EBN notebook

The GATE frame: critical appraisal with pictures

INTRODUCTION

Epidemiological evidence about the accuracy of diagnostic tests, the power of prognostic markers, and the efficacy and safety of interventions is the cornerstone of evidence-based health care.[1] Practitioners of evidence-based health care require critical appraisal skills to judge the validity of this evidence. The Evidence-Based Medicine (EBM) Working Group members are international leaders in teaching critical appraisal skills, and their users' guides for appraising the validity of the healthcare literature[2] have long been the basis of teaching programmes worldwide. However, we found that many of our students took a reductionist ''paint by numbers'' approach when using the Working Group's guides. Students could answer individual appraisal questions correctly but would have difficulty assessing overall study quality. We believe this is due to a poor understanding of epidemiological study design. So over the past 15 years of teaching critical appraisal, we have modified the EBM Working Group approach and developed the Graphic Appraisal Tool for Epidemiological studies (GATE) frame to help our students conceptualise the whole study as well as its component parts. GATE is a visual framework that illustrates the generic design of all epidemiological studies (figure 1). We now teach critical appraisal by ''hanging'' studies and the EBM Working Group's appraisal questions on the GATE frame.

This notebook outlines the GATE approach to critical appraisal, illustrated throughout using the Heart and Estrogen/progestin Replacement Study (HERS), a randomised, double blind, placebo controlled trial of the effect of daily oestrogen plus progestin on coronary heart disease (CHD) death in postmenopausal women.[3] A detailed critical appraisal of HERS using a GATE-based checklist is available online.[4]

### HANGING THE STUDY AND NUMBERS ON THE GATE FRAME

The GATE frame incorporates a triangle, circle, square, and arrow (figure 1), labelled with the acronym PECOT (or PICOT).

The triangle (figure 2) represents the population studied: ''P'' for population or participants. We divide the triangle into 3 overlapping levels: (i) the whole triangle represents the source population from which participants were selected; (ii) the lower 2 levels combined represent the eligible population (ie, those who meet the study eligibility criteria); and (iii) the lowest level—the tip of the triangle—represents those who agreed to take part (ie, the study participants). In the HERS study, all 3 levels were well described (figure 2), although the number of people screened who met the eligibility criteria was not provided.

The circle, divided into 2 sections by an interrupted vertical line (figure 3), represents 2 groups of participants being compared in the study population. These are the exposure (E) group, which is often called the intervention (I) group in a trial, and the comparison (C) group. In HERS, 2763 study participants were randomly allocated to either E (hormone replacement therapy [HRT]) or the comparison (identical placebo). To include >2 groups in the circle, add more vertical division lines. For example, some studies may compare 2 doses of a drug [E_1 and E_2] with placebo or alternative therapy [C]).

The study outcomes (O) are represented by a square (figure 4). This is typically divided into 4 sections and is the generic 2×2 table of epidemiological studies with dichotomous exposures (E and C) and dichotomous outcomes (yes and no). The top row (a + b) of the square represents the participants from E and C who experience a
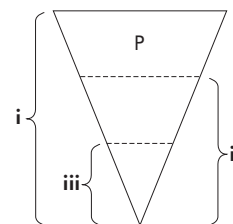
i. Source Population:
from multiple sources.
<u>68 561 screened</u> by phone.

ii. Eligible Population:
post-menopausal, <80 yrs, with CHD. No MI < 6 months or HRT < 3 months TG > 3.39 mmol/l. <u>Number not given.</u>

iii. Participant Population:
All eligibles invited. Of 3463 attending 2nd screen. <u>2763 randomised.</u>

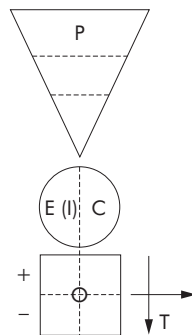**Figure 2** Population (P)—the HERS example.
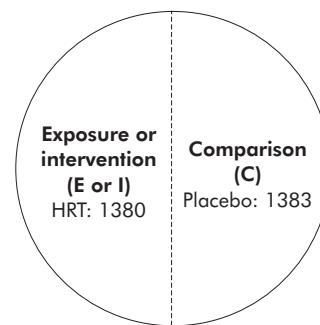
Figure 1 The GATE frame.

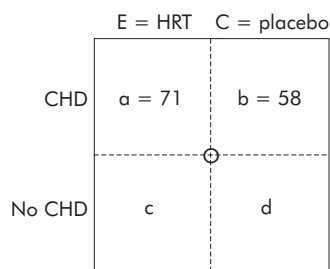**Figure 3** Exposure (E) & comparison (C)—the HERS example.

Figure 4 Outcomes (O)—the HERS example (fatal CHD).

specified study outcome. In HERS, 71 women (a) in the HRT group and 58 women (b) from the placebo group died from CHD during the study follow up period. The bottom row (c + d) represents those participants who did not experience this outcome. Few studies explicitly state the number of participants in c + d, but ideally these data should be given or be possible to calculate. In the HERS study, it is stated that there was 100% follow up for mortality, so it is possible to calculate c (1380 − 71 = 1309) and d (1383 − 58 = 1325). Any number of categorical exposure and outcome groups can be incorporated into the GATE frame by adding additional vertical and horizontal division lines. Outcomes measured continuously (eg, blood lipids in HERS) can be represented by removing the horizontal division line in figure 4 and presenting mean concentrations (eg, mean high density lipoprotein cholesterol = 1.40 mmol/l in the HRT group and 1.27 mmol/l in the placebo group).

Study time (T) is represented by horizontal and vertical arrows (figure 5). A horizontal arrow is used for study outcomes measured at 1 point in time (ie, prevalence or cross-sectional measures) such as the assessment of blood lipids in the HERS study at 1 year after randomisation to HRT or placebo. A vertical arrow is used to describe outcomes measured over a period of time (ie, incidence or longitudinal measures). For example, the measurement of CHD events is over an average of 4.1 years of follow up in HERS.

FRAMING VALIDITY QUESTIONS WITH GATE

After hanging a study on the GATE frame (figures 2–5), appraisers should have a good understanding of what question the study addressed and how the investigators addressed it. Appraisers should have documented the characteristics of participants (including the source population and eligibility criteria), the exposure and comparison definitions, the outcome criteria, and the time period at or over which outcomes were measured. In addition, the numbers of people included, excluded, and lost to follow up at each phase of the study should have been annotated on the
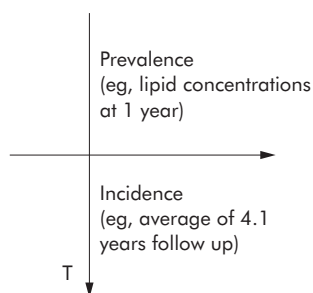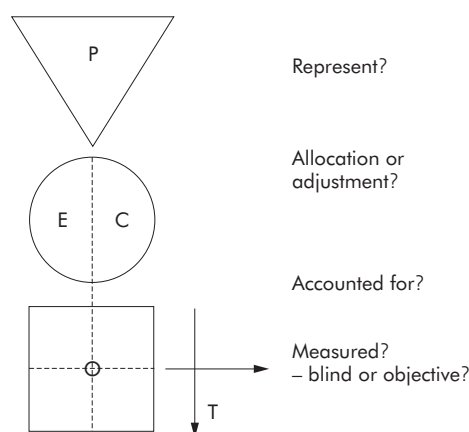

Figure 5 Study time (T)—the HERS example.

Figure 6 Study validity (RAAMbo).

GATE frame. Appraisers should now be well prepared to appraise the study for its validity. Our approach involves rearranging versions of the EBM Working Group's user guides questions[2] on the GATE frame. Only the main validity issues are discussed here; more detail is available from online GATE checklists.[4]

We link the acronym **RAAMbo** to the GATE frame (figure 6) to help appraisers address the key validity issues in epidemiological studies.

A study report should provide sufficient detail to allow the appraiser to determine whom the participants **REPRESENT**. This requires information on the 3 levels outlined in figure 2 (ie, source population, eligible population, and participant population). Representativeness is more important for some questions (eg, prognosis) than others (eg, relative treatment effects) and is the key criterion for determining the external validity or generalisability of study findings.

The method of **ALLOCATION** to exposure and comparison groups is particularly important for intervention studies. Randomised allocation is the best way to avoid imbalances between the groups that may influence the occurrence of outcomes (known as confounding or a "mixing of effects"). In non-randomised studies, influence of imbalances between the exposure and comparison groups can be reduced by **ADJUSTMENT**. This is typically done by stratification of the groups being compared into subgroups (eg, dividing each of the exposure and comparison groups into subgroups of smokers and non-smokers) or by using multivariate statistical methods.

All participants should be **ACCOUNTED** for at the completion of a study, and the numbers in the tip of the triangle (study participants) should equal the numbers in the circle (exposure and comparison groups), which should in turn equal the numbers in the square (those with and without the specified study outcome). Also, in good quality studies, a high proportion of participants remain in the exposure (or comparison) group to which they were initially allocated, with high compliance (most remain on allocated exposure), low contamination (most do not receive other exposures), and low loss to follow up. However, contamination, reduced compliance, and loss to follow up are difficult to eliminate entirely, and if the degree differs between the exposure and comparison groups, it can be an important source of bias (ie, a differential error). Blinding of participants and others associated with participants to exposure status is an effective method of reducing differential errors.

The other major validity issue to address in epidemiological studies is the accuracy of outcomes **MEASURED**. As most outcome assessments are to some extent subjective, there is potential for error in their measurement. As discussed above, **BLINDING** of participants and study staff to exposure status reduces differential errors. Also, the more **OBJECTIVE** the outcome measure (eg, all cause mortality, automatic test, standardised measurements, or strict diagnostic algorithms), the less likely there will be a differential or non-differential error in measurement. So generally, outcome measures should be blinded or objective.

When the RAAMbo appraisal criteria suggest (as usual!) some flaws in the study design or conduct, we need to make a judgment on the study's validity. This requires an assessment of the likely net impact of the flaws. We recommend that the appraiser consider the direction and degree of impact each flaw will have on the study numbers discussed in the previous section and whether the combined impact of the flaws is likely to substantially change the overall effect estimates discussed in the next section. We find that visualising the potential combined impact of these flaws using the GATE frame facilitates the process of making a judgment on the overall quality of the study.

CALCULATING OCCURRENCE AND EFFECT ESTIMATES IN THE GATE FRAME

Occurrence estimates

All epidemiological studies are designed for 1 task: to calculate the occurrence (or ''risk'') of health related outcomes in populations. There are 2 measures of occurrence: the incidence of health related events and the prevalence of health related states. Occurrence is calculated by measuring specified health outcomes in a population (a, b, c, or d in the GATE square) and dividing by the number of persons in that population (exposure or comparison group in the GATE circle).

Incidence measures of occurrence count the number of health related events (eg, heart attacks) that occur over the study time period, with the time period indicated by a vertical arrow in GATE. Prevalence measures of occurrence count the number of persons with a defined health status (eg, diabetes) at 1 point in time, indicated by the horizontal arrow in GATE.

If the appropriate numbers for exposure (E), comparison (C), a, b, c, d, and time (figures 3–5) are keyed into GATE MS Excel checklists,[4] which have embedded calculators, the exposure group occurrence (EGO) and comparison group occurrence (CGO) (generic versions of the terms experimental event rate [EER] and control event rate [CER] used for intervention studies) are automatically calculated, as illustrated in boxes 1 and 2. Although occurrence of outcomes was calculated using a and b as the relevant outcomes (eg, CHD deaths), some analyses (eg, survival analyses and negative likelihood ratios) calculate the occurrence based on those who do not have the study outcome (ie, c and d).

Effect estimates

Measures of occurrence (or risk) in the exposure and comparison groups are compared to assess the ''effect'' of the exposure (compared with the comparison) on outcomes. The standard measures of effect are risk ratios (eg, relative risks, likelihood ratios, and odds ratios), risk difference or absolute risk difference (eg, absolute risk reduction [ARR] or increase [ARI]), and numbers needed to treat (NNT) (or generically, numbers needed to expose) as shown in box 3.
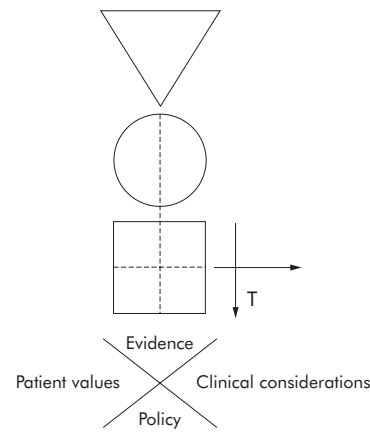

Figure 7 Applying the evidence.

The online GATE checklists automatically calculate these effect estimates and the associated 95% confidence intervals.[4]

FRAMING THE STEPS OF EVIDENCE-BASED PRACTICE WITH GATE

Critically Appraised Topics (CATs)[5] are tools for modelling the 5 steps of evidence-based practice,[6] and our online GATE checklists[4] are designed to document these steps. We frame the first 4 steps using GATE. Step 1 involves ''*asking a focused question,*'' and as there are 5 components to most epidemiological studies (ie, PECOT or PICOT), there are 5 components to a question addressing epidemiological evidence. Similarly, when ''*accessing evidence*'' (step 2 of evidence-based practice), the key search terms can be framed by the same 5 components, although typically search terms only use combinations of the P, E, and O components. Step 3 (*critical appraisal*) has been discussed in detail above.

The X below the GATE frame in figure 7 illustrates the fourth step of evidence-based practice, ''*the application of evidence in practice.*'' We call this the ''X factor'' or ''eXpertise factor'' because an expert practitioner is one who can integrate the evidence with the other key issues (ie, patient values, clinical considerations—ranging from comorbid conditions to patient circumstances—and policy issues) that

---

Box 1. Calculating incidence measures of occurrence in the GATE frame

Incidence = (number of people *developing* an outcome ÷ the number in a population) *during* time T
Exposure group occurrence [EGO] or experimental event rate [EER]
$= a \div E$ during time T
$= a \div (a+c)$ during time T
$= (a \div (a+c))/T$
Example: incidence of CHD death in HRT group from HERS[3]
EGO (or EER)
= [71 events ÷ 1380 women] over 4.1 years
= 51.5 events per 1000 women over 4.1 years
= 12.5 events per 1000 women per year
Comparison group occurrence [CGO] or control event rate [CER]
$= b \div C$ during time T
$= b \div (b+d)$ during time T
$= (b \div (b+d))/T$
Example: incidence of CHD death in placebo group from HERS[3]
CGO (or CER)
= [58 events ÷ 1383 women] over 4.1 years
= 41.9 events per 1000 women over 4.1 years
= 10.2 events per 1000 women per year

Box 2. Calculating prevalence measures of occurrence in the GATE frame

Prevalence = (number of people with an outcome ÷ the number in a population) at time T

Exposure group occurrence [EGO] or experimental event rate [EER]
= a ÷ E at time T
= a ÷ (a+c) at time T

Example: prevalence of smoking in HRT group at time of randomisation
EGO (or EER) = 13 smokers per 100 women (13%)*

Example: average HDL cholesterol concentrations in HRT group at 1 year
EGO = sum of HDL cholesterol concentrations in women in HRT group
÷ the number of women in HRT group
= 1.40 mmol/l*

Comparison group occurrence [CGO] or control event rate [CER]
= b ÷ C at time T
= b ÷ (b+d) at time T

Example: prevalence of smoking in placebo group at time of randomisation
CGO (or CER) = 13 smokers per 100 women (13%)*

Example: mean HDL cholesterol concentrations in placebo group at 1 year
CGO = sum of HDL cholesterol concentrations in women in placebo group
÷ the number of women in placebo group
= 1.27 mmol/l*

*Only percentages and means presented in HERS paper.[3]

Box 3. Calculating measures of ''effect'' in the GATE frame

Risk ratio or relative risk [RR]
= the ratio of occurrence* (risk) in the exposure group to the occurrence in the comparison group
= EGO ÷ CGO (or EER ÷ CER in intervention studies)

Example: the occurrence (risk) of CHD death in the HRT group relative to the risk in the placebo group at the conclusion of HERS

RR = 12.5 ÷ 10.2 events per 1000 women per year (from box 1)
= 1.23 (95% CI 0.87 to 1.74)†

Risk difference [RD] or absolute risk difference [ARD]
= the difference in occurrence* (or risk) between the exposure group and the comparison group
= EGO − CGO or
= EER − CER

Example: the difference in occurrence between the HRT treated group and the comparison group in HERS

RD = 12.5 − 10.2 events per 1000 women per year (from box 1)
= 2.3 events per 1000 women per year (95% CI −1.61 to 6.25)†

Number needed to expose [NNE] (or number needed to treat [NNT]) to reduce (or increase) events by 1 during time (T)
= 1 ÷ (EGO − CGO) or
= 1 ÷ (EER − CER)

Example: the number needed to be treated with HRT for 1 year to cause 1 additional CHD death in HERS

NNE or NNT‡
= 1 ÷ 2.3 events per 1000 women per year
= 435 women for 1 year (95% CI 160 to ∞ to −620)

ie, 435 women in HERS needed to be treated with HRT for 1 year to cause 1 additional CHD death

---

*Incidence or prevalence.
†The GATE checklists[4] calculate 95% CIs for effect estimates.
‡NNE is the generic form of NNT (numbers needed to treat).
§When the CI for the risk difference crosses 0, the CI for an NNT (or NNE) passes through infinity.[7]

---

must be considered when making good healthcare decisions. (Our students suggested we needed an X so we would have all 4 symbols used in a PlayStation® game [triangle, circle, square, and cross]. We thank Chris Hoffman, an orthopaedic surgeon from Wellington, New Zealand, for suggesting how to use an X in the GATE frame.).

CONCLUSIONS

The GATE frame is a graphic representation of the generic structure of all epidemiological studies. We have found that hanging studies on the GATE frame helps students understand epidemiology and can facilitate the critical appraisal of epidemiological studies, especially making overall judgments about study quality. There is only 1 epidemiological study design. The ''different'' designs described in the epidemiological literature are simply variations on this generic design. When you understand the GATE frame, you will understand basic epidemiology.

ROD JACKSON, MBChB, PhD
SHANTHI AMERATUNGA, MBChB, PhD
JOANNA BROAD, MPH
JENNIE CONNOR, MBChB, PhD
ANNE LETHABY, MA
GILL ROBB, MPH
SUE WELLS, MBChB, MPH
*University of Auckland
Auckland, New Zealand*

PAUL GLASZIOU, MBBS, PhD
CARL HENEGHAN, BM BCH
*Centre for Evidence-Based Medicine
Oxford, UK*

1 Sackett DL, Rosenberg WM, Gray JA, *et al.* Evidence based medicine: what it is and what it isn't. *BMJ* 1996;**312**:71–2.
2 http://www.cche.net/usersguides/main.asp (accessed 19 November 2005).
3 Hulley S, Grady D, Bush T, *et al.* Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. Heart and Estrogen/progestin Replacement Study (HERS) Research Group. *JAMA* 1998;**280**:605–13.
4 http://www.epiq.co.nz (accessed 19 November 2005).
5 Sackett DL, Straus S, Richardson WS, *et al. Evidence-based medicine: how to practice and teach EBM*, 2nd edition. Edinburgh: Churchill Livingstone, 2000:87.
6 Sackett DL, Straus S, Richardson WS, *et al. Evidence-based medicine: how to practice and teach EBM*, 2nd edition. Edinburgh: Churchill Livingstone, 2000:3.
7 Altman DG. Confidence intervals for the number needed to treat. *BMJ* 1998;**317**:1309–12.

This notebook also appears in Evidence-Based Medicine and ACP Journal Club.