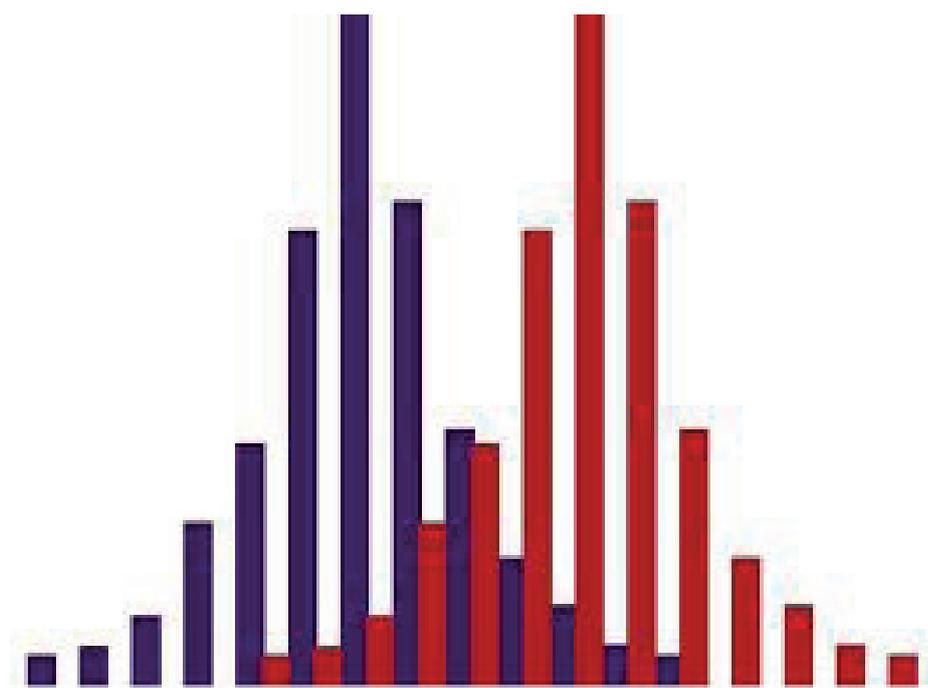


EVIDENCE-BASED PRACTICE IN CHILD AND ADOLESCENT MENTAL HEALTH

John Hamilton & Füsün Çuhadaroğlu-Çetin



John Hamilton MD, MSc

Senior Physician, The
Permanente Medical Group,
Inc., Sacramento, California,
USA

Conflict of interest: none
reported.

Füsün Çuhadaroğlu-Çetin
MD

Professor of Child and
Adolescent Psychiatry,
Hacettepe University Medical
School, Ankara, Turkey

Conflict of interest: none
reported.

Acknowledgments. Professor
Eric Youngstrom, University of
North Carolina, Professor Peter
Szatmari, McMaster University,
and Dr Eric Daleiden,
PracticeWise LLC, were very
helpful reviewers in correcting

This publication is intended for professionals training or practicing in mental health and not for the general public. The opinions expressed are those of the authors and do not necessarily represent the views of the Editor or IACAPAP. This publication seeks to describe the best treatments and practices based on the scientific evidence available at the time of writing as evaluated by the authors and may change as a result of new research. Readers need to apply this knowledge to patients in accordance with the guidelines and laws of their country of practice. Some medications may not be available in some countries and readers should consult the specific drug information since not all dosages and unwanted effects are mentioned. Organizations, publications and websites are cited or linked to illustrate issues or as a source of further information. This does not mean that authors, the Editor or IACAPAP endorse their content or recommendations, which should be critically assessed by the reader. Websites may also change or cease to exist.

©IACAPAP 2012. This is an open-access publication under the [Creative Commons Attribution Non-commercial License](#). Use, distribution and reproduction in any medium are allowed without prior permission provided the original work is properly cited and the use is non-commercial. Send comments about this book or chapter to jmreyATbigpond.net.au

Suggested citation: Hamilton J, Çuhadaroğlu-Çetin F. Evidence-based practice in child and adolescent mental health. In Rey JM (ed), *IACAPAP e-Textbook of Child and Adolescent Mental Health*. Geneva: International Association for Child and Adolescent Psychiatry and Allied Professions 2012.

This chapter shows you how to use data from those studies with less bias and less random error in order to make good clinical decisions to help youth seeking clinical treatment for mental health problems. Specifically, this chapter shows you how to:

- Think about bias and random error as undesirable
- Understand those statistical concepts most relevant to aligning your clinical work with good evidence
- Search databases electronically available at no cost to anyone with an Internet connection
- Use good diagnostic interview techniques to show that the youth does or does not meet criteria for well-defined syndromes
- Find the treatments most likely to be effective for those syndromes.

In brief, this chapter shows clinicians how to align what they do with *what works*: actions that will facilitate a higher proportion of youth to improve their functioning more substantially and more quickly than alternative actions. But to get at the truth about what works, it is first necessary to do battle with two great enemies of the truth: *bias and error*.

BIAS AND RANDOM ERROR

The problem of bias

Bias interferes with validity – the degree to which a study measures what it intends to measure – by creating systematic (non-random) deviations from the underlying truth (Guyatt & Rennie, 2002). Bias is different from *random error*. Random error is diminished by increasing sample size, whereas bias can remain in a sample of any size. Consider the targets (a) through (d) in Figure A.6.1 where the bull's eye in the center marks the true effect. In study (a), there was little or no systematic deviation from the true effect (little or no bias), as well as little scatter (random error). In (b), there was also little bias but much random error, and in (c) much bias but little random error. In (d) there was much bias and random error.

and adding to earlier drafts. The final responsibility for the manuscript, including errors, belongs to the authors. You may contact the corresponding author at john.hamilton@kp.org

Error

All experiments or clinical decisions are subject to either random error, systematic error or both. **Random errors** are due to unknown and unpredictable changes in measuring instruments or in the environmental conditions. **Systematic errors** (or bias), by contrast, are inaccuracies that are consistently in the same direction. They may occur because there is something wrong with the measurement instrument (e.g., a rating scale, diagnostic criteria) or because the instrument is wrongly used by the experimenter (e.g., a clinician who systematically over-estimates – or under-estimates – suicide risk).

Figure A.6.1 Examples of bias and of random error in data samples (Kabai, 2011a).

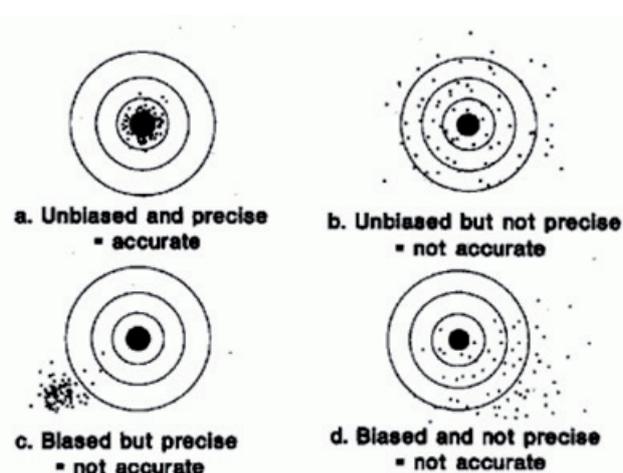
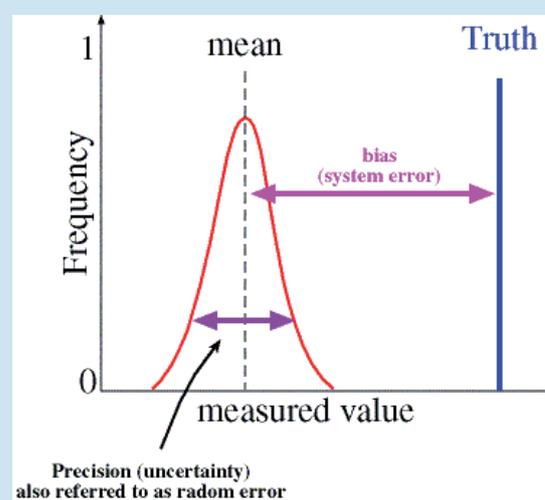


Figure A.6.2 Schematic representation of random error and bias (Kabai, 2011b).



Finding evidence with low bias and low random error

The most useful evidence in guiding clinical decisions is evidence *with minimal bias and minimal random error*. Figure A.6.2 is a schematic representation of random error and bias. Table A.6.1 shows common sources of bias and strategies for reducing each while Table A.6.2 shows common sources of random error and strategies for reducing it.

Table A.6.1 Common sources of bias and strategies to reduce them.

Source of bias	Strategy to reduce the source of bias
• Treatment and control patients differ in prognosis	• Random allocation to treatment or control
• Placebo effects of treatment	• Patients blinded to active treatment or control
• Caregivers add other treatments	• Caregivers blinded to treatment or control
• Assessors of outcome are biased	• Assessors of outcome blinded to treatment or control status of cases they assess
• Loss to follow-up	• Follow-up of all or at least a high proportion of cases entering the study

Bias

Bias is an ever-present risk in research, even in well-designed RCTs. For example, positive outcomes are reported much more commonly in trials funded by the pharmaceutical industry (78%), compared to those without industry sponsorship (48%) or those financed by a competitor (28%) (Kelly et al, 2006).

Table A.6.2 Two common sources of random error and strategies for reducing each.

Sources of Random Error	Strategy to Reduce It
• Imprecise measuring instruments	• Improve precision of instruments
• Small sample size	• Increase sample size

USEFUL EVIDENCE-BASED PRACTICE CONCEPTS

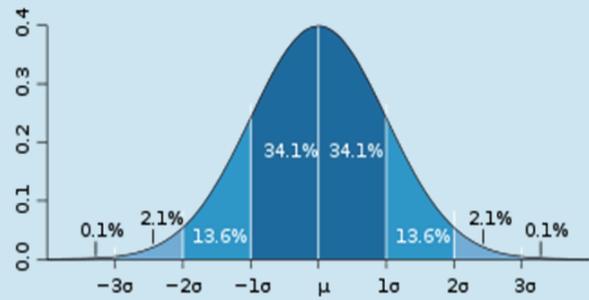
Mean

A few basic statistical concepts are all that is needed to understand many core ideas used in evidence-based approaches. First, recall the basic *normal* or Gaussian curve which describes the distribution of many aspects of nature, and which can be derived mathematically using probability theory. The most common value is the same as the *mean* value, represented here by the Greek letter *mu* (μ).

Standard deviation (SD)

SD is a measure of variability: it reflects how much variation or "scatter" there is from the mean value (average). A small SD relative to the mean indicates a distribution where the data points are closely clumped together. A large SD relative to the mean indicates a distribution where data points are spread out over a large range. SD is sometimes described as the *typical amount that cases differ from each other*. Two out of three cases are expected to fall within plus or minus one SD of the mean.

Figure A.6.3 Plot of a normal (Gaussian) distribution having a mean value represented by the Greek letter μ and a standard deviation, σ , represented by each colored band. Note that slightly over one-third (34.1%) of the population of this distribution lies within one standard deviation above the mean; slightly over one-third (34.1%) also lies within one standard deviation below the mean. Only about 2% of the population lies within 2 or more standard deviations above the mean, and only about 2% lies within 2 or more standard deviations below the mean (Wikipedia).



Effect size

Effect size in a research study measures how big the effect of a treatment is, relative to the differences between cases in general. It is the standardized mean difference between the two groups in a study. Technically, this is measured as the ratio of the difference due to treatment compared to what is called the *pooled* standard deviation. *Pooled* refers to combining figures from both the control and experimental group to create a pooled standard deviation.

$$\text{Effect size} = \frac{[\text{Mean of experimental group}] - [\text{Mean of control group}]}{[\text{Pooled standard deviation}]}$$

Figure A.6.4 shows a typical representation of data from an experimental treatment in child psychiatry with an effect size of 0.73. Note that the two curves overlap considerably.

How to calculate the pooled standard deviation can be found at statistical sites (e.g., <http://www.leeds.ac.uk/educol/documents/00002182.htm>). Effect size places the emphasis on the most important aspect of an intervention – the size of the effect – rather than its statistical significance, which is determined by both effect size and sample size.

Example of effect size

A study (McCracken et al, 2002) sought to examine whether an atypical antipsychotic (risperidone) was more effective than placebo in reducing aggression, tantrums or self-injurious behavior in children suffering from autism aged 5 to 17 years. Scores after 8 weeks on a measure of irritability were 11.3 (SD 7.4) for the treatment group and 21.9 (SD 9.5) for the control group. What is the effect size?

Effect size = $(11.3 - 21.9) / 9.5$, that is 1.1. This is a large, highly significant effect size. Effect sizes above 0.5 are often considered clinically significant, but smaller effects can still have clinical significance if the treatment can easily be applied to large populations (e.g., the use of aspirin to prevent heart attacks, where the effect size is small but when this occurs over large populations, the public health consequences are significant).

Absolute risk reduction (ARR)

Absolute risk reduction is a measure used to compare two different alternatives: *how much does one treatment reduce the risk of a specified "bad" outcome (like a depressed youth staying depressed) compared to either an alternative treatment or placebo?* For example, in the Treatment for Adolescents with Depression Study (TADS), at 12 weeks, 65.2% of adolescents receiving placebo were not rated as improved or very much improved compared with only 39.4% of adolescents receiving fluoxetine. Therefore, the ARR of not improving for youth with major depressive disorder who took fluoxetine compared to placebo was $65.2 - 39.4 = 25.8\%$. In other words, fluoxetine reduced the risk of not improving considerably: 25.8%, or about 1 out of 4.

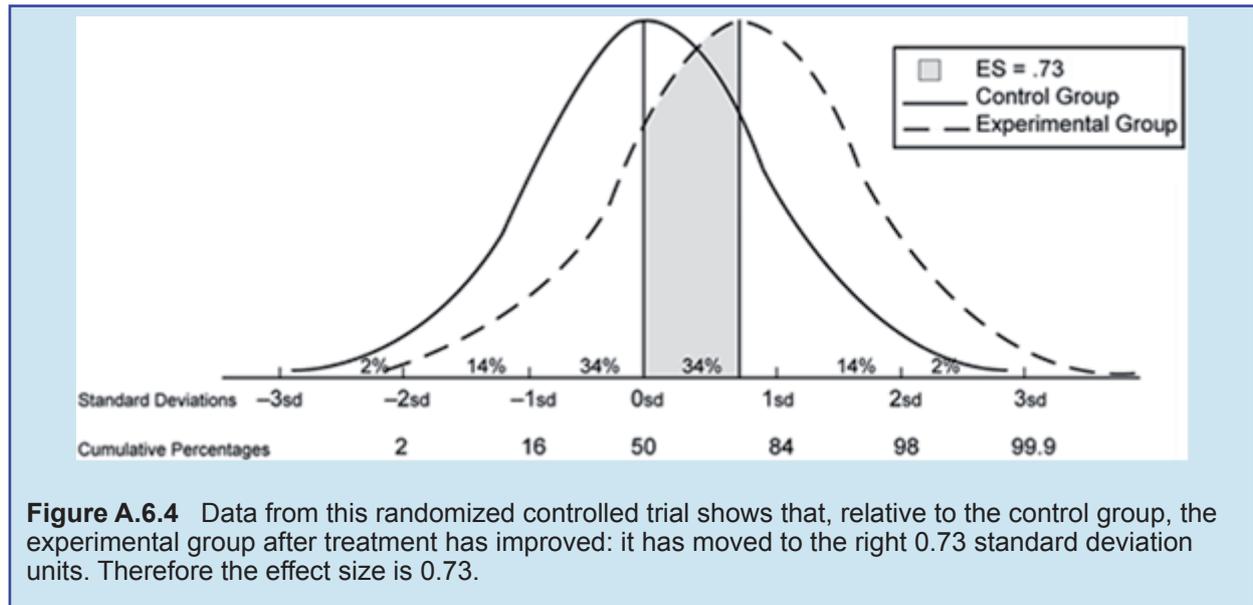


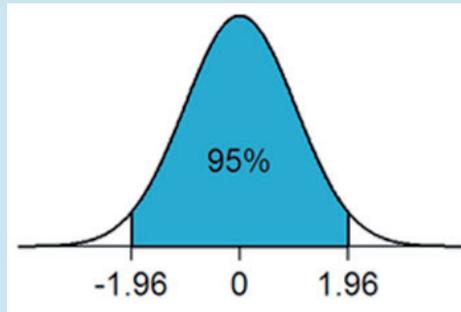
Figure A.6.4 Data from this randomized controlled trial shows that, relative to the control group, the experimental group after treatment has improved: it has moved to the right 0.73 standard deviation units. Therefore the effect size is 0.73.

Absolute risk reduction is helpful for two reasons. First, as discussed below, the ARR can be used to compute the *number needed to treat* (NNT). The formula is simply $NNT = 1/ARR$ with the answer rounded up to the next whole number. Second, the ARR does not exaggerate the treatment effect as does the *relative risk reduction* (RRR) which is simply computed as the risk in the treatment group divided by the risk in the control group. In TADS the RRR when comparing outcome for youth receiving fluoxetine to youth receiving placebo is the risk of depressive symptoms being rated as not significantly improved in the treatment group receiving fluoxetine divided by the risk for youth not being significantly improved in the placebo group, or $39.4 / 65.2 = 0.60$, generally expressed as 60%. Note that this figure is quite different from the absolute risk reduction as calculated in the paragraph above (25.8%). In general, ARR is a more useful statistic than the RRR.

Confidence interval (CI)

Rather than beginning with a definition, a specific example might be more useful to gain a feel for this important concept. Suppose the Minister of Health tells you she is concerned about depressed teenagers and she asks you to determine the *average* (i.e., mean) level of symptoms of depression in 13 year old children in the large city where you live. You accept the challenge and begin the project by arranging for 13-year olds from all over the city to complete an inventory of depressive symptoms. The mean score is 39.8 on the 49 completed questionnaires. You calculate the standard deviation and it is 9.0. The Minister of Health wants you to estimate, based on your data from the 49 returned questionnaires, how close is your mean to the *true* value, also called the *population* value. That is, if it were possible to assess *all* 13-year old children in the city with this depression scale, this would presumably be very close to the *true* value. You can see, therefore, that the *true* or *population* value is a useful concept, but in reality, it is often not practical to test all the children. Simply by chance you may have ended up with a higher proportion of unhappy children or of unusually happy children in your sample than is present in the city in general, even if you managed to avoid being biased in your sampling (like choosing only students from the same school, or only students in a certain class).

Figure A.6.5 An important fact associated with the "1.96" in the formula for calculating 95% confidence intervals: 95% of the area of a normal distribution is within 1.96 standard deviation units of the mean.



You tell the minister the mean score in your sample is 39.8, as a rough estimate. She, however, wants to know more, and asks, "How rough an estimate is that? What is the highest true value? And the lowest?" You answer, "Yes, I can give you a range or interval but it depends on how often you want me to be right. Do you want me to be right like 80% of the time that I give you these estimates? Or 95%? Or even 99%? The more you insist that the answer I give you does contain the true answer for the population, the wider the interval will be." The minister says, "OK. I want your answers on these kinds of estimates to be correct 95% of the time." The minister has just asked for the 95% confidence interval of your calculated mean.

A statistician can now help. She tells us that a good estimate for calculating confidence intervals in normal distributions is as follows, where N is the number of subjects generating the data (49 in this case) and $\text{SQRT } N$ means square root:

$$\text{Confidence interval, upper bound} = \text{Calculated mean} + 1.96 \times \text{SD} / \text{SQRT } N$$

$$\text{Confidence interval, lower bound} = \text{Calculated mean} - 1.96 \times \text{SD} / \text{SQRT } N$$

Therefore the confidence intervals are:

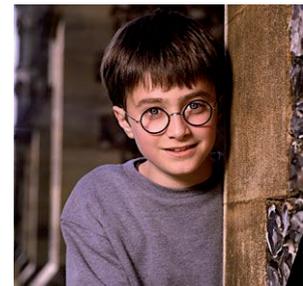
$$CI_{\text{upper}} = 39.8 + 1.96 \times 9.0 / \text{SQRT } 49 = 39.8 + 1.96 \times 9.0 / 7 = 39.8 + 2.52 = 42.32$$

$$CI_{\text{lower}} = 39.8 - 1.96 \times 9.0 / \text{SQRT } 49 = 39.8 - 1.96 \times 9.0 / 7 = 39.8 - 2.52 = 37.28$$

Now you have an answer for the Minister.

In summary, CIs are expressed with three numbers, the first being the mean, our best estimate of the true value. The other two numbers are of course the numbers stating the lower bound of the interval and the upper bound. The more certain we want to be, the wider the confidence interval will be: wide confidence intervals *make sure* the correct number is in there somewhere. *Confidence intervals are important because statistical analysis of studies is based on the central idea that, despite making observations on a limited sample of subjects, the inferred truth to be drawn from the study will apply to the population of all such subjects.* The main purpose of a confidence interval is to indicate the precision or imprecision of the study sample as an estimate of the true population value. Confidence intervals are therefore useful and even necessary whenever an inference is made from the results of one study to the wider world. In the first example, there were only 49 children completing the depression questionnaire, but the Minister hopes to generalize the findings to all 13 year old children in the city. Also note that even if you and the Minister chose your sample wisely – e.g., from very different schools in different

Confidence intervals can make you a data wizard





How high should be the hurdle researchers require for a youth to be considered “improved” when they calculate NNT?

Higher hurdles will raise NNT because fewer youth will get over them.

parts of the city to avoid bias – the problem of random variation in the level of depression symptoms remains. Merely by chance, you may have come across children with notably higher or notably lower depression levels than exist in the entire population of children of that age in the city, even though there was no significant systematic bias in your sampling method.

Number needed to treat

The number needed to treat, often abbreviated NNT, can be used to summarize in a single number how effective a specific treatment is compared to placebo. NNT is defined as *the number of people we must treat in order to prevent one additional bad outcome. A bad outcome is defined by specific criteria.* Therefore low NNTs (like 3 or 4) indicate an effective treatment, because it is common for the treatment to *convert* a patient to a successful outcome who would not have improved without the treatment. Typical successful outcomes are improved mood or improved functioning, often defined with a cutoff point on a scale.

Note, though, that NNT depends very much on the *hurdle* that needs to be cleared in order to be considered “better” or “cured.” *How high or low this hurdle is will impact on the NNT hugely.* Always note, therefore, what hurdle authors have chosen when reading a paper and noting an NNT.

High NNTs (like 25, 30 or even higher) suggest either a treatment that is not very effective or an outcome that is difficult to achieve. For example, if we require that a successful outcome in a depressed adolescent is an outcome without a single depressive symptom, then it will be difficult to achieve this outcome and the corresponding NNT will be higher than if we had chosen a more modest outcome.

Number needed to harm

Number needed to harm is abbreviated NNH and summarizes how frequently a specific side-effect occurs with a specific treatment. NNH is defined as the number of people we must treat in order that a single person is harmed by the treatment who would not have been harmed if he had received only a placebo intervention. For example, with medications, an undesirable effect or side effect occurs that would not have occurred if the patient had received a placebo. Therefore high NNHs indicate a safe treatment: many individuals must receive the treatment before a single individual is harmed by the treatment. Decisions to recommend a specific treatment will therefore require balancing the benefits (NNT) with the risks (NNH).

Mann et al (2004) conducted a meta-analysis to estimate the effectiveness of acamprostate in achieving abstinence among alcoholics. They reported that across studies 36.1% of participants taking acamprostate had achieved abstinence at 6 months compared with 23.4% of those on placebo. What is the NNT?

In this case $NNT = 1 / (0.234 - 0.361) = 7.9$. That is, eight patients (it is customary to round to the next whole number) will need to be treated with acamprostate for one additional patient to abstain from alcohol at 6 months compared to those not receiving treatment.

Number needed to harm (NNH)

NNH equals one divided by the rate in non-exposed minus rate in exposed individuals.

In the Treatment for Adolescents with Depression Study (TADS) Team (2004) study the authors reported that 11.9% of participants treated with fluoxetine alone had reported some harm-related event, compared with 5.4% for placebo. What is the NNH?

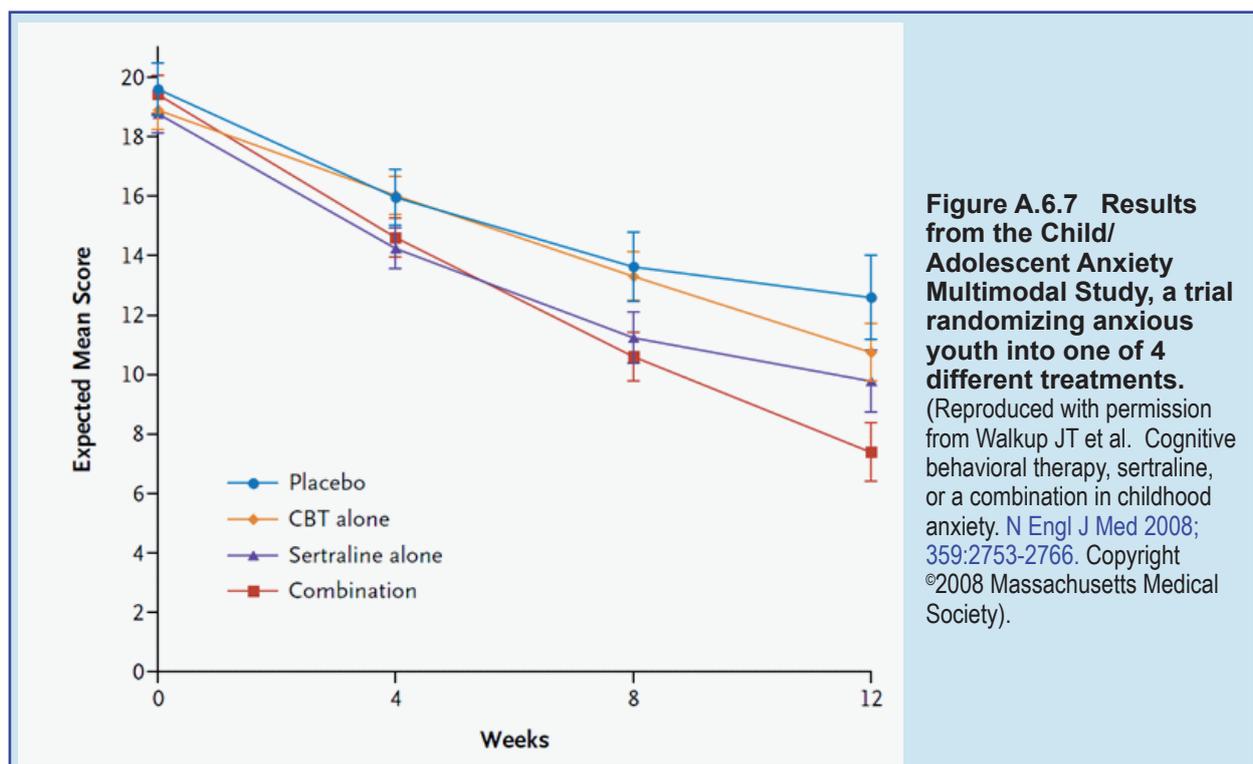
In this case $NNH = 1 / (0.119 - 0.054) = 15.4$. That is, fifteen depressed patients will need to be treated with fluoxetine for one additional patient to suffer a harm-related event attributed to the medication.

UNDERSTANDING A STUDY BY LOOKING AT ITS BASIC STATISTICS

Now consider the results from a study examining the response of children and adolescents with an anxiety disorder using these core statistical concepts. Called the Child/Adolescent Anxiety Multimodal Study (CAMS), this study recruited children and adolescents from multiple sites with common anxiety disorders – generalized anxiety, social phobia, or separation anxiety disorder – and randomized these youth into receiving one of four possible treatment strategies: CBT alone, sertraline alone, combined CBT and sertraline, and placebo.

The graph in Figure A.6.7 shows Pediatric Anxiety Rating Scale (PARS) scores for the 4 groups, each receiving a different treatment. Scores greater than 13 are consistent with an anxiety disorder. Combining sertraline and CBT, based on these data, is the most effective treatment. The *error bars* on the PARS figures show how much *scatter* there is in those numbers. The top of an error bar is most easily understood as the mean plus the standard error of the mean (calculated as the standard deviation / square root of the number of participants), and the bottom of the error bar is the mean minus the standard error of the mean. More complex methods of calculating *standard error* represented by the error bars also exist.

In this case, the clinical magnitude of the impact of treatment on outcome was evaluated by calculating effect sizes for each of the 3 intervention groups compared to the placebo group: how different were each of the 3 intervention groups on the PARS compared to the placebo group? Based on improvements on the scale the authors used to measure anxiety in youth, the PARS, the effect size at 12 weeks was 0.86 for combination therapy, 0.45 for sertraline alone and 0.31 for CBT alone.



The authors also calculated the number needed to treat. Recall that to calculate an NNT, we must choose a “hurdle” for each subject to “clear”. Higher hurdles are more difficult to clear and therefore will generate higher NNTs; lower hurdles will generate lower NNTs. The authors chose as a hurdle that the youth must be rated as either improved or very much improved on the PARS by the end of treatment at 12 weeks. Using this criterion, the authors calculated the NNT for sertraline alone as 3.2 and for cognitive behavioral therapy alone as 2.8, and for the combination of both treatments the NNT was 1.7. This low NNT is encouraging because it demonstrates that effective treatments exist for many youth with anxiety disorders.

The authors also used statistics to convey to readers how certain they are that their results were not a lucky fluke caused by random events that together resulted in improvements in those youth receiving sertraline and CBT. To assess this possibility, the authors calculated the confidence intervals for each effect size and for each number needed to treat. The confidence intervals for 95% certainty around the calculated effect size of 0.86 were 0.56 to 1.15 for youth receiving the combination of sertraline and CBT compared to youth receiving placebo, which is a good, robust effect in child psychiatry. Effect sizes of 0.8 and higher indicate a large effect, 0.3 and below a small effect according to Cohen's (1988) conventions.

RELIABILITY AND VALIDITY OF DIAGNOSTIC METHODS

THE FOUNDATION OF EVIDENCE-BASED PRACTICE

Aligning the diagnostic methods in your practice with the diagnostic methods used by researchers is important because then youths in your practice who you define as having, for example, major depressive disorder, will tend to have similar constellations of depressive symptoms to those youth whom researchers studied when testing treatments for major depressive disorder. You can then confidently import their results about what helps such youth into your own practice.

So, how reliable are the procedures you use to identify syndromes in youth? Reliable means first that the test when used repeatedly yields consistent results. For example, if a well-defined, structured diagnostic interview produces identical or very similar results when re-administered a short time after the first administration, this suggests the test is reliable (test-retest reliability). Reliable also means it yields the same or very similar results when used by different observers on the same youth (inter-rater reliability)?

And how valid are the diagnostic systems you use to identify syndromes in youth? A valid diagnostic system is one that truly measures what it claims to measure. For example, a valid diagnostic system to assess the severity of depression is one that actually measures the severity of depression. Evidence supporting the validity of a diagnostic system for the severity of depression might include data showing that self-reports of depression severity from adolescents are highly correlated with reports from trained interviewers who interview the same adolescents using that diagnostic system. Diagnostic agreement between structured diagnostic interviews of the type often used in research studies and clinical evaluations vary widely by disorder and is low to moderate for most disorders (Rettew, 2008).

Reliability

Reliability is the consistency achieved by a measuring instrument (e.g., a diagnostic test or questionnaire) across individuals, times and clinicians. Reliability is measured using the kappa (κ) statistic for categorical variables (e.g., diagnosis) and intraclass correlation coefficient (ICC) for continuous variables (e.g., scale scores). Common values for both range from 0 (not better than chance) to 1 (perfect agreement). Reliabilities above 0.5 are acceptable and above 0.7 are excellent.

Validity

Validity is the extent by which a diagnostic system measures what it claims to measure. Assessment of validity is often indirect, relying on the capacity of a measure to agree with other measures of the same construct, or to predict course, outcome, response to treatment etc. Correlation coefficients and factor loadings are both common ways of quantifying at least some types of validity.

TABLE A.6.3 Examples of reliable measurement instruments* (see also Chapter A.5)

Disorder	Instrument
Anxiety disorders including generalized anxiety disorder and separation anxiety disorder	<ul style="list-style-type: none"> • Screen for Child Anxiety Related Emotional Disorders (SCARED) • See also Chapters F.1 and F.2
Depressive disorders including major depressive disorder and dysthymia	<ul style="list-style-type: none"> • Hamilton Rating Scale for Depression (HAMD-21) • Johns Hopkins Depression Checklist for Children (HDCL-C) • See also Chapter E.1
Attention deficit/hyperactivity disorder	<ul style="list-style-type: none"> • Swanson Scale for ADHD (SWAN ADHD) • The SNAP-IV Teacher and Parent Rating Scale • Vanderbilt ADHD Diagnostic Parent Rating Scale • See also Chapter D.1
Obsessive-compulsive disorder	<ul style="list-style-type: none"> • Children's Yale-Brown Obsessive Compulsive Scale (CY-BOCS) • See also Chapter F.3

*These instruments were developed and tested in Europe and the US. Syndromes in countries far-removed from these geographic areas may be substantially different. In addition, remember that the results of a single instrument are only one aspect of coming to a conclusion, or even an initial hypothesis, about diagnosis; many other sources of data such as the interview with the youth, his parents, as well as historical information and medical and psychiatric records are also useful.

In summary, a reliable and valid diagnostic system is an essential foundation to clinical practice. Unfortunately, unstructured clinical interviews have a poor track record for reliability and validity in diagnosing youth with psychiatric disorders. Diagnostic interviewers proceeding with no structure tend to jump to conclusions, often fail to assess important co-morbidities, and produce diagnostic results that are not reliable. This is a major problem but one that can fortunately be solved.

Reliable and valid diagnostic instruments

Creating a reliable and valid diagnostic system is a difficult challenge. The most useful aids to create such a system are specific diagnostic instruments for specific classes of disorders. Diagnostic aids are often available free on the Internet, are highly useful in defining disorders and their severity and, importantly, serve as a baseline which may allow you to measure change in symptoms during treatment. Some examples of what is available for some of the most common disorders can be found in Table A.6.1.

An example of using diagnostic tools to assess an adolescent

Correctly "translating" symptoms described by young persons, their parents and teachers into a psychiatric syndrome or, at least, part of a syndrome (*subsyndromal condition*) is important. This step takes the unique complaints of an individual child and family, translates them into a recognized disorder, and tries to avoid the many possible sources of error in the process. Consider, for example, Sam, a 14 year old boy whose parents complain he is constantly irritable and having trouble sleeping. He has falling grades and withdraws socially. His constellation

of symptoms may fit into more than one syndrome. Adolescents may become irritable when feeling anxious or oppositional but irritability also suggests the possibility of a mood disorder; his social withdrawal similarly has many possible origins. Fortunately, widely available and inexpensive broad-based assessment tools are useful in getting the diagnosis right (see Table A.6.1).

Whereas these instruments help us zoom in and define more precisely symptom levels in the domains suggested by the clinical history, we also need a way to zoom out and screen Sam for other, less obvious syndromes such as being a victim of trauma or abuse, or using substances, or having a behavioral disorder he has neglected to tell us about.

Screening questions from a well-recognized comprehensive assessment of psychopathology can be helpful here. The Kiddie-SADS (K-SADS) is a widely used semi-structured diagnostic interview procedure generously placed in the public domain for non-profit use and ethical research (Kaufman et al., 1997). Note that the K-SADS is not a structured interview – one in which the interviewer repeats written questions verbatim – but rather a semi-structured interview in which the interviewer has the choice of several possible questions trying to probe the same domain. Interviewers do *not* have to ask every question, but rather stop when they have sufficient information. Here are some K-SADS screening questions designed to probe for possible alcohol abuse:

- *How old were you when you had your first drink?*
- *What's your favorite thing to drink?*
- *Do you have a group of friends you usually drink with, or do you usually drink alone?*
- *Where do you usually drink? At home? Parties? A friend's house? The street? Bars?*
- *Are there special times when you are more likely to drink than others? School dances or other parties?*
- *How old were you when you started to drink regularly, say two drinks or more per week?*
- *In the past six months has there been at least one week in which you had at least two drinks?*

What is a good way to screen for trauma?

Here is a sampler of how the semi-structured interview K-SADS asks about trauma, but only after a “warm-up” period to gather basic information and put the youth at ease. Note that proper use of the K-SADS requires training and the list below includes only some of its questions about trauma (Kaufman et al, 1997)

1. Traumatic Events

Probe: I am going to ask you about a number of bad things that often happen to children your age and I want you to tell me if any of these things have ever happened to you. Be sure to tell me if any of these things have ever happened, even if they only happened one time.

- a. Car Accident: Have you ever been in a bad car accident? What happened? Were you hurt? Was anyone else in the car hurt?
- b. Other Accident: Have you ever been in any other type of bad accidents? What about a biking accident? Other accidents? What happened? Were you hurt?
- c. Fire: Were you ever in a serious fire? Did your house or school ever catch on fire? Did you ever start a fire that got out of control? What happened? Did anyone get severe physical injuries? Was there a lot of damage?

The K-SADS continues from this point with questions about witnessing a disaster like a bad storm or a violent crime, domestic violence, or physical or sexual abuse.

The authors of the K-SADS suggest a clear threshold for further assessment: either the youth or parent reports the youth has 2 or more drinks per week for at least 4 weeks. The explicit thresholds of the K-SADS for "counting" a symptom as present or absent are useful, as are its suggested questions. They are clear and direct, yet offer the interviewer multiple ways of getting the information needed. The specificity of the K-SADS questions, their clear thresholds, and the presence of screening questions are useful for clinicians trying to improve the reliability and validity of their clinical interviews. Using the entire interview requires specific training and may be unrealistic in many clinical settings, but its probes, suggested questions and thresholds are useful.

Once a good – reliable and valid – diagnostic evaluation of the adolescent has been completed, the task of choosing a treatment is next.

CHOOSING A TREATMENT

The concept of a pyramid of evidence

A *pyramid of evidence* is a useful way of summarizing what is the *best* evidence – that is, closest to a precise statement of the truth. Sources of evidence with less bias and less random error are placed higher up, towards the top of the pyramid as the most credible. Pyramids are a good shape to represent the transition from the hundreds or even thousands of sources of “lower level evidence” such as case studies, non-controlled trials and even expert opinion rightly placed close to the base of the pyramid – to the relatively few systematic reviews of randomized controlled trials and meta-analyses that synthesize all the most credible evidence – rightly placed at the top of the pyramid. Since there are many case reports and other studies likely to contain bias and random error and relatively few that thoughtfully synthesize all credible evidence, the broad base and narrow peak shape of a pyramid is especially appropriate (Figure A.6.8).

Pyramids are a good shape to represent different kinds of evidence, since they are broad at the base and narrow at the top. Our goal is to get to the top, where the view is good, minimizing distorting biases and random error.

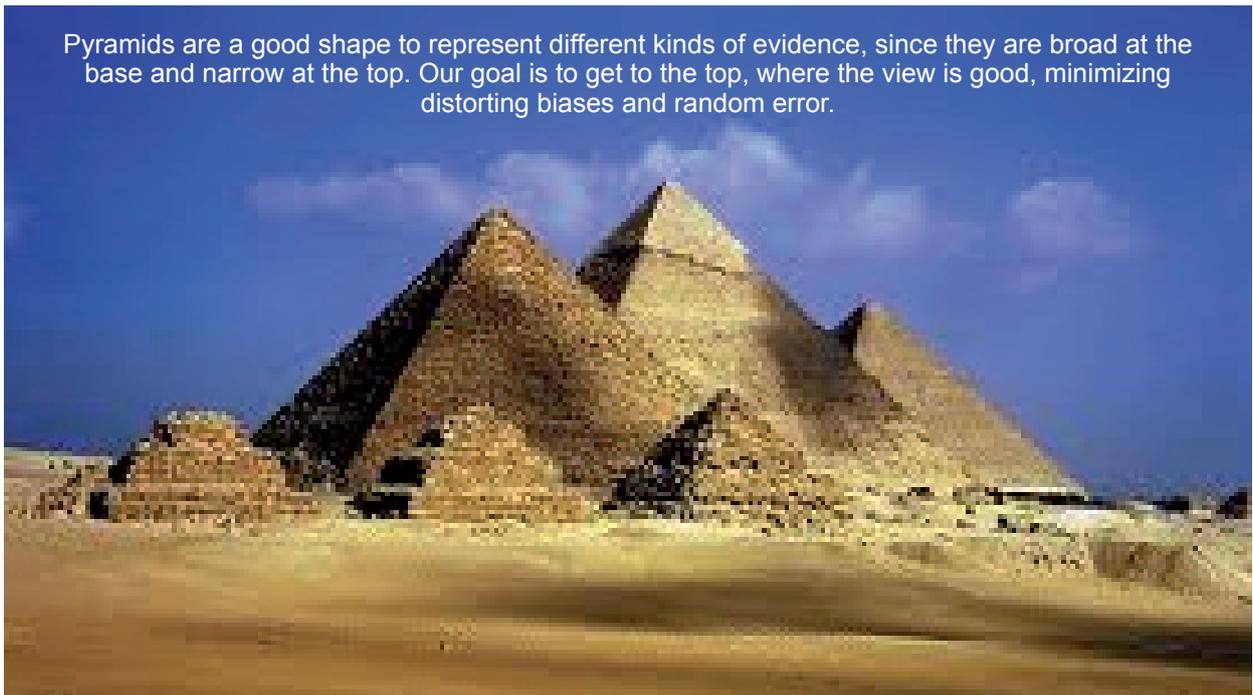
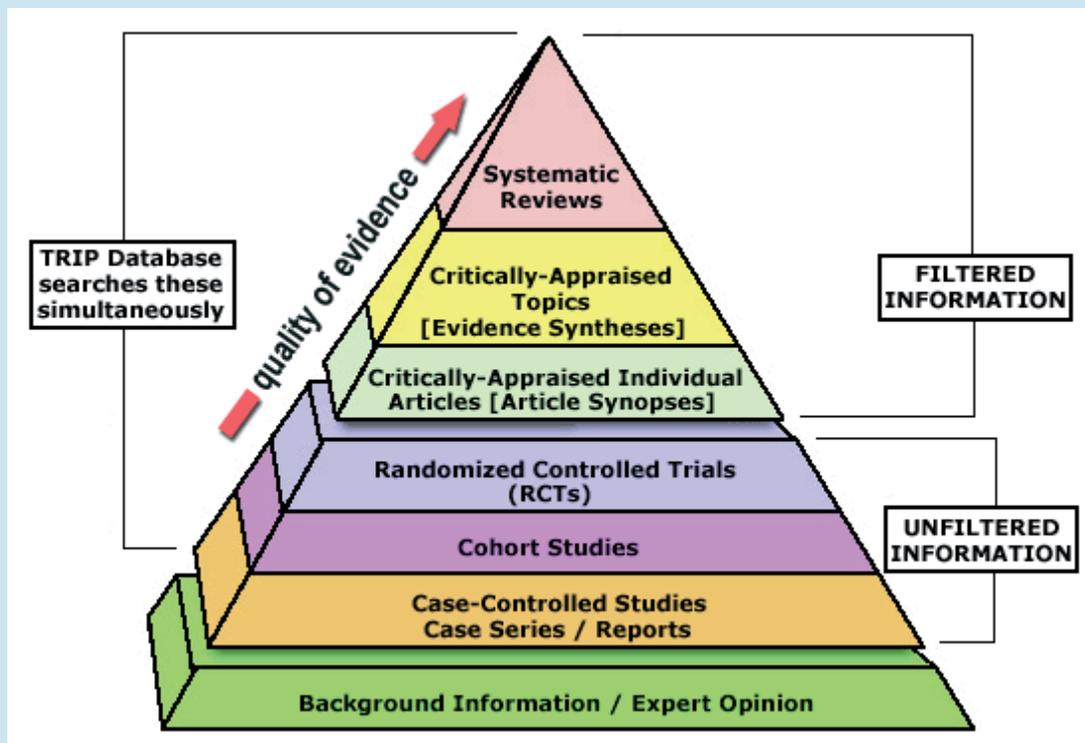


Figure A.6.8 A pyramid of evidence where higher-up on the pyramid means more freedom from bias and random error.



Source: University of California, San Diego, library system and its TRIP database which uses 'filters' (software programs) to select published evidence most free from bias and random error.

Background information

At the bottom, background information on the subject, such as the neurophysiology of depression or autism, for example, and the opinion of experts with experience provide only weak evidence. Background information only sets the stage for what might happen: many possible dramas can occur using the same stage set. If, for example, imaging studies show changes in blood flow in frontal cortical areas in depression, this does not necessarily help us choose the most effective treatment for depression.

Expert opinion

Similarly, *expert opinion* can be dead wrong. The expert may be employed by a drug company and feel pressure to favor a specific drug, or may feel shame about negative outcomes, or pride about positive outcomes and subconsciously slant the results in a positive direction, or simply be loyal to a certain treatment because it is familiar. The main point is that a single individual's opinion unsupported by systematic data gathered using recognized methodology represents weak evidence. Familiar humor states this conclusion as "In God we trust. Everyone else, bring data."

Cases series

Case series and case reports are also near the bottom of the evidence pyramid. A single practitioner assessing a single case, carrying out the treatment, and then assessing the results is highly vulnerable to biases. Case series produce

weak data because there is no comparison group, although their data may be useful to investigators generating hypotheses for future studies.

Case-control studies

Case-control studies can be useful in showing an association between exposure to a risk factor and an outcome. In case-control studies, the outcome is known, such as adolescent suicide. The investigator then creates a control group of individuals without the outcome but with similar characteristics as the individuals with the known outcome and determines the rate of exposure to a risk factor in the controls compared to those with the known outcome. Using this methodology, the accessibility and availability of a gun in the home has been shown to be associated with increased risk of adolescent suicide (Brent et al, 1993). This is useful evidence when advising parents of a potentially suicidal adolescent.

Cohort studies

Cohort studies can be useful in demonstrating prospectively an association between exposure to a specific risk factor and outcome. Here sampling is by exposure, not by outcome. Children exposed to certain kinds of abuse, for example, are known to be at higher risk of developing depression. Some cohort studies use very large databases, even including all children in an entire country, and therefore generate great statistical power (the ability to detect even a small effect).

Randomized controlled trials

Individual randomized controlled trials (RCTs) are the principal building blocks of higher levels of the pyramid. With their unique advantages of a control group chosen at random and often deliberate blinding of subjects, treating clinicians, and assessors as to assignment to the active intervention or control group, RCTs have unique advantages. Higher levels of the pyramid are designed to wring out limitations on RCT data, such as eliminating data from poorly done RCTs, while combining data from multiple well-done RCTs to generate a larger number of subjects studied in order to minimize random error. Beware of ignoring unpublished RCTs.

Critical appraisal

The critical appraisal of individual articles eliminates data from trials with significant flaws. Perhaps such a flawed trial did not adequately blind raters, for example, or lost a large proportion of the sample to follow-up. If so, then the results may contain significant bias and the studies are excluded. The results of critically-appraised individual articles on the same topic, now known to be free of obvious sources of bias, can then be pooled to increase sample size in order to reduce random error. Consider using specific convenient checklists of criteria when appraising a study, such as the Consolidated Standards for Reporting Trials (CONSORT) checklist, if and when you have time to do so.

The flow diagram in Figure A.6.9 (page 20) shows how some critical steps in running an RCT are worth keeping in mind as you read a paper. At times, for example, an author may describe a study as a “randomized trial” even though it is clear in the methods section that assignment was not truly random. There are many potential sources of bias in doing an RCT, but most practitioners simply don't have the will, the time or the statistical and research expertise to read through a published RCT. Critical appraisal implies having a critical look at each step of

CONSORT stands for Consolidated Standards of Reporting Trials.

[Click here](#) to access the CONSORT checklist of information to include when reporting a randomized trial.

a study to ensure that no bias is passed through to the final results. Attending a course in EBM and participation in a journal club for people interested can help to improve critical appraisal skills.

Evidence synthesis and systematic reviews

Evidence synthesis pools data from individual RCTs that have been critically-appraised for possible sources of bias. Pooling data tends to decrease random error to produce a more precise result.

Finally, systematic reviews are an especially organized method of evidence synthesis relying on agreed-upon standards of synthesis. With this methodology, perhaps best exemplified by the Cochrane Library's electronic library of Systematic Reviews, the method used to combine data and draw conclusions is well-established prior to beginning the review. This methodology is also transparent, in that the reader can review how the results were obtained. A meta-analysis is a type of systematic review that considers data from several studies to arrive at a single estimate of the effect size or NNT.

USING THE INTERNET AND THE PYRAMID TO FIND EFFECTIVE TREATMENTS

Answering questions relevant to your practice using the Internet

Try to remain curious when doing clinical work: ask questions relevant to caring for the patients you see that day. This means asking questions relevant to the specific case and trying to answer them by using the Internet to look through the published literature, focusing on evidence free from bias and random error.

The first step is to ask a question. To achieve this, begin the question with "What is the evidence that...?" This ensures that the question can be answered: you will either find there is no evidence, in which case the answer is "none," or you will find at least some evidence.

The goal of a search is a practical result, good enough to inform care without claiming to be definitive, within the time and resources available. Technology allows searching databases for valid and relevant results. Software has made sensitive and specific searching easy on [PubMed](#), the Internet site of the National Library of Medicine in the United States. PubMed is maintained by the United States government at no cost to users and no subscription is necessary. [PsycINFO](#), the Internet site for psychological studies, is maintained by the American Psychological Association and offers *abstracts* at no cost to users. [CENTRAL](#), the Cochrane Library's database of controlled trials, is based in the United Kingdom. [CENTRAL](#) is the abbreviation used for The Cochrane Central Register of Controlled Trials; it contains summaries of nearly 700,000 controlled trials both from PubMed and [EMBASE](#), a database especially strong in European journals. EMBASE is accessible only to subscribers and is maintained by Elsevier, a corporation based in Amsterdam. PubMed is a good place to start for questions related to medication and includes studies that were not RCTs. PsycINFO is good for questions related to psychosocial issues. CENTRAL is good to search exclusively for results from RCTs.

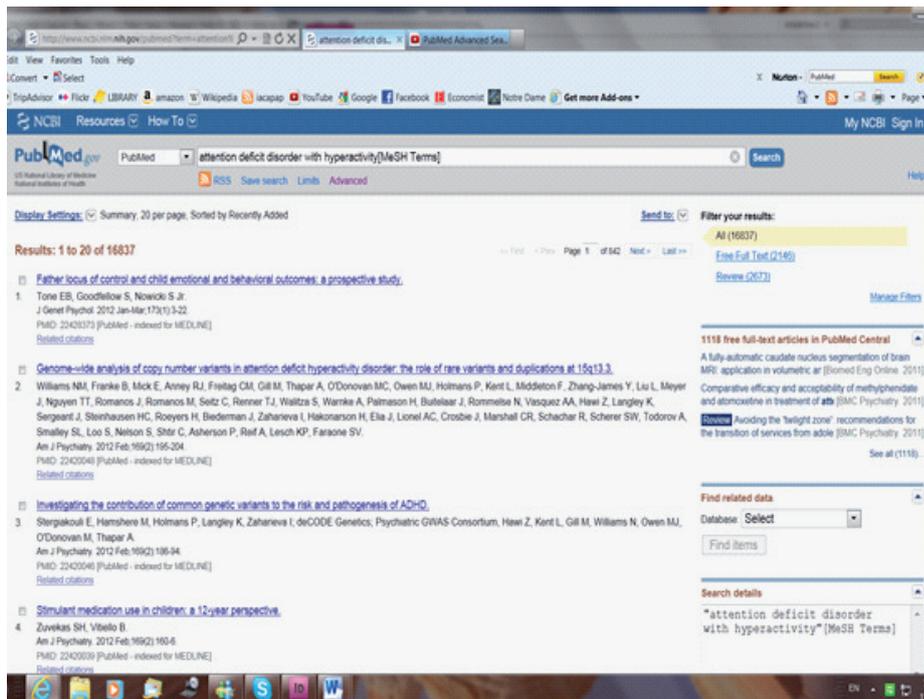


The logo of PubMed, a free source of evidence. It has filters to find systematic reviews, useful as summaries of high-quality evidence

Terry: a 10-minute search for a solution for a boy with inattentive type ADHD

A child psychiatrist is trying to help a 10-year-old boy with inattentive type ADHD (ADHD-I). Based on SWAN scores (see Table A.6.3) from his teacher, he required methylphenidate 15 mg three times daily to achieve a substantial improvement (a teacher rating of less than 9 on the first nine SWAN questions, the inattentive items). But he lost so much weight on this regimen that it had to be stopped. Now his parents are increasingly frustrated and are seeking a solution: "Isn't there anything else?" His child psychiatrist wants an agent other than a stimulant or atomoxetine – both suppress appetite. Our doctor recalls how helpful guanfacine has been in the past treating ADHD boys with ODD and hyperactive/impulsive symptoms. But is guanfacine effective in ADHD-I? What is the evidence that guanfacine is effective in 10-year boys with ADHD-I? If it is effective, what is the effect size?

1. As a reasonable place to begin, go to [PubMed](#). Click on “Advanced”, select in the drop down menu “MeSH term” (MeSH stands for Medical Subject Headings), type “attention deficit hyperactivity disorder” in the search box and click “Search”. The result should look like the figure below.



There are too many resulting “hits” to be able to read through them yourself – over 16,000! We need a way to limit the results to a manageable number by choosing only those results from the top of the pyramid in order to see the *best* evidence. Here's how.

2. Begin at the top of the pyramid: search for any *meta-analyses* published. Go back to the search page and, after repeating the previous step, type “guanfacine” in the search box “AND” and “All Fields” and click “Search”. This results in 94 “hits”. Now click on “Limits” and select “Meta-Analysis” in the new menu and click “Search”; 2 hits result but this boy does not belong to the population studied by either article: one concerns adults,

Searching PsycINFO

PsycINFO is an excellent database choice for searches of constructs like “bullying” or “identity” because it has a thesaurus of index terms. Searching the term “bullying”, for example, shows the following Index Terms listed in a column to the left of the abstracts, along with the number of hits for each index term.

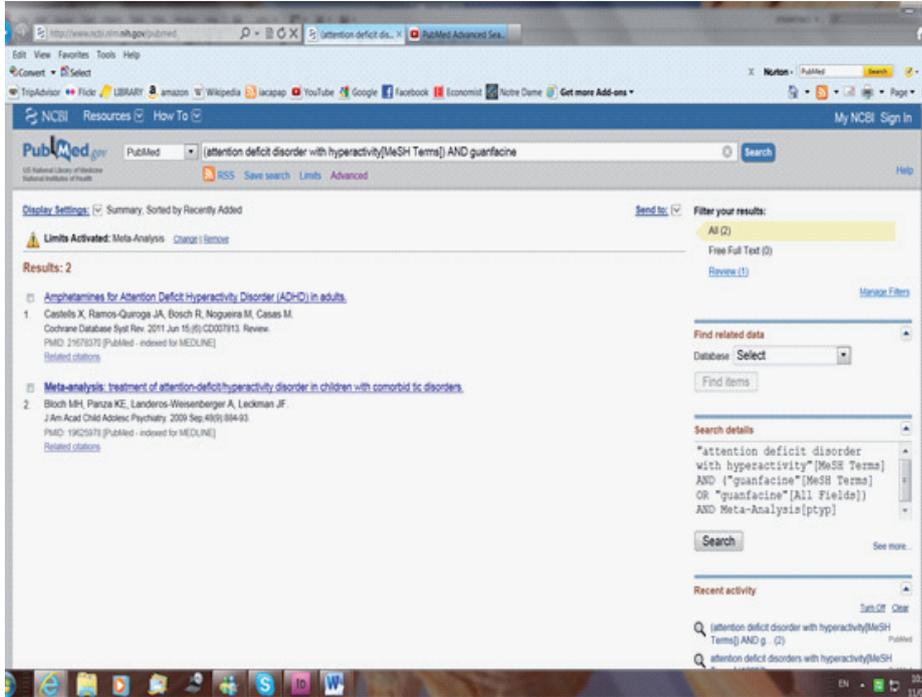
Index Term:

- Bullying (2,808)
- Victimization (1,175)
- Aggressive Behavior (845)
- Peer Relations (543)
- Schools (437)
- 5 more...

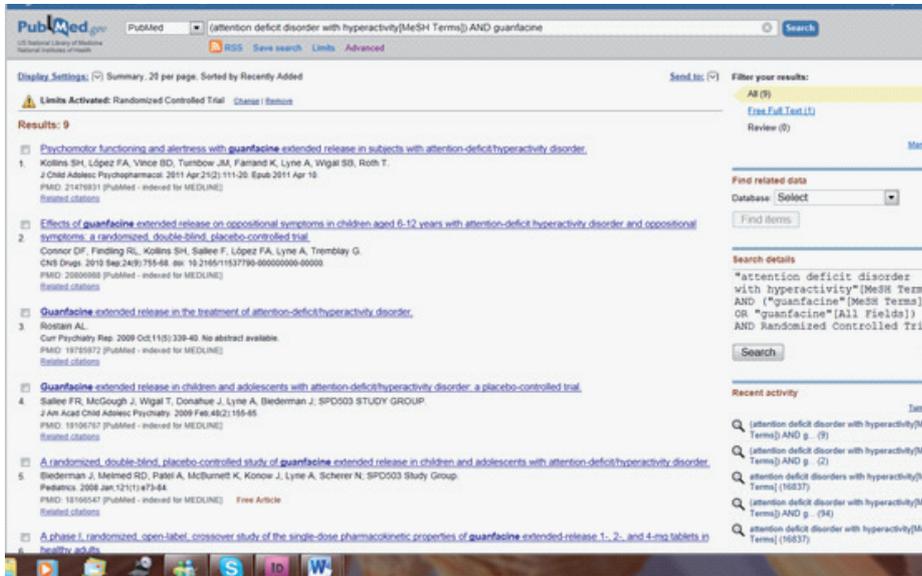
Clicking on the hypertext on the website brings you to those abstracts which are available at no charge. Use the “Advanced Search” tab to search “bullying” and also choose “Only Show Content Where...” to include the dropdown menu choice “Treatment Outcome/Randomized Controlled Trial.” This reduces 2808 hits to only 10 (October 28, 2011).

For an exercise, try finding all randomized trials studying the effect of relaxation training in adolescents.

the other, children with ADHD and tics, and our patient has ADHD without tics (see Figure below). We have become too selective and need to broaden the search.



3. Move down a step in the pyramid to RCTs by changing the "Limits" box from "Meta-analysis" to "Randomized controlled trials", and 9 articles appear (see Figure below).



Articles 4 and 5 are the only ones directly relevant to a 10 year old boy, and only article 5 is available in full text at no charge (Biederman et al, 2008). This article was funded by pharmaceutical companies, a potential source of bias, and concerns extended-release guanfacine, which is expensive and not available to this boy. A quick Internet search, however, finds that guanfacine's half-life in children and adolescents is 13 to 14 hours, suggesting that it is possible to mimic the extended-release preparation with an affordable preparation – guanfacine itself – taken at least twice daily (Strange, 2008).



Click on the picture to view a video tutorial on how to search using PubMed (2:27)

How big was guanfacine's effect on inattentive symptoms in youth with the combined type ADHD among the children studied? The authors note "Effect sizes in the GXR treatment groups calculated *post hoc* were 0.58, 1.19, and 1.34 for the 0.05 to 0.08 mg/kg, 0.09 to 0.12 mg/kg, and 0.13 to 0.17 mg/kg groups, respectively." However, these effect sizes were for the changes in ADHD total scores (inattentive and hyperactive/impulsive symptoms). Fortunately, the authors clarify "All groups of children taking GXR showed significant improvement on both the hyperactivity/impulsivity and inattentiveness subscales of the ADHD-RS-IV [a rating scale] compared with placebo. Mean changes from baseline in hyperactivity/impulsivity in the placebo and GXR 2-mg, 3-mg, and 4-mg groups were -4.06, -6.94, -7.09, and -9.46, respectively. *Mean changes from baseline in inattentiveness were -4.78, -8.46, -8.71, and -9.51, respectively.*[italics added]" Even though no effect size is offered concerning changes in inattentiveness alone, these data suggest that the effect size will be at least equal to the effect size for inattentive and hyperactive/impulsive symptoms combined.

It is reasonable to conclude that the 10-year old boy with inattentive type ADHD may respond with a significant reduction in symptoms to guanfacine titrated to a dose of 1.0 to 1.5 or 2.0 mg twice daily.

Internet strategies to find information about a broad subject area

Rather than looking for evidence on a specific clinical question about one youth, imagine that you are writing a guideline for your country on how to treat children and adolescents with schizophrenia. One relevant question might be, "What is the evidence there are effective treatments for youth with schizophrenia?" Begin by finding the most current, relevant data from randomized trials. Begin searching PubMed for treatment studies in children and adolescents with schizophrenia. Here's how to do it (note that the number of hits will change with time as new studies are added).

1. *Find the medical subject heading for your search topic.* Enter "Schizophrenia" in the PubMed search box after using the dropdown menu to change the setting to "MeSH". This means you are now searching the Medical Subject Headings used by the National Library of Medicine in the United States. These Medical Subject Headings are used to categorize every article published in the medical literature, making it possible to retrieve all articles on the same subject. In this case, multiple (12) MeSH terms are retrieved, commonly described as 12 hits. Each contains the word "schizophrenia" either in the title or the description.
2. *Send the relevant terms to the Search Builder box, combining them with "OR."* Only 6 appear relevant since the final 6 concern an unusual protein. After selecting these 6, click on "Add to Search Builder".
3. *Add a truncated version (tw) of the central concept as a text word to the search with an "OR".* In this case, add "Schizophren*" (tw) to the search box with the "OR" operator. This step will retrieve any article containing any word beginning with the letters "Schizophren" in the title or abstract. This additional term can help find any articles which may have not been correctly classified under one of the MeSH terms as containing information about schizophrenia, as well as articles only using such terms

Searching the Cochrane Library's CENTRAL database

Searching CENTRAL for "bullying" yields 36 hits, all RCTs studying bullying. One done in Australia, for example, in 2009 shows promise in adolescent boys, and did not appear in the PsycINFO search for RCTs of "bullying." (Berry K and Hunt CJ (2009), Evaluation of an intervention program for anxious adolescent boys bullied at school. *Journal of Adolescent Health* 45:376-382). CENTRAL is useful in retrieving RCTs from around the world and making abstracts available at no cost to everyone.

as "schizophrenic" or "schizophrenics"

4. Search PubMed with this string of "OR" terms.

125889 hits occur. It is time to now get more specific and begin carving this large set of references into only the most relevant.

5. Click on "Limits" and check the box "Children 0 to 18"

22545 hits remain

6. Click on "Limits" and check the boxes "Meta-analysis" and "Randomized controlled trial".

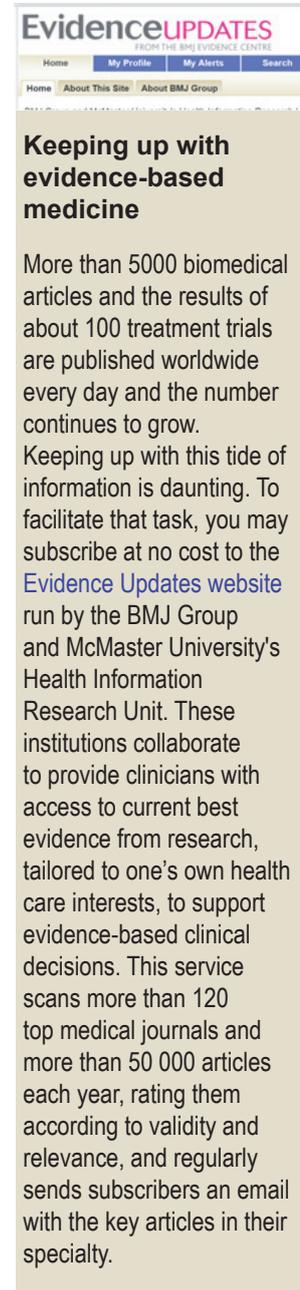
945 hits remain. This may be a reasonable number to look through for a major position paper for an entire country. Two or three individuals could scan the titles and abstracts and choose the most relevant. Alternatively, you could go back to "Limits" and choose the box "Published in last 5 years", further reducing the number of references. Note also that many references only concern adults, not children or youth, and can be dropped immediately.

Here is an alternative if you don't have the time to look through so many hits:

1. Return to step 1 and note that one of the MeSH terms returned is "Schizophrenia, childhood." Send only the term "Schizophrenia, childhood"[MeSH] to the search box, ordering the search engine to only search that single MeSH term. 1434 hits appear.
2. Now apply the limits "Meta-analysis" or "Randomized controlled trial". 8 hits remain. These hits are mandatory reading, but the limits seem to have eliminated too many hits.
3. Return to step 2 with 1434 hits and apply the limits "Last 10 years." 59 hits remain. Scanning these hits reveals a series of highly relevant articles with topics like "Effects of clozapine and olanzapine on cortical thickness in childhood-onset schizophrenia," and "Asymmetry loss is local rather than global in childhood-onset schizophrenia." Note the dramatic falloff in the number of hits, suggesting that – as the PubMed sites notifies us – "Schizophrenia, childhood" as a MeSH term is primarily an historical term; yet the 59 hits retrieved from the last 10 years appear mostly highly relevant and interesting for a position paper.

The take-home message is to *keep search strategies transparent and clear, and always monitor how many hits result; these data allow you to be continually adjusting your search to your purposes, time, and interest.*

Keep track of how many hits a strategy produces in the context of the purposes of your search: are the results manageable and adequate for your purposes? If you sense they are not manageable (too many hits and too little time), continue to carve out new results with further limits. If you sense they are not sufficient, change your initial search strategy to be more sensitive. In this example, adding the term "Schizophren*" (tw) added more than 20,000 hits. Another alternative to increase sensitivity is to click on the "Related citations" hypertext below the citations which seem especially useful; the software reports a series of citations similar in subject matter to the citation you find useful.



EvidenceUPDATES
FROM THE BMJ EVIDENCE CENTRE

Home My Profile My Alerts Search

Home About This Site About BMJ Group

Keeping up with evidence-based medicine

More than 5000 biomedical articles and the results of about 100 treatment trials are published worldwide every day and the number continues to grow.

Keeping up with this tide of information is daunting. To facilitate that task, you may subscribe at no cost to the [Evidence Updates website](#) run by the BMJ Group and McMaster University's Health Information Research Unit. These institutions collaborate to provide clinicians with access to current best evidence from research, tailored to one's own health care interests, to support evidence-based clinical decisions. This service scans more than 120 top medical journals and more than 50 000 articles each year, rating them according to validity and relevance, and regularly sends subscribers an email with the key articles in their specialty.

CLASSIC EBM AND THE 'ANSWERABLE QUESTION'

"EBM classic" is our phrase for the approach developed originally by David Sackett and colleagues at McMaster University in Canada and later at many other sites. Their work was fundamental to the approach discussed here. We distinguish our approach, however, which is tailored to fit the reality of doing work in child mental health, from "EBM classic," which was originally developed in the context of internal medicine. It's important to know about "EBM classic" as the bedrock precursor of what is discussed here.

"EBM classic" can be summarized as a five-step model:

1. Posing an "answerable" question about a clinical situation, usually beginning "What is the evidence that..." a specific *patient or problem*, treated with a specific *intervention* yields in *comparison* to another intervention or placebo yields what *outcomes*? These are sometimes called PICO questions (Patient/Intervention/Comparison/Outcomes).
2. Searching for evidence to answer the PICO question.
3. Critically appraising the evidence by asking questions about how well the studies were done.
4. Integrating clinical expertise and the patient's values with the evidence to make a decision about treatment.
5. Evaluating the result.



Click on the picture to access the Centre for Evidence Based Medicine website, which has information and training on EBM

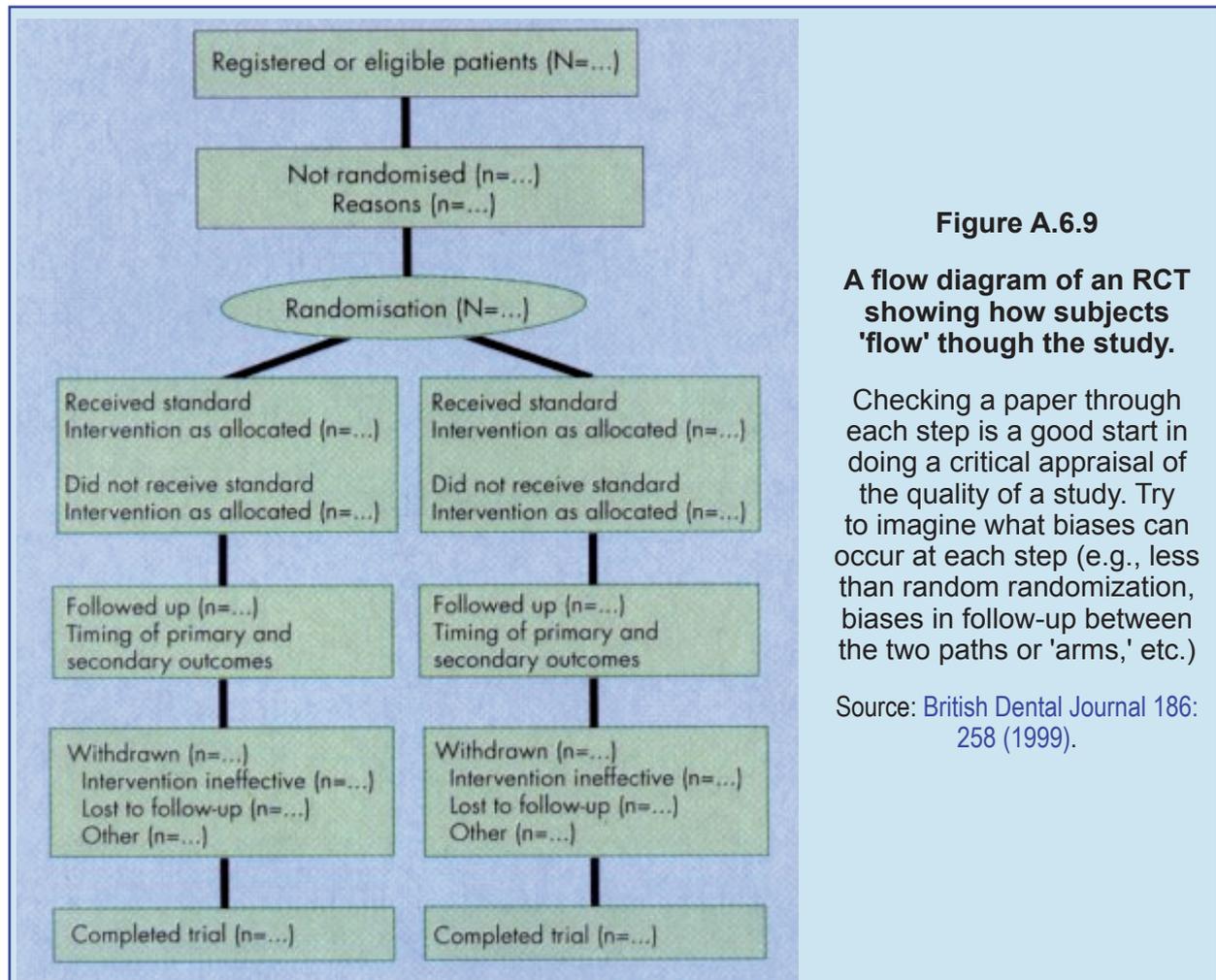


Figure A.6.9

A flow diagram of an RCT showing how subjects 'flow' through the study.

Checking a paper through each step is a good start in doing a critical appraisal of the quality of a study. Try to imagine what biases can occur at each step (e.g., less than random randomization, biases in follow-up between the two paths or 'arms,' etc.)

Source: *British Dental Journal* 186: 258 (1999).

"EBM Classic" also takes on such important issues as the sensitivity and specificity of diagnostic tests and how to apply the evidence (Sackett et al, 2000).

EVIDENCE-BASED PRACTICES AND LOCAL CULTURE

Culture – left undefined for our purposes – is an important factor coloring how the ideas discussed in this chapter are received by clinicians and patients alike. Cultural components affecting our practices include beliefs and values, child rearing practices, the role of the extended family, and the impact of the Internet.

Beliefs and values

Beliefs regarding the treatment of mental illnesses are important, and it is crucial that clinicians understand the culture where they work in order to enable the acceptance of effective interventions. In societies where there is more severe stigmatization of the mentally ill, it can be very difficult to convince a child and family to even attend an initial psychiatric interview. Even if they attend, it may be very difficult for them to accept the diagnosis as well as the suggested treatment. In some cultures mental illnesses are believed to be spiritual and therefore spiritual treatments are the first choice. Families may comply with the treatment only if the spiritual healer approves. In one such case a Turkish spiritual healer had prepared a *muska* – a magical necklace of wrapped paper and cloth to get rid of the illness – for a psychotic teenager who the healer had sent for hospitalization. In the process of his treatment the adolescent opened the *muska*. He found only one scrawled word: *Haldo!*

Yet clients often look for magical, immediate cures. It is very difficult at times to explain, and have them accept, the time required for good treatment, or the advisability to use a medicine in spite of its potential side effects. In cultures where people first ask their neighbors or relatives about the problem and rely on their personal or third party experiences, EBT suggested by a clinician may not always fit with the neighbors' suggestions and often acquires "not-to-be used" status (the neighbor's child got well with another medicine, thus this may not be the right one).

Assumptions in child rearing practices

Assumptions like "boys normally show high motor activity" can be a reason for a family to refuse medication to treat ADHD, because the family thinks the drug slows down the child and makes him "abnormal." Another common assumption is that "children should not be given pills for a long time." This often disrupts medication treatment. Or, "Grandma says he is just like his father as a child, and his father is doing OK. Are you sure he needs treatment?"

Role of the extended family

In traditional cultures, grandparents or elder brothers or sisters of the parents may be in charge of treatment decisions. They may refuse to allow the parents to administer the suggested treatment. On some occasions parents who believe in EBT are faced with the conflict of administering the treatment secretly. Family members may also interfere with the dosage prescribed, based on their personal opinions.

The influence of the Internet

Parents or adolescents checking the Internet may well refuse treatment. Since the Internet may misinform as well as inform, patients and their families often find more misinformation than guidance. For example, they may Google a word and read whatever is displayed, often the information found emphasizes side-effects, ignores benefits, and opposes medications for mental illness. Providing parents and teens with information about Internet sites with fewer biases and more accurate and complete information may improve adherence to treatment.

To summarize, a key role of the clinician is to understand both the cultural system of beliefs and values of her patients as well as the fundamental ideas of epidemiology and clinical research that make up the first part of this chapter. The successful combination of both, as applied to the patient and family in the office, is the essence of evidence-based practice.

REFERENCES

- Biederman J, Melmed R, Patel A et al (2008). A randomized, double-blind, placebo-controlled study of guanfacine extended-release in children and adolescents with attention-deficit/hyperactivity disorder. *Pediatrics* 121: e73-84
- Brent D, Perper J, Moritz G et al (1993). Firearms and adolescent suicide. A community case-control study. *American Journal of Diseases of Children*, 147:1066-1071.
- Guyatt GH, Rennie D (2002). *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice* Chicago: AMA Press
- Kabai P (2011b). *Bias and Random Error*. http://www.behav.org/ecol/wildlife/w_05_design.htm
- Kabai P (2011a). *More on Bias (systematic) and Random Error*. http://apollo.lsc.vsc.edu/classes/remote/lecture_notes/measurements/bias_random_errors.html
- Kaufman J, Birmaher B, Brent D et al (1997). *Kiddie-SADS-Present and Lifetime Version (K-SADS-PL)*. Pittsburgh: Department of Psychiatry, University of Pittsburgh School of Medicine.
- Kelly R, Cohen L, Semple R (2006). Relationship between drug company funding and outcomes of clinical psychiatric research. *Psychological Medicine*, 36:1647-1656.
- McCracken JT, McGough J, Shah B (2002). Risperidone in children with autism and serious behavioral problems. *New England Journal of Medicine*, 347:314-321.
- Sackett DL, Straus SE, Richardson WS et al (2000). *Evidence-Based Medicine: How to Practice and Teach EBM*, 2nd edition. New York: Churchill Livingstone.
- Strange BC (2008). Once-daily treatment of ADHD with guanfacine: patient implications. *Journal of Neuropsychiatric Disease and Treatment*, 4:499-506.
- Walkup JT, Albano AM, Piacentini J et al (2008). Cognitive behavioral therapy, sertraline, or a combination in childhood anxiety. *New England Journal of Medicine*, 359:2753-2766
- Wikipedia (2011). *Standard Deviation*. http://en.wikipedia.org/wiki/Standard_deviation
-