Table 1. Categorization of Clinical Evidence in Drug Labels[a]

Quality	Clinical Validity[b]	Clinical Utility[c]
Convincing	High-quality longitudinal cohort study Validated clinical decision rule Systematic review or meta-analysis of well-designed longitudinal cohort studies with homogeneity Systematic review or meta-analysis of randomized clinical trials or uncontrolled interventional trials (or substudies of these trials)[c] Randomized clinical trial or uncontrolled interventional trial[c]	Systematic review or meta-analysis of randomized clinical trials showing consistency in results At least one large randomized clinical trial
Adequate	Systematic review of lower-quality studies Case-control study with good reference standards Unvalidated clinical decision rule Substudy of large randomized clinical trial[d]	Systematic review of randomized clinical trials with heterogeneity One or more clinical trials without randomization Systematic review of cohort studies with consistent results
Incomplete	Single case-control study using nonconsecutive cases or lacking consistently applied reference standards Cohort or case-control study in which the reference standard is defined by the test or is not used systematically or in which the study is not masked Case report[d] or case series Unpublished and/or non-peer-reviewed research, clinical laboratory, or manufacturer data Consensus guidelines No relevant studies or inability to classify study design based on description in drug label[d]	Systematic review of nonrandomized clinical trials, uncontrolled interventional trials, cohort studies, or case-control studies with heterogeneity Single cohort or case-control study Case series Unpublished and/or non-peer-reviewed research, clinical laboratory, or manufacturer data One or more uncontrolled interventional trials[d] No relevant studies or inability to classify study design based on description in drug label[d]

[a] Table adapted from Teutsch et al.[17]

[b] Clinical validity of a genetic test is defined by the Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group as "its ability to accurately and reliably predict the clinically defined disorder or phenotype of interest."

[c] Clinical utility of a genetic test is defined by the EGAPP Working Group as the "evidence of improved measurable clinical outcomes, and its usefulness and added value to patient management decision-making compared with current management without genetic testing."

[d] These study designs were added to allow for classification of evidence that did not fit into the EGAPP categorization framework.

tify the study design as incomplete evidence for clinical validity of the drug X–*G6PD* association. We also rated the evidence of clinical validity as incomplete if the drug label did not contain or cite any evidence that addressed the clinical validity of the drug-biomarker association.

For evidence in drug labels to support the clinical utility of biomarker testing, we required that it demonstrated improved patient outcomes when biomarker testing was incorporated into treatment decisions compared with when no testing was performed. For targeted therapies in which the medication was developed specifically to interfere with or manipulate a certain biomarker to have its intended effect (eg, trastuzumab was developed specifically to target the *ERBB2* [formerly *HER2* or *HER2/neu*] biomarker), we did not require the presence of a nontesting group for the evidence to support clinical utility but instead focused our assessment on studies that evaluated the drug's efficacy in the intended patient population (ie, those who screened positive for the targeted biomarker). We erred on the side of considering evidence as convincing when characterizing the quality of a particular study design or when determining whether the evidence addressed the clinical validity and/or utility for each drug-biomarker combination (Table 1).

Completeness of Citation of Supporting Studies

We graded the completeness of citation of supporting studies in drug labels as full, partial, or none. A grade of full was conferred for labels that included a sufficient citation of supporting studies in a separate references section or by name (eg, PREDICT-1 study for abacavir) such that the studies can be located in the PubMed/MEDLINE database. A grade of partial was given for labels that did not include citations for supporting studies but described the study design and/or results. For clinical va-

lidity, these results include variations in levels of a drug based on status of the associated biomarker. A grade of none was reserved for labels that made no description of the supporting studies. When the completeness of description of supporting studies in drug labels differed between clinical validity and clinical utility, we used the higher grade for our analysis.

Treatment Recommendations Contained Within the Label

We evaluated whether the label recommended incorporation of the biomarker test result in therapeutic decision making. These recommendations were categorized as being based directly on a drug's mechanism of action, as indicated by mention in the Indications or Mechanism of Action sections of the drug label (eg, trastuzumab's targeting of *ERBB2* overexpression or busulfan's treatment of chronic myelogenous leukemia, a condition characterized by a translocation in the Philadelphia chromosome biomarker, were interpreted as implicit recommendations for biomarker testing before treatment initiation), being based on drug-biomarker associations (dosage adjustment, contraindication or avoidance, and follow-up laboratory testing), or being absent. A recommendation was considered to be present even when it advocated for alterations in therapy of another drug taken concurrently as a result of drug-drug interactions as long as the interaction was due to polymorphisms in the given biomarker.

Statistical Analysis

We used descriptive statistics to summarize the supporting evidence and recommendations contained in the drug labels for each drug-biomarker combination. We then performed a Fisher exact test to determine whether biomarkers for targeted therapies (ie, medications developed specifically to interfere with or manipulate a biomarker to have its intended effect) dem-

Table 2. Features of Evidence and Recommendations in Labels Stratified by Therapeutic Area

Quality of Cited Evidence	Therapeutic Areas, No. (%)						
	All (N = 119)	Oncology (n = 37)	Neuropsychiatry (n = 33)	Gastroenterology (n = 9)	Infectious Disease (n = 9)	Cardiovascular Disease (n = 8)	Other (n = 23)
Clinical validity							
Convincing	43 (36.1)	24 (64.9)	2 (6.1)	2 (22.2)	5 (55.6)	1 (12.5)	9 (39.1)
Adequate	6 (5.0)	4 (10.8)	1 (3.0)	0	0	1 (12.5)	0
Incomplete	70 (58.8)	9 (24.3)	30 (90.9)	7 (77.8)	4 (44.4)	6 (75.0)	14 (60.9)
Clinical utility							
Convincing	18 (15.1)	14 (37.8)[a]	0[b]	0	2 (22.2)	0	2 (8.7)
Adequate	0	0	0	0	0	0	0
Incomplete	101 (84.9)	23 (62.2)	33 (100)	9 (100)	7 (77.8)	8 (100)	21 (91.3)
Completeness of study citation							
Full	13 (10.9)	3 (8.1)	0	0	5 (55.6)	2 (25.0)	3 (13.0)
Partial	70 (58.8)	28 (75.7)	16 (48.5)	7 (77.8)	3 (33.3)	5 (62.5)	11 (47.8)
None	36 (30.3)	6 (16.2)	17 (51.5)	2 (22.2)	1 (11.1)	1 (12.5)	9 (39.1)
Presence of testing recommendation							
On the basis of mechanism of action	34 (28.6)	26 (70.3)	0	2 (22.2)	1 (11.1)	0	5 (21.7)
On the basis of drug-biomarker associations	27 (22.7)	3 (8.1)	14 (42.4)	1 (11.1)	1 (11.1)	1 (12.5)	7 (30.4)
Absent	58 (48.7)	8 (21.6)	19 (57.6)	6 (66.7)	7 (77.8)	7 (87.5)	11 (47.8)

[a] P < .001 compared with all other disease groupings (14 of 37 [37.8%] vs 4 of 82 [4.9%]).

[b] P < .001 compared with all other disease groupings (0 of 33 [0%] vs 18 of 86 [20.9%]).

onstrated a different proportion of convincing data that supported clinical utility compared with nontargeted therapies and to compare the quality of cited evidence in labels with and without testing recommendations. Because numerous cancer agents were specifically developed to be targeted therapies, whereas most neuropsychiatric drug-biomarker combinations are discovered after approval and are related to adverse events, we performed a prespecified Fisher exact test to determine whether biomarkers for oncology drugs and neuropsychiatry drugs demonstrated a different proportion of convincing evidence supporting clinical utility in their labels compared with other biomarkers.

Results

We identified 119 drug-biomarker combinations, representing 107 drugs and 39 unique biomarkers (eTable in the Supplement). Most of these drug-biomarker combinations (75 [63.0%]) are intended to reduce the occurrence of adverse drug events, whereas the remainder (44 [37.0%]) relate to the drugs' efficacy. The most common clinical specialties covered by these biomarkers were oncology (37 [31.1%]), neuropsychiatry (33 [27.7%]), gastroenterology (9 [7.6%]), infectious disease (9 [7.6%]), and cardiovascular disease (8 [6.7%]) (**Table 2**).

Evidence of Clinical Validity and Utility
Forty-three drug labels (36.1%) provided convincing evidence of the clinical validity of the biomarker (ie, the ability of the biomarker to predict the phenotype of interest), whereas 18 (15.1%) provided convincing evidence that the use of the bio-

marker has clinical utility (ie, the biomarker's ability to improve clinical outcomes). **Table 3** contains examples of our grading approach. Seventy-six drug labels (63.9%) did not provide convincing evidence of clinical validity or utility. Biomarkers for cancer drugs were much more likely to demonstrate convincing evidence supporting clinical utility in their labels compared with all other biomarkers (14 of 37 [37.8%] vs 4 of 82 [4.9%], P < .001), whereas neuropsychiatry biomarkers were less likely to demonstrate convincing clinical utility evidence in their labels than the remaining biomarkers (0 of 33 vs 18 of 86 [20.9%], P < .001). Targeted therapies consisted mainly of oncology drugs (26 of 34 [76.5%]). Seventeen targeted therapies (50.0%) contained convincing data that supported clinical utility in their labels compared with 1 of 85 (1.2%) nontargeted therapies (P < .001). Abacavir, whose label recommends it not be used in patients who screen positive for the *HLA-B*5701* allele, was the only nontargeted therapy whose biomarker was supported by convincing clinical utility evidence.

Thirteen labels (10.9%) contained a full citation of supporting studies, whereas 36 (30.3%) made no mention of the scientific literature that supported the biomarker it discussed.

Treatment Recommendations Contained Within the Label
Sixty-one labels (51.3%) made recommendations about how clinical decisions should be based on the results of a biomarker test, with 34 (28.6%) being based directly on the drug's mechanism of action and 27 (22.7%) being based on drug-biomarker associations. Among biomarkers with neither convincing clinical utility nor validity data, 24 of 76 labels (31.6%) still contained testing recommendations. The drug label for the

Table 3. Examples of Evidence Evaluation for Clinical Validity and Utility

Drug Name	Biomarker	Excerpt From Drug Label of Studies Supporting Clinical Validity and/or Clinical Utility	Study Design/Grade	
			Validity	Utility
Abacavir	HLA-B*5701	Validity and utility: "CNA106030 (PREDICT-1), a randomized, double-blind study, evaluated the clinical utility of prospective HLA-B*5701 screening on the incidence of abacavir hypersensitivity reaction in abacavir-naive HIV-1-infected adults (n = 1650). In this study, use of pre-therapy screening for the HLA-B*5701 allele and exclusion of subjects with this allele reduced the incidence of clinically suspected abacavir hypersensitivity reactions from 7.8% (66/847) to 3.4% (27/803)..."[a]	Randomized clinical trial/ convincing	Randomized clinical trial/ convincing
Clopidogrel	CYP2C19	Validity: "The relationship between CYP2C19 genotype and Plavix treatment outcome was evaluated in retrospective analyses of Plavix-treated subjects in CHARISMA (n=4862) and TRITON-TIMI 38 (n=1477), and in several published cohort studies. In TRITON-TIMI 38 and the majority of the cohort studies, the combined group of patients with either intermediate or poor metabolizer status had a higher rate of cardiovascular events (death, myocardial infarction, and stroke) or stent thrombosis compared to extensive metabolizers. In CHARISMA and one cohort study, the increased event rate was observed only in poor metabolizers."[b] Utility: no relevant data	Systematic review or meta-analysis of randomized clinical trials/ convincing	No relevant studies/ incomplete
Carbamazepine	HLA-B*1502	Validity: "Retrospective case-control studies have found that in patients of Chinese ancestry there is a strong association between the risk of developing SJS/TEN with carbamazepine treatment and the presence of an inherited variant of the HLA-B gene, HLA-B*1502. The occurrence of higher rates of these reactions in countries with higher frequencies of this allele suggests that the risk may be increased in allele-positive individuals of any ethnicity."[c] Utility: no relevant data	Case-control study with good reference standards/ adequate	No relevant studies/ incomplete
Fluorouracil	DPD	Validity: "One case of life-threatening systemic toxicity has been reported with the topical use of Efudex in a patient with DPD enzyme deficiency..."[d] Utility: no relevant data	Case report/ incomplete	No relevant studies/ incomplete

Abbreviations: CHARISMA, Clopidogrel for High Atherothrombotic Risk and Ischemic Stabilization, Management and Avoidance; HIV-1, human immunodeficiency virus 1; SJS/TEN, Stevens-Johnson syndrome and toxic epidermal necrolysis; TRITON-TIMI 38, Trials to Assess Improvement in Therapeutic Outcomes by Optimizing Platelet Inhibition with Prasugrel-Thrombolysis in Myocardial Infarction.

[a] Ziagen [package insert]. Research Triangle Park, NC: GlaxoSmithKline; 2008. http://www.accessdata.fda.gov/drugsatfda_docs/label/2008/020977s017,020978s020lbl.pdf. Accessed September 19, 2013.

[b] Plavix [package insert]. Bridgewater, NJ: Bristol-Myers Squibb/Sanofi

Pharmaceuticals Partnership; 2010. http://www.accessdata.fda.gov/drugsatfda_docs/label/2010/020839s042lbl.pdf. Accessed September 19, 2013.

[c] Tegretol [package insert]. East Hanover, NJ: Novartis Pharmaceuticals Corporation; 2007. http://www.accessdata.fda.gov/drugsatfda_docs/label/2007/016608s098lbl.pdf. Accessed September 19, 2013.

[d] Efudex [package insert]. Costa Mesa, CA: ICN Pharmaceuticals Inc; 2004. http://www.accessdata.fda.gov/drugsatfda_docs/label/2004/16831slr047_efudex_lbl.pdf. Accessed September 19, 2013.

psychotropic drug iloperidone, for instance, recommended consideration of dose adjustments based on patients' CYP2D6 status despite lack of clinical utility and validity data. Similarly, among labels that made recommendations, only 18 of 61 (29.5%) provided convincing clinical utility data. Labels that made testing recommendations were more likely to provide convincing clinical utility data compared with labels that made no recommendations (36 [30.3%] vs 0 [0%], P < .001).

Discussion

Overall, our analysis revealed deficiencies in the evidence provided in drug labels that supports the use of many pharmacogenomic biomarkers, with fewer than one-sixth of labels containing or citing convincing evidence for clinical utility and almost two-thirds even lacking convincing data for clinical validity. This deficiency was especially prominent among biomarkers of nontargeted therapies, which constituted more than 70% of our study sample.

The limited amount of convincing evidence for the clinical utility of pharmacogenomic biomarkers in the labels we reviewed is perhaps not surprising given the difficulty of demonstrating that a test alters clinical outcomes rather than simply predicts a disorder or phenotype. Nevertheless, the primary goal and potential of pharmacogenomics and personalized medicine is to change patient outcomes. In our opinion, it is

premature to include testing recommendations in labels when such utility data are neither described nor cited in this resource, even if it exists elsewhere. It could be argued that highly-quality data indicating that individuals with certain polymorphisms metabolize a drug differently may be relevant to patient care and therefore should be included in drug labels as the basis for biomarker testing. However, other than predictive value, the inclusion of this type of information does not provide guidance as to what health care professionals should actually do when they get the test results. The case of CYP2C19 testing for clopidogrel illustrates this well. Polymorphisms of this gene are associated with poorer drug metabolism and a higher risk of thrombotic events, prompting testing recommendations for these variants to be added to the drug's label in 2010. However, the consequences of the clinical actions that might rationally be taken based on this information have still not been convincingly evaluated. If this drug's dose were increased or an alternative drug with a less favorable risk-benefit profile were chosen on the basis of what is actually an uninformative biomarker result, it may lead to worse treatment outcomes while contributing to increasing expenditures at a time when our health care system is least able to afford it.

As a result, we believe that testing recommendations supported by clinical validity alone adds confusion, not clarity, to the clinical decision-making process, especially if the evidence is not clearly explicated or cited alongside the guidance. There may also be legal implications regarding the in-

clusion of testing recommendations in drug labels. Could prescribers be liable for adverse events that may have been predicted by biomarker testing, regardless of the evidence base for the recommendations? If not, at what level of evidence should the reasonableness standard to test apply?

A multipronged approach should be used to address the current situation. At minimum, an explicit statement about the quality of clinical utility evidence for each testing recommendation should be presented in the labels rather than charging health care professionals with the task of extrapolating quality based on the data presented. Alongside this statement should be complete references to help patients and health care professionals easily access the full studies for further assessment. More stringently, the FDA could issue regulations to only include information about pharmacogenomic biomarkers if compelling clinical utility information has been generated, although exceptions should be made for drug-biomarker combinations associated with efficacy or safety end points of particular significance, such as clopidogrel and *CYP2C19*.

The perceived lack of incentive for biomarker test developers and pharmaceutical manufacturers to conduct robust studies that describe the clinical utility of biomarker testing, which may demonstrate the lack of need for biomarker testing or reduce the number of patients eligible for a particular drug, poses another challenge. However, establishing the clinical utility of a biomarker to target a particular high-risk patient population with high-quality studies may increase drug sales and biomarker test orders, as was the case with abacavir.[19,20] To provide further incentive, the FDA could waive user fees and/or prioritize the review of a subsequent drug application for manufacturers who conduct robust clinical utility trials in support of their approval applications. In addition, governmental agencies can directly support trials that investigate drug-biomarker combinations of particular clinical significance, as has been done for drugs for which manufacturers have lacked specific incentive to do so.[21]

In the meantime, physicians and other prescribers should be aware of the relative lack of evidence to support many treatment recommendations pertaining to biomarkers that are contained within the labels and should instead scrutinize the primary literature that supports these recommendations before taking clinical action. Our finding that more than two-thirds of labels that make testing recommendations do not contain convincing evidence to support clinical utility of the biomarker further reinforces the need for skepticism.

Our study has several limitations. For each medication, we evaluated the earliest accessible label that contained mention of each associated biomarker's polymorphisms to assess the evidence available to physicians at the time the biomarker was first included in the label. The initial labels of some of the drugs in our study were not accessible, which may have resulted in our evaluation of subsequent labels with updated pharmacogenomic evidence and recommendations. This discrepancy may have resulted in an overestimate of the strength of the cited evidence. Moreover, studies that found a lack of clinical validity or utility of drug-biomarker associations may not be included in the drug labels, which are drafted by manufacturers. However, because the FDA reviews and regulates the content of the labels, it should still require that major studies of such nature be included in subsequent labeling revisions. In addition, we assessed the quality of clinical evidence only of information presented within the drug labels and additional full studies adequately referenced in these documents; we did not evaluate other sources of evidence, including additional primary literature or FDA medical reviews. Our rationale for doing this was 2-fold: drug labels should be self-sufficient in presenting time-constrained prescribers with the evidence base and references that support incorporation of biomarker testing into clinical practice, and the content presented in the labels is often used to inform other widely used tertiary-level drug information resources. The results of our analysis are consistent with the 3 systematic reviews of drug-biomarker combinations that the EGAPP Working Group has conducted, and our analysis applied its evaluative framework: selective serotonin reuptake inhibitors and *CYP450*s,[22] *UGT1A1* and irinotecan,[23] and *EGFR* and cetuximab and panitumumab.[24]

In addition, although the FDA has generally required less robust evidence for approval of oncology drugs with the goal of accelerating access for patients, we applied the same methods in our evaluation of labels for biomarkers associated with these drugs to generate a standardized description of the evidence base available. Attention to the wide variation of clinical trial evidence submitted to the FDA to support the successful approval of novel agents,[25] coupled with calls for more rigorous oncology trials to better reflect developments in treatment paradigms and clinical outcomes of different cancer types,[26] further supports the need to take a uniform approach to understanding the existing scientific landscape. Despite the decreased rigor required for approval of certain cancer medications, our analysis revealed that biomarkers for oncology drugs still had a much higher level of convincing evidence base in their labels than biomarkers for drugs that treat other conditions. Future discussions and analyses should explore whether the current EGAPP Working Group rating criteria should be customized for different types of diseases or different prevalence of adverse effects. For example, drug-biomarker combinations that are designed to address rare adverse effects may be better evaluated with well-designed observational studies to generate evidence of clinical utility rather than relying exclusively on randomized clinical trials.

Conclusions

There is reason to be enthusiastic about the potential of biomarkers to enhance clinical care, and our analysis identified many examples of tests that have convincing evidence of their ability to meaningfully improve health care outcomes in patients with common conditions. However, other less evidence-based labeling recommendations highlight the need for clearer guidance on their optimal use. Until this problem is addressed, physicians are left with the challenging task of navigating a sea of guidance with varying foundations of clinical support in pursuit of practicing clinically sound and cost-conscious medicine, a challenge that will likely increase in cadence with the growth of the pharmacogenomics field.

REFERENCES

1. Stanek EJ, Sanders CL, Frueh FW. Physician awareness and utilization of Food and Drug Administration (FDA)-approved labeling for pharmacogenomic testing information. *J Pers Med*. 2013;3:111-123.

2. Davies KS. Physicians and their use of information: a survey comparison between the United States, Canada, and the United Kingdom. *J Med Libr Assoc*. 2011;99(1):88-91.

3. Micromedex Mobile App General and Technical FAQs. Truven Health Analytics, 2014. June 13, 2014. http://micromedex.com/support/frequently-asked -questions/micromedex-20-mobile-app-iphone. Accessed June 13, 2014.

4. Uhl K, Kennedy DL, Kweder SL. Risk management strategies in the Physicians' Desk Reference product labels for pregnancy category X drugs. *Drug Saf*. 2002;25(12):885-892.

5. Stanek EJ, Sanders CL, Taber KA, et al. Adoption of pharmacogenomic testing by US physicians: results of a nationwide survey. *Clin Pharmacol Ther*. 2012;91(3):450-458.

6. Ong FS, Deignan JL, Kuo JZ, et al. Clinical utility of pharmacogenetic biomarkers in cardiovascular therapeutics: a challenge for clinical implementation. *Pharmacogenomics*. 2012;13(4):465-475.

7. Kimmel SE, French B, Kasner SE, et al; COAG Investigators. A pharmacogenetic versus a clinical algorithm for warfarin dosing. *N Engl J Med*. 2013; 369(24):2283-2293.

8. Verhoef TI, Ragia G, de Boer A, et al; EU-PACT Group. A randomized trial of genotype-guided dosing of acenocoumarol and phenprocoumon. *N Engl J Med*. 2013;369(24):2304-2312.

9. Pirmohamed M, Burnside G, Eriksson N, et al; EU-PACT Group. A randomized trial of genotype-guided dosing of warfarin. *N Engl J Med*. 2013;369(24):2294-2303.

10. Stergiopoulos K, Brown DL. Genotype-guided vs clinical dosing of warfarin and its analogues: meta-analysis of randomized clinical trials. *JAMA Intern Med*. 2014;174(8):1330-1338.

11. National Health Expenditure Projections 2012-2022. Centers for Medicare and Medicaid Services. June 21, 2014. http://www.cms.gov /Research-Statistics-Data-and-Systems/Statistics -Trends-and-Reports/NationalHealthExpendData /downloads/proj2012.pdf. Accessed June 21, 2014.

12. PricewaterhouseCoopers. The new science of personalized medicine: translating the promise into practice. 2009. http://www.pwc.com/us/en /healthcare/publications/personalized-medicine .jhtml. Accessed June 11, 2014.

13. US Food and Drug Administration. Table of pharmacogenomic biomarkers in drug labeling. http://www.fda.gov/drugs/scienceresearch /researchareas/pharmacogenetics /ucm083378.htm. Accessed April 13, 2013.

14. US Food and Drug Administration. Drugs@FDA: FDA approved drug products. http://www .accessdata.fda.gov/scripts/cder/drugsatfda/. Accessed September 15, 2013.

15. Pharm GKB. The Pharmacogenomics Knowledgebase. http://www.pharmgkb.org/. Accessed May 3, 2013.

16. DailyMed. http://dailymed.nlm.nih.gov /dailymed/about.cfm. Accessed December 3, 2013.

17. Teutsch SM, Bradley LA, Palomaki GE, et al; EGAPP Working Group. The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) initiative: methods of the EGAPP working group. *Genet Med*. 2009;11(1):3-14.

18. National Center for Biotechnology Information. US National Library of Medicine, 2014. https://www.ncbi.nlm.nih.gov/pubmed/. Accessed June 14, 2014.

19. Stratified medicine and pharmacogenomics. Priority medicines for Europe and the world 2013 update. World Health Organization, 2013. http://www.who.int/medicines/areas/priority _medicines/Ch7_4Stratified.pdf. Accessed June 14, 2014.

20. Lai-Goldman M, Faruki H. Abacavir hypersensitivity: a model system for pharmacogenetic test adoption. *Genet Med*. 2008; 10(12):874-878.

21. Furberg CD, Wright JT Jr, Davis BR, et al; ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group. The Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial. Major outcomes in high-risk hypertensive patients randomized to angiotensin-converting enzyme inhibitor or calcium channel blocker vs diuretic: the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT). *JAMA*. 2002;288(23):2981- 2997.

22. Matchar DB, Thakur ME, Grossman I, et al. Testing for cytochrome P450 polymorphisms in adults with non-psychotic depression treated with selective serotonin reuptake inhibitors (SSRIs). *Evid Rep Technol Assess (Full Rep)*. 2007;(146):1-77.

23. Palomaki GE, Bradley LA, Douglas MP, Kolor K, Dotson WD. Can *UGT1A1* genotyping reduce morbidity and mortality in patients with metastatic colorectal cancer treated with irinotecan? an evidence-based review. *Genet Med*. 2009;11(1):21- 34.

24. Lin JS, Webber EM, Senger CA, Holmes RS, Whitlock EP. Systematic review of pharmacogenetic testing for predicting clinical benefit to anti-EGFR therapy in metastatic colorectal cancer. *Am J Cancer Res*. 2011;1(5):650-662.

25. Downing NS, Aminawung JA, Shah ND, Krumholz HM, Ross JS. Clinical trial evidence supporting FDA approval of novel therapeutic agents, 2005-2012. *JAMA*. 2014;311(4):368-377.

26. Hirsch BR, Califf RM, Cheng SK, et al. Characteristics of oncology clinical trials: insights from a systematic analysis of ClinicalTrials.gov. *JAMA Intern Med*. 2013;173(11):972-979.