THE CARLAT REPORT PSYCHIATRY

A CME Publication

AN UNBIASED MONTHLY COVERING ALL THINGS PSYCHIATRIC

Steve Balt, MD
Editor-in-Chief
Volume 12, Number 7&8
July/August 2014
Special Double Issue:
worth 2 CME Credits
www.thecarlatreport.com

IN THIS ISSUE

Focus of the Month:
Research in Psychiatry

Learning objectives for this issue:
1. Determine sources of bias in medical research articles. **2.** Effectively read and interpret a research paper. **3.** Describe the FDA drug approval process. **4.** Detail some of the methods by which associations are found in clinical data. **5.** Explain the NIMH's research domain criteria (RDoC). **6.** Summarize some of the current findings in the literature regarding psychiatric treatment.

Investigating Bias in Research

Rebecca Twersky-Kengmana, MD
Clinical Assistant Professor of Psychiatry, Albert Einstein College of Medicine, New York, NY

Dr. Twersky-Kengmana has disclosed that she has no relevant relationships or financial interests in any commercial company pertaining to this educational activity

Unbiased medical research is essential to the process of informed clinical decision-making. Unfortunately, some degree of research bias is unavoidable, so the best approach to conducting a study is to attempt to minimize bias, wherever possible.

Bias in research can be defined as any "condition that produce(s) results which depart from the true values (in a study) in a consistent direction" (Riegelman RK. *Studying a Study and Testing a Test.* 4th ed. Philadelphia: Lippincott Williams & Wilkins;2000). Bias is also referred to as "systematic error."

It's important to understand what

Continued on page 2

In Summary

Bias in medical research

- is related to the unintentional skewing of results due to human failings such as poor trial design or unconscious wishes for desirable outcomes
- can occur anywhere in the research process from initial concept to publication
- is inherent, but often preventable, in study design, study implementation, data analysis, and post study/publication

Q&A With the Expert

Research Domain Criteria (RDoC)
Bruce Cuthbert, PhD

Director, Division of Adult Translational Research and Treatment Development
National Institute of Mental Health (NIMH)

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

TCPR: Dr. Cuthbert, how did the Research Domain Criteria (RDoC) initiative begin?

Dr. Cuthbert: Around 2008 there was a feeling that, compared to other areas of medicine, we were not moving ahead in the prevention and treatment of psychiatric disorders. If you look at areas like leukemia, HIV, stroke, and heart disease, we have seen vast decreases in morbidity and mortality over the last 20 to 30 years, but the same is not true for mental disorders. Our vision statement is that we envision a world in which mental illnesses are prevented and cured. So we are trying to work toward transforming understanding and treatment of mental illness through research, and paving the way for prevention, recovery, and cure—not simply managing disorders.

TCPR: There have been concerns that the *DSM* system is no longer useful for psychiatric research. Can you explain?

Dr. Cuthbert: The *DSM* criteria have not kept up with developments in genetics,

Continued on page 4

Investigating Bias in Research
Continued from page 1

bias is *not*. Bias is not fraud, as it does not involve intentional prejudice. Bias is also not random error, which creates deviation in results strictly by chance, and can be mitigated with a large enough study sample.

Put simply, bias has to do with the unintentional skewing of results due to human failings such as poor trial design or unconscious wishes for desirable outcomes.

Bias can occur anywhere in the process from initial concept to publication. For convenience, we will examine four stages in the research process—study design, study implementation, data analysis, and post study/publication—and the biases inherent in these stages (although there is a certain amount of overlap).

Biases in Study Design

To avoid bias, any study needs to be fully and clearly conceptualized from the outset.

Conditions and subjects need to be well defined to avoid ***definition bias.*** Examples of definition bias are ambiguity about what constitutes severe versus moderate anxiety, or what score on a symptom rating scale qualifies a subject for a trial.

Similarly, outcome measures should be as objective as possible. This can be tricky. Rating scales *seem* objective, but when used properly they may require extensive subjective judgment on the part of the rater.

The Hamilton Rating Scale for Depression (HAM-D), for instance, is often used to rate severity of depression. In the 1960 article introducing the scale, its creator, Max Hamilton, wrote that it should be "used for quantifying the results of an interview, and its value depends entirely on the skill of the interviewer in eliciting the necessary information" (Hamilton M, *J Neurol Neurosurg Psychiat* 1960;23:56–62).

Relevant outcome measures should be defined before the study is conducted. However, it's sometimes impossible for the reader to determine what these are. For instance, a study's primary outcome (eg, depression severity) may be negative, while a secondary outcome (eg, anxiety symptoms) gives statistically significant results.

Reporting only the secondary outcome may be an attempt to salvage a serendipitous—and sometimes useful—conclusion from what would otherwise be a negative and costly study, but it wasn't what the researchers intended to study.

Selection bias occurs when the study and control groups differ in ways that might influence the outcome of the study. This can be controlled for by matching study and control groups to ensure that a particular variable, such as average age or health status, is the same in both groups. Unfortunately, the matched variable can no longer be studied as a determinant of outcome.

In *pairing,* a type of matching, individual subjects are paired with individual controls who have exactly the same characteristics, apart from the ones the study is assessing.

Alternatively, some studies use a *crossover* study design, in which the subject is used as his or her own control. For example, the same patient can be assessed on no medication, and later, on the study medication. This may be a problem because the subject may know when he or she is taking a medication (versus placebo). There may also be carryover effects from one phase to the next. This can be avoided through the use of a *washout* period.

Self-selection bias can occur whenever volunteers are recruited for a study, since volunteers are often not representative of the general population. A related bias, frequently found in medical research, is known as ***Berkson's bias,*** in which subjects are selected from an "enriched" population. For example, an epidemiological study of patients recruited from a specialty mood disorders clinic may identify characteristics that are not widespread among the more general population of psychiatric outpatients.

Biases in Study Implementation

Interviewer bias refers to a systematic skewing of the way data is elicited, recorded, or interpreted (Pannucci CJ & Wilkins EG, *Plast Reconstr Surg* 2010;126(2):619–625). For example, an interviewer might ask, "Have you ever been diagnosed with social phobia?" as opposed to questions like "Do you ever get anxious in crowds?" The best remedy for interviewer bias is to use a standardized measurement tool and to ensure the interviewer is blind to the subject's history.

Bias can also be introduced by the subjects themselves. ***Recall bias*** occurs when the subjects in one arm of a study are more likely to remember past events than the subjects in another arm. This type of bias commonly occurs in case-control studies of traumatic outcomes. For example, women who give birth to babies with congenital malformations may be more likely to remember taking psychotropic drugs during pregnancy than women with normal babies.

Similarly, ***reporting bias*** is common when sensitive information is being elicited, such as a patient's sexual history. A good study design should try to reduce recall and reporting biases, for example by assessing specific symptoms, side effects, or historical details, rather than asking open-ended questions.

Performance bias, in which there are selective differences between the care that is given to different groups, is a

Continued on page 3

Investigating Bias in Research
Continued from page 2

particular problem in studies of psychotherapy. Some therapies, such as CBT, are fairly easy to standardize (a process called manualization). But with other psychotherapies, it's harder to control for variations in treatment, and differences in therapists' styles and levels of experience can heavily influence outcomes.

Biases in Data Analysis

Confounding factors can be a significant source of bias in data analysis. For example, a study might conclude that schizophrenia is a risk factor for lung cancer. This conclusion misses the role that smoking plays in both illnesses. To correct for this, an adjustment must be made that separately analyzes subjects who smoke and those who do not.

A major issue in psychiatric clinical trials is how to deal with subjects who drop out of the study, also called ***attrition bias***. There are two ways researchers can deal with the lost data.

In a *per-protocol* analysis, drop-outs from the active arm of a trial are treated as though they had never been given the active intervention. In other words, only the subjects who finished per-protocol are included. Thus, a per-protocol analysis only looks at those subjects in the active arm of a trial who found an intervention tolerable and potentially helpful, not those who dropped out for any reason, so it usually overestimates the effect of the intervention.

In contrast, an *intention-to-treat* analysis is designed to measure everybody who enters the trial (ie, everyone who was intended to be treated). When subjects drop out before the end of the trial, their final results may be estimated based on measurements already taken.

Alternatively (and more commonly), researchers can take the most recent assessment and assume that it will also be the subject's final measurement. This is called last-observation carried forward (LOCF). LOCF is much more conservative, making it the more acceptable approach, even though it is, strictly speaking, an inaccurate representation of the data.

Biases in Post Study/Publication

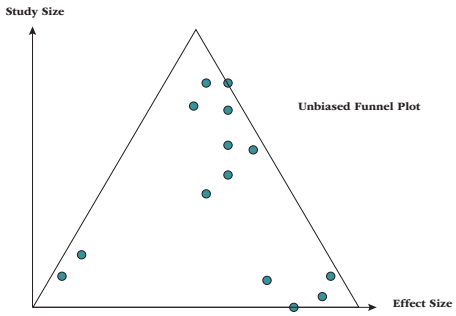
Publication bias is the tendency not to publish small trials that don't demonstrate statistical significance. It's understandable that this happens—if a study's researchers have nothing to show for their efforts, they may choose not to devote the time and effort to writing up their results.

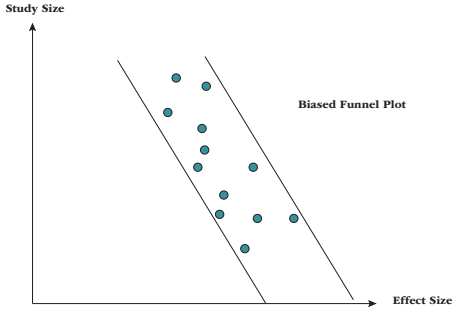
Stage of Study	Types of Bias
Design	definition, selection, self-selection, outcome measures
Implementation	interviewer, recall, reporting, performance
Analysis	confounding factors, attrition
Publication	publication, grey literature, outcome reporting

For example, researchers found in 2008 that out of 74 antidepressant trials submitted to the FDA, 38 were positive and 36 were negative. Interestingly, nearly all (37) of the positive trials were published in peer-reviewed journals, while only six of the negative trials were published, sometimes spun to *sound* positive (Turner EH et al, *NEJM* 2008;358:252–260). The average reader wouldn't know this, and may be practicing evidence-based medicine *without* all the evidence (see also "Overview of the FDA Drug Approval Process" in this issue).

When multiple published studies are compared, as in a meta-analysis, publication bias is easy to detect, by using a funnel diagram. The X-axis is a measure of treatment effect (eg, effect size, very low to very high), while the Y-axis is a measure of the size of the trials.

When there is no publication bias, the graph resembles an upside down





funnel: smaller studies are more subject to chance, so the reported effect of treatment can be widely variable, whereas larger studies tend to report comparable effect sizes near the true value of the effect. When publication bias *does* exist, the smaller trials are missing, because they haven't been published, and the results are skewed.

Sometimes researchers report only results that are significant, ignoring those that were insignificant or unfavorable, something known as ***outcome reporting bias***. Recent grass-roots initiatives like AllTrials (www.alltrials.net) demand that the protocols and Clinical Study Reports (CSRs) for all drug trials be made available to independent researchers, to determine whether authors cherry-pick their data or simply refuse to publicize trials that are negative—often called the "file-drawer effect" (See *TCPR*, Nov. 2013, for more on AllTrials.)

Another useful resource is trialsjournal.com, an open-access online journal that aims to publish results of studies regardless of outcome or significance of findings.

Related to publication bias is ***grey literature bias,*** which occurs when non-peer-reviewed sources like dissertations and posters give different results from those found in the published literature. Poster presentations at conferences can be good ways to keep up with new research, but their results are not peer-reviewed and must be considered preliminary.

TCPR'S VERDICT: Most readers of medical literature are not also professional researchers, and may not be familiar with the various biases in clinical research, both intentional and unintentional. Hopefully, readers can become more aware of the potential pitfalls in interpreting the literature to avoid drawing their own erroneous conclusions.

Expert Interview
Continued from page 1

neuroscience, and behavioral science. The system for reviewing research grant applications has been based on *DSM*-defined disorders, but researchers increasingly want to study certain mechanisms, such as working memory, that cut across disorders. Thomas Insel, the director of the NIMH, proposed that we establish a research classification system looking at dimensional aspects of behavior and of corresponding neural or biological systems like fear circuits or working memory.

We are trying to include neuroscience to help us to a more precise diagnosis, but that doesn't mean we are throwing out ideas about phenomenology.

Bruce Cuthbert, PhD

TCPR: From a researcher's point of view, how is RDoC an improvement over *DSM*?

Dr. Cuthbert: RDoC is a framework for researchers to study behaviors, neural systems, and genetics, either to look at mechanisms that cut across disorders or to look at subtypes or dimensions within a given disorder. For instance, until recently you have had to submit a research grant organized around a *DSM* category, for example, schizophrenia vs. controls. But if you say you want to study schizophrenia as compared to bipolar disorder, that would be much harder to get approved in review. RDoC has freed up researchers to directly study those concepts of interest to them without being constrained by the *DSM* system.

TCPR: How is an office-based clinician going to benefit from RDoC?

Dr. Cuthbert: The RDoC project was never intended to be a new distinct diagnostic nosology to compete with the *DSM* or the ICD. Rather, RDoC is a research project about psychopathology intended to build a research literature that will tell us how we can build a *DSM* or an ICD that is based upon genetics, neuroscience, and behavioral science. If I am a clinician in my office, RDoC can help me think about a disorder somewhat differently. Instead of saying, "Okay, this patient has social phobia, that patient has schizophrenia." I am actually looking at specific aspects of that patient's dysfunction and disability and treating that, rather than treating a diagnosis.

TCPR: This is what many psychiatrists do intuitively, is it not?

Dr. Cuthbert: Yes, in many ways we know that this is what clinicians do now. Often with a complex syndrome like schizophrenia, you are treating specific aspects, such as paranoid features, cognitive problems, or social problems. RDoC would simply provide a framework to study these aspects and to think about them systematically.

TCPR: Down the road, how do you anticipate this will change psychiatric diagnosis?

Dr. Cuthbert: It is a little early to say. I think this is the type of question that the current research is designed to answer. For instance, I could see us still diagnosing schizophrenia and bipolar disorder and looking for subtypes. But I could also see us diagnosing a broader syndromal psychotic spectrum disorder, and then specifying features such as severe cognitive dysfunction or mood disruption. Overall, we consider this as psychiatry's approach to "precision medicine." As in other areas of medicine like cancer, cystic fibrosis, or diabetes, we are increasingly realizing that broad clinical phenotypes are not valid at a more fine-grained molecular or genetic level. And that is very important for prevention, as well. For instance, if a person showed difficulties on a very sensitive cognitive task before ever developing a symptom of a disorder, we might have an intervention that is suitable—not necessarily a psychotropic drug, but possibly a behavioral or neuroplasticity intervention to prevent more severe disease.

TCPR: There is a common perception that there is antagonism between the *DSM* and RDoC. Why is this?

Dr. Cuthbert: I know there is a perceived antagonism, but in fact, NIMH and the framers of the *DSM*-5 and WHO officials have very similar goals. *DSM*-5 is committed to having what they call a "living document," by which they mean they will continually update *DSM* categories or subcategories as the data emerge from new research. We believe that we have complementary interests.

TCPR: *DSM* diagnostic criteria are the centerpiece of most drug development and pharmaceutical research today. How might the RDoC affect that area?

Dr. Cuthbert: Over the last several years, we have seen a large-scale withdrawal of the pharmaceutical industry from CNS drug development, especially for mental disorders. The usual reason given is that our disorders are so heterogeneous and vaguely defined that they lack good targets, unlike their productive targets in cancer, diabetes, and other areas. So we are trying to provide a system based on actual genetics and neuroscience that would offer them more precise targets.

TCPR: And this would be similar to how they develop drugs for other conditions?

Dr. Cuthbert: I think the industry is realizing that in an age of precision medicine they are going to have to revamp their business models. An example is the drug ivacaftor (Kalydeco) for cystic fibrosis. This is a rationally-engineered drug developed for one particular mutation of the *CFTR* gene that is present in only four percent of patients with cystic fibrosis. It was approved by the FDA after a very short clinical trial, which showed it was extremely effective for the patients who had that mutation, but didn't do much of anything for the other 96 percent. This drug is also spurring feverish research to look at all the other genotypes and engineer new drugs that would target those. [Editor's note: as of July 2014, the price of Kalydeco is over $300,000 per year.]

TCPR: Have you introduced any RDoC-based research initiatives to discover new psychiatric drugs?

Dr. Cuthbert: We have one contract mechanism in place called the Fast-Fail Trials Initiative, to screen new drugs for further testing (http://1.usa.gov/TGC0lY). Rather than looking at broad symptom categories and performing a clinical trial to see if the drug affects symptoms, we require investigators to show that their drug engages a particular receptor or affects neurotransmitter signaling (eg, through PET or fMRI imaging), or that it alters brain function in some way (eg, by cognitive testing). That way, if a drug fails, we

Continued on page 7

How to Read a Research Article

Joshua Sonkiss, MD
Medical Director, Behavioral Health Unit
Fairbanks Memorial Hospital
Fairbanks, AK

Dr. Sonkiss has disclosed that he has no relevant relationships or financial interests in any commercial company pertaining to this educational activity.

As a physician, you're expected to practice evidence-based medicine. But how can anyone keep up with the latest research? While there are lots of secondary sources of information (including *The Carlat Report*), reading original research articles allows you to reach your own conclusions about each study. But it can also be daunting.

In this article I'll discuss a focused approach for identifying and evaluating research most relevant to your practice.

Step 1: Decide What to Read

Scores of new papers appear every day, and no one can read them all. Many clinicians' eyes glaze over at the thought of reading journal articles, so I recommend that you focus on articles relevant to your own clinical cases. This primes your mind for new information and helps with recall.

If you start with a concise clinical question about a real patient, online sources like PubMed's Clinical Queries page (www.ncbi.nlm.nih.gov/pubmed/clinical) can make it easy to find relevant articles. Search engines like Google can also guide you toward primary literature, but be aware that some search results are heavy in promotional content and its associated biases.

Once you've identified a paper, just skimming the abstract doesn't cut it. There's no way to evaluate a study's caliber—or how well the results apply to your patients—without reading the actual paper. Worse yet, glancing over abstracts can lure you into accepting authors' sometimes biased conclusions at face value (See Box).

Step 2: Get your Hands on It

You can always plunk down cash, but getting full articles is expensive if you don't have a system.

If you work at a hospital or university, you probably have access to a medical library where mainstream journal articles are available for free, while less-common publications can be ordered. Some medical libraries "lend" journal articles to physicians in their communities even if they aren't university-affiliated.

PubMed's search page (www.pubmed.gov) has direct links to many free articles, and the National Library of Medicine has a page dedicated to finding full-text resources (http://1.usa.gov/1q3bvnw).

Another excellent resource is Google Scholar (scholar.google.com), which scours the Internet for PDF versions of full-text articles, and is a powerful search engine in its own right, comparable to PubMed. Finally, many professional organizations offer online access to their journals as a benefit of membership.

Step 3: Understand the Design

Once you've selected an article, you'll need to identify the study design. For an in-depth review of study designs, check out *Clinical Epidemiology: The Essentials* by Robert and Suzanne Fletcher (5th edition, Lippincott Williams Wilkins;2012).

There are many variations and hybrids, so take a close look at the "methods" section of papers you read. Most published research in psychiatry can be categorized as one of the following types:

- **Case reports:** Someone writes up an interesting case they've seen. Case reports generate hypotheses but don't test them. They are highly susceptible to bias, in part because they often describe the joint occurrence of uncommon events. Case reports rarely describe treatment failures. By definition, they only describe one patient (ie, "N=1"), so case reports almost never provide a basis for altering clinical practice.
- **Case series:** Someone writes up a small number of similar cases. Case series have no control group and don't test hypotheses, and they suffer the same susceptibility to bias as case reports. However, they reveal patterns among similar patients, and may lead to new hypotheses or suggestions for managing unusual or refractory conditions.
- **Case-control studies:** Researchers select cases *with* versus *without* a particular outcome, then ask subjects about prior exposures. For example, people with or without a current diagnosis of schizophrenia may be asked about exposure to cannabis. Case-control studies are highly susceptible to recall bias. They give an estimate of risk called the odds ratio.
- **Cohort studies:** Groups of people are followed prospectively to see how many people either *with* or *without* a particular exposure develop an outcome of interest. For example, people who do and don't smoke cannabis are followed up after 10 years to see how many in each group developed schizophrenia.

  Cohort studies allow calculation of relative risk, but they are prone to misclassification and susceptibility bias. Cohort studies can also evaluate

Misleading Conclusions

In an infamous example of biased conclusions, the author of a widely-reported meta-analysis dismissed antidepressant efficacy in all but the most severe cases of depression (Kirsch I et al, *PLoS Med* 2008;5(2):e45).

Another research team reanalyzed the same data and reached a very different conclusion: antidepressants are effective in all but the mildest cases of depression (Vöhringer P & Ghaemi N, *Clin Ther* 2011;33(12):B49–B61).

Reading only the abstracts, one might be inclined to think that the different conclusions represent solely the biases of the authors. In reality, an analysis of each paper shows that different statistical methods were used in each, and these details can sometimes lead to different conclusions.

Continued on page 7

Overview of the FDA Drug Approval Process

Glen Spielmans, PhD
Associate Professor of Psychology
Metropolitan State University, St. Paul, MN

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

After a psychiatric drug is approved by the Food and Drug Administration (FDA), a marketing juggernaut often follows, trying to convince us that the newly approved drug offers substantial benefits for treating a mental disorder.

But how does the FDA determine whether to approve a drug? What follows is a breakdown of the process.

Phases of Clinical Trials

When determining whether to approve a psychiatric drug, the FDA evaluates evidence from a series of trials, conducted in "phases." In a **phase I** trial, a drug is tested in normal volunteers and/or people with a relevant psychiatric diagnosis to examine potential adverse events and how the drug is absorbed, distributed, excreted, and metabolized.

Phase II studies look for initial evidence of efficacy. These are typically small and are frequently single-blind (ie, the researchers know what's being given but the subjects do not) or open-label (ie, everyone knows what's being given and the goals of the study). These studies should also provide an idea of the appropriate dose range.

If a drug progresses to **phase III**, it typically undergoes a series of double-blind, placebo-controlled trials. These may also involve an active treatment comparison, such as an already approved medication for treating the condition, although most psychiatric trials still compare the new drug to placebo.

Upon completion of phase III trials—or sometimes as the data are being gathered—the FDA may convene an "advisory panel" of external experts who interpret and provide advice upon on the relevant evidence regarding a drug's efficacy and safety. Typically, the FDA heeds the advice of such panels.

Phase III trials are key to determining FDA approval. Two positive studies that are "adequate and well-controlled," as defined by the FDA, are needed to demonstrate acceptable efficacy for approval. A *positive* study shows a statistically significant advantage for the new drug over placebo on the primary outcome measure, whereas a *negative* study finds no such benefit on the primary outcome.

This comes with a catch. Because only two positive studies are required—seemingly irrespective of the number of negative studies—pharmaceutical companies often conduct a slew of studies to secure regulatory approval. In fact, for several antidepressants, including citalopram (Celexa), sertraline (Zoloft), and bupropion SR (Wellbutrin), the number of negative trials actually *outnumbered* the number of positive trials. This has led one FDA historian to note that the FDA efficacy requirements can be placed anywhere between a "scintilla and a preponderance (http://1.usa.gov/QNemmn)."

Also, while a statistically significant benefit in multiple studies is generally required, keep in mind that a statistically significant effect (the drug's effect is greater than zero) is not necessarily the same as a *clinically* significant effect (drug's effect is substantial). (See accompanying articles in this issue for more on these distinctions.)

Outcome Measures

In psychiatric drug trials, the primary outcome is nearly always a clinician-rated measure of symptom severity. Some of the most common include the Hamilton Rating Scale for Depression (HAM-D) and the Positive and Negative Syndrome Scale (PANSS) for schizophrenia. Patient-rated scales, if they're even included in the trial at all, are generally not considered important in the approval process.

Broader constructs like the patient's everyday functioning or quality of life are rarely considered as critical outcomes. Consider the three FDA-approved atypical antipsychotics for treatment-resistant depression: aripiprazole (Abilify), olanzapine/fluoxetine (Symbyax), and quetiapine (Seroquel XR). On a quality-of-life measure in placebo-controlled trials, aripiprazole yielded only a very small benefit, which was not statistically significant when patients who violated study protocol were removed from the final analysis. Further, there was only a marginal benefit when patients rated their own subjective response to the drug.

No self-report depression scales were administered in trials of quetiapine or olanzapine/fluoxetine, and subjects showed no overall benefit on quality-of-life measures with these drugs. Thus, even though three drugs gave rise to statistically significant (though modest) benefit on clinician-rated measures, these benefits did not carry over to important outcomes relevant to the patient (Spielmans GI et al, *PLoS Med* 2013;10:e1001403).

Safety

Assessing the safety of new medications is another challenge altogether. Some adverse events or side effects, such as weight gain, can be observed easily and reliably, but many adverse events are more difficult to measure. In most clinical trials, participants answer nonspecific open-ended questions regarding the experience of side effects, rather than completing a structured checklist of particular adverse events. This might underestimate their frequency.

Consider the case of SSRI-related sexual dysfunction. The controlled trials upon which the FDA approved these medications found low rates of sexual side effects: only 5% to 10% (or fewer) of all subjects.

Subsequent trials, which systematically assessed the potential onset of sexual adverse events and which included only patients *without* prior sexual dysfunction, found that over ***half*** of participants taking SSRIs reported at least one such event (Serretti A et al, *J Clin Psychopharmacol* 2009;29:259–266).

This problem is now acknowledged on the drug label for all SSRIs, which says that "…estimates of the incidence of untoward sexual experience and performance cited in product labeling are likely to underestimate their actual incidence."

Other adverse events may also be similarly underreported. One study found much higher rates of adverse events via a structured checklist completed by patients, as opposed to the treating psy-

Continued on page 9

Expert Interview
Continued from page 4

know why it failed: either it didn't engage the target we thought, or it didn't really change anything in the brain even though it reached the molecular target. The idea is to fail quickly and move onto the next prospective drug rather than spending many years and millions of dollars in animal models and then failing in a large clinical trial.

TCPR: Switching gears for a moment, there is a criticism that psychiatry is too biologically oriented—and that human behaviors and emotions transcend mere biology. With RDoC, might we exclude people who would benefit from psychiatric care simply because they don't have particular biomarkers or genotypes?

Dr. Cuthbert: That is a very good question. I will give you a four-part answer. First, a good interview will always be an important part of psychiatric care, so we are not disavowing that. Second, we are trying to include neuroscience to help us to a more precise diagnosis, but that doesn't mean we are throwing out ideas about phenomenology. We are not *just* proposing biological tests like MRIs or EEGs; we may also develop very precise psychometric tasks to measure things like reward-related behavior and reports of pleasurable activities. We advocate for looking at the whole range of how people actually function.

TCPR: And the third and fourth points?

Dr. Cuthbert: The third point is that RDoC includes a focus on environmental effects on neurodevelopment. Just because we include biology in a diagnosis doesn't mean that many of the etiological factors in a person's psychosocial milieu or environment are ignored. And finally, we are trying to move toward primary prevention: developing measures that can pick up incipient problems before they become manifest. You can't do that unless you have these measures that are presymptomatic. We don't know if down the line, there may be some subtypes of patients for whom no predictive biosignature exists. But our view is, "let's do the research to find out."

TCPR: Can you explain the RDoC matrix?

Dr. Cuthbert: It is a four-dimensional matrix—but we visually present it as a 2 x 2. The first dimension is divided into five major domains of functioning. They include:

1. Aversive properties—those systems that respond to aversive/negative situations. We call it "negative valence."
2. Appetitive properties, or a "positive valence"—working toward rewards, habits, responding to reward, and so forth.

Continued on page 11

How to Read a Research Article
Continued from page 5

non-randomized treatment effects, for instance, whether cannabis smokers who take antidepressants may be more likely to develop schizophrenia than those who do not.

Some cohort studies obtain information from registries of patient data—for example, all members of an HMO; all VA patients; or all individuals born in Denmark between 1980 and 1989. These are helpful because they deal with "real-world" patients, although they are not randomized. Electronic medical records (EMRs) make registry studies much more feasible.

- **Randomized controlled trials:** Subjects are carefully selected and then randomized to treatment or placebo groups. RCTs evaluate the efficacy of treatment in the short-term, but they are costly.

  Some RCTs are "open-label," meaning that subjects and researchers know what's being given, while "double-blind" trials mean no one knows who's receiving treatment and who's receiving a placebo. Blinding can be difficult to accomplish, for example, in studies where one treatment arm receives psychotherapy. All RCTs are subject to selection bias.

- **Systematic review:** This is a review of research designed to answer a specific clinical question, for example, "what is the most effective approach for treating psychotic depression?" The Cochrane Collaboration (www.cochrane.org) produces many high-quality systematic reviews. Systematic reviews impose strict inclusion criteria on the studies they analyze, but publication bias poses a major problem.

- **Meta-analysis:** Researchers use statistical methods to produce a weighted average of treatment effect sizes derived from multiple studies. The weight accorded to each study depends on sample size and quality. Some systematic reviews incorporate meta-analytic methods. Again, publication bias can strongly influence results.

Step 4: Identify Biases

Bias refers to anything that systematically and unexpectedly influences research results. It affects all research, not just the studies carried out by the pharmaceutical industry. There are many classifications of bias, and study designs differ in their susceptibility to each.

Understanding how bias affects internal validity—meaning your confidence that a study accurately identifies cause-and-effect relationships—is indispensable to critically appraising research. (See "Bias in Research" in this issue for more about bias.)

Step 5: Think About Random Error

Pure chance can throw off study results and render them invalid. In general, effect size and study power determine how likely the results may have arisen purely from chance. Appendectomy for acute appendicitis, for instance, has such a large effect size that its benefit is almost certainly not due to chance.

A larger number of subjects—otherwise known as more power—reduces the role of chance in clinical trials, but

Continued on page 10

Worksheet: Finding Associations in the Data

Relative Risk

This is the ratio of the rate (or probability) of an event in an "exposed" group, to the rate of the same event in an "unexposed" group, typically used in cohort studies.

For example: Do depressed residents make more medication errors? One hundred psychiatry residents were followed for one month to determine whether a diagnosis of depression ("exposure") led to more medication errors ("event rate").

Of the 100 residents, 25 met criteria for depression and 75 did not. Six of the depressed residents made an error (event rate = 6/25 = 0.24), while 10 of the non-depressed residents made an error (event rate = 10/75 = 0.13).

$$RR = \frac{\text{event rate in people exposed}}{\text{event rate in people NOT exposed}} = \frac{6/25}{10/75} = \frac{0.24}{0.13} = 1.8$$

Thus, depressed residents had an 80% (or 1.8 times) higher risk of making an error than non-depressed residents (Fahrenkopf AM et al, *BMJ* 2008;336:488).

Odds Ratio

This is the ratio of the odds that an outcome will occur given a particular exposure, to the odds of the outcome in the absence of the exposure. Most commonly used in case-control studies.

For example: Is borderline personality disorder (BPD) predictive of recurrent suicidality in adolescents evaluated for suicidal behavior? In a case-control study, 77 adolescents (cases) with recurrent suicidal behavior within a six month period were compared with 186 adolescents (controls) without recurrent suicidality. A total of 205 had a diagnosis of BPD ("exposure"), 70 of whom presented with recurrent suicidality ("outcome") while 135 did not. Of the remaining 58, only 7 had recurrent suicidality.

$$OR = \frac{\text{odds of outcome in people exposed}}{\text{odds of outcome in people NOT exposed}} = \frac{70/135}{7/51} = \frac{0.518}{0.137} = 3.8$$

Thus, the odds of recurrent suicidal behavior in the six months after an initial onset of suicidal behavior are 3.8 times greater for adolescents with a diagnosis of BPD than those without (Greenfield B et al, *J Can Assoc Child Adol Psychiatry* 2008;17(4):197–201).

If the "event" is rare, then the odds ratio is a good approximation of relative risk, but if the event is common, the odds ratio will overstate the risk (see Sedgwick P, *BMJ* 2014;348:g1407)

Statistical Significance of RR and OR

Relative risks and odds ratios are usually reported with a "confidence interval" (CI), which shows the range of ratios in which the *actual* ratio is likely to lie, with some high degree of certainty (usually 95%). In the Greenfield et al study mentioned previously, for instance, the OR was reported as: 3.8, 95% CI: 1.6–8.7. Because the CI does not cross 1.0 (which would indicate *no* increased odds of the outcome), this elevation in odds is statistically significant.

Number Needed to Treat (NNT)

This is the number of patients who need to be treated to produce one good outcome or prevent one additional bad outcome.

For example: Does a two-day treatment of intranasal ketamine reduce next-day MADRS scores to a greater extent than placebo? Eighteen people with depression received intranasal ketamine. As a control, the same 18 subjects were given placebo on a different date (ie, a "crossover" design). At 24 hours after ketamine, eight (44%) of the subjects had a response (a >50% reduction in MADRS score); after placebo, only one (6%) had a response.

$$NNT = \frac{1}{\text{absolute risk reduction (ARR)}}$$

$$ARR = \text{difference in response rates} = 0.44 - 0.06 = 0.38$$

$$NNT = \frac{1}{0.38} = 3 \quad \text{(NNTs are always given as whole numbers, and are rounded up)}$$

Continued on page 9

Worksheet: Finding Associations in the Data
Continued from page 8

This means that three people need to be treated in order to see one response due to the drug (Lapidus KAB et al, *Biol Psychiatry* 2014;online ahead of print).

In clinical trials, often only the differences in outcome scores are reported (in this study, a difference in MADRS score of 7.6 points, which was statistically significant). However, the NNT helps us to identify how many patients would need to be treated with ketamine for one patient to respond.

Low NNTs are better (an ideal drug has NNT=1). And even though the FDA does not use NNT to approve drugs, all FDA-approved medications for bipolar disorder, for example, have NNT<10 (Ketter TA et al, *Acta Psychiatr Scand* 2011;123(3):175–189).

Keep in mind that an NNT=10 means that one patient will respond while nine people will not improve as a result of the drug. However, these nine may benefit from a placebo effect or simply through the natural evolution of symptoms. Using the drug is still acceptable if the risk and/or cost of the intervention is relatively small (see Citrome L, *Curr Psychiatry* 2008;6(3):66-71).

Effect Size

This is a measure of the size of the difference between two groups. Most commonly found in meta-analyses.

For example: How effective are antidepressants for chronic low back pain (CLBP)? Nine RCTs were evaluated, with a total of 504 patients. Measurements of pain severity were standardized across all studies, and rates of improvement were measured for antidepressants and for placebo.

Effect size can be calculated in several ways. The most common is known as Cohen's d (or simply d), also called a standardized mean difference (SMD).

The numerator is the difference in outcome scores (eg, a measure of pain or depressive symptoms), while the denominator is simply the standard deviation (SD) of these scores. If scores vary greatly, SD will be high and SMD will be low.

$$\text{SMD} = \frac{(\text{mean outcome on drug}) - (\text{mean outcome on placebo})}{\text{standard deviation}}$$

When these numbers were calculated for the nine antidepressant trials in CLBP, and then averaged, the result was 0.41 (Salerno SN et al, *Arch Int Med* 2002;162(1):19–24).

A higher SMD (or d) means a more effective intervention. When Cohen first developed criteria in 1988, he arbitrarily defined small, medium, and large effect sizes as 0.20, 0.50, and 0.80, respectively.

Incidentally, the effect size for antidepressants in depression has been estimated to be approximately 0.31 (Kirsch I et al, *PLoS Med* 2008;5(2):e45). Of course, effect sizes vary according to the severity of illness in subjects, the quality of the studies included, and many other factors.

Overview of the FDA Drug Approval Process
Continued from page 6

chiatrist's observations (Zimmerman M et al, *J Clin Psychiatry* 2010;71(4):484–490). The psychiatrists in this particular study used open-ended questioning to assess for potential adverse events—quite similar to the methods used in psychopharmacology clinical trials. Even though serious adverse events are probably reported a bit more accurately in trials, as participants are likely more motivated to report them, the rates of adverse events on drug labels should be considered as low estimates.

Recommendations for Change

To ensure that the FDA approves only those drugs with a high likelihood of benefit and low potential for harm, the following changes may be beneficial:

- Negative trials should be more strongly considered when the FDA assesses drug efficacy. The current system requires two positive studies from a pool of several other trials that may be negative.
- Measures of quality of life or daily functioning should be taken more seriously in the drug review process. If a drug does not improve one's overall quality of life, then the drug may not be worthy of approval.
- Self-reports of patient symptoms should be considered. Though clinician-rated measures are useful, the information provided directly by subjects in a clinical trial may be more relevant to a patient's experience with a drug.
- The clinical significance of drug benefits should be considered. A drug that offers a statistically significant, but very small, benefit may not be worthy of approval, particularly if its adverse event profile is problematic.
- Adverse events should be assessed systematically using structured checklists. This would help to provide a more accurate picture of drug side effects.

For more on the FDA's approval process for psychiatric drugs, see Spielmans GI and Kirsch I, Annu Rev Clin Psychol *2014;10:741–766.*

TCPR'S VERDICT: Federal law as enforced by the FDA does not set a particularly high bar for psychiatric medicines to win approval. At least a minimal degree of efficacy compared to placebo and evidence of safety are required. Improvements to the process may result in better and safer medications but fewer new drugs may meet more stringent standards.

Research Updates IN PSYCHIATRY

Section Editor, Glen Spielmans, PhD

Glen Spielmans, PhD, has disclosed that he has no relevant financial or other interests in any commercial companies pertaining to this educational activity.

PREGNANCY

New Study Supports Cardiac Safety of SSRIs in Pregnancy

Concerns about the safety of psychiatric medications during pregnancy are common among psychiatrists and patients alike. In many cases, one must weigh the risks of a medication to mother or child against the risks inherent in untreated mental illness. Recent research, however, lends support to the growing data about the safety of antidepressant medications in pregnancy.

In a study conducted between 2000 and 2007, researchers examined medical records of close to a million pregnant women in 46 U.S. states and evaluated cardiac outcomes of their liveborn infants. A total of 64,389 women in the study sample used antidepressants during the first trimester. From these women, 580 infants (90.1 per 10,000) were born with a cardiac defect, as compared to 6,403 (72.3 per 10,000) born to women who had not taken antidepressants. The relative risk of any cardiac defects with the use of SSRIs in the first trimester was 1.25 (95% CI, 1.13 to 1.38).

While this suggests an elevated risk, the researchers pointed out that antidepressants are prescribed for other conditions (pain disorders, fatigue, smoking, etc). When they adjusted the analysis for women with a diagnosis of depression and for other confounding variables such as known or suspected risk factors for heart defects, the relative risk was reduced to 1.06 (95% CI, 0.93 to 1.22)—a non-significant elevation.

Researchers did not find a dose-response relationship either in terms of when the medication was started (looking back as far as three months before the first missed menstrual period) or in terms of the highest dose dispensed. Interestingly, they found no significant association between two anticipated risks, namely, paroxetine and right ventricular outflow tract obstruction (RR 1.07, 95% CI, 0.59 to 1.93) or sertraline and ventricular septal defects (RR 1.04, 95% CI, 0.76 to 1.41) (Huybrechts KF et al, *NEJM* 2014;370:2397–2407).

TCPR's Take: Past epidemiologic studies have shown an association between first-trimester antidepressant use and congenital cardiac defects, but this study found that the risk of cardiac malformations was no higher for depressed women who took SSRIs (or other antidepressants) than for those who did not, after controlling for other potential risk factors. Limitations of the study include the use of prescription records, which do not indicate whether women actually took the medication; and the exclusion of miscarriages, stillbirths, or termination of pregnancy, all of which may be a consequence of fetal cardiac malformation.

How to Read a Research Article

Continued from page 7

introduces more heterogeneity into the population.

Researchers use statistical methods to determine the likelihood their results arose from chance. If that probability falls below an arbitrary threshold, such as $p<.05$ in most RCTs, the results are said to be statistically significant. Statistical methods go beyond the scope of this article, but *Clinical Epidemiology: The Essentials* provides an excellent introduction.

Step 6: Appraise the Study

Just because results are touted as statistically significant doesn't mean they'll help you help your patients. Validity refers to how legitimate the results of the study actually are, and can be evaluated by the FRISBEE mnemonic, created by Duke University's residency program. The last E, "equivalence," may be the most important, as it pertains to *external* validity, or how well study results will generalize to patients in your practice (Xiong GL & Adams J, *Curr Psychiatry* 2007;6(12):96).

Although FRISBEE was developed for appraising RCTs, similar concepts can be applied to other study designs. Appraisal worksheets for treatment and diagnostic studies are available on Duke University's website (bit.ly/1rMWqGc).

Throwing FRISBEEs at Research

Follow-up: Did the study include high drop-out rates or poor follow-up? If so, why?

Randomization: Were subjects randomly allocated to treatment and placebo groups?

Intent-to-treat analysis: How well did the study account for drop-outs? Did researchers analyze data from all subjects who entered the trial?

Similar baseline: Were treatment and placebo groups similar at the start of the trial? (Many research papers include a table comparing the two groups on a variety of measures. They should be comparable.)

Blinding: Were subjects, researchers and other health care personnel blind to treatment? Keep in mind that if the same rater assessed for both side effects and clinical benefit, the rater may be "unblinded" due to knowledge of the patient's side effects.

Equal treatment: Were the groups treated equally apart from the intervention being studied?

Equivalence to your patient: Are subjects in the study similar to your patient? Subjects in clinical trials usually have few medical or psychiatric comorbidities; how do they compare with the patients in your practice?

CME Post-Test

This CME post-test is intended for participants only seeking AMA PRA Category 1 Credit™. For those seeking ABPN self-assessment (MOC) credit, a 25 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 August 31, 2015. 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 two (2) *AMA PRA Category 1 Credit*™ or 2 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. What term best defines the bias that occurs when the study and control groups are chosen so that they differ in ways that might affect the outcome of the study (Learning Objective #1)?
   [ ] a) Definition bias [ ] b) Selection bias [ ] c) Recall bias [ ] d) Performance bias

2. Which of the following best describes outcome reporting bias (LO #1)?
   [ ] a) When non-peer-reviewed sources like dissertations and posters give different results from those found in the published literature
   [ ] b) When researchers don't properly deal with the data from participants who dropped out before a study was complete
   [ ] c) When there are selective differences between the care that is given to different groups
   [ ] d) When researchers report only results that are significant, ignoring those that were insignificant or unfavorable

3. Which of the following is a review of research designed to answer a specific clinical question, for example, "what is the most effective approach for treating psychotic depression?" (LO #2)?
   [ ] a) Systematic review [ ] b) Case study [ ] c) Randomized controlled trial [ ] d) Cohort study

4. In general, which of the following determine how likely the results of a study may have arisen from chance (LO #2)?
   [ ] a) Effect size and interviewer bias [ ] b) Effect size and study power
   [ ] c) Odds ratio and validity [ ] d) Odds ratio and study power

5. How many positive phase III studies that are "adequate and well-controlled," as defined by the FDA, are needed for a drug to be approved (LO #3)?
   [ ] a) One [ ] b) Two [ ] c) Three [ ] d) Four

6. Which of the following best defines relative risk (LO #4)?
   [ ] a) A measure of the size of the difference between two groups
   [ ] b) How many patients who need to be treated to prevent one additional bad outcome
   [ ] c) The ratio of the rate (or probability) of an event in an "exposed" group, to the rate of the same event in an "unexposed" group
   [ ] d) The ratio of the odds that an outcome will occur given a particular exposure, to the odds of the outcome in the absence of the exposure

7. Which of the following statements is more accurate (LO #5)?
   [ ] a) Historically, the system for awarding federal grant money has been based on *DSM* diagnostic categories
   [ ] b) Historically, the system for awarding federal grant money has been based on mechanisms that cut across disorders

8. In the Huybrechts et al study of cardiac risks of antidepressants in pregnancy, what was the unadjusted relative risk of any cardiac defects with the use of SSRIs in the first trimester (LO #6)?
   [ ] a) 1.06 [ ] b) 1.25 [ ] c) 1.75 [ ] d) 2.25

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

Expert Interview
Continued from page 7

3. Cognitive systems.
4. Systems for social processes.
5. Arousal and regulatory systems, like serotonin systems and the ascending reticular activating system.

[Editor's note: more information on the RDoC matrix can be found at http://1.usa.gov/Uhkl59.]

TCPR: And the second dimension?

Dr. Cuthbert: The second dimension includes what we call "units of analysis." These are different ways that you can study the functional domains of the first dimension. They include genetics; molecular processes and cellular processes; measures of circuits; physiological measures, like heart rate, skin conductance, or serum cortisol; behavioral measures, like an assessment battery; and

Continued on page 12

THE CARLAT REPORT: PSYCHIATRY

P.O. Box 626
Newburyport, MA 01950

PRSRT STD
US Postage
PAID
Nashville, TN
Permit 989

This Month's Focus:
Research in Psychiatry

Next month in *The Carlat Psychiatry Report:* Continuing education and maintenance of certification

Expert Interview
Continued from page 11

self-reports, which are defined to include questionnaires as well as structured diagnostic interviews. We are not trying to reduce everything down to the molecular level, but rather relate measurements at all these different units of analysis to each other to get a comprehensive understanding of how a particular concept, such as working memory works at multiple levels. This is integrative, not reductionistic.

TCPR: And the third dimension?

Dr. Cuthbert: The third dimension is neurodevelopment. This is important because we now understand most psychiatric disorders to be developmental disorders. And it is vital to understand how the functional domains and the units of analysis described above evolve with the maturing organism.

TCPR: And finally the fourth dimension?

Dr. Cuthbert: The fourth dimension is environmental influences. When something happens in the environment at a certain stage of neurodevelopment it can influence the functioning of all of the above parameters.

TCPR: So clinically speaking, we have to look at a patient as a combination of all four dimensions, to guide us towards better therapeutics.

Dr. Cuthbert: Yes, although clinicians shouldn't have to master the whole matrix. It is simply a framework so that we, the researchers, can help deliver more precise diagnoses and more precise ways to deliver clinical care.

TCPR: Thank you, Dr. Cuthbert.