THE CARLAT REPORT
PSYCHIATRY

A CME Publication

AN UNBIASED MONTHLY COVERING ALL THINGS PSYCHIATRIC

IN THIS ISSUE

Focus of the Month:
Biomarkers in Psychiatry

Learning objectives for this issue:
1. Describe the latest research on biomarkers in psychiatry. **2.** Describe how to evaluate methodology in current biomarker studies. **3.** Discuss the current evidence related to the GeneSight test. **4.** Summarize some of the current findings in the literature regarding psychiatric treatment.

Using Psychiatric Biomarkers in Your Practice

Steve Balt, MD
Psychiatrist in private practice in the San Francisco Bay area

Dr. Balt discloses his spouse is employed as a sales representative for Otsuka America.

Treatment of the psychiatric patient is as much an art as it is a science. Diagnosing and deciding on treatments in psychiatry is based on a series of human interactions. We observe and we listen to patients, we ask questions, we consider their responses, and we synthesize all this information in order to render a judgment. Such judgments are subjective. We think this patient has bipolar disorder, but could it be schizoaffective? We think lithium will

In Summary

- There are three types of biomarkers: diagnostic; prognostic; and theranostic (predictive).

- The MDDScore purports to diagnose depression and the VeriPsych assay purports to diagnose schizophrenia.

- The ATR score uses quantitative EEG to predict response to particular antidepressants.

Q&A With the Expert

How to Evaluate the Methodology of Biomarker Studies

James Coyne, PhD

Clinical health psychologist, Emeritus Professor of Psychiatry, University of Pennsylvania School of Medicine and Professor of Health Psychology, University of Groningen, the Netherlands

Dr. Coyne has disclosed that he has no relevant financial or other interest in any commercial companies pertaining to this educational activity.

TCPR: Dr. Coyne, there is a lot of fanfare in the media about biomarker studies. But they often end up getting debunked. How can we become smarter consumers of this literature?
**Dr. Coyne:** That's a great question. I think that biomarker research holds great promise. If we had biomarkers that worked it would remove a lot of the uncertainty from the clinical encounter, but the science is difficult and we are pretty far from finding a reliable biomarker in psychiatry.
TCPR: What are some of the key problems in these studies?
Dr. Coyne: The first is that you have to look carefully at who is recruited, so that you know if the biomarkers will work in patients you are likely to see. For example, are the patients from outpatient clinics or from inpatient psychiatric units?
TCPR: We're talking about the issue of generalizability.
Dr. Coyne: Absolutely. As an example, let's look at the closest we've come to a reliable biomarker in psychiatry, which is the dexamethasone suppression test (DST). I
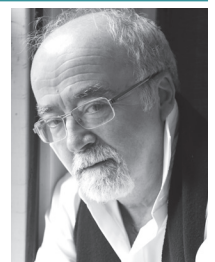
Using Psychiatric Biomarkers in Your Practice
Continued from page 1

work best, but perhaps we should start with lurasidone? Choosing the "right" treatment for a given patient involves a large number of factors, including personal preference—both the clinician's and the patient's.

Is there a better way? Of course there is—most other medical specialists can order labs that provide objective evidence used to guide decisions. These are called "biomarkers." Psychiatrists have been searching for biomarkers for quite a while. Some of you will remember the dexamethasone suppression test (DST) from the 1960s—which was billed as a biomarker for depression. The biological hypothesis was that people with depression had excessive production of the stress hormone cortisol. Normally, giving healthy people an injection of the steroid dexamethasone would suppress cortisol production. But if depressed patients were churning out too much cortisol, dexamethasone would not suppress

cortisol as much as it should—leading to a "positive" DST. Unfortunately, after years of research, the test didn't pan out, because it had a low sensitivity, 49-67% by some reports, and as low as 14% in depressed outpatients (Health and Public Policy Committee, ACP, *Ann Int Med* 1984;100:307–308).

Recently, we've seen a spate of potential biomarkers for various psychiatric disorders. Before examining the evidence, there is a major inherent limitation to all of this research, which is that we have no true gold standard diagnostic test in psychiatry against which we can judge the accuracy of a new test. Other specialties have such gold standards (such as biopsies for cancer or cardiac catheterization for coronary artery disease). At this point, our only gold standard is the interview.

Recently introduced biomarkers in psychiatry come in three flavors. They can be diagnostic (does the patient have a disease?); prognostic (will the patient develop a disease?); or predictive or "theranostic" (will the patient respond to a particular therapy?) (Weickert CS, et al, *Disease Markers* 2013;35(1):3–9). While we don't have room to review each and every potential biomarker out there, we will focus on some that have received the most attention. See the chart on page 3 for a list of all biomarkers being marketed (or close to being marketed) in psychiatry.

Diagnostic Biomarkers

MDDScore. The MDDScore, marketed by Ridge Diagnostics, is a blood test that purports to diagnose depression. It measures the levels of nine proteins in a single sample of the patient's blood. These values are fed into an algorithm and a score is reported, ranging from one to nine, where anything greater than or equal to five represents "90% likelihood of depression," according to their patient materials.

Two studies have evaluated the accuracy of the test. In the first study, blood samples were taken from 70 depressed patients and 43 non-depressed patients in a clinical population, and the algorithm predicted a diagnosis of depression with 91% sensitivity and 81% specificity (Papakostas GI, Shelton RC

et al, *Molec Psychiat* 2011;18:332–339). This sounds impressive, but all we can really say is that the depressed patients differed from healthy controls on certain measures of inflammation, stress response, and metabolism. We don't know whether patients with other conditions such as bipolar disorder or schizophrenia share the same profile, nor do we know the formula. Notably, the study excluded patients on antidepressants, antipsychotics, or NSAIDs in the two weeks prior to the blood test, and those patients with "serious and/or unstable medical disorders." (For more, see Sensitivity and Specificity: A Refresher on page 6.)

A follow-up study was published earlier this year, which also demonstrated impressive sensitivity (96%) and specificity (86%) on a similarly selected group of 68 depressed and 86 non-depressed patients. However, while they measured the same nine proteins, they didn't use the same algorithm as in the earlier study. In fact, they "trained" the algorithm—a fancy way of saying they ran it over and over on a subgroup (N=102) of the total subjects, tweaking the algorithm each time to make it positive for the depressed patients and negative for the non-depressed patients, and then "validated" it on the remaining 52 patients. They also made adjustments for gender and body mass index (BMI) and normalized cortisol levels, which the initial study did not do.

In other words, the "MDDScore" seems to measure something of interest in patients with depression, but we don't know how—or whether—it differentiates depression from other disorders, nor whether it can guide treatment (eg, which SSRI or SNRI, or maybe psychotherapy). And with the two replication studies published by the manufacturer, both with proprietary algorithms, it's hard to know what's exactly being tested when you order this test. At a cost of $826.45 ($385 if the patient pays within 30 days), but with coverage by "most insurance companies," according to Ridge Diagnostics, we're not so sure how this test adds any information that improves upon treatment as usual.

VeriPsych. The VeriPsych assay, introduced in 2010, measures the

Using Psychiatric Biomarkers in Your Practice

levels of 51 proteins in a single blood sample with the goal of diagnosing schizophrenia. Selection of the 51 proteins was similar to the process used in the MDDScore, with a few differences. Basically, researchers measured the levels of 181 proteins from serum samples of 250 acutely psychotic, treatment-naïve patients and 230 controls. They found significant differences in the levels of 22 of these proteins. Nine additional proteins were added to the assay because they have been "associated with schizophrenia" in earlier research, while an additional 20 proteins were selected because they're expressed differently in patients with bipolar disorder; they included these proteins in hopes of differentiating between the two disorders.

As with the MDDScore, they developed a mathematical "decision rule" based on levels of these proteins in a test sample of 577 schizophrenic patients and 229 controls. The resulting formula correctly identified schizophrenia in this population with a sensitivity and specificity of 83% each (Schwarz E, Izmailov R et al, *Biomarker Insights* 2010;5:39–47).

The VeriPsych test was initially offered at a cost of $2,500. In a 2011 interview, the Chief Medical Officer of Myriad RBM, developer of the test, reported that "hundreds of patients" had been tested and that it was covered by "numerous insurance carriers" (Kaplan A. Blood Tests for Diagnosis of Schizophrenia and Depression *Psychiatr Times*. 2011;28(8)). However, in January 2013 the company suspended the test. The high cost may have been one factor, as well as the fact that the test failed to differentiate between schizophrenia and other psychiatric conditions (source: http://bit.ly/1eWorX2). Meanwhile, many of the same scientists have moved to the SchizDx project (http://schizdx.pera.com/) to continue work on the test, so stay tuned.

Predictive Biomarkers

ATR Score. The ATR score is based on using quantitative EEG (qEEG) to predict response to particular antidepressants (See *TCPR*, November 2009 for earlier coverage of the test). The score incorporates three frontal EEG features (theta power, alpha power, and alpha2 power) measured at baseline and one week after a patient starts medication. The scores are given on a scale of 0 to 100, which corresponds to probability of response.

Early studies showed that the ATR score predicted response to escitalopram (Lexapro) with an accuracy of 74%, while other measures like serum drug levels and genetic polymorphisms were not predictive (Leuchter AF, Cook IA et al, *Psychiatry Res* 2009;169:124–131). In a separate study, the ATR score showed a different result in responders to fluoxetine than in those who responded to placebo, indicating a treatment-specific biological change (Hunter AM et al, *J Clin Neurophysiol* 2011;28(5):478–482).

Despite years of research and having teamed up with a medical device company (Aspect Medical, now a division of Covidien) the ATR has not yet reached clinical practice. If and when it does, the value of enhanced (though not perfect) treatment selection must be balanced against the significant capital investment and technical expertise it will undoubtedly require. So its true utility may be limited.

Others. Not a month goes by without a new biomarker study published in a major psychiatric journal. But most need to be replicated and, as of yet, none can be easily ordered from your office. For instance, a study showed last year that reduced expression and increased methylation of a gene called *SKA2* was linked to suicidal behavior (Guintivano J et al, *Am J Psychiatry* 2014;171(12):1287–1296). Another group found that suicidal ideation was associated with differential expression of six RNA transcripts in peripheral blood (SKA2 was not among those that made the cut) (Le-Niculescu H et al, *Mol Psychiatry* 2013;18(12):1249–1264).

Brain imaging can also hint at

Some of the Most Common Psychiatric Biomarkers Currently on the Market					
Name of Biomarker	**How Does It Work?**	**Source Material?**	**Why Would I Use It?**	**Who Makes It?**	**Approximate Cost?**
MDDScore	Measures levels of nine proteins, proprietary algorithm	Blood sample	Diagnosis of major depressive disorder (MDD)*	Ridge Diagnostics	$825
VeriPsych	Blood test; measures levels of 51 proteins, proprietary algorithm	Blood sample	Diagnosis of schizophrenia*	SchizDx (European Union consortium)	N/A ($2,500 when previously available)
ATR Score	Quantitative EEG (qEEG)	EEG	Predicts response to antidepressants	Covidien	N/A
NEBA	Ratio of theta to beta brain waves	EEG	Diagnosis of ADHD*	NEBA Health	$325
GeneSight	Genetic test	Cheek swab	Genotype of various genes involved in drug metabolism, with medication recommendations	Assurex Health	$3,800
Genecept	Genetic test	Saliva sample or cheek swab	Genotype of various genes involved in drug metabolism and possibly disease states	Genomind	$750**
florbetapir-F18	Detects amyloid plaques in brain (http://bit.ly/1Hj2X6q)	PET scan	Assists in diagnosis of Alzheimer's disease	Amyvid; Eli Lilly	variable

Biomarker tests in *italics* are not presently available
* = recommended for diagnostic use in conjunction with clinical evaluation
** = cost to patient self-purchasing test

was an assistant professor at Michigan just after Barney Carroll, who established the DST, left. So I learned quite a bit about that research. He would get severely depressed women who were inpatients, in order to get a homogeneous sample. They were on multiple medications and he would get them off of all the medications. He would give them the steroid dexamethasone in the evening and the next morning he would check cortisol levels. Normally, dexamethasone should cause morning cortisol levels to decrease. But there were some women who were quite severely depressed whose cortisol levels remained unchanged.

TCPR: So that is what is called "failing the dexamethasone suppression test?"

Dr. Coyne: Exactly. And at first the data seemed so compelling for this biomarker that people were suggesting that we stop using the inexact label "depression" and instead start calling the disorder hypercortisolemia. The problem is that the psychiatry residents started testing each other, and if they gave dexamethasone suppression tests to themselves the night before they were doing a grand rounds presentation they looked like one of these patients, or if they took up running they would look like one of these patients. And so obviously it wasn't very specific, but in this narrow range of patients it seemed to work at first.

But it failed when they tried to cross validate it in other samples of depression. And there were also issues with how people did the DST test, but the bottom line is once they got away from that very specific patient population they could not reliably validate the test. There were too many false positives and false negatives.

> "So you have a blood test that lights up when a patient has depression. But will it also light up for bipolar depression, or anxiety, or ADHD, or even general stress?"
>
> James Coyne, PhD

TCPR: So the DST serves as an interesting kind of cautionary tale for the future development of biomarkers. How do we use that lesson and apply it more generically to these other biomarker tests?

Dr. Coyne: It is very easy to find a biological abnormality that some depressed people, but not all, have. Anything from measuring electromyography of facial muscles when people are asked to imagine different things, to blood tests, to genomic expression, but these various biological measures correlate only modestly with each other. And a lot of people in whom clinically we're absolutely convinced need treatment for depression won't have the abnormality, and on the other hand, there are a lot of people who have the abnormality, but don't seem to need treatment. So we are not yet at the point where we are able to overrule a diagnostic interview—nor do we have a test that adds significant value to the interview.

TCPR: Now when you talk about these false positives and false negatives, I know a lot of these studies will report high specificity or high sensitivity rates, and I think people are captivated by those numbers. What are the potential pitfalls here?

Dr. Coyne: The biggest problem is that researchers will publish their first results based on an exploratory study. These first studies are not truly testing the biomarker, but rather are records of how they got the best possible results for the test in their particular sample in the first place. For example, let's say you have an idea for a blood test for depression and you think there are 10 possible metabolites that might correlate with depression. So you recruit 50 people who are clinically depressed (based on an interview), and 50 people with no depression, and you draw everyone's blood. Then you statistically start fishing for the combination of those 10 metabolites that do the best job of picking out just the people who are depressed. So let's say you come up with five metabolites that work well. If you publish those results, you'll be able to say "Look, we came up with a blood test of five metabolites and it has 90% sensitivity and 90% specificity for depression."

TCPR: So what's wrong with that claim?

Dr. Coyne: Because you've picked a certain sample and you've tried to fit your blood tests to that sample. If you choose a completely different sample of depressed patients, it is very unlikely that those five metabolites are going to pick out the depressed patients. And this is exactly what typically happens with these tests. The first paper is promising, but when the researchers try to replicate the test with a different sample, the results are much less impressive.

TCPR: This sounds like a "moving the goal post" problem. In the first studies you adjust the position of the goal post to make sure the football clears it. But in later studies when you can't move the goal post you stop hitting your field goals.

Dr. Coyne: Right. Another problem is a specificity issue. These studies almost always compare "pure" depressed patients with normal patients. So you have a blood test that lights up when a patient has depression. But will it also light up for bipolar depression, or anxiety, or ADHD, or even general stress? Unless you do a study comparing patients with all these diagnoses you won't know. So in the clinic, if you were to start testing a bunch of real patients, you wouldn't know how to interpret a positive result.

TCPR: So what should we be looking for in studies for biomarkers?

Dr. Coyne: At a bare minimum you have to have two samples: the discovery sample, which is where you are creating the test; and a separate validation sample where you see if the test you created works, without changing it or tweaking it to fit the data. And beyond that you want to recruit three types of patients: those with the diagnosis of interest; those with some other, related diagnosis; and healthy controls.

TCPR: What about sample size? Is 60 enough? 100? How do we determine that?

Dr. Coyne: You have to have some estimate ahead of time of how big the effect is that you are looking for. There are tables that we can look up and decide what size sample we'd need to find a certain sized difference between groups. For instance, I could

The GeneSight Genetic Test: A Review of the Evidence

Assurex Health recently sent me an email inviting me to dine at Legal Seafood to learn about "Clinical Applications of Psychiatric Pharmacogenetics." I didn't go, but increasingly I am hearing from colleagues about their experiences at these dinner programs: "What do you think about this GeneSight test? The data looked pretty impressive at this dinner."

Clearly, the field of pharmacogenetic testing is growing up when companies can afford this kind of promotional money nation-wide. There are now several companies marketing such tests—including Assurex, Genomind, Genelex and others.

We covered this topic last fall (*TCPR*, October 2014) and concluded that there was not enough good evidence for using genetic testing in routine practice. In the following, I'll zero in specifically on GeneSight to review the data—so that you'll be more informed if you choose to dine courtesy of Assurex. Good food and wine tend to dull your critical faculties, and you don't want a company to hypnotize you into adopting a very expensive test unless it will really help your patients.

The Context of Pharmacogenetic Testing

Humans vary genetically—not only in eye and hair color but in more obscure ways, such as how we metabolize drugs. While the majority of our patients metabolize drugs normally, a small percentage do not. Using the most important enzyme, 2D6, as a benchmark, roughly 5–10% of Caucasians are poor metabolizers, and 1% are ultra-rapid metabolizers (Ingelman-Sundberg, M, *Pharmacogenomics* J 2005;5(1):6–13). Of the remainder, the majority are "extensive" (or normal) metabolizers. You should know that these frequencies vary among ethnic groups—for example, only about 1% of Asians are slow metabolizers.

Before I discuss GeneSight, an important but often overlooked point is that the recent GeneSight studies were preceded by decades of disappointing clinical studies in this field. Two large reviews of the literature on pharmacogenetics in mood disorders—one in 2007 (*Genet Med* 2007;9(12):819–825) and one in 2013 (*CNS Spectr* 2013;18(5):272–284)—could find no association between metabolizer status and response of depressed patients to SSRIs. Tricyclics are a different story, with evidence-based guidelines recommending dosage adjustments based on P450 testing (Hicks JK et al, *Clin Pharmacol Ther* 93(5):402–408, 2013). But most psychiatrists are not using the new genetic tests for tricyclics, which we rarely prescribe.

The take home point is that there's a long history of negative or inconclusive studies in the field—so GeneSight's evidence had better be pretty convincing before we decide to change our clinical practice based on it.

The GeneSight Test

The basic GeneSight test as evaluated in their clinical trials analyzes patient DNA for genes encoding three metabolic enzymes and two serotonin-related proteins. The three enzymes are 2D6, 2C19, and 1A2—all of which are located in the liver and are involved in metabolizing various medications. The other two molecules are SLC6A4 (the serotonin transporter gene) and HTR2A (the serotonin 2A receptor gene). (GeneSight has expanded its testing panel since the clinical trials—you can find the current list on its website, www.genesight.com).

The metabolic enzymes are clearly relevant for metabolism—no surprises there. But why did they include the serotonin genes in the assay? Presumably because they can theoretically affect the antidepressant response. However, there's no scientific consensus that we have found the right genes yet. The most definitive meta-analysis found

no consistent evidence of a relationship between serotonin genes and response to any antidepressant (*Am J Psychiatry* 2013;170(2):207–217). So GeneSight's adding these genes appears to serve little use other than to make their assay appear more robust than a simple test of P450 enzymes.

The testing process is quite simple: you use cotton cheek swabs to collect the DNA, the samples are overnighted to Assurex, and results are provided within 36 hours. As a doctor, you'll be sent a report classifying a list of 38 psychiatric drugs into three possible categories: green bin ("use as directed"), yellow bin ("use with caution") and red bin ("use with caution and with more frequent monitoring"). An easy example is Paxil, which is metabolized primarily by 2D6. Paxil will presumably show up in the red bin in two cases: if your patient is a poor metabolizer (because Paxil doses could go too high) or if he is an ultra-rapid metabolizer (the dose could be too low).

It all makes sense theoretically. But what about in actual practice?

The GeneSight Evidence

Three clinical studies have been conducted thus far. See the chart on this page for a brief summary of the main findings. There have been two open label studies, and one randomized, "double-blind" study (I'll explain why I put quotes around that in a second). The methods of the studies are similar. Patients who are taking medications for depression are recruited from a clinic. They are assigned to one of two groups. In the guided group, patients are given the GeneSight test, the prescriber sees the results, and is free to make changes in the patients' meds based on the results of the test. In the unguided group, patients get the test, but the results are sealed until after the study is over. Patients are periodically evaluated

Study	Hamm Clinic, 2012	LaCrosse Clinic, 2013	Pine Rest, 2013
Methodology	Open-label, non-randomized	Open-label, non-randomized	Randomized, blinded to patients and raters, but not to prescribers
Number of patients	44 (22 in guided, 22 unguided)	165 (72 in guided, 93 unguided)	49 (25 in guided, 24 in unguided)
Ham-D improvement	Guided: 30.8%, Unguided: 18.2%	Guided: 46.9%, Unguided: 29.9%	Guided: 30.8%, Unguided: 20.7%
Ham-D response rate	Not reported	Guided: 43.1%, Unguided: 26.9%	Guided: 36%, Unguided: 20.8%
Statistically significant?	Yes	Yes	No

References: Winner JG et al, *Discover Medicine* 2013;16(89):219–227, Hall-Flavin DK et al, *Pharmacogenet Genomics* 2013;23(10):535–548, Hall-Flavin DK et al, *Transl Psychiatry* 2012;2:e172

Using Psychiatric Biomarkers in Your Practice

new biomarkers. Researchers at Emory University have used PET scans to show that low activity of the right anterior insular cortex predicts a good response to psychotherapy, while high activity predicts a response to escitalopram (McGrath CL et al, *JAMA Psychiatry* 2013;70(8):821–829). And the list goes on and on.

The "Holy Grail" of Psychiatry?

While diagnostic biomarkers might someday revolutionize psychiatric diagnosis that day is not yet here. The experts still emphasize the value of the clinical interview and advise the use of biomarker tests as confirmatory or supportive evidence. However, one potential advantage of biomarkers is to identify new targets for medications and other physiological therapies, as in other areas of "precision medicine."

The operative word here is "might." Despite decades of research dating back to the DST, no biomarker test has revolutionized psychiatric care. Even the most commonly used biomarkers—pharmacogenetics—are questionable in their ability to improve psychiatric care over treatment as usual (see the companion article on GeneSight in this issue on page 5).

Sensitivity and Specificity: A Refresher

SENSITIVITY: the ability of a test to correctly identify the presence of a feature

$$\frac{\text{True positives}}{\text{True positives} + \text{False negatives}} = \frac{\text{True Positives}}{\text{Everyone who } \textit{actually} \text{ has the feature}}$$

SPECIFICITY: the ability of a test to correctly identify the absence of a feature

$$\frac{\text{True negatives}}{\text{True negatives} + \text{False positives}} = \frac{\text{True negatives}}{\text{Everyone who } \textit{doesn't} \text{ have the feature}}$$

For sensitivity and specificity calculations, it's important to know what the feature is being tested for. In the case of diagnostic biomarkers, it's a diagnosis which calls into question what is the "gold standard" in making that diagnosis, and whether it can account for the heterogeneity in patients who carry that diagnosis. For a predictive biomarker the feature is usually a biochemical or genetic measure, which is usually much easier to detect (you either have it or you don't), but then the question becomes how good is this measure in predicting outcomes?

The GeneSight Genetic Test: A Review of the Evidence

with standard depression scales, and the study length was eight weeks in the open label trials, 10 weeks in the randomized trial. The main outcome is whether patients assigned to the guided group improve more than those assigned to the unguided group.

Open Label Results

The open label studies found a statistically significant effect of using GeneSight to guide treatment. Open label studies are easy to conduct, and they're great for generating hypotheses, but we shouldn't rely on them to make clinical decision. Remember gabapentin? Open label studies found it apparently effective for bipolar disorder—but subsequent double-blind studies did not.

GeneSight's open label studies are vulnerable to various possible biases that might render the results meaningless. Here are some potential problems.

1. Patients were not assigned to groups randomly, but were chosen based on conversations with doctors and researchers. One potential source of bias: doctors may have preferentially assigned more complicated patients to the guided group on the theory that they would be most helped by genetic testing. If so, we wouldn't really know if the results are applicable to all the patients we see, or just some undefined subset.
2. Patients assigned to the guided group knew they were getting a cutting-edge genetic test that could predict which medication would work best for them. Patients in the other group knew the test results would not be used for their treatment. Clearly, those in the guided group would be more optimistic about their treatment, which is a key component of the non-specific placebo effect.
3. Prescribers knew which patients were being guided by the test, potentially leading to the "cheerleader effect."
4. The symptom raters knew which patients were in which group, potentially leading to biased ratings, since the raters may have vested interests in GeneSight being successful.

The bottom line is that the open label studies can tell us very little other than that GeneSight appears to have potential—and that it's time to do the more expensive randomized double-blinded tests.

Randomized Results

The randomized double blind study was not "double" blind in the way drug trials are. The idea behind the double blind is that neither the researchers nor the patients know which group they are assigned to, so that there is no possibility of various biases or placebo effects creeping in. The GeneSight double blind study was blinded only to patients and symptom raters, and not to prescribers, who might have been cheerleading their patients to wellness.

Nonetheless, the results of that study showed a numerical superiority of guided treatment, but it was not statistically significant. For example, the difference in the Ham-D scores had a p value was 0.29—which means that there was a 29% chance that this difference was due to chance. The usual cut-off for statistical significance is 5%, so this result was not close to being significant. It's possible that if they had recruited more patients, they might have found a significant difference. But for now we have to conclude that there is no convincing evidence that the GeneSight test helps us prescribe more effective medications for our patients

However, there is a bit of a silver lining for GeneSight in this study. In a subanalysis, the author focused on the 13 patients who entered the study taking antidepressants that were classified in the red bin—in other words, "use with caution and with more frequent

CME Post-Test

This CME post-test is intended for participants seeking AMA PRA Category 1 Credit™. For those seeking ABPN self-assessment (MOC) credit, a 13 question pre- and post-test must be taken online. For all others, to earn CME or CE credit, you must read the articles and log on to www.TheCarlatReport.com to take the post-test. You must answer at least four questions correctly to earn credit. You will be given two attempts to pass the test. Tests must be taken by May 31, 2016. As a subscriber to *TCPR,* you already have a username and password to log on www.TheCarlatReport.com. To obtain your username and password or if you cannot take the test online, please email info@thecarlatreport.com or call 978-499-0583.

The Carlat CME Institute is accredited by the Accreditation Council for Continuing Medical Education to provide continuing medical education for physicians. Carlat CME Institute is also approved by the American Psychological Association to sponsor continuing education for psychologists. Carlat CME Institute maintains responsibility for this program and its content. Carlat CME Institute designates this enduring material educational activity for a maximum of one (1) *AMA PRA Category 1 Credit*™ or 1 CE for psychologists. Physicians or psychologists should claim credit commensurate only with the extent of their participation in the activity.

Below are the questions for this month's CME post-test. This page is intended as a study guide. Please complete the test online at www.TheCarlatReport.com. Note: Learning objectives are listed on page 1.

1. The dexamethasone suppression test (DST) was developed as a biomarker for _____. (Learning Objective #1)

     [ ] a) Panic disorder   [ ] b) Major depression   [ ] c) Obsessive-compulsive disorder [ ] d) Bipolar disorder

2. Early studies showed that the ATR score predicted response to escitalopram (Lexapro) with an accuracy of _____. (L.O. #1)

     [ ] a) 42%   [ ] b) 74%   [ ] c) 87%   [ ] d) 90%

3. The best studies of biomarkers for a particular diagnosis should recruit the following samples of patients: (L.O. #2)

     [ ] a) Those with the diagnosis and healthy controls

     [ ] b) Those with the diagnosis, healthy controls, and those with a different diagnosis

     [ ] c) Those with varying severities of the diagnosis

     [ ] d) Healthy controls and those with a different, though related diagnosis

4. Poor specificity is often a problem with biomarker studies, which means that the test: (L.O. #2)

     [ ] a) Is positive in the diagnosis of interest, but also positive in other disorders.

     [ ] b) Is negative in the diagnosis of interest, but positive in other disorders.

     [ ] c) Is positive in the diagnosis of interest, and negative in other disorders.

     [ ] d) Is negative in essentially all disorders tested.

5. The basic GeneSight test analyzes patient DNA for genes encoding two serotonin-related proteins and three types of: (L.O. #3)

     [ ] a) Proteases   [ ] b) Metabolic enzymes   [ ] c) Digestive enzymes   [ ] d) Dopamine-related proteins

PLEASE NOTE: WE CAN AWARD CME CREDIT ONLY TO PAID SUBSCRIBERS

Expert Interview

say I want a 90 percent probability that if there is something present in this sample on this variable that I could find it. And a lot of these studies are done with too small a sample. If you have 20 or 30 patients then you probably have less than a 50/50 chance of finding something even if it's there.

TCPR: So if you have a very small study and they find a statistical effect, does that mean you have more or less confidence in the result? It sounds like you should be more confident, because with small sample sizes you'd have to have a really big effect for it to be significant.

Dr. Coyne: But the paradox is that small sample study results are more dependable when they yield negative findings then when they claim positive findings because positive findings are more likely to be due to chance. And then throw in a strong preference for publishing positive results, even if they are false or exaggerated. The history of the literature is that large effects from small studies don't replicate. I would operate on the assumption that most discoveries with small samples aren't going to hold up.

TCPR: So to sum up, we should be looking at several aspects of biomarker studies of depression. We are looking for participants in the study that are legitimately diagnosed with depression, and that are generalizable to the kinds of real patients that we see. We are looking for enough participants so that you would expect to find a difference that is not due to chance alone. And we are looking for a design that doesn't just compare depressed patients with healthy patients but also looks at the marker in related diagnoses.

Dr. Coyne: And you also have to be aware that there is a lot of incentive out there to have a false discovery because the journals are going to be excited, the media is going to be excited, and potentially funders are going to be excited. A lot of people want to rush to get a start-up company going.

TCPR: Thank you, Dr. Coyne.
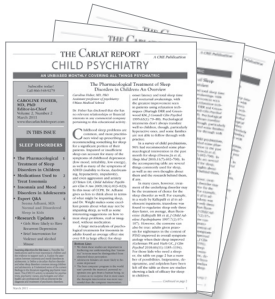
THE CARLAT REPORT: PSYCHIATRY

P.O. Box 626
Newburyport, MA 01950

This Month's Focus:
Biomarkers in Psychiatry

The GeneSight Genetic Test: A Review of the Evidence

monitoring." Six of these were randomized to the guided group, and their doctors used the results to adjust the meds of all of these patients, who ended the study with a 33.1% Ham-D improvement. On the other hand, the seven patients who were assigned to the unguided group were less likely to have their meds changed, and they improved by only 0.8% on the Ham-D. This is intriguing, and raises the possibility that GeneSight might be useful for some patients. But remember—this is based on a subanalysis of 13 patients from an already tiny study.

Bottom Line

If we were to hold the GeneSight test to the usual standards we require for making medication decisions, we'd conclude that there's very little reliable evidence that it works. On the other hand, some of you will probably want to try it out, especially for those patients who have insurance that will cover the cost of the test. If you do order it, reserve it for patients who are most likely to benefit, including patients who have failed to respond to multiple medications (which could be caused by ultra-rapid metabolism, causing drug levels to be too low), and patients who have had lots of side effects (potentially caused by slow metabolism, causing drug levels to he too high).

DR. CARLAT'S VERDICT: GeneSight—a little evidence, a lot of wishful thinking.