See related article by Thase et al

Does Pharmacogenomic Testing Meaningfully Improve Antidepressant Treatment Outcomes When Looking Only at Patients Taking Phase I Hepatically Metabolized Drugs? A Little

Joseph F. Goldberg, MD[a,*]

When a clinical intervention trial has a negative primary outcome, it is often reasonable for the study investigators to conduct exploratory post hoc analyses to generate hypotheses about patient subgroups or clinical parameters that might influence outcomes in future studies. Such "post hocking" can help to fine-tune the response signal of a modestly significant parent study and perhaps enrich the identifiable features of patients most likely to benefit from a particular intervention. However, the danger of embracing as factual any post hoc findings that were not planned (and adequately powered) from the outset of a study is the risk for making a type I error (ie, thinking a true association exists when really it does not). To minimize that risk, statisticians impose methods of varying rigor to correct *P* values (reflecting a set α level) for multiple comparisons (often, by dividing a *P* value by the number of tests performed—a so-called Bonferroni correction, although this can sometimes be overly stringent, especially if one's intention is to explore possible leads rather than draw definitive conclusions).

Investigators get excited and are often inspired by leads and suggestions from post hoc analyses because such results can shed light on what worked and what did not. (Imagine examining the winning and losing horses after a race is over before then placing bets on the next race.) Non-investigator clinician-readers of the literature are sometimes presented with post hoc study findings as being factually embraceable without clear provisos about their tentative nature—when uncorrected *P* values lack the same statistical and often clinical relevance of planned analyses that were adequately statistically powered. How much does that matter? Well, in the world of psychiatric genetics and pharmacogenomics, where multiple candidate genes are thought to exert very small effects and the distribution of gene variants in the disease state versus controls is usually minimal, genome-wide association studies usually require a *P* value with 8 or more zeros to the right of the decimal.

In their *Journal* article, Thase and colleagues[1] perform such an "after-the-race-has-been-run" analysis on the previously reported negative Genomics Used to Improve DEpression Decisions (GUIDED) study[2] in which randomization to a proprietary combinatorial pharmacogenomics test failed to improve depressive symptoms better than usual care in a large group of treatment-resistant depression (TRD) subjects. That parent study performed over 2 dozen post hoc analyses (as registered on ClinicalTrials.gov: NCT02109939) from which nominally significant (uncorrected *P* values) advantages were identified for pharmacogenomically guided versus usual care for "response" (26.0% versus 19.9%) or "remission" (15.3% versus 10.1%). These findings translated to dismally high numbers needed to treat (NNTs) of 17 and 19, respectively.[3]

In the current post hoc analysis from that original dataset, Thase et al[1] examined only subjects who at the time of study entry were taking antidepressants that had known relevant pharmacokinetic interactions (eliminating 40% of the original intent-to-treat sample, posing a drastic change to the study design and incurring a commensurate drop in statistical power). In principle, the spirit of this design modification is entirely reasonable, given that the main established value of pharmacogenomic testing is to not predict drug efficacy but, rather, drug tolerability based on a combination of mainly pharmacokinetic single nucleotide polymorphisms (SNPs). (In other words, if someone falls within the approximately 2%–10% of the population who poorly metabolize substrates for key cytochrome P450 [CYP] isoenzymes, they may incur more trouble with side effects when taking antidepressants metabolized by those pathways.)

Surprisingly, Thase et al chose not to report on potential differences in tolerability (or dropout due to intolerances) based on CYP genotype variants—the very thing most germane to pharmacokinetic SNPs and an obviously important parameter given that drug intolerances contribute more than a little to antidepressant effectiveness (staying in treatment) as opposed to efficacy (judging whether a drug beats a placebo under the best of circumstances). Oddly, in the original parent study, neither the number of reported adverse drug effects nor the proportion of patients who encountered adverse effects differed between the pharmacogenomically guided versus usual-treatment groups.[2] Perhaps this will be material for a future post hoc analysis.

[a]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York

*Corresponding author: Joseph F. Goldberg, MD, 128 East Ave, Norwalk, CT 06851 (joseph.goldberg@mssm.edu).

J Clin Psychiatry 2019;80(6):19com13119

To cite: Goldberg JF. Does pharmacogenomics testing meaningfully improve antidepressant treatment outcomes when looking only at patients taking phase I hepatically metabolized drugs? a little. *J Clin Psychiatry.* 2019;80(6):19com13119.

To share: https://doi.org/10.4088/JCP.19com13119

By just how much does their refinement of the original study actually fine-tune the signal? When reporting the change from baseline in 17-item Hamilton Depression Rating Scale (HDRS-17) scores (the primary outcome of the original parent study), the authors proclaim "$\Delta = 5$, $P = .029$"—that is, a delta of 5 percentage points, not HDRS-17 points—and the reported significance of the uncorrected *P* value remains nominal (remember, this is a post hoc analysis uncorrected for multiple comparisons). Remission (arguably the most important outcome when treating major depression) occurred in 18.2% of GUIDED patients and 10.7% of usual-care patients. I calculate the latter observation to yield an NNT of 13.3 (a shade better than the 19 from the parent study). Put differently, an extra 3% of patients over the original study group achieved remission when focusing exclusively on those for whom pharmacogenomic testing would seem to be the most relevant. One might have expected a much bigger effect than that after eliminating "dilution" in the original study group, and there are still far fewer than 8 zeros to the right of the decimal point for the *P* value.

Does the information presented in the article meaningfully help us choose a medication for the next TRD patient we see? The overall remission rate by 8 weeks remained under 20% with or without pharmacogenomic testing. Bothersome is the authors' hyperbolic positioning of percentages in the data presentation: for instance, when they report outcomes of patients who switched medications during the study. Although GUIDED subjects' remission rates rose only to 20.3% (up from 18.2% on the first go-around for this enriched subgroup), this rate is lauded as being numerically more than an 80% improvement over usual treatment, but still quite meager in terms of absolute values. This is reminiscent of the original parent study promulgating a "50% improvement in remission" with pharmacogenomically guided versus usual care[2]—that is, a 15% remission rate being "50% more than" a 10% remission rate, not a 50% rate of remission.

Especially troubling to many psychopharmacologists may be the authors' statement in the Discussion that "many non-genetic factors contribute to medication failure….However, any impact of non-genetic factors should affect both study arms equally due to balanced randomization."[1] How can one know if randomization was actually successful without examining those factors, which include (but are not limited to) chronicity, age at onset, baseline anxiety, depressive subtypes, comorbid personality disorders, psychosis, social isolation, education, employment status,[4] or histories of trauma?[5] Even more fundamentally, little to no information is provided about the clinical psychopharmacology expertise of the study treaters, a parameter that one would hope surely must account for at least some element of antidepressant choice and treatment outcome. Macaluso and Preskorn[6] observed that a psychopharmacologist knowledgeable about CYP interactions would largely arrive at the same recommendations summarized in the printout of a commercial pharmacogenomics testing summary. Collectively, these limitations make it hard to interpret the awfully low response and remission rates in both the pharmacogenomic and treatment-as-usual arms for both the original parent study and the current post hoc analysis. In a TRD patient taking phase I hepatically metabolized antidepressants, does pharmacogenomic testing buy an additional 13th remitter over and above the care of a well-trained psychopharmacologist or as compared to a random prescriber with no special knowledge and expertise?

A final point regarding TRD in the post-STAR*D (Sequenced Treatment Alternatives to Relieve Depression) era involves acknowledging the inescapable limitations of monoaminergic antidepressants, among which there is no clear next-best treatment option after nonresponse to multiple agents and with which even expert care produces only modest benefits. Selective serotonin reuptake inhibitors show remarkably modest effect sizes, serotonin-norepinephrine reuptake inhibitors are not much better,[7] and no particular drug therapy combinations, however cleverly engineered, have as yet broadly been shown to make a substantial impact in TRD. Very few interventions of any kind have been systematically studied specifically in depression unresponsive to many drug trials (eg, > 5), including esketamine. Practitioners, by and large, are left reshuffling the same old deck of monoaminergic cards with no clear therapeutic breakthroughs. Emerging new technologies, such as glutamate-modulating drugs and neurosteroids, inspire optimism for the future. For now, though, the GUIDED data and the study's post hoc analyses leave us with the sober reality that meaningful improvement remains elusive for a substantial majority of TRD patients, regardless of knowing someone's genetic predilection for metabolizing certain antidepressants especially quickly or slowly.

Published online: December 3, 2019.
Potential conflicts of interest: None.
Funding/support: None.

REFERENCES

1. Thase ME, Parikh SV, Rothschild AJ, et al. Impact of pharmacogenomics on clinical outcomes for patients taking medications with gene-drug interactions in a randomized controlled trial. *J Clin Psychiatry*. 2019;80(6):19m12910.
2. Greden JF, Parikh SV, Rothschild AJ, et al. Impact of pharmacogenomics on clinical outcomes in major depressive disorder in the GUIDED trial: a large, patient- and rater-blinded, randomized, controlled study. *J Psychiatr Res*. 2019;111:59–67.
3. Goldberg JF, Rosenblat JD, McIntyre RS, et al. Letter to the editor: clinical versus statistical significance of pharmacogenomic-guided antidepressant therapy: what's really being measured and marketed? *J Psychiatr Res*. 2019;114:208–209.
4. Shelton RC, Trivedi MH. Moderators of antidepressant response in major depression. *J Clin Psychiatry*. 2011;72(10):e32.
5. Williams LM, Debattista C, Duchemin AM, et al. Childhood trauma predicts antidepressant response in adults with major depression: data from the randomized international study to predict optimized treatment for depression. *Transl Psychiatry*. 2016;6(5):e799.
6. Macaluso M, Preskorn SH. Knowledge of the pharmacology of antidepressants and antipsychotics yields results comparable with pharmacogenetic testing. *J Psychiatr Pract*. 2018;24(6):416–419.
7. Papakostas GI, Thase ME, Fava M, et al. Are antidepressant drugs that combine serotonergic and noradrenergic mechanisms of action more effective than the selective serotonin reuptake inhibitors in treating major depressive disorder? a meta-analysis of studies of newer agents. *Biol Psychiatry*. 2007;62(11):1217–1227.