CHAPTER 2
ACCESSING, CRITIQUING AND READING RESEARCH LITERATURE

Learning Objectives
By the end of this chapter, the participant will:

1. Define evidence-based medicine and differentiate among various research methods.
2. Define the terms "meta-analysis," "confidence intervals" and "odds ratio,"
3. Interpret an odds ratio graph.

When we provide maternity care, we try to do the most good for the women we care for with the least risk and lowest cost. When we do not have the complete information to make clinical decisions, our care is hampered. When we have complete information but we do not use it properly or consistently, inappropriate care may be provided. **The goal of the ALARM International Program is to promote care based on the best available research evidence** and to encourage participants to develop skills in obtaining, evaluating and incorporating evidence into daily clinical practice.

> "Half of what you are taught as medical students will in 10 years have been shown to be wrong. And the trouble is, none of your teachers knows which half."
> Dr. Sydney Burwell, Dean of Harvard Medical School

What is Evidence-Based Medicine?

Evidence-based medicine (EBM) or evidence-based practice is the integration of the best research evidence with clinical expertise and patient values:

- By best research evidence, we mean clinically relevant research, from the basic sciences or especially from patient-centered clinical research. This research may focus on the accuracy and precision of diagnostic tests (including the clinical examination), the power of prognostic markers, or the efficacy and safety of therapeutic, rehabilitative and preventive regimens. New evidence from clinical research may validate or invalidate previously accepted diagnostic tests and treatments. It may introduce new tests or treatments that are more powerful, more accurate, more efficacious and safer.
- By clinical expertise, we mean the ability to use our clinical skills and past experience to rapidly identify each patient's unique state of health and form a diagnosis, while taking into consideration risks and benefits of potential interventions for each individual.
- By patient values we mean the unique preferences, concerns and expectations each patient brings to a clinical encounter and that must be integrated into clinical decisions if they are to serve the patient.

When these three elements are integrated, clinicians and patients form a diagnostic and therapeutic alliance that optimizes clinical outcomes and quality of life (Straus et al, 2000; p.1).

How do we actually practice EBM?

The complete practice of EBM comprises five steps (Straus et al, 2000; p. 3-4):

Step 1 – converting the need for information (about prevention, diagnosis, prognosis, therapy, causation, etc.) into an answerable question
Step 2 – tracking down the best evidence with which to answer that question
Step 3 – critically appraising that evidence for its validity (closeness to the truth), impact (size of the effect) and applicability (usefulness in our clinical practice)
Step 4 – integrating the best evidence with our clinical expertise and with our patient's unique biology, values and circumstances

Step 5 – evaluating our effectiveness and efficiency in executing steps 1– 4 and seeking ways to improve them both for next time

Step 1 - Asking answerable clinical questions

To obtain the evidence we need to practice effectively, we have to ask an answerable question. Let us consider a clinical scenario (example):

A 20-year-old primigravida presents for a routine prenatal visit at 28 weeks gestation. She is healthy and has had no complications in this pregnancy. You review her laboratory results that confirm she is HIV positive. In your discussion of her HIV status with her, she asks what this will mean for her unborn baby. You discuss mother to child transmission (MTCT) of the virus, and she asks if there is anything, you can do to prevent this. You have heard about antiretroviral (ARV) prophylaxis for MTCT but you do not know if it is effective.

In forming our question, we can use the acronym PICO to ensure that we have all of the important elements of the question. We need to identify the Patient (P), the Intervention of interest (I), the Comparison of interest (C) and the Outcome of interest (O). Refer to Table 1.

Table 1: PICO

*P*atient	Pregnant woman 28 weeks gestation
*I*ntervention	ARV prophylaxis for MTCT of HIV prevention
*C*omparison	No ARV prophylaxis
*O*utcome	Reduced MTCT of HIV

Question: Is ARV prophylaxis in the last trimester of pregnancy effective in preventing MTCT of the HIV?

Step 2 – Tracking down the best evidence

Types of Evidence

Evidence can be qualitative or quantitative. **Qualitative evidence** is "the organization and interpretation of non numerical information for the purpose of discovering important underlying dimensions and patterns of relationships" Polit D et al, 1995; p 630). **Quantitative evidence** uses numerical data, analyzed through statistical procedures for the purpose of describing phenomena, relationships and significance of the results. Evidence presented in the ALARM International Program consists primarily of quantitative data.

Quantitative research trials can range from weak to strong in terms of scientific rigor. The strongest quantitative evidence comes from randomized controlled trials (RCTs). In a RTC, each participant has an equal chance of being in any group in the study. There is no bias in assigning some participants to one intervention and others to another intervention. Based on the group results, the outcomes and value of an intervention can be determined.

Observational studies test a hypothesis using "cases" and "controls" and allow nature to take its course without controlling for confounders or bias. Two types of observational studies are **cohort** and **case control**. A **cohort study** classifies subjects or individuals according to exposure to the intervention or treatment (cases) or to nothing (controls). It then follows subjects over time to see how they respond to the intervention or treatment. The outcomes of the "cases" are then compared to those of the "controls". A **case control study** is a retrospective study that compares a group of individuals with the outcome of interest, such as a condition or disease, and then looks back in time to investigate "exposures" that may have caused the outcome or condition or disease.

Evidence collected from prospective research trials is more powerful than that collected from a retrospective analysis of outcomes. In a prospective trial, the researcher has a hypothesis or idea and goes forward in time collecting data to observe the results of an intervention. Variables can be anticipated, and the study can be designed to control for those variables. A retrospective trial is not as powerful because it looks back in time at events that have occurred. The researcher cannot control for variables and must often rely upon charts that are incomplete.

Therefore, a prospective RCT, which controls for known and unknown variables between those receiving and those not receiving a given intervention, is the most powerful way to discover if the intervention has a significant impact. Figure 1 demonstrates the levels of evidence according to the strength of study design.

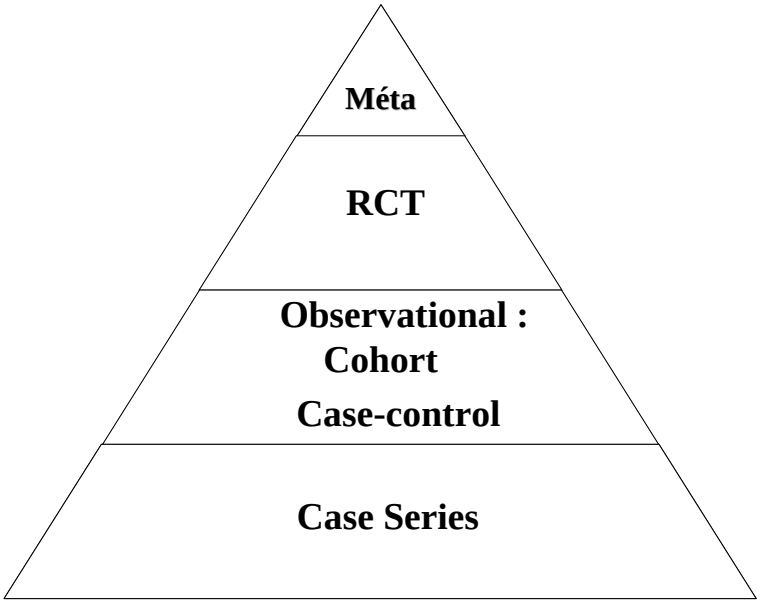
Meta-analysis

Trials are more likely to show a difference between the intervention and control group, if it truly exists, when the numbers studied are large and when the difference in outcome is great.

In obstetrics, most serious outcomes are rare. This means it can be very difficult or impossible to carry out RCTs that are powerful enough (have enough subjects) to demonstrate a difference between a treatment and control group, even if it exists. Trials that have been done are often too small alone to provide statistically significant results. Meta-analysis is one tool that may allow useful information to be obtained from these studies.

A meta-analysis is a systematic evaluation of a collection of several studies that are similar in design, study populations and outcomes examined. By combining data appropriately, the answer to an important question may be found in the cumulative information in the medical, nursing and midwifery literature. The systematic review of the medical literature, which is done by the researchers who retrieve and synthesize the information and make it more readily available for our use, is a secondary benefit of meta-analysis.

Figure 1 - Levels of evidence



Statistical tests are used to summarise findings and can be used to establish whether research results are valid and reliable.

Available websites for obtaining information include:

- http://www.google.com or www.scholar.google.com
  - These search engines can be very useful in getting an overview of what can be found on a topic. They will locate articles recently published on a topic
- http://www.emedicine.com
  - This is a very handy site for health care providers in getting quick access to material. Alternatively, you can type in your search topic "postpartum haemorrhage" followed by emedicine in your search engine, such as Google, and it will bring you directly to the emedicine website.
- http://www.ncbi.nlm.nih.gov
  - This is the website of the National Library of Medicine and the National Institute of Health in the USA
- http://www.cochrane.org
  - This is the website where you will find the Cochrane reviews. The abstract is available at this website for all of the systematic reviews that have been done. Developing countries can also access a selection of systematic reviews from the Cochrane Library including commentaries of relevance to a developing country environment, videos, etc. in English, Spanish and Chinese. This is available free of charge to developing countries from:

    WHO HRP, CH 1211 Geneva 27, Switzerland
    Fax 0941 22 791 -4171/ Tel -3380 (J Khanna)
    Email RHL@who.ch
    http://www.rhlibrary.com

Step 3 – Critically appraising that evidence for its validity (closeness to the truth), impact (size of the effect) and applicability (usefulness in our clinical practice)

The skills of critically appraising the research literature are beyond this course. There are good, reliable resources of information that have undergone extensive peer review and appraisal. This would include a well-completed meta-analysis, such as the Cochrane reviews, and guidelines from various sources that have been based on reviews of the research literature. Some of these clinical practice guidelines can be found at:

- http://www.sogc.org
  - The website of the Society of Obstetrics and Gynaecology of Canada (SOGC)
- http://www.rcog.uk.org
  - The website of the Royal College of Obstetrics and Gynaecology in the United Kingdom

It must always be remembered that much of the clinical care provided is not supported by 'good' evidence simply because trials have not been done. Whenever possible, we must use the information from sound systematic reviews to guide us in the appropriate and compassionate practice of medicine.

Step 4 and Step 5

Although Step 4, "Integrating the best evidence with our clinical expertise and with our patient's unique biology, values and circumstances," and Step 5, Evaluating our effectiveness and efficiency in executing steps 1–4" and seeking ways to improve them both for next time" are steps that health care providers need to incorporate into their daily practice, they go beyond the scope of this course. However, with the information and skills from the course you will be able to begin or to continue in the practice of EBM.

EBM Favourites

Following are definitions for some key terms used in the ALARM International Program.

- Absolute Risk Reduction, ARR: the absolute arithmetic difference in rates of bad outcomes between experimental and control participants in a trial
- Relative Risk Reduction, RRR: the proportional reduction in rates of bad outcomes between experimental and control participants in a trial

- Numbers Needed to Treat, NNT (or NNP: numbers needed to prevent): the number of patients who need to be treated to achieve one additional favourable outcome
- Relative Risk, RR: the proportion of rates of bad outcomes between experimental and control patients in a trial (used in RCTs, cohort studies, and meta-analysis of RCTs)
- Odds Ratio, OR: the ratio of the odds of having the target disorder in the experimental group relative to the odds in favour of having the target disorder in the control group (used in case control studies where only the odds (and not risk) can be calculated.

Relative Risk

The relative risk (RR) statistic is defined as the probability of the outcome of interest in the exposed (or treated) group divided by the probability of disease in the unexposed or untreated group. The RR can be calculated based on findings from clinical trials or based on results from cohort studies by comparing outcomes in two different groups.

The RR helps to determine if there is an increase or decrease in risk associated with the exposure and outcome variables. The risk of an outcome is simply the rate of the outcome in the group studied. The RR compares the risk in the exposed and unexposed groups or in the treated and untreated groups.

Calculating Risk

Using the data presented in Table 2, the risk of the outcome (HIV transmission) in the treatment group is 12 infants out of a total of 100 infants (12/100), which is to say the risk is 0.12 or 12%.

The risk of the outcome (HIV transmission in the placebo group is 25 infants out of a group of 100 infants, (25/100), which is to say the risk is 0.25 or 25%.

Relative Risk (RR) = $[a / (a + b)] / [c / (c + d)]$
= [12/100] / [25/100]
= .48

This result of RR = 0.48 means that the likelihood of MTCT of HIV is decreased if an ARV treatment is used; i.e. it is 0.48 times as likely or nearly half as likely that MTCT will occur if an ARV is used.

In this example, the 95% CI can be calculated to be 0.26 to 0.90. Because this value does NOT contain the value 1, we would consider that it is statistically significant.

95% Confidence interval

95% confidence intervals (CIs) around a statistical outcome can be calculated to help us understand the degree of precision of the finding, as well as to determine if the finding is statistically significant. The CI is a measure of statistical significance, generally calculated as the least and greatest value within which (in the case of RR) the RR results of this same experiment would fall 95% of the time. If the null value (usually 1, but can be 0) is between the lower and higher CI, then the result is not statistically significant; otherwise the result is statistically significant at level of $P<0.05$. Other useful information about the 95% CI is that as the sample size increases the 95% CI becomes more precise (narrower), and as the variation in the sample increases the 95% CI becomes less precise (wider).

Let us look at our hypothetical case of rates of MTCT of HIV with ARV in pregnancy vs. placebo. Suppose we found a study, which had randomized women to antenatal ARV in pregnancy vs. placebo. The results are presented in Table 2.

Table 2 - Relative Risk

	Infant infections	Non-Infected Infants	All Births
ARV Treatment	*12 (a)*	*88 (b)*	*100 (a + b)*
Placebo	*25 (c)*	*75 (d)*	*100(c + d)*
Total	*a + c*	*b + d*	*200 (a+b+c+d)*

Using the EBM Favourites

Using our example from above, of ARV treatment during pregnancy in HIV-positive mothers

- Absolute Risk Reduction (ARR) = 25% - 12 % = 13%
- Relative Risk Reduction (RRR) = 13/25 = 0.52 (52%)
- Numbers Needed to Treat (NNT) = 1/ARR = 1/0.13 = 8.3 (8) mothers have to be treated to prevent one infant from getting HIV

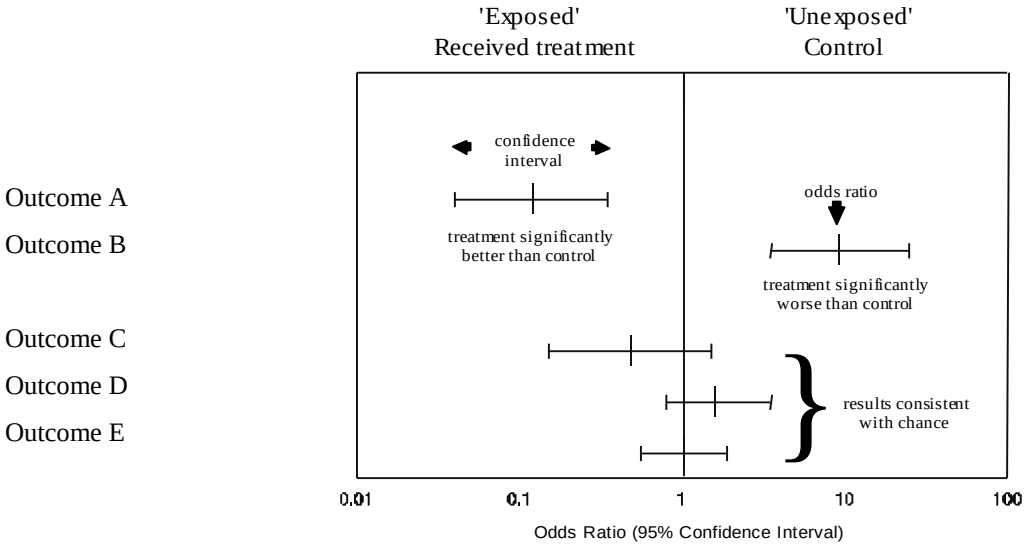
Sometimes a comparison is made of the odds of having an outcome between two groups. This is particularly true in case control studies, where a group with a known condition is compared to a group without a condition and the likelihood of having been exposed to a risk is assessed. In this case, the odds of exposure are compared between the groups, and an **odds ratio (OR)** is calculated. If we were calculating odds in our example above, the odds of having the outcome in the treatment group would be 12/88 (12 infants with the outcome to 88 without the outcome) and in the placebo group would be 25/75. When the outcome is very small and the sample size is large, the RR and the OR are very close in value.

Reading Data Summary Graphs

Individual findings and aggregate findings are sometimes represented graphically in medical literature. The RR or OR (another commonly used statistic for comparing outcomes in study populations) can be presented as a point estimate with the associated 95% CI on a "forest plot" that uses a horizontal logarithmic scale. A vertical line drawn at "1" indicates no difference in the outcome between the two groups. Values less than one will be represented to the left of the vertical line, and those greater than one will be represented on the right side of the vertical line. The data presentation is usually constructed so that the results that are less than 1 are an improvement in outcome.

In figure 2, the outcomes of interest for a single trial are shown with their individual OR and CI.

Figure 2 – Odds Ratio and Confidence Interval

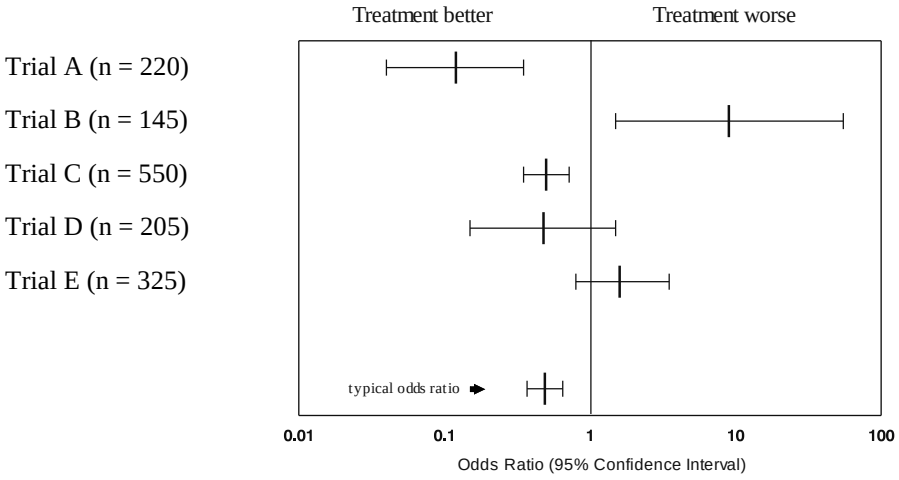


Statistical significance vs. clinical significance in a single study

Measurements of confidence, i.e. statistical significance, do not eliminate the possibility that the results of an experiment are due to chance; they just indicate how likely it is that such a result is due to chance. The judgment of the clinician is required to interpret whether or not a result is clinically significant regardless of the statistical expression of probability used.

In figure 3, the results of a single outcome of interest are combined to derive a total point estimate (typical OR) from several trials that have reported on this outcome.

Figure 3 - Effect of intervention on outcome of interest



The CIs displayed for each study will vary according to study size and will be wider than the CI around the total; this is because the total is calculated based on the combination of results for all subjects in all of the studies included. In the example above, Trial C has the narrowest CI given its larger size. At times, a trial may fail to demonstrate a difference that truly exists because of a lack of power (i.e. not enough subjects, given the effect size). This might be the case in Trial D in the above example. Again, CIs that cross the vertical line at 1 means that the probability of this result occurring by chance alone is greater than 5% (P >0.05).

The results displayed in a meta-analysis fall into one of three categories:

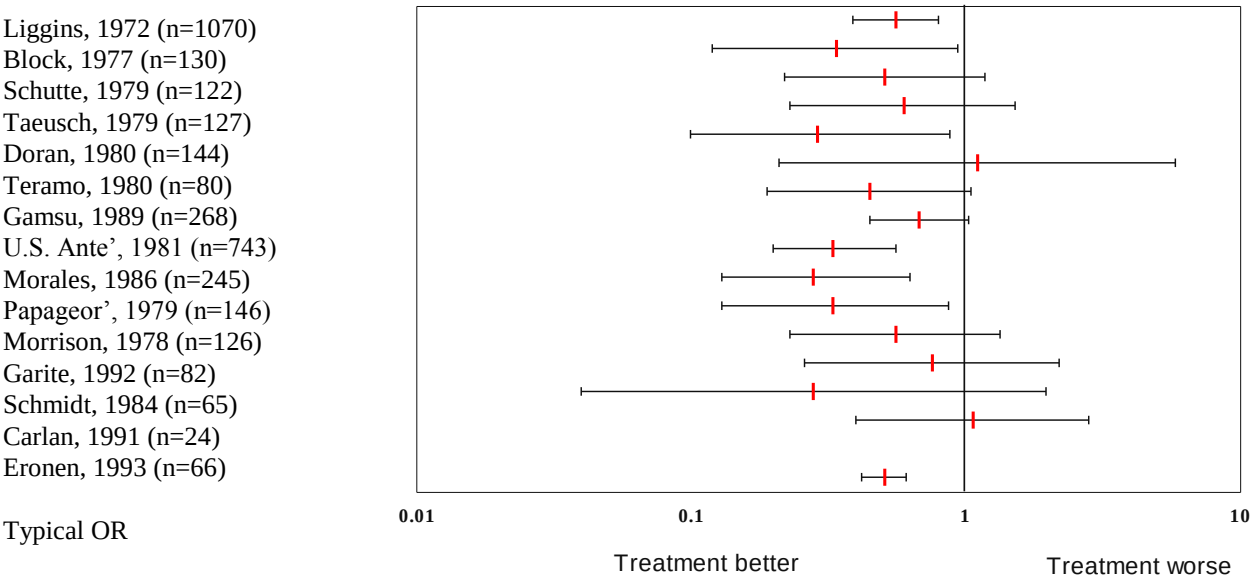
- The point estimate lies to the left of the vertical axis (1) and the CI does not cross 1. This indicates that the outcome for the treated group is less likely to occur than in the control group, and the result is statistically significant. This is shown in Trials A and C and in the summary finding.
- The point estimate is at or near 1 and the CI line crosses 1, such as occurs in Trial D and E. This indicates that there is no statistically significant difference in outcomes between the groups.
- The point estimate lies to the right of the vertical axis and the CI line is completely to the right of 1. This indicates that the outcome is more likely to occur in the treated group than in the control group, such as in Trial B.

Even though several studies may not all achieve statistical significance individually, perhaps because some have insufficient numbers or the effect size is small, the meta-analysis display may show most of these studies demonstrate the same trend. Under this circumstance, when the total and its CI demonstrates a significance difference, one may be more confident that this difference truly exists.

A good example of an important meta-analysis in obstetrical management is that of the use of antenatal corticosteroids for preterm birth. Several studies were conducted to determine the impact on fetal lung maturity and the occurrence of neonatal respiratory distress syndrome.

Figure 4 demonstrates how the results of a meta-analysis are illustrated to identify the lead author, year of study publication and number of subjects in each trial.

Figure 4 - Results of a meta-analysis



Statistical significance vs. clinical significance in a study.

Measurements of confidence (i.e. statistical significance) do not eliminate the possibility that the results of an experiment are due to chance; they just indicate how likely it is that such a result is due to chance. Once the practitioner is relatively confident that the research findings are "real," i.e. they are not chance occurrences, then judgment is required to interpret whether or not a result is clinically significant regardless of the statistical expression of probability used. For example, a study could show with very low probability of chance that using a new blood pressure medication will lower the blood pressure by 1 mm of mercury compared to the standard therapy. Because

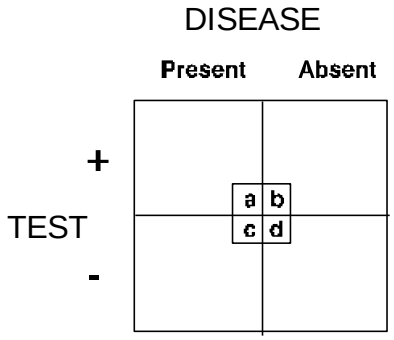
this decrease in blood pressure is not clinically important, based on this finding alone, it is unlikely that a practitioner would change prescribing patterns, even though the study was able to demonstrate a decrease that was **statistically significant**.

Critical Appraisal in Diagnosis and/or Screening

Several indexes are used in order to answer question: Does this (valid) evidence demonstrate an important ability of this test to distinguish accurately patients who do and do not have a specific disorder?

1. **Sensitivity** is the likelihood that the diagnostic test will indicate the presence of disease when the disease is actually present. (True positive rate). [a/(a+c)]
2. **Specificity** is the likelihood that the diagnostic test will indicate the absence of disease when the disease is actually absent. (True negative rate). [d/(b+d)]
3. **Positive predictive value** is the likelihood that a positive test result actually means that the disease is present. [a/(a+b)]
4. **Negative predictive value** is the likelihood that a negative test result actually means that the disease is absent. [d/(c+d)]
5. **Bayes' Theorem:** The predictive value of a test will depend on the prevalence of the disease. With high prevalence, the positive predictive value will increase and vice versa (i.e. in a high prevalence setting, a positive test result is more likely to indicate a true positive than in a low prevalence setting)

Figure 5 - 2x2 table used to estimate the sensitivity, specificity, positive predictive value and negative predictive value of a diagnostic and/or screening tool



For example, in the clinical setting of doing a culture for Group B Strep (GBS):

- Sensitivity of the test (positive culture) is the chance that if the woman had GBS it would be picked up by the test.
- Specificity is the chance that the test will indicate no GBS (negative culture) when in fact the woman does not have it.
- Positive predictive value is the chance that a positive culture represents GBS colonization.
- Negative predictive value is the chance that a negative culture actually rules out GBS.

Table 3 - Results of systematic review of serum ferritin as a diagnostic test for iron deficit anemia (example)

Target disorder (Iron deficit anemia)

Diagnostic test result (serum ferritin)	Anemia	No anemia	Total
Test positive	731 a	270 b	1001 a+b
Test negative	78 c	1500 d	1578 c+d
Total	809 a+c	1770 b+d	2579 a+b+c+d

Sensitivity = a/(a+c) = 731/809= 90%
Specificity = d/(b+d) =1550/1770= 85%
Prevalence= (a+c)/(a+b+c+d) = 809/2579=31%
Positive predictive value, PPV = a/(a+b) = 73%
Negative predictive value, NPV = d/(c+d) = 95%

Grades of Evidence and Classification of Recommendations

Since 1976, the Canadian Task Force on the Periodic Health Examination has used explicit analytic criteria to guide its evaluation of the effectiveness of health care interventions. The criteria are summarized in a classification system that the SOGC and many other organizations have adopted. Graded strength is placed on published medical evidence based on the quality of its design. Greatest weight is placed on the features of study design and analysis that eliminate or minimize biased results. Recommendations are based on the level of evidence. The strongest recommendations (A and E) are reserved for interventions supported or negated by high quality studies (Type I or RCTs). Type II evidence is generally associated with B and D recommendations. In 2003, the Task Force on Preventive Health Care modified the grades to reflect the ongoing evolution of methodology and reporting. Recommendations after this change will include a redefinition of Grade C and the addition of an "I" grade. C grade recommendations are reserved for cases where evidence of adequate quality and quantity may exist but it is conflicting in that the effectiveness of the action remains unclear. Grade C signals a situation where other factors, such as values and individual patient characteristics, may play an even larger role than when evidence is clear-cut. An "I" grade implies that the existing body of evidence is of insufficient quantity or quality (or both) to support a specific recommendation (Canadian Task Force on Preventive Health Care, 2003).

Summary of grades of evidence and classification of recommendations

Quality of Evidence

TABLE 2

QUALITY OF EVIDENCE ASSESSMENT[74]	CLASSIFICATION OF RECOMMENDATIONS[74]
The quality of evidence reported in these guidelines has been described using the Evaluation of Evidence criteria outlined in the Report of the Canadian Task Force on the Periodic Health Exam. I: Evidence obtained from at least one properly randomized controlled trial. II-1: Evidence from well-designed controlled trials without randomization. II-2: Evidence from well-designed cohort (prospective or retrospective) or case-control studies, preferably from more than one centre or research group. II-3: Evidence obtained from comparisons between times or places with or without the intervention. Dramatic results in uncontrolled experiments (such as the results of treatment with penicillin in the 1940s) could also be included in this category. III: Opinions of respected authorities, based on clinical experience, descriptive studies, or reports of expert committees.	Recommendations included in these guidelines have been adapted from the ranking method described in the Classification of Recommendations found in the Canadian Task Force on the Periodic Health Exam. A. There is good evidence to support the recommendation that the condition be specifically considered in a periodic health examination. B. There is fair evidence to support the recommendation that the condition be specifically considered in a periodic health examination. C. There is poor evidence regarding the inclusion or exclusion of the condition in a periodic health examination, but recommendations may be made on other grounds. D. There is fair evidence to support the recommendation that the condition not be considered in a periodic health examination. E. There is good evidence to support the recommendation that the condition be excluded from consideration in a periodic health examination.

Key Messages

1. Health care providers have the responsibility of analyzing, researching and judging the generalizability of the results.
2. Health care providers have the responsibility of incorporating the best available research evidence into their daily clinical practices.
3. The appropriate analysis of research literature translates into appropriate clinical practice.

Suggestion for Applying the Sexual and Reproductive Rights Approach to this Chapter

Health care providers should provide evidence-based care to women and their families. Have a good understanding of the literature so that you may provide this information to your clients in simple language so that they understand the reasoning behind why certain procedures are performed and so that they may become part of the decision-making process in the management of their care.

Resources:

- Canadian Task Force on Preventive Health Care. New grades for recommendations from the Canadian Task Force on Preventive Health Care. CMAJ 2003; (3)169:207–8.
- Norman G and Streiner D. *Biostatistics: The Bare Essentials*. Toronto: Mosby; 1999.
- Polit D and Hungler B. *Nursing Research*. Philadelphia: Lippincott; 1995, p 630.

- Straus SE, Scott Richardson W, Rosenberg W. Haynes RB, Sackett DL. Evidence-based Medicine: 1st Edition.2000.
- Streiner DL Norman GR and Blum H. *PDQ Epidemiology.* Toronto: B.C. Decker Inc. 1989.
- Streiner DL Norman GR and Blum H. *PDQ Statistics.* Toronto: B.C. Decker Inc. 1989.
- Thacker MD, Peterson HB, Stroup DF. *Metanalysis for the obstetrician-gynecologist*. Am J Obstet Gynecol, May 1996; 174:1403–7.
- Understanding the odds-ratio diagrams in the Cochrane Database of Systematic Reviews (CDSR), available on line at: http://cochrane.kfinder.com/MetaViewer/metaviewer_help.html
- Woolf SH, Battisa RN, Anderson GM, Logan AG, Wang Eel. *Canadian Task Force on the Periodic Health Exam. Ottawa: Canada Communication Group*. 1994 p. xxxvii