

GATE frame: ピクチャによる批判的吟味

診断検査の検査精度、予後マーカーの検出力、そして介入の効能および安全性に関する疫学的エビデンスは、エビデンスに基づく医療の基礎である(1)。エビデンスに基づく医療の実践者には、このエビデンスの妥当性を判断するための批判的吟味のスキルが求められる。Evidence-Based Medicine (EBM) Working Group のメンバーは、批判的吟味のスキルの伝授における国際的な指導者であり、Working Group によって発行される医学文献の妥当性評価のためのユーザーガイド(2)は、世界中で長年にわたり教育プログラムの基礎として位置づけられてきた。しかしながら、Working Group のユーザーガイドを使用する学生の多くが単純化志向で、“指示通りにやる(paint by number)”といったアプローチをとっているにすぎない。こういった学生は、吟味に関する質問に正確に答えることはできても、全体的な研究の質の評価となると行き詰まってしまう。これは、疫学的研究デザインの理解が不足していることに起因すると考えられる。

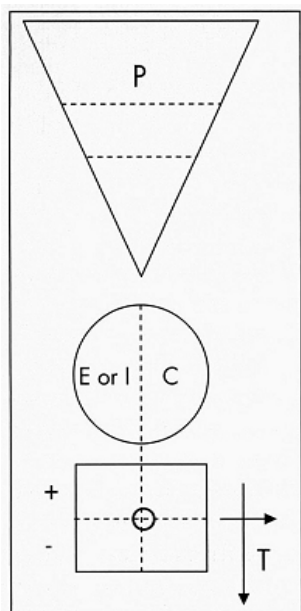


Figure 1. The GATE frame. P = population; E or I = exposure or intervention; C = comparison; O = outcome; T = study time.

われわれは、過去15年間に及ぶ批判的吟味の指導を踏まえ、学生が研究の各構成要素だけではなく研究全体を概念化できるよう支援するためにEBM Working Group のアプローチを修正して **Graphic Appraisal Tool for Epidemiological studies (GATE)**を開発した。GATE はあらゆる疫学研究(図1)の全体的なデザインを図解する視覚的枠組みである。こうしてわれわれは、批判的吟味について指導するときは、まず、研究ならびにEBM Working Group の評価クエスチョンを GATE frame に“掲示(hanging)”するようにしている。

この論説では、批判的吟味のための GATE アプローチを概略し、そのための例として、エストロゲンおよびプロゲステンの連日服用が閉経後の女性の冠動脈疾患(CHD)による死亡に与える影響を調べた盲検ランダム化プラセボ比較試験である Heart and Estrogen/progestin Replacement Study (HERS)を取り上げた(3)。GATE ベースのチェックリストを活用した HERS の批判的吟味の詳細説明については、オンラインで閲覧可能である(4)。

GATE frame に研究および数値を掲示する

GATE frame には三角形、円、四角形そして矢印が表示され(図1)、それぞれに頭文字からなる PECOT (または PICOT) のラベルが貼付されている。

三角形(図2)は population(対象集団)(P)を指す。この三角形は、3つの重複したレベルに分割される。1) 三角形全体は、選択された参加者が所属していた集団を指す。2) その下位にある2つのレベルは適格集団(すなわち、研究の適格基準に合致した人々)を指し、3) 最下位のレベル(三角形の先端)は参加に同意した人々(研究参加者)を指す。HERS では、3つのレベル全てについて良く記載されていたが(図2)、スクリーニングされた人々のうち適格基準を満たした人の数については説明がなかった。

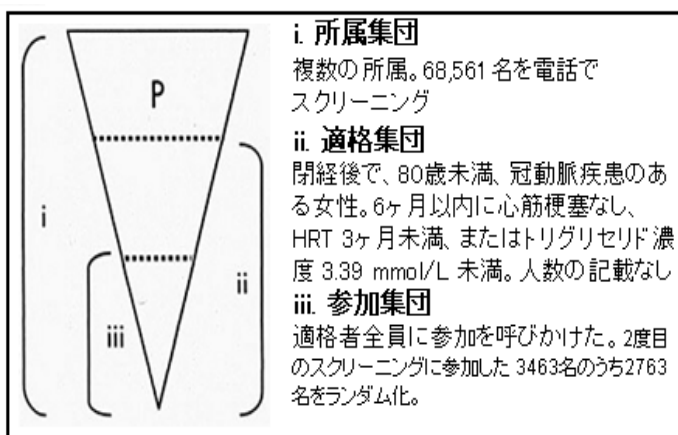


Figure 2. 集団 HERS 例
HRT = ホルモン補充療法

破線状の縦線で2つに区切られた円(図3)は、研究対象集団において比較対象となっている2つの集団を指す。この2つの集団とはすなわち曝露(E)群(試験の枠組みでは介入(I)群と呼ばれることが多い)と対照群(C)のことを指す。

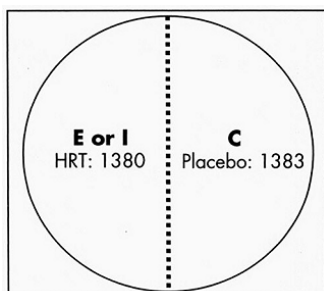


Figure 3. Exposure and comparison: The HERS example.

HERS では、2763 名の研究参加者が無作為に曝露(E)(ホルモン補充療法 [HRT]) (n = 1380) または対照(C)(全く見分けのつかないプラセボ) (n = 1383) に割り付けられている。

この円の中に2つ以上の集団を示す場合は、円を区切るための破線状の縦線を付け加えるといふ。たとえば、ある薬剤の2種類の服用量を(E₁ および E₂)、プラセボまたは代替療法(C)と比較する研究もあるだろう。

研究のアウトカム(O)は四角形によって示される(図4)。この四角形は通常4つのセクションに分割されており、二値曝露(E および C)と二値アウトカム(「はい」または「いいえ」)を有する疫学研究を表した一般的な 2 x 2 表である。GATE frame には、新たに縦線および横線で区切ることによってカテゴリカルな曝露とアウトカムのグループをいくつでも含めることができる。連続的に測定したアウトカムについては(HERS における血中脂質など)、図4の横向きの分割線を取り除き、平均値(HDL コレステロールの平均値が HRT 群では 1.40 mmol/L、プラセボ群では 1.27 mmol/L)を示すとよい。

四角形の上の行(a + b)は、所定の研究アウトカムを経験するEおよびCに属する参加者を示す。HERS では、HRT 群における71名の女性(a)、およびプラセボ群に属する58名の女性(b)が研究のフォローアップ期間中にCHDが原因で死亡している。四角形の下行(c + d)は、これらのアウトカムを経験しなかった参加者を示している。c + dに属する参加者の人数を明示的に示している研究は少ないものの、これらのデータが提供されているか、もしくは計算が可能であることが望ましい。HERS では、死亡率が100%フォローアップされているという記載があることから、c(1380 - 71 = 1309)およびd(1383 - 58 = 1325)を算出することが可能である。

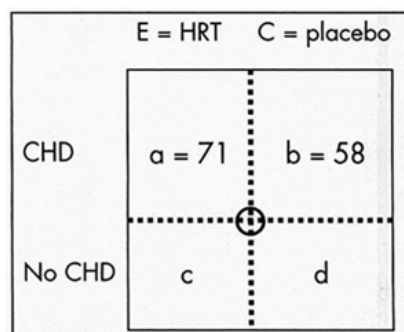


図4. アウトカム: HERS の例
(致命的な冠動脈疾患 [CHD])

研究における時間(T)は横向きの矢印と縦向きの矢印によって示される(図5)。

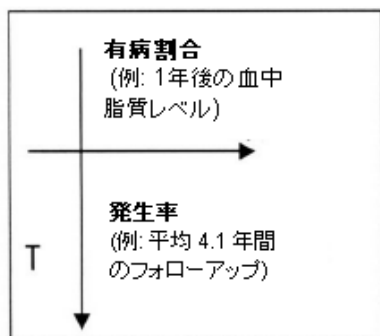


Figure 5. Study time: the HERS example.

横向きの矢印は、HERS 研究において HRT またはプラセボに無作為に割付けられてから1年後の血中脂質を評価する場合などのように、ある一時点で測定された研究アウトカム用の矢印である(有病割合または横断的指標)。

縦向きの矢印はある一定の期間にわたって測定されたアウトカム用の矢印である(発生率または縦断的指標)。たとえば、HERS では平均して4.1年間に及ぶフォローアップの中でCHD イベントの測定が行われている。

GATE を利用した妥当性クエスチオンの形成

GATE frame に研究を掲示することで(図2~5)、評価者は研究において取り上げられているクエスションや、そのクエスションに対し研究者がどういった対応を行ったのかについて理解を深めることができるはずである。評価者は、この時点で参加者の特徴(参加者の所属集団および適格基準を含む)、曝露と対照の定義、アウトカムの基準、そしてアウトカムが測定された時期または期間について記録することが望ましい。さらに、研究の各段階で組み込まれた人の数、除外された人の数、ならびにフォローアップから逸脱した人の数を、注釈として GATE frame に書き込んでおくべきである。これで評価者は研究の妥当性を吟味する準備が整ったことになる。われわれのアプローチでは、EBM Working Group のユーザーガイドに示されるクエスション(2)を GATE frame 上で再編成する。なお、ここでは、妥当性に関わる主要な問題についてのみ説明し、さらなる詳細はオンライン上の GATE チェックリストから入手可能である(4)。

疫学研究における、妥当性に関わる主要な問題について検討するには、まず頭文字である RAAMbo (Represent (代表)、Allocation (割付け)、Adjustment (調整)、Accounted (説明)、Measured (測定)、Blinding (盲検化)、Objective (客観的))と GATE frame の関連付けを行う(図6)。研究報告は、評価者にとって参加者がどういった集団を Represent (代表)しているのかがわかるように十分な情報が提供されていることが望ましい。そのためには、図2に示した3つのレベル(所属集団、適格集団、および参加者集団)に関する情報が必要となってくる。代表性の重要性は、クエスションによって左右され(例えば、予後では重要だが、相対的治療効果においてはそれほど重要ではない)、研究結果の外的妥当性または一般化可能性を評価する上での重要な基準でもある。

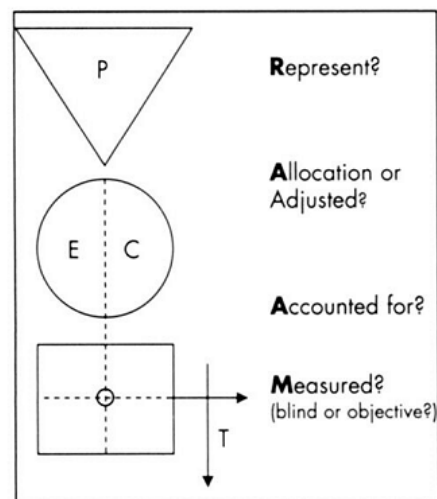


Figure 6. Study validity: RAAMbo.

介入研究においては、曝露群または対照群への割付け(Allocation)の方法が特に重要となってくる。曝露群と対照群との間でアウトカムの発生に影響を及ぼすような不均衡(交絡または「効果の混同」ともいう)が生じないようにするためには、ランダム化割付けの実施が最善策となる。非ランダム化試験における曝露群と対照群の間の不均衡の影響は、調整(Adjustment)によって緩和することができる。そのためには、比較の対象となっているグループをサブグループに層別化するか(曝露群と対照群とをそれぞれ喫煙者と非喫煙者のサブグループに分割するなど)、または多変量統計手法を使うのが一般的である。

研究の終了時点で全参加者について説明がつく (Accounted) ようになっているべきである。また、三角形の先端に属する人数 (研究参加者) が、円の中の人数 (曝露群および対照群) と等しく、円の中の人数は四角形の人 (所定の研究アウトカムが確認された人と確認されなかった人) と等しくなっていることが望ましい。さらに、質の高い研究においては、参加者の大部分が最初に割り付けられていた曝露群 (または対照群) にとどまり、コンプライアンスも良好で (大多数が割り付けられた曝露にとどまる)、コンタミネーションが少なく (他の曝露にさらされる例がほとんどない)、フォローアップからの脱落も少ない。しかし、コンタミネーション、コンプライアンス不良、およびフォローアップからの脱落を完全に排除することは難しく、その度合いが曝露群と対照群との間で異なっている場合は、これがバイアスの重要な原因 (差分誤差) となることがある。参加者および参加者と関係のある人たちに曝露ステータスについてブラインド化することが差分誤差を軽減する効果的な方法であるといえる。

疫学研究において検討すべき妥当性に関わるもう一つの重要な問題としては、Measured (測定) されたアウトカムの正確さがあげられる。アウトカムの評価は多かれ少なかれ主観的な側面があるため、測定に誤差が生じる可能性がある。前述のように、参加者および研究スタッフの曝露ステータスについてのブラインド化 (Blinding) が差分誤差を軽減できる。また、アウトカム指標が Objective (客観的) であればあるほど (あらゆる原因による死亡率、自動テスト、標準化された測定手法、または厳密な診断アルゴリズム)、測定において差分誤差や非差分誤差が生じる可能性が低くなる。そのため、一般的にアウトカム指標はブラインド化されているか、もしくは客観的である必要がある。

RAAMbo 評価基準によって (例のごとく!) 研究のデザインまたは実施になんらかの欠陥があることが明らかになった場合は、研究の妥当性に関する判断を下す必要がある。そのためには、その欠陥にどのような正味インパクトがありそうなものを評価する必要がある。このとき、評価者は、各欠陥が前項で説明した研究の数値に対し、どういった方向性や程度のインパクトを持ち、欠陥全体としてのインパクトが、次項で説明する全体的な推定効果を大きく左右するかどうかについて考慮するとよいだろう。GATE frame を使ってこのような潜在的欠陥の全体としてのインパクトを視覚化することが、研究の全体的な質を評価しやすくなると考えられる。

GATE frame において発現の推定値および推定効果を計算する

発現の推定値

いかなる疫学研究でも、ある1つのタスクを遂行できるようなデザインとなっている。それはすなわち、集団における健康関連アウトカムの発現 (またはリスク) の算出である。発現の指標としては、健康関連イベントの発生率と、ある健康関連状態にある人の割合という2通りの指標がある。発現を算出するには、集団における所定の健康アウトカムを測定し (GATE の四角形における a、b、c、または d)、それを当該集団の人数 (GATE の円の曝露群または対照群) で割る。

発現を発生率として示す場合は、研究期間中に発生した健康関連イベント (心臓発作など) の件数をカウントする。このとき、時間は GATE の縦向きの矢印で示される。発現を有病割合として示す場合は、ある1時点において特定の健康状態 (糖尿病など) を有する人々の数をカウントする。この1時点は GATE の横向きの矢印で表される。

曝露、対照、a、b、c、d、および時間を示す適切な数値を (図3~5) 計算機能が組み込まれた GATE Microsoft Excel チェックリストにキー入力すると (4)、Box .1 および Box. 2 に示されるように、曝露群での発現 (EGO) ならびに対照群での発現 (CGO) が自動計算される (EGO および CGO は、介入研究という実験群でのイベント発生率 (EER) および対照群でのイベント発生率 (CER) の一般名称である)。アウトカムの発現は、関連アウトカムとして a および b を用いて算出されたものだが (例、CHD による死亡)、分析によっては (例、生存率分析や陰性尤度比)、研究アウトカムを有さない人々の数 (即ち、c および d) に基づいて発現が算出されることがある。

発生率 = 研究期間(T)中の(アウトカムを発症した人の数 ÷ 集団における数)

曝露群での発現(EGO)または実験群でのイベント発生率(EER)

$$\begin{aligned} &= a \div T \text{ 中の } E \\ &= a \div T \text{ 中の } (a + c) \\ &= (a \div [a + c]) / T \end{aligned}$$

例: HERS の HRT 群における CHD による死亡の発生率

$$\begin{aligned} \text{EGO (または EER)} \\ &= 4.1 \text{ 年間で } (71 \text{ イベント} \div \text{女性 } 1380 \text{ 名}) \\ &= 4.1 \text{ 年間で、女性 } 1000 \text{ 名あたりのイベント } 51.5 \text{ 件} \\ &= 1 \text{ 年につき 女性 } 1000 \text{ 名あたりのイベント } 12.5 \text{ 件} \end{aligned}$$

対照群での発現(CGO)または対照群でのイベント発生率(CER)

$$\begin{aligned} &= b \div T \text{ 中の } C \\ &= b \div T \text{ 中の } (b + d) \\ &= (b \div [b + d]) / T \end{aligned}$$

例: HERS のプラセボ群における CHD による死亡の発生率

$$\begin{aligned} \text{CGO (または CER)} \\ &= 4.1 \text{ 年間で } (58 \text{ イベント} \div \text{女性 } 1383 \text{ 名}) \\ &= 4.1 \text{ 年間で、女性 } 1000 \text{ 名あたりのイベント } 41.9 \text{ 件} \\ &= 1 \text{ 年につき 女性 } 1000 \text{ 名あたりのイベント } 10.2 \text{ 件} \end{aligned}$$

Box 1. GATE frame において発現を発生率として算出する場合

有病割合 = (T) 時点での(アウトカムを有する人々の数 ÷ 集団における人数)

曝露群での発現(EGO)または実験群でのイベント発生率(EER)

$$\begin{aligned} &= a \div T \text{ 時点での } E \\ &= a \div T \text{ 時点での } (a + c) \end{aligned}$$

例: ランダム化割付け時点での、HRT 群における喫煙者の割合

$$\text{EGO} = \text{女性 } 100 \text{ 名中喫煙者 } 13 \text{ 名 (13\%)}$$

例: HRT 群における 1 年後の平均 HDL コレステロール値

$$\begin{aligned} \text{EGO} &= \text{HRT 群の女性の HDL コレステロール値の合計} \\ &\div \text{HRT 群の女性の人数} \\ &= 1.40 \text{ mmol/L} * \end{aligned}$$

対照群での発現(CGO)または対照群でのイベント発生率(CER)

$$\begin{aligned} &= b \div T \text{ 時点での } C \\ &= b \div T \text{ 時点での } (b + d) \end{aligned}$$

例: ランダム化割付け時点での、プラセボ群における喫煙者の割合

$$\text{CGO} = \text{女性 } 100 \text{ 名中喫煙者 } 13 \text{ 名 (13\%)}$$

例: プラセボ群における 1 年後の平均 HDL コレステロール値

$$\begin{aligned} \text{CGO} &= \text{プラセボ群の女性の HDL コレステロール値の合計} \\ &\div \text{プラセボ群の女性の人数} \\ &= 1.27 \text{ mmol/L} * \end{aligned}$$

* HERS の論文ではパーセンテージと平均値のみが提示されていた(3)。

Box 2. GATE frame において発現を有病割合として算出する場合

推定効果

曝露群と対照群における発現(またはリスク)の指標を比較することによって、曝露がアウトカムに与える影響を(対照群と比較しながら)評価する。効果を表す標準的指標としては、Box 3 に示すようにリスク比(相対リスク、尤度比、オッズ比)、リスク差または絶対リスク差(絶対リスク減少[ARR]または増加[ARI])、そして治療必要数(NNT)(曝露必要数ともいう)があげられる。オンラインの GATE チェックリストでは、こういった推定効果や、これに関連する 95%信頼区間が自動計算される(4)。

リスク比または相対リスク(RR)

= 曝露群における発現 *(リスク)と対照群における発現の比
= EGO ÷ CGO(または、介入研究の場合は EER ÷ CER)

例: HRT 群における CHD による死亡の発現(リスク)と HERS の終了時点でのプラセボ群におけるリスクの比
RR = 12.5 ÷ 10.2 件(女性 1000 名あたりの 1 年間のイベント発生件数)(Box 1 より)
= 12.3(95% CI 0.87~1.74) †

リスク差(RD)または絶対リスク差(ARD)

= 曝露群と対照群の間の発現 *(リスク)の差
= EGO - CGO または
= EER - CER

例: HERS における HRT 群と対照群との間の発現の差
RD = 12.5 - 10.2 件(女性 1000 名あたりの 1 年間のイベント発生件数)(Box 1 より)
= 2.3 件(女性 1000 名あたりの 1 年間のイベント発生件数)(CI 1.61~6.25) †

研究期間中(T)にイベント発生件数を 1 件減らす(または増やす)ための曝露必要数(NNE)または治療必要数(NNT)

= 1 ÷ (EGO - CGO) または
= 1 ÷ (EER - CER)

例: HERS において CHD による死亡率を 1 件増やすための 1 年あたりの HRT による治療必要数(NNT)。
NNE または NNT ‡
= 1 ÷ 2.3 件(女性 1000 名あたりの 1 年間のイベント発生件数)
= 1 年あたり女性 435 名(CI 160~∞~620) §

(CHD による死亡率を 1 件増やすのに、HERS における 435 名の女性に対し 1 年間にわたって HRT による治療を行う必要があった)

* 発生率または有病割合

† GATE チェックリストは、推定効果の 95% CI を算出する。

‡ NNE とは、NNT(治療必要数)のジェネリック名称である。

§ リスク差の CI が 0 をまたがる場合は、NNT(または NNE)の CI は無限を通過する(7)。

Box 3. GATE frame にて効果の指標を算出する場合

GATE を使用してエビデンスに基づく診療のためのステップを形成する

Critically appraised topics (CAT) (5) は、エビデンスに基づく診療の 5 つのステップをモデル化するためのツールであり、オンライン GATE チェックリスト(4) はこれらのステップを記録するためのものである。最初の 4 つのステップは GATE を使って形成する。5 つめのステップ(ステップ 1~4 を実行するユーザーのスキルを評価するステップ)は「メタステップ(meta-step)」であると考えられる。GATE frame による CAT において、われわれはユーザーの診療行為を評価するためのステップ 5b を追加した。

ステップ 1 では、“特定事項を尋ねるクエスチョン(asking a focused question)”が提起される。ほとんどの疫学研究は 5 つの構成要素(PECOT または PICOT)を持つため、疫学的エビデンスに関するクエスチ

ンもまた5つの構成要素からなる。同様に、“エビデンスにアクセス (*accessing evidence*)”する際も(エビデンスに基づく診療のステップ2) 同じく5つの構成要素から主要検索用語を形成することができるが、検索用語の多くは5つの構成要素のうち P、E、O の組み合わせからなる。ステップ3 “批判的吟味 (*critical appraisal*)”については上記にてすでに詳細を説明済みである。

図7の GATE frame の下にある ×印はエビデンスに基づく診療の4つ目のステップである“実践でのエビデンスの適用 (*the application of evidence in practice*)”を示す。これを、「x-ファクター」、または「専門ファクター」と呼ぶ。というのも、医療に関する優れた意思決定を行う際に考慮しなければならない他の重要事項(患者の価値観、合併症に始まり患者の置かれている状況に至るまでの臨床的配慮、ならびに政策上の問題)とエビデンスを統合することができるのは医療を実践する専門家に他ならないからである。(GATE frame 内での x-ファクターの使用法について示唆してくれたニュージーランド Wellomgton の整形外科医 Chris Hoffman 氏に謝意を表す。これはそもそもわれわれの学生が、プレイステーションで使用されているシンボル全て(三角形、円、四角形、十字)を使用するには x が必要であると提案したものである。)

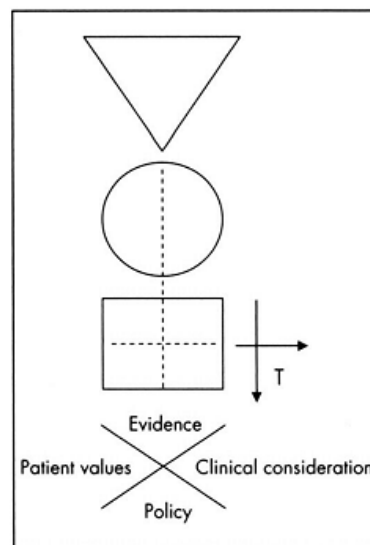


Figure 7. Applying the Evidence.

結論

GATE frame は、あらゆる疫学研究の一般的構造をグラフィック化したものである。GATE frame に研究を掲示することが疫学に関する学生の理解を促し、疫学研究の批判的吟味、特に研究の質についての総合的判断の円滑化につながる事が確認されている。疫学研究デザインは1つしか存在しない。疫学文献において異なるデザインの説明があったとしても、それはこのジェネリックデザインのバリエーションに過ぎない。GATE frame の理解は、基礎疫学の理解を意味する。

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, England, UK

References

1. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ*. 1996;312:71-2. 8555924
2. 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. 9718051
4. www.epiq.co.nz (accessed 19 November 2005).
5. Sackett DL, Straus S, Richardson WS, Rosenberg W, Haynes RB. Evidence-Based Medicine: How to Practice and Teach EBM. 2d ed. Edinburgh: Churchill Livingstone; 2000:87.
6. Sackett DL, Straus S, Richardson WS, Rosenberg W, Haynes RB. Evidence-Based Medicine: How to Practice and Teach

EBM. 2d ed. Edinburgh: Churchill Livingstone; 2000:3.

7. Altman DG. Confidence intervals for the number needed to treat. BMJ. 1998;317:1309-2. 9804726