

Geometrically Evident: Framing Studies Using the Graphic Appraisal Tool for Epidemiology (GATE)

ANDRÉS MARTIN, M.D., M.P.H., AND VINOD SRIHARI, M.D.

A picture is worth a thousand words.

Educators in evidence-based medicine (EBM) have noted that the core set of epidemiological concepts outlined in standard sources (Guyatt and Rennie, 2002; Straus et al., 2005) are sometimes put to use as oversimplified checklists for the appraisal of research reports. This can paradoxically cloud the usefulness of tools that were specifically designed to spur critical thinking. Extended in-depth appraisal efforts can also get lost in the minutiae of sorting formulae and calculating numbers. Given this state of affairs—one coinciding with medical education's ongoing incorporation of EBM principles—a team led by New Zealand epidemiologist Rod Jackson developed the Graphic Appraisal Tool for Epidemiology (GATE; Jackson et al., 2006). The GATE was designed as a way combat, by visual means, a formulaic approach to appraising the scientific literature. The GATE is available

on the *Journal's* website at www.jaacap.com via the Article Plus feature.

By prompting the reader to explore the structure of clinical studies by “hanging” their different elements onto its “frame,” GATE teaches core principles of epidemiological study design while also serving as a practical tool for applying the results of studies to particular patients. Furthermore, the GATE is free and available online as an Excel workbook. The elegance of the tool can best be appreciated in its spreadsheet format, with features such as color coding to identify errors in data entry, automatic calculations of summary statistics and confidence intervals, and pop-up windows with hints on the EBM process. Quite apart from these user-friendly aspects, we believe that the real didactic value of the GATE is in conveying core epidemiology and study design concepts.

In this column, we define and deconstruct the GATE frame, exemplify its use by applying it to a widely known study in child and adolescent psychiatry (Treatment for Adolescents With Depression Study [TADS], 2004), note the core appraisal and epidemiology principles that are highlighted through its use, and propose applications for the GATE in clinical, teaching, and research settings.

FIGURING IT OUT: WALKING THROUGH THE GATE

The GATE uses five basic geometric shapes: the triangle, circle, square, arrow, and cross (Fig. 1). The first four follow the classic EBM acronym upon which all epidemiological studies are based, PICOT (for population, intervention, comparison, outcome, and time). We define each of the components below and provide Treatment for Adolescents With Depression Study (TADS)-based examples for each in the third column of Table 1. A more detailed EBM-based discussion of TADS can be found in previous papers by

Accepted July 3, 2006.

Dr. Martin is with the Yale Child Study Center and Dr. Srihari is with the Department of Psychiatry, Yale University School of Medicine, New Haven, CT.

The authors first learned about the GATE from its main developer, Rod Jackson, M.B.Ch.B., Ph.D., while attending the 11th Oxford Workshop in Teaching Evidence-Based Practice at the University of Oxford's Centre for Evidence-Based Medicine, September 5–9, 2005. They express their gratitude to Professor Jackson for allowing them to share the GATE materials with the broader child and adolescent psychiatric community. The GATE continues to be modified and is available to download at <http://www.epiq.co.nz> as well as in the Article Plus materials for this article on the Journal's Web site.

Supported in part by the Albert J. Solnit endowment and by a gift from Amy Arkin to the Yale Child Study Center. The comments of Drs. John March and Assistant Editor John Hamilton on an earlier draft manuscript are gratefully acknowledged.

Article Plus (online only) materials for this article appear on the Journal's Web site: www.jaacap.com.

Correspondence to Dr. Andrés Martin, Yale Child Study Center, 230 South Frontage Road, PO Box 207900, New Haven, CT 06520-7900; e-mail: andres.martin@yale.edu.

0890-8567/06/4512-1521©2006 by the American Academy of Child and Adolescent Psychiatry.

DOI: 10.1097/01.chi.0000237707.50884.c2

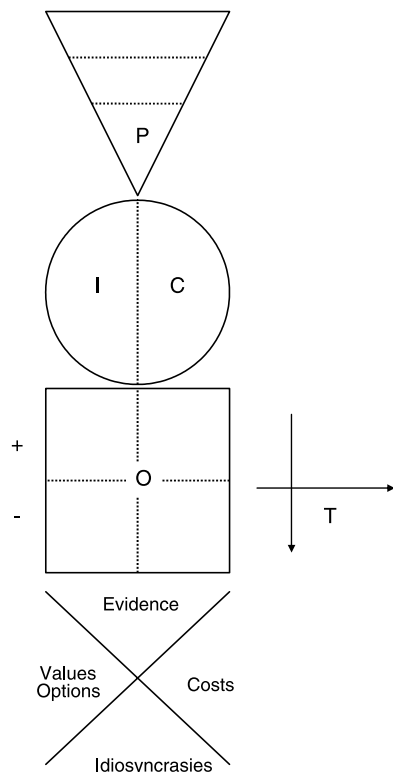


Fig. 1 The Graphic Appraisal Tool for Epidemiology (GATE) frame. P = population; I = intervention; C = comparison; O = outcome; T = time.

its authors (March et al., 2005; Treatment for Adolescents With Depression Study [TADS], 2004) and by us (Srihari and Martin, 2006).

Triangle

The triangle represents the population studied (P). The triangle can be subdivided into three overlapping levels: the entire area represents the source population from which all of the participants were selected and encompasses the other two triangles. The middle triangle represents the eligible population: those subjects meeting the study’s inclusion and exclusion criteria. The lowermost and smallest triangle represents the participant population: those individuals who were actually enrolled in the study.

Circle

The circle, divided in two by a vertical line, separates the study participants according to the exposure of interest into an intervention (I) group on the left and a comparison (or control [C]) group on the right. Although the traditional epidemiological study involves only two groups, the GATE can be easily adapted to accommodate

multiple contrasts, as in the case of TADS, in which four different arms were involved (fluoxetine [FLU] alone, cognitive-behavioral therapy [CBT] alone, their combination [FLU + CBT], or pill placebo [PLA]). For visual depiction, additional vertical lines can be drawn in the circle; when using the Excel version of GATE, additional worksheets can be added for each comparison.

Square

The square represents the outcomes (O) of the study and consists of the generic 2 × 2 table at the heart of all epidemiological designs: its columns once again represent the relevant exposures (I or C), its rows the dichotomous outcomes of interest (yes or no). For a multiple comparison study such as TADS, more than one table is necessary to fully examine the results. For simplicity of exposition, only one of the major study comparisons is presented in detail here (FLU versus PLA). In addition to showing response rates and actual number of subjects in each of the four cells, the simple calculation of the number needed to treat (NNT) is also provided in the example. Note that the NNT is calculated as the inverse of the difference between response rates, rounded to the higher nearest integer; in other words, 1/(response rate intervention – response rate comparison). Although confidence intervals (CIs) around response rates or NNTs can be manually derived, calculations can be cumbersome and are automated in the GATE spreadsheet (95% CIs are provided in parentheses in the Table 1 example). When considering continuous outcomes (e.g., change scores in the Children’s Depression Rating Scale), the spreadsheet can also be used to calculate effect sizes.

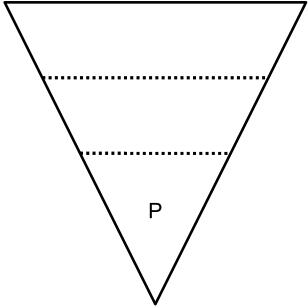
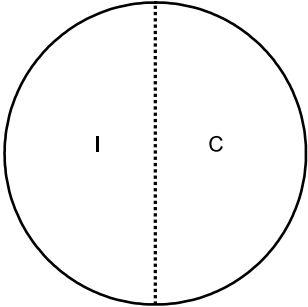
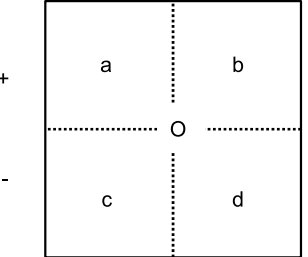
Arrows

Study time (T) is represented by arrows. These can be depicted horizontally to indicate a cross-sectional observation at a given point in time (such as the prevalence of suicidal ideation in 29% of subjects at baseline) or vertically to indicate change over time (e.g., different time trends in suicidal ideation during the 12 weeks of the study for each of the interventions).

FROM GEOMETRIC EVIDENCE TO VISUAL EPIDEMIOLOGY

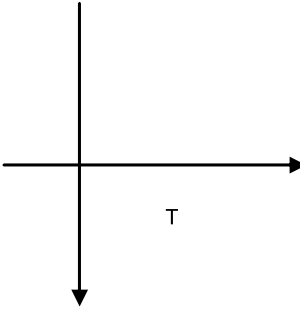
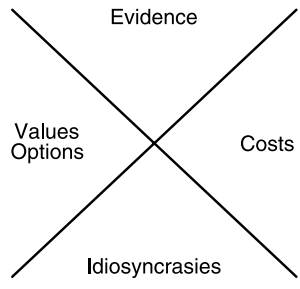
Once a study has been adequately “hung” on the GATE frame, a reader can approach the task of critical appraisal armed with the necessary information. The

TABLE 1
The GATE Frame Deconstructed

GATE Component <i>Acronym: PICOT</i>	Definition	TADS Example	Epidemiological Relevance <i>Acronym: RAAAMBO</i>
<p>Triangle</p> 	<p>Source Population: From which all participants are selected</p> <p>Eligible Population: Subjects meeting the study's inclusion and exclusion criteria</p> <p>Participant Population: Individuals actually enrolled</p>	<p>2,804 subjects screened by telephone from various sources</p> <p>1,088 subjects signed consent for evaluation of eligibility</p> <p>439 subjects were randomized and yielded analyzable data</p>	<p>Representative (<i>R</i>)? Generalizability or external validity: how applicable is the study?</p> <p>Possible threats to external validity: Efficacy in a controlled environment (such as a clinical trial with rigid eligibility criteria) may translate poorly into 'real world' conditions (effectiveness).</p> <p>Possible remedies: More liberal eligibility criteria may introduce some 'noise' (e.g., comorbidity), but yield more 'realistic' study samples.</p>
<p>Circle</p> 	<p>Participants are divided according to the exposure of interest into: Intervention (<i>I</i>) group(s) and Comparison (or Control, <i>C</i>) group</p>	<p>Fluoxetine alone (FLU) 109 subjects randomized</p> <p>Cognitive behavioral therapy alone (CBT) 111 subjects randomized</p> <p>Combination treatment (FLU + CBT) 107 subjects randomized</p> <p>Pill placebo (PLA) 112 subjects randomized</p>	<p>Allocation (<i>A</i>)? Are subjects in the <i>I</i> and <i>C</i> groups comparable?</p> <p>Possible threats to internal validity: Selection bias</p> <p>Confounding, or mixing of effects</p> <p>Possible remedies: Randomized allocation balances for confounds, both known and unknown.</p> <p>Adjustment (<i>A</i>): Stratified or multivariate analyses.</p>
<p>Square</p> <p>I C</p> 	<p>Columns represent the exposure of interest (<i>I</i> or <i>C</i>)</p> <p>Rows represent the outcome of interest: Yes (+) or no (-) for dichotomous endpoints, such as a priori defined response or non-response.</p> <p>Mean values for each column can be used for continuous endpoints.</p>	<p>FLU Responders, 61% [51–70] (cell a, 67 subjects) Non-responders, 42% (cell c, 42 subjects)</p> <p>PLA Responders, 35% [26–44] (cell b, 40 subjects) Non-responders, 72% (cell d, 72 subjects)</p> <p>Number Needed to Treat (NNT) = $1/(0.61-0.35)$ = 3.8 = 4 [3–8]</p> <p>Mean improvement in CDRS-R score: FLU = 22.6 points PLA = 19.4 points Effect Size = 0.68</p>	<p>Are all subjects Accounted for (<i>A</i>)? Number of participants should be same as <i>I</i> + <i>C</i>, and same as number analyzed in the 2×2 tables. Subjects lost to follow-up, non-completers, or those "contaminated" in their exposure, may account for discrepancies.</p> <p>Possible threats to internal validity (<i>possible remedies</i>): Systematic errors (bias), such as:</p> <ol style="list-style-type: none"> 1. Withdrawal <i>Sensitivity analysis</i> 2. Measurement (<i>M</i>) <i>Blinding (B) of subjects and/or assessors</i> <i>Use of Objective (O) and reliable measures</i>

(Continued)

TABLE 1
(continued)

GATE Component <i>Acronym: PICOT</i>	Definition	TADS Example	Epidemiological Relevance <i>Acronym: RAAAMBO</i>
<p>Arrows</p> 	<p>Study Time</p> <p>A horizontal arrow is used to depict a given point in time (e.g., point prevalence, or cross-sectional rates of a side effect)</p> <p>A vertical arrow is used to depict temporal effects (e.g., time trends)</p>	<p>Clinically significant suicidal ideation was present in 29% of participants at baseline.</p> <p>FLU + CBT showed the best improvements in suicidal ideation over 12 weeks' time.</p>	<p>Prevalence refers to the proportion of subjects in a population with a given characteristic. It is a unitless measure (e.g., number with major depression per every 1,000 adolescents)</p> <p>Incidence refers to <i>new cases</i> identified in a given period of time. By definition it includes a 'per time' denominator (e.g., number of new cases of major depression per every 1,000 adolescents identified <i>per year</i>)</p>
<p>Cross</p> 	<p>'X' or 'eXpertise' factor</p> <p>Some of the key aspects to take into account when incorporating best evidence into clinical practice</p>	<p>Patients and families may have preconceived notions about medication or psychotherapy</p> <p>Options: apart from CBT and medication, can other approaches be considered? (e.g., interpersonal, supportive or psychodynamic psychotherapies)</p> <p>Idiosyncrasies: a given community may not have any practitioner adequately trained in CBT, making fluoxetine a more viable treatment option.</p> <p>Costs: does insurance allow for medication and/or psychotherapy coverage? If not, can the family afford these or can they be eligible for alternative funding options?</p> <p>Evidence: see above, as well as TADS (2004) and Srihari and Martin (2006) for more detailed analyses</p>	<p>Acronym: VOICE</p> <p>Values: aspects of importance to the patient/family</p> <p>Options: alternative treatments to consider</p> <p>Idiosyncrasies: unique characteristics of patient (e.g., comorbidity) or practitioner (e.g., locally available resources)</p> <p>Costs: is the treatment not only available, but affordable as well? How can larger scale costs of treatment (or lack thereof) inform policy decisions?</p> <p>Evidence: incorporating and particularizing EBM into clinical practice</p>

Note: Numbers in brackets represent 95% confidence intervals around the relevant point estimate. PICOT = Population, Intervention and Comparison, Outcome, and Time; RAAAMBO = Representative, Allocation, Adjustment, Accounted for, Measurement, Blinding, and Objective; CDRS-R = Children's Depression Rating Scale-Revised; GATE = Graphic Appraisal Tool for Epidemiology (Jackson et al, 2006); TADS = Treatment of Adolescents with Depression Study (TADS, 2004); VOICE = Values, Options, Idiosyncrasies, Costs, and Evidence.

core of critical appraisal is to determine the study's validity and whether the threats to it are serious enough to call its conclusions or applicability into question. EBM provides the tools to make this task as objective

and systematic as possible, and the GATE facilitates this process and makes it intuitively appealing. As illustrated in the final column of Table 1, and as further described below, the acronym RAAAMBO can be linked to the

GATE elements to tie key validity issues into the critical appraisal process.

First, a study should provide enough information to determine how representative (*R*) its findings are likely to be. The external validity, generalizability, or applicability of a study is related to how similar or different the patients included in it are to those in actual clinical settings. A study may be methodologically sound or internally valid, but if it cannot be generalized to a wider population, its efficacy may not translate into broader (real life) effectiveness.

Next, the allocation (*A*) of subjects between intervention and comparison groups needs to be explicitly understood. If subjects in these two groups differ in ways that may independently affect the outcome of interest, then to state that the intervention is responsible for the change can be compromised through confounding or the mixing of effects. Given that not all potentially confounding factors can be known with certainty at the outset, randomized allocation, whenever possible, is the ideal method for intervention studies such as clinical trials. When randomization is not possible, adjustment (*A*) through stratification or multivariate statistical analyses can be used instead. In stratification, data are analyzed across a confounding variable (e.g., children with or without a comorbid anxiety disorder can be analyzed separately if anxiety is deemed to confound treatment response). Finally, given that even an optimally allocated study (e.g., TADS) is likely to cause some residual confounding, multivariate analyses that adjust for age, gender, treatment site, and the like are almost invariably conducted.

Third, the accounting (*A*) of all subjects is essential and something that becomes clear through the use of GATE. This is especially useful for studies that, unlike TADS, may not provide a clear tally of subject flow (see Treatment for Adolescents With Depression Study [TADS], 2004; Fig. 1). A study's balance sheet can be ascertained by seeing whether all of the subjects at the lower tip of the triangle (i.e., the study participants) equal the total number in the circle (i.e., those assigned to the I and C groups). Although the totals in the square should ideally add up to the same total, this is not always the case because subjects will often be lost to follow-up, become "contaminated" (i.e., accidentally receive the alternative exposure), or demonstrate poor compliance (thus becoming "unexposed" to the intervention). The GATE forces us to account for each and

every subject and to assess which ones are lost along the way. More important, if losses are proportionally greater in one group compared with another, a systematic error, or withdrawal bias, may have been introduced.

Another common type of bias concerns measurement (*M*), a situation that is particularly germane to child and adolescent psychiatry, in which instruments are often based on subjective reports and can give varying results according to informant, site, and other characteristics. Protections against measurement bias include blinding (*B*) participants and assessors to exposure status, and the use of instruments that are as objective (*O*) and reliable as possible (i.e., yielding similar results across time or informants).

Science mirrors life, and no study is flawless. In the final analysis, the GATE is best used not to punch holes in a study design but rather to provide a textured scientific picture: to put into bolder relief the epidemiologically relevant strengths and weaknesses of the study. Do a study's blemishes add up to its being deemed internally invalid and thus not useful? Alternatively, is a study sturdy enough but of limited generalizability to the patients under our care? Is the size of the effect large and precise enough for us to pay attention? When looking at published reports in the context of their unique strengths and weaknesses, fatally flawed studies become the exception. GATE encourages a more objective statement on the overall quality of a study and directs the reader to evaluate just how much the results can be relied on as an answer to the clinical question that inspired the critical appraisal in the first place. Can the results be relied on to make clinical decisions and direct further studies? To address these questions is to engage in critical appraisal in the full sense of the term.

X Factor

A commonly held misconception about the EBM approach is that it "fails to appreciate the uniqueness of each child and takes the art out of clinical practice" (Hamilton, 2005). The bottom figure in the GATE is an explicit way of countering such a view by incorporating evidence within a broader clinical context, one that takes into account the individuality and experience of all parties involved.

The "X" or "eXpertise" factor relies on the practitioner incorporating appraised evidence into a larger clinical context. The acronym VOICE can be used

around the “X” to emphasize how the voice of the individual patient (as much as of the practitioner) is of central importance to the practice of EBM. The acronym stands for *v*alues that guide preferences (e.g., is a patient’s family distrustful of mood-altering psychotropic medications?); *o*ptions (e.g., is the practitioner adequately trained in alternate treatments such as CBT?); *i*diosyncrasies (e.g., psychiatric or substance abuse comorbidity, or atypical pharmacokinetic drug handling); *c*osts (e.g., what is the relative cost-effectiveness for the individual to consider different treatment approaches?). As can be seen, the final E (standing for *e*vidence) is but one among a range of critical considerations subsumed under the rubric of clinical expertise.

Readers who use the spreadsheet version of GATE may notice a discrepancy in the elements around the “X”, namely, patient preferences, epidemiological evidence, policy issues, and clinical considerations in the Excel version. We prefer our acronymically organized VOICE instead, but the underlying constructs are fundamentally the same in both versions.

APPLIED GEOMETRY: READING, TEACHING, AND REVIEWING

Our emphasis here has been on a randomized, clinical trial, given that treatment studies are the ones most likely to be “framed” by child and adolescent psychiatrists. The GATE, however, as an epidemiological “map” for clinical studies generalizes well to other commonly used study designs (cross-sectional, case control, cohort). Make no mistake: “hanging” a study on the GATE frame takes time, especially when learning how to navigate its spreadsheet version, no matter how user-friendly. Despite this initial investment of time, or precisely because of it, we believe that using the GATE can be an invaluable learning experience.

As a tool for self-directed reading and learning, GATE brings “Epidemiology 101” into the world of clinical decision making. It guides the reader wading through reports that are complex or statistically multi-layered with basic concepts such as the role of bias and the play of chance. The authors have also found GATE spreadsheet templates to be a handy place to collect

relevant information about a pivotal trial (often published in multiple separate reports) and store the appraisal for future reference.

As a didactic tool, the GATE can be used to structure the discussion in courses that require presentations from the literature. For example, in following the EBM-based approach to journal clubs proposed by March et al. (2005), we have used the GATE as a way to engage trainees in a process of selectively interrogating the literature for answers to their clinical questions. Simultaneously, trainees sharpen their critical appraisal skills. Walking through the GATE with colleagues and trainees can be an effective way of modeling and supporting self-directed, problem-focused didactic approaches that are effective with adult learners (Kaufman, 2003).

Finally, the GATE may be ideally suited for use by reviewers of submitted manuscripts or grant proposals. Seasoned reviewers are intimately acquainted with study design and methodology and can readily identify and address threats to validity. By contrast, rookie reviewers may benefit from going through GATE exercises repeatedly, until the underpinnings of epidemiology and study design become second nature.

Disclosure: The authors have no financial relationships to disclose.

REFERENCES

- Guyatt GH, Rennie D (2002), *Users’ Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*. Chicago: AMA Press
- Hamilton J (2005), Clinicians’ guide to evidence-based practice. *J Am Acad Child Adolesc Psychiatry* 44:494–498
- Jackson R, Ameratunga S, Broad J et al. (2006), The GATE frame: critical appraisal with pictures. *Evid Based Med* 11:35–38
- Kaufman DM (2003), Applying educational theory in practice. *BMJ* 326:213–216
- March JS, Chrisman A, Breland-Noble A et al. (2005), Using and teaching evidence-based medicine: the Duke University child and adolescent psychiatry model. *Child Adolesc Psychiatr Clin N Am* 14:273–296
- Srihari V, Martin A (2006), Gained in translation: evidence-based medicine meets pediatric psychopharmacology. *Child Adolesc Psychiatr Clin N Am* 15:263–287
- Straus SE, Richardson WS, Glasziou P, Haynes RB (2005), *Evidence-Based Medicine: How to Practice and Teach EBM*, 3rd ed. Edinburgh: Churchill Livingstone
- Treatment for Adolescents With Depression Study (TADS) (2004), Fluoxetine, cognitive-behavioral therapy, and their combination for adolescents with depression: Treatment for Adolescents With Depression Study (TADS) randomized controlled trial. *JAMA* 292:807–820