

Abandoning personalization to get to precision in the pharmacotherapy of depression

Roy H. Perlis

Center for Experimental Drugs and Diagnostics, Department of Psychiatry and Center for Human Genetic Research, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

Effectiveness studies and analyses of naturalistic cohorts demonstrate that many patients with major depressive disorder do not experience symptomatic remission with antidepressant treatments. In an effort to better match patients with effective treatments, numerous investigations of predictors or moderators of treatment response have been reported over the past five decades, including clinical features as well as biological measures. However, none of these have entered routine clinical practice; instead, clinicians typically personalize treatment on the basis of patient preferences as well as their own. Here, we review the reasons why it has been challenging to identify and deploy treatment-specific predictors of response, and suggest strategies that may be required to achieve true precision in the pharmacotherapy of depression. We emphasize the need for changes in how depression care is delivered, measured, and used to inform future practice.

Key words: Antidepressants, major depression, precision medicine, risk stratification, personalized medicine, biomarkers, treatment matching

(*World Psychiatry* 2016;15:228–235)

After decades of effort to identify predictors of antidepressant treatment response, including more than 100 publications reporting genetic predictors, the approach to treating major depressive disorder remains one of trial and error. Initial management strategies vary widely across providers and health systems¹. Next-step treatment is marked by even greater variation². A recent survey of psychopharmacologists, for example, revealed roughly equal split between within- and across-antidepressant class switch following non-response to initial treatment³. This trial-and-error approach clearly matters to patients: a survey of Danish patients found that they would pay up to \$280 to avoid a single medication change⁴.

At the same time, pharmacogenomics has already made some clinical inroads in antidepressant prescribing. Among the more than 100 medication labels approved by the US Food and Drug Administration (FDA) that include information on genetic variation, at least 10 pertain to antidepressant pharmacotherapies or medications commonly used to augment antidepressants⁵. Multiple marketed assays are intended to guide antidepressant treatment; while none have yet pursued FDA approval, such diagnostic tests are available commercially from the laboratories that developed them. And clinical guidelines

for the use of pharmacogenomic testing are available from US and international agencies⁶. Still, very few patients receive such testing, and its utility remains unclear, in part because of a relative lack of randomized controlled studies indicating benefit.

In this paper, we focus on the scientific challenges that have contributed to the persistence of artisanal prescribing of antidepressants even in the face of growing enthusiasm for the concept originally described as personalization, then stratification, and most recently precision medicine⁷. We also review the obstacles to translation of pharmacogenomic tools to common clinical practice. Finally, we address strategies that could be helpful in ensuring that the next decade does bring significant progress towards achieving true precision in the pharmacotherapy of depression.

WHAT ARE THE CHALLENGES IN PERSONALIZING ANTIDEPRESSANT TREATMENT?

Personalization is not precision

In oncology, the concept of matching treatments to patients to achieve and maintain remission is well established: there are particular tumor profiles that

respond differentially to particular interventions. For major depressive disorder, while remission certainly remains a key goal, other considerations are also important: in addition to safety, clinicians may consider key symptoms to target and key adverse effects to avoid.

To this end, psychiatrists (and primary care physicians) already personalize treatment, albeit in a more artisanal and less scientific fashion than oncologists. A systematic approach to this process has been described by Preskorn⁸. Essentially, some medications are excluded on the basis of safety: for example, medications like bupropion that lower seizure threshold might be avoided in individuals at high risk for seizures. Others are avoided on the basis of adverse effects: in an obese patient, medications that commonly increase appetite, such as mirtazapine, would be excluded from initial consideration. Among the remaining options, some clinicians simply pick their favorite; others follow guidelines approved by their employer or payers, perhaps based on which medications are available at lowest cost; and others provide an individual patient with a few choices and discuss adverse effect profiles for each. The difficulty here is that, while most clinicians likely follow some variant of these approaches, there is no agreed-upon or evidence-based framework for such practices.

The evidence base for next-step interventions is even more modest. A particular challenge is the emphasis on randomized controlled trials, that tends to favor more recent industry-supported studies. Consider the case of augmentation: the strongest evidence base supports certain second-generation antipsychotics, simply because older strategies (for example, bupropion, buspirone or pramipexole) involve off-label use of medications long since generic. So, even on the basis of evidence-based personalization, the clinician cannot be strongly informed by treatment guidelines that tend to simply count large-scale positive trials.

In summary, clinicians already personalize, but in a haphazard and inconsistent way. Unfortunately, the very resistance to more systematic treatment approaches, like algorithms and guidelines, on the basis of the need to personalize actually hinders efforts at personalization: there is no agreed-upon standard on which to improve. Faced with an algorithm, many clinicians insist on the need to tailor treatment depending on particular clinical features, even in the absence of strong evidence that such features are truly predictive. The missing ingredient here may be humility: most clinicians likely rate themselves as above average in terms of ability to identify efficacious treatments, but clearly some are not. Ironically, one of the advantages of biologically-based treatment selection would be the ability to introduce more systematic approaches while preventing narcissistic injury to clinicians.

Treatment-specific effects are modest

Beyond a general resistance to external guidance on prescribing is the larger problem that treatment-specific differences in efficacy appear likely to be quite modest. While antidepressants are more effective than placebo, the magnitude of this difference is generally small, at least in the outpatient context⁹. This does not mean that prediction cannot be useful, just that some such prediction is

actually pertaining to placebo response and thus by definition not treatment-specific. As discussed below, such non-specific predictors may still be useful in stratifying treatment intensity, if not specific treatment choice.

Data needed to compare active treatments are lacking

The regulation of medications in the US does not require active comparator studies: there is no obligation (or even expectation) that a new drug be superior to an existing one. So, not surprisingly, such studies are rarely done, and when they are, they are likely to be engineered to yield results which are misleading at best, with an active comparator group included only for “assay sensitivity” which may not even be analyzed in comparison with the study medication.

In the rare cases where straightforward comparator studies are done, they have tended to be a poor investment for the sponsor: treatment differences are likely very small on average, and the substantial placebo response places a floor effect on the performance of comparator drugs (unless comparators are actually worse than placebo, a phenomenon rarely encountered in psychopharmacology¹⁰).

Moreover, where large comparator studies are done, the data are typically held by the industry sponsor. Until recently, large pharmaceutical companies have been reluctant to share DNA or genotypic data in conjunction with treatment response, even where they did agree to sharing for association studies of disease. Presumably, the risk of finding a predictor of non-response, or perhaps the perceived need to involve the FDA in reporting genomic data pertaining to marketed drugs, has outweighed the scientific interest in such work.

Power is accordingly poor to find real effects

Combining a small effect with a small sample size represents a recipe for an underpowered study – one where the

risk for both false positive and false negative findings is high¹¹. Worse, because of the problem of “winner’s curse”, even when true effects are detected, they are likely to be overestimated – thus the pattern all too familiar in psychiatric pharmacogenomics in which initial exciting findings subsequently prove either to be false positives or of less importance than anticipated^{12,13}.

Standard statistical approaches to finding predictors of differential response between two or more interventions rely on a test for a treatment-by-predictor interaction, which is substantially less powerful than tests for main effect. Notably, such a test has greatest power when a predictor has opposite effects in two groups: for example, biomarker A is associated with greater-than-average response to fluoxetine, but worse-than-average response to bupropion. Biologically, this scenario seems implausible: more likely, biomarker A is associated with greater-than-average response to fluoxetine but no difference with bupropion. In this scenario, a test for interaction is even less powerful.

GETTING TO PRECISION

Begin using what we know rather than seeking a silver bullet

Efforts at personalization may have suffered from their ambition, with an unwillingness to employ more basic or mundane socio-demographic predictors in pursuit of a single powerful biomarker. In reality, multiple studies suggest that readily available patient-level features may at least help to set prior probability of response.

Phenomenology

Among the earliest putative predictors were the depressive subtypes, melancholic and atypical depression. An extensive literature explored these sets of symptoms in terms of phenomenology and associated peripheral markers. This literature illustrates some of the challenges in identifying response predictors.

Melancholic depression in general is highly correlated with depression severity, such that, while it is associated with poorer outcomes in general, these outcomes may better be explained by considering total severity. This underscores the importance of ensuring that putative predictors represent the easiest or most straightforward means of measuring a phenomenon. The value of total severity in this regard is further discussed below.

Atypical depression has been difficult to establish as a strong predictor of outcome because of problems in distinguishing individual symptoms from a true subtype. Empirical evidence suggested that, while reversed neurovegetative signs – hypersomnia rather than insomnia, or hyperphagia rather than loss of appetite – are common, they do not necessarily represent a distinct subgroup. That is, many patients may experience one or the other. Worse, as symptoms may fluctuate over time and across episodes, the determination of whether a patient meets criteria for this subtype likely depends upon when in the episode course the patient is assessed.

More recently, Fava et al¹⁴ suggested additional depressive subtypes on the basis of questionnaires included in the baseline assessment for clinical trial participants – in particular, emphasizing the notion of hostile (irritable) and anxious depression. Both of these strongly predict poorer outcomes across multiple studies¹⁵. However, in addition to correlation with each other, they are also correlated with total severity, and like the other clinical subtypes may fluctuate within an episode.

Anxious depression in particular received some support from the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study, where it predicted poorer treatment response¹⁶. A subsequent replication effort in the Genome-based Therapeutic Drugs for Depression (GENDEP) study, however, did not provide further support¹⁷. This non-replication may suggest the importance of considering reference populations when attempting to derive predictors.

One of the most robust recent predictors of outcome was described by Uher

et al¹⁷ using results of factor analysis in lieu of the traditional depressive subtypes. They found that an interest-activity symptom dimension at baseline – which captured poor interest, decreased activity, indecisiveness, and anhedonia – was strongly associated with poor outcome both in GENDEP and in the larger STAR*D study. This association persisted despite control for overall severity and type of antidepressant.

As one of the best-validated predictors of outcome other than total severity, it would seem that the interest-activity factor could represent a good starting point for stratification. That it has not been so applied relates in part to the unwillingness of most clinical practices to employ systematic assessment of symptoms, notwithstanding the imposition of the Patient Health Questionnaire (PHQ-9) in primary care settings. This obstacle is discussed further below.

Notably, efforts to identify predictors of differential treatment response (often described as moderators of response¹⁸) also date back to the dawn of structured psychotherapies. These investigations often focus on specific scales quantifying the target of a particular kind of intervention. For example, the Coping Self-Efficacy Scale was a predictor of response to cognitive-behavioral therapy, delivered either by telephone or in person¹⁹.

Another strategy attempts to integrate socio-demographic and clinical features to predict treatment resistance in major depressive disorder. From among a larger panel of variables, symptoms predictive of treatment resistance included insomnia and decreased energy, along with elements of history such as trauma exposure, post-traumatic stress disorder, and even mild psychotic-like symptoms. In an independent validation cohort also drawn from the STAR*D study, but from different sites, specificity for treatment resistant depression exceeded 0.91, although sensitivity was lower at 0.26²⁰. This study also produced a risk visualization tool (<http://trdrisk.mghcedd.org>), intended to promote development of similar efforts integrating clinical and genomic data.

Employing any of these simple predictors would in no way preempt the use of biological predictors as they are identified. Indeed, even a simple baseline model would be a valuable basis for comparison with newer models – a starting point to be improved on by adding biological or other predictors. In this context, frameworks such as net reclassification improvement²¹ may be more useful for understanding how the addition of a new marker improves prediction, compared to standard metrics such as area under receiver operating characteristic curve²².

Genetic and genomic predictors

Among the potential biological predictors of outcome, cytochrome P450 (CYP450) variation has been understood to influence blood levels of multiple drugs for two decades or more. Unlike most genetic associations, the functional implications of the key variations have been described – that is, particular alleles are known to increase or decrease enzyme activity in a predictable way²³.

The central challenge to the use of CYP450 testing for antidepressant prescribing stems from the lack of a clear relationship between blood levels and either efficacy or adverse effects. At the extremes, some relationship is intuitive: individuals with undetectable blood levels will not respond to true drug effects (although they may still respond to placebo); individuals with supra-therapeutic blood levels should be more likely to experience adverse effects. However, for most antidepressants, even a simple dose-response relationship has been difficult to establish.

Given the clearer relationship of efficacy (and toxicity) of tricyclics to blood levels, it is unsurprising that this is the class of antidepressants with the strongest evidence that CYP450 testing is likely to inform dosing. Unfortunately, despite the substantial efforts expended to develop and promote guidelines for CYP450-informed dosing⁷, this class has largely been superseded by other antidepressants on the basis of equivalent efficacy and wider therapeutic index (i.e., greater

margin of safety). So, the intervention where precision medicine in depression treatment may be most feasible is now also the one least clinically useful. The term in decision analysis for this scenario is a dominating choice: in most if not all circumstances, the cost-effectiveness of CYP450-guided tricyclic treatment will be less than that of simply prescribing a generic non-tricyclic.

For selective serotonin reuptake inhibitors (SSRIs) and serotonin-norepinephrine reuptake inhibitors (SNRIs), the impact of CYP450 variation is not fully understood. Most SSRIs and SNRIs are substrates for one of the common CYP450 enzyme systems, so it is possible to make predictions about changes in blood levels. What those levels mean, though, is not so clear: with the possible exception of modest data regarding fluoxetine^{24,25} and venlafaxine²⁶, higher doses within the therapeutic range have generally not been shown to be more efficacious than lower doses. The evidence of poor tolerability at higher doses is rarely studied directly, particularly as it relates to CYP450 status: one study suggested that non-wild-type metabolizers of the CYP450 2C19 substrate citalopram experienced poorer tolerability with this treatment²⁷.

Further, even in circumstances where drug blood levels are important, CYP450 variation is only one contributor to such levels. Numerous environmental factors, including diet and other medications (as well as other, unmeasured genetic variation), may be important. One illustration of these effects was a study of venlafaxine-treated patients that examined the plasma ratio of venlafaxine to its metabolite desvenlafaxine in order to define individuals who were “functionally” poor metabolizers. Overall, 27% of individuals appeared to be poor metabolizers, even though only 4% were CYP450 2D6 poor metabolizers genotypically²⁸.

As venlafaxine is the pro-drug for desvenlafaxine, individuals who are poor metabolizers at CYP450 2D6 might be hypothesized to be less likely to respond to treatment (as they will have very low effective levels of the active drug). Indeed, in the four venlafaxine studies, poor metabolizers were less likely to

achieve remission than wild-type metabolizers²⁹.

Other biomarkers

Efforts to identify predictors of differential antidepressant treatment response based on blood or other peripheral measures date back to the dawn of psychopharmacology. The dexamethasone suppression test (initially considered to diagnose depression, and later employed to guide treatment) presents a useful cautionary tale of a diagnostic tool deployed in psychiatry without sufficient consideration of its utility, or even what exactly it predicted^{30,31}.

An example of a prototypical predictor might be C-reactive protein (CRP), a marker of inflammation associated with cardiovascular disease. In the GENDEP study, a notable treatment-by-predictor interaction – exactly the sort that could potentially be informative for treatment selection – was identified with CRP. Specifically, symptomatic improvement was greater with escitalopram treatment among individuals with CRP levels lower than 1 mg/L, while it was greater with nortriptyline treatment among individuals with CRP levels higher than 1 mg/L. Still, given the poorer safety profile of tricyclic antidepressants, the modest difference in efficacy (three points on the Montgomery-Åsberg Depression Rating Scale) may not be sufficient to justify preferential use of nortriptyline even in the latter patient subset. While some frameworks for defining clinical significance exist – see, for example, the calculator at depressiontools.org³² – the necessary effect size for utility of a given predictor depends critically on its context.

Numerous other minimally invasive markers are under active investigation for response prediction. Functional neuroimaging has been perhaps the most studied, with intriguing but not definitive results – not surprising given the relatively small cohorts studied. Similarly, quantitative electroencephalography has been applied to predict either overall treatment outcome or differential response. In a representative small study,

a measure of frontal recordings at baseline and week 1 was associated with speed and probability of response to escitalopram over 13 weeks³³, consistent with a prior pilot study using fluoxetine³⁴. The pilot study, importantly, included a placebo arm where no such an association was identified. Still, as noted earlier, the absence of any comparison drug makes the specificity of this effect unclear. One other notable aspect of these studies is the inclusion of a post-baseline (week 1) time point in the biomarker: prediction of outcome based on short-term treatment exposure, while not a standard strategy in psychopharmacology, may be easier than relying solely on baseline measures.

Educate clinicians and patients

In addition to patient education, preliminary experience with genomic testing suggests the necessity and value of clinician education³⁵, in terms of how results are presented to patients and families. These tests typically yield probabilistic results, very different from the dichotomous outcome yielded by many other tests in medicine, though common in other areas such as cancer, where estimates of survival are the coin of the realm. In one pilot pharmacogenomic study of antidepressant response, only 1/4 of consented patients were able to indicate an understanding of such testing³⁶.

A particular concern in psychopharmacology is the misinterpretation of CYP450 results as contraindicating a medication or class of medications. In light of the relative paucity of good therapeutic options, particularly for patients who do not remit with first-line treatments, ruling out a medication unnecessarily can be highly consequential. In reality, non-wild-type metabolizers simply require more cautious and informed titration: those who are poor metabolizers require lower doses of substrate drugs, while those who are ultrarapid metabolizers may require doses exceeding the FDA labeling, though still with careful titration. While simply avoiding

substrate drugs is a basic heuristic that may be reasonable when selecting initial treatments, such a heuristic can actually be detrimental as the range of reasonable options narrows. To this end, the tendency to present CYP450 results with color-coding – listing substrates in red, or with a stop sign, for example – may be unhelpful.

For both medications and diagnostic tools, clinician education can be mandated by the FDA within the approval process as part of the risk evaluation and mitigation strategy³⁷. Similar education may be required for some interventions aimed at personalization of antidepressant prescribing, if only to limit the consequences of misinterpretation of test results.

Aim for stratification, not treatment-matching

Even where we cannot identify medication-specific predictors, distinguishing high- or low-risk groups may still be extremely useful. Three examples include greater depression severity, the interest-energy factor identified by Uher et al¹⁷, and the treatment resistance risk score described earlier: to date, each of these appears to be a predictor of poorer outcome in general, rather than a feature that identifies an optimal treatment. So, while presence of greater risk may not help with selection of an individual treatment – venlafaxine versus fluoxetine, for example – it may instead indicate that a particular patient requires more intensive treatment in general. Individuals at high risk for treatment resistance could be triaged to more aggressive interventions – combination treatment, or incorporation of cognitive-behavioral therapy – or even more aggressive assessments, like specialist consultation or application of more intensive diagnostic tools.

Our approach to initial non-response needs to change

Protocols and randomization

Ironically, moving towards more truly personalized medicine may require mov-

ing away from traditional means of personalization by enrolling patients in protocol-driven treatment, much as is the case with cancer chemotherapies at academic medical centers. While clinicians maintain the importance of artisanal personalization, we are aware of no empirical data to indicate that such strategies improve upon uniform or standardized treatment selection (much less random selection among a small number of similar options). As much as it pains the expert psychopharmacologist to recognize this point, in general the clinician is at equipoise among multiple next-step strategies. Recent survey results reinforce this point³. But if this inconvenient reality were acknowledged and disclosed (“There are several reasonable next steps, we’re going to let the computer select the first one to try”), it is possible that different strategies could be investigated.

Systematic measurement of outcomes

A related problem remains clinicians’ reluctance to incorporate systematic measurement of outcomes – any outcomes – into their practices. The reasons for this resistance are manifold: the measures can be time consuming, they are rarely well integrated into clinical workflow, and they fail to capture the breadth of depressive symptomatology clinicians feel they need. While less often acknowledged, such measurement is likely to also create a bias to action: that is, identifying symptoms creates more requirement to act on these symptoms, or potentially liability for not acting on them.

Many health systems have elected to invest in the PHQ-9, a depression screening tool with limited utility for outcome measurement (a role it was never designed to fill). More recently, led in part by movements aiming for more patient-centered care followed by financial support from the Patient Centered Outcomes Research Institute, enthusiasm has grown for patient-reported outcomes – particularly measures of functional status and quality of life.

It seems reasonable to measure the improvement yielded by psychiatric interventions in a systematic way. To the clinicians who argue that the PHQ-9 captures only a limited amount of the benefit they provide, a reasonable response is to agree, and ask what better measures can be employed. Whatever psychosocial or pharmacological interventions do for depression, it should be possible to measure it. Either less intrusive and better integrated measures need to be found, or more resources need to be provided for clinicians to incorporate such measures. Notwithstanding the massive hyperbole currently attached to ambulatory monitoring devices, cell-phone-based survey tools may help to fill this void³⁸ – provided better platforms can be developed to safely and efficiently integrate these data for use by clinicians.

Use of electronic medical records and other large data sets

Yet another opportunity to improve precision in antidepressant treatment comes from the increasing availability of large clinical data sets, i.e., electronic medical records, with or without linkages to biobanks. These data sets provide a rich trove of clinical detail, typically far exceeding what is available from the health claims data sets employed for pharmacovigilance studies and health services research³⁹. Compared to standard clinical trials, the patients and outcomes are likely to be more generalizable, as the biases inherent in patient recruitment are avoided. When biological materials – DNA or plasma, for example – are available, these resources also allow highly efficient *in silico* biomarker studies.

We have previously demonstrated the utility of electronic medical records for defining antidepressant treatment outcomes⁴⁰, and applying these metrics to characterize clinical⁴¹ and genetic⁴² predictors of non-response. A less well appreciated benefit of such large cohorts is the ability to study relatively rare but serious adverse effects, such as lithium-associated renal failure⁴³. These designs

also facilitate investigation of quantitative drug effects, such as antidepressant-associated weight gain⁴⁴ or QT-interval prolongation⁴⁵.

Still, some important caveats apply to approaches using electronic medical records or national health registries. First, treatment assignment is not randomized, so the risk for confounding – particularly confounding by indication, in which the indication for a particular treatment confounds the result – is substantial (for an illustration of the impact of such confounding, see the study by Gallagher et al⁴¹, in which treatment with non-steroidal anti-inflammatory drugs was associated with poorer antidepressant treatment outcome until the indication – e.g., pain – was controlled for). Statistical methods can help to control bias, but the risk for confounding cannot be entirely eliminated. Second, clinical care typically includes less precise measures of outcome as well as other relevant clinical covariates. In some cases proxy measures may suffice (hospitalization; treatment changes), but traditional clinical trial outcomes such as remission and response are more challenging to characterize. Indeed, one observation from studies based on electronic medical records⁴⁰ (and consistent with some mood disorder cohort studies⁴⁶) is the extent to which episodic definitions of depression likely underestimate chronicity and persistence of residual symptoms relative to clinical cohorts.

Randomized trials of precision medicine will be needed. . . or will they?

Despite the utility of alternative approaches, randomized controlled trials remain the gold standard for investigating new interventions, pharmacological or otherwise. Even for pharmacotherapy, there has been continued innovation in the design and conduct of such studies. But for diagnostic tests, the optimal design of randomized trials remains subject to debate. For example, if subjects are to be randomized to assay-guided treatment compared to treatment as

usual, how constrained or algorithmic should the treatment as usual be? Should clinicians stay unblinded, or should they receive a “dummy” or uninformative report? If the latter, is it ethical to delay reporting results (or even to report misleading results), and will clinicians be able to distinguish an uninformative from a placebo report? Design of a treatment-as-usual arm is particularly challenging as the inclination is to reduce heterogeneity by making this intervention more structured and algorithmic. However, as we have noted, standard of care is far from algorithmic at present, so this sort of comparator is artificial and itself likely to improve outcomes⁴⁷.

A further, practical problem is deciding who should pay for these studies. If the tools are developed by a for-profit entity, it is reasonable to require that the entity fund such studies. However, this barrier may be prohibitive for smaller companies or less costly tests. The shifting regulatory structure in the US, in which the FDA has allowed marketing of laboratory-developed tests without pre-market review, but has indicated that it intends to increase oversight of this pathway⁴⁸, is likely to increase the pressure to conduct such randomized trials, if not the available resources.

To date, there is one small randomized trial investigating a pharmacogenetic assay for antidepressant prescribing, relying on a panel of CYP450 variants (2D6, 2C19 and 1A2) as well as some pharmacodynamic common variants. Among 51 outpatients with major depressive disorder, followed for eight weeks, the magnitude of improvement was numerically but not statistically significantly different between the treatment-as-usual (19%) and the assay-guided treatment (31%) groups ($p=0.3$). One of two unblinded cohort studies using the same assay did identify significantly greater improvement in the assay-guided treatment group⁴⁹. In the absence of blinding or consideration of the impact of individual predictors, estimates of the benefit associated with specific variants await randomized trials.

In the meantime, electronic medical records or claims data may help to un-

derstand the potential impact of putative predictors of response. One approach uses cost-effectiveness analysis to examine the effect of a predictor, based on other assumptions about treatment costs and outcomes. In an illustration of this approach, we previously developed a model based on STAR*D data⁵⁰ considering a predictor of differential SSRI response. Under some assumptions, even a moderate difference between treatments was not cost-effective simply because using an alternate antidepressant was a dominating (better) choice. On the other hand, for a low-cost test, when the likelihood of an informative test is high, even relatively modest effect sizes could be cost-effective.

A major limitation in all such models is the need to make numerous assumptions about costs, probabilities and utilities. Their value is primarily in clarifying the circumstances where precision medicine may be most likely to be beneficial in antidepressant prescribing, as a means of designing future interventions.

Another perspective on cost-effectiveness comes from investigation of insurance claims databases in which some patients have already received pharmacogenomic testing. For the assay with the negative randomized trial described earlier, this investigation found that individuals receiving medications indicated to be “less desirable”, based on an algorithm incorporating multiple variants, incurred greater past-year health care costs⁵¹. Whether such high-cost individuals represent an optimal population for deploying precision medicine is an intriguing, but as yet untested, hypothesis.

A direct but unrandomized assessment of cost-effectiveness comes from another study of health claims data that compared a cohort of 111 individuals who had received a commercial test combining CYP450 and pharmacodynamic variants with a propensity-score matched cohort who did not receive testing⁵². While not a substitute for randomization, this method allows some control of confounding by matching an unexposed (untested) group as closely as possible to the tested group. That study found, after matching and

adjustment, that outpatient treatment costs were 9.5% lower among tested patients. It also identified improvement in medication adherence among the tested group. Still, like other reports of pharmacogenomic testing outcomes, the absence of analysis by individual variant precludes an understanding of the elements of the assay most important for prediction.

CONCLUSIONS

Personalized medicine is already a reality in the treatment of depression, but precision medicine is not – that is, while clinicians routinely attempt to match treatments to patients, these strategies are neither systematic nor empirically supported. Making the transition to precision medicine will, first, require a commitment to the systematic practice of medicine: following algorithms or guidelines, and measuring patient outcomes to guide decision making. If physicians trained to rely on the art of medicine cannot make this transition, it is likely that nurse-clinicians and pharmacists will make it for them. Second, it will require a willingness to begin to study and deploy risk stratification tools that may not be perfect, but rather better than the current standard of care. A further benefit of these two steps will be an acceleration of the ability to develop and investigate new personalization strategies, because it will become more straightforward to identify biomarkers and study them in large clinical cohorts.

Evidence from effectiveness studies and clinical cohorts indicate that many patients remain poorly served by existing antidepressant treatments. Aiming for more precise treatment matching may help to ensure optimal outcomes even while the field strives for better treatment options.

REFERENCES

1. Johnson CF, Dougall NJ, Williams B et al. Patient factors associated with SSRI dose for depression treatment in general practice: a primary care cross sectional study. *BMC Family Pract* 2014;15:210.

2. von Wolff A, Meister R, Harter M et al. Treatment patterns in inpatient depression care. *Int J Methods Psychiatr Res* 2016;25:55-67.
3. Goldberg JF, Freeman MP, Balon R et al. The American Society of Clinical Psychopharmacology Survey of Psychopharmacologists' Practice Patterns for the Treatment of Mood Disorders. *Depress Anxiety* 2015;32:605-13.
4. Herbild L, Bech M, Gyrd-Hansen D. Estimating the Danish populations' preferences for pharmacogenetic testing using a discrete choice experiment. The case of treating depression. *Value Health* 2009;12:560-7.
5. National Research Council of the National Academies. *Toward precision medicine: building a knowledge network for biomedical research and a new taxonomy of disease*. Washington: National Academies Press, 2011.
6. US Food and Drug Administration. *Table of pharmacogenomic biomarkers in drug labeling 2014*. www.fda.gov.
7. Hicks JK, Swen JJ, Thorn CF et al. Clinical Pharmacogenetics Implementation Consortium guideline for CYP2D6 and CYP2C19 genotypes and dosing of tricyclic antidepressants. *Clin Pharmacol Ther* 2013;93:402-8.
8. Preskorn SH. *Outpatient management of depression: a guide for the primary-care practitioner*, 2nd ed. Caddo: Professional Communications, 1999.
9. Fountoulakis KN, Möller HJ. Antidepressant drugs and the response in the placebo group: the real problem lies in our understanding of the issue. *J Psychopharmacol* 2012;26:744-50.
10. Pande AC, Crockatt JG, Janney CA et al. Gabapentin in bipolar disorder: a placebo-controlled trial of adjunctive therapy. *Gabapentin Bipolar Disorder Study Group*. *Bipolar Disord* 2000;2:249-55.
11. Ioannidis JP. Why most published research findings are false. *PLoS Med* 2005;2:e124.
12. Lee CS, Cheng AT. Variant GADL1 and response to lithium in bipolar I disorder. *N Engl J Med* 2014;370:1859-60.
13. Chen CH, Lee CS, Lee MT et al. Variant GADL1 and response to lithium therapy in bipolar I disorder. *N Engl J Med* 2014;370:119-28.
14. Fava M, Uebelacker LA, Alpert JE et al. Major depressive subtypes and treatment response. *Biol Psychiatry* 1997;42:568-76.
15. Perlis RH, Uher R, Ostacher M et al. Association between bipolar spectrum features and treatment outcomes in outpatients with major depressive disorder. *Arch Gen Psychiatry* 2011;68:351-60.
16. Fava M, Rush AJ, Alpert JE et al. Difference in treatment outcome in outpatients with anxious versus nonanxious depression: a STAR*D report. *Am J Psychiatry* 2008;165:342-51.
17. Uher R, Perlis RH, Henigsberg N et al. Depression symptom dimensions as predictors of antidepressant treatment outcome: replicable evidence for interest-activity symptoms. *Psychol Med* 2012;42:967-80.
18. Kraemer HC, Stice E, Kazdin A et al. How do risk factors work together? Mediators, moderators, and independent, overlapping, and proxy risk factors. *Am J Psychiatry* 2001;158:848-56.
19. Stiles-Shields C, Corden ME, Kwasny MJ et al. Predictors of outcome for telephone and face-to-face administered cognitive behavioral therapy for depression. *Psychol Med* 2015;45:3205-15.
20. Perlis RH. A clinical risk stratification tool for predicting treatment resistance in major depressive disorder. *Biol Psychiatry* 2013;74:7-14.
21. Pencina MJ, D'Agostino RB Sr, Demler OV. Novel metrics for evaluating improvement in discrimination: net reclassification and integrated discrimination improvement for normal variables and nested models. *Stat Med* 2012;31:101-13.
22. Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 2007;115:928-35.
23. De Gregori M, Allegri M, De Gregori S et al. How and why to screen for CYP2D6 interindividual variability in patients under pharmacological treatments. *Curr Drug Metab* 2010;11:276-82.
24. Fava M, Alpert J, Nierenberg A et al. Double-blind study of high-dose fluoxetine versus lithium or desipramine augmentation of fluoxetine in partial responders and nonresponders to fluoxetine. *J Clin Psychopharmacol* 2002;22:379-87.
25. Fava M, Rosenbaum JF, McGrath PJ et al. Lithium and tricyclic augmentation of fluoxetine treatment for resistant major depression: a double-blind, controlled study. *Am J Psychiatry* 1994;151:1372-4.
26. Berney P. Dose-response relationship of recent antidepressants in the short-term treatment of depression. *Dialogues Clin Neurosci* 2005;7:249-62.
27. Mrazek DA, Biernacka JM, O'Kane DJ et al. CYP2C19 variation and citalopram response. *Pharmacogenet Genomics* 2011;21:1-9.
28. Preskorn SH, Kane CP, Lobello K et al. Cytochrome P450 2D6 phenocconversion is common in patients being treated for depression: implications for personalized medicine. *J Clin Psychiatry* 2013;74:614-21.
29. Lobello KW, Preskorn SH, Guico-Pabia CJ et al. Cytochrome P450 2D6 phenotype predicts antidepressant efficacy of venlafaxine: a secondary analysis of 4 studies in major depressive disorder. *J Clin Psychiatry* 2010;71:1482-7.
30. Baldessarini RJ, Arana GW. Does the dexamethasone suppression test have clinical utility in psychiatry? *J Clin Psychiatry* 1985;46:25-9.
31. Perlis RH. Translating biomarkers to clinical practice. *Mol Psychiatry* 2011;16:1076-87.
32. Uher R, Tansey KE, Malki K et al. Biomarkers predicting treatment outcome in depression: what is clinically significant? *Pharmacogenomics* 2012;13:233-40.
33. Cook IA, Hunter AM, Gilmer WS et al. Quantitative electroencephalogram biomarkers for predicting likelihood and speed of achieving sustained remission in major depression: a report from the biomarkers for rapid identification of treatment effectiveness in major depression (BRITE-MD) trial. *J Clin Psychiatry* 2013;74:51-6.
34. Hunter AM, Cook IA, Greenwald SD et al. The antidepressant treatment response index and treatment outcomes in a placebo-controlled trial of fluoxetine. *J Clin Neurophysiol* 2011;28:478-82.
35. Roberts JS, Chen CA, Uhlmann WR et al. Effectiveness of a condensed protocol for disclosing APOE genotype and providing risk education for Alzheimer disease. *Genet Med* 2012;14:742-8.

36. Rose D, Russo J, Wykes T. Taking part in a pharmacogenetic clinical trial: assessment of trial participants understanding of information disclosed during the informed consent process. *BMC Med Ethics* 2013;14:34.
37. US Food and Drug Administration. Risk evaluation and mitigation strategies 2015. www.fda.gov.
38. Marcano Belisario JS, Jamsek J, Huckvale K et al. Comparison of self-administered survey questionnaire responses collected using mobile apps versus other methods. *Cochrane Database Syst Rev* 2015;7:MR000042.
39. Kohane IS. Using electronic health records to drive discovery in disease genomics. *Nat Rev Genet* 2011;12:417-28.
40. Perlis RH, Iosifescu DV, Castro VM et al. Using electronic medical records to enable large-scale studies in psychiatry: treatment resistant depression as a model. *Psychol Med* 2012;42:41-50.
41. Gallagher PJ, Castro V, Fava M et al. Antidepressant response in patients with major depression exposed to NSAIDs: a pharmacovigilance study. *Am J Psychiatry* 2012;169:1065-72.
42. O'Dushlaine C, Ripke S, Ruderfer DM et al. Rare copy number variation in treatment-resistant major depressive disorder. *Biol Psychiatry* 2014;76:536-41.
43. Castro VM, Roberson AM, McCoy T et al. Stratifying risk for renal insufficiency among lithium-treated patients: an electronic health record study. *Neuropsychopharmacology* 2016;41:1138-43.
44. Blumenthal SR, Castro VM, Clements CC et al. An electronic health records study of long-term weight gain following antidepressant use. *JAMA Psychiatry* 2014;71:889-96.
45. Castro VM, Clements CC, Murphy SN et al. QT interval and antidepressant use: a cross sectional study of electronic health records. *BMJ* 2013;346:f288.
46. Perlis RH, Dennehy EB, Miklowitz DJ et al. Retrospective age at onset of bipolar disorder and outcome during two-year follow-up: results from the STEP-BD study. *Bipolar Disord* 2009;11:391-400.
47. Guo T, Xiang YT, Xiao L et al. Measurement-based care versus standard care for major depression: a randomized controlled trial with blind raters. *Am J Psychiatry* 2015;172:1004-13.
48. US Food and Drug Administration. Laboratory developed tests 2015. www.fda.gov.
49. Hall-Flavin DK, Winner JG, Allen JD et al. Using a pharmacogenomic algorithm to guide the treatment of depression. *Transl Psychiatry* 2012;2:e172.
50. Perlis RH, Patrick A, Smoller JW et al. When is pharmacogenetic testing for antidepressant response ready for the clinic? A cost-effectiveness analysis based on data from the STAR*D study. *Neuropsychopharmacology* 2009;34:2227-36.
51. Winner J, Allen JD, Altar CA et al. Psychiatric pharmacogenomics predicts health resource utilization of outpatients with anxiety and depression. *Transl Psychiatry* 2013;3:e242.
52. Fagerness J, Fonseca E, Hess GP et al. Pharmacogenetic-guided psychiatric intervention associated with increased adherence and cost savings. *Am J Manag Care* 2014;20:e146-56.

DOI:10.1002/wps.20345