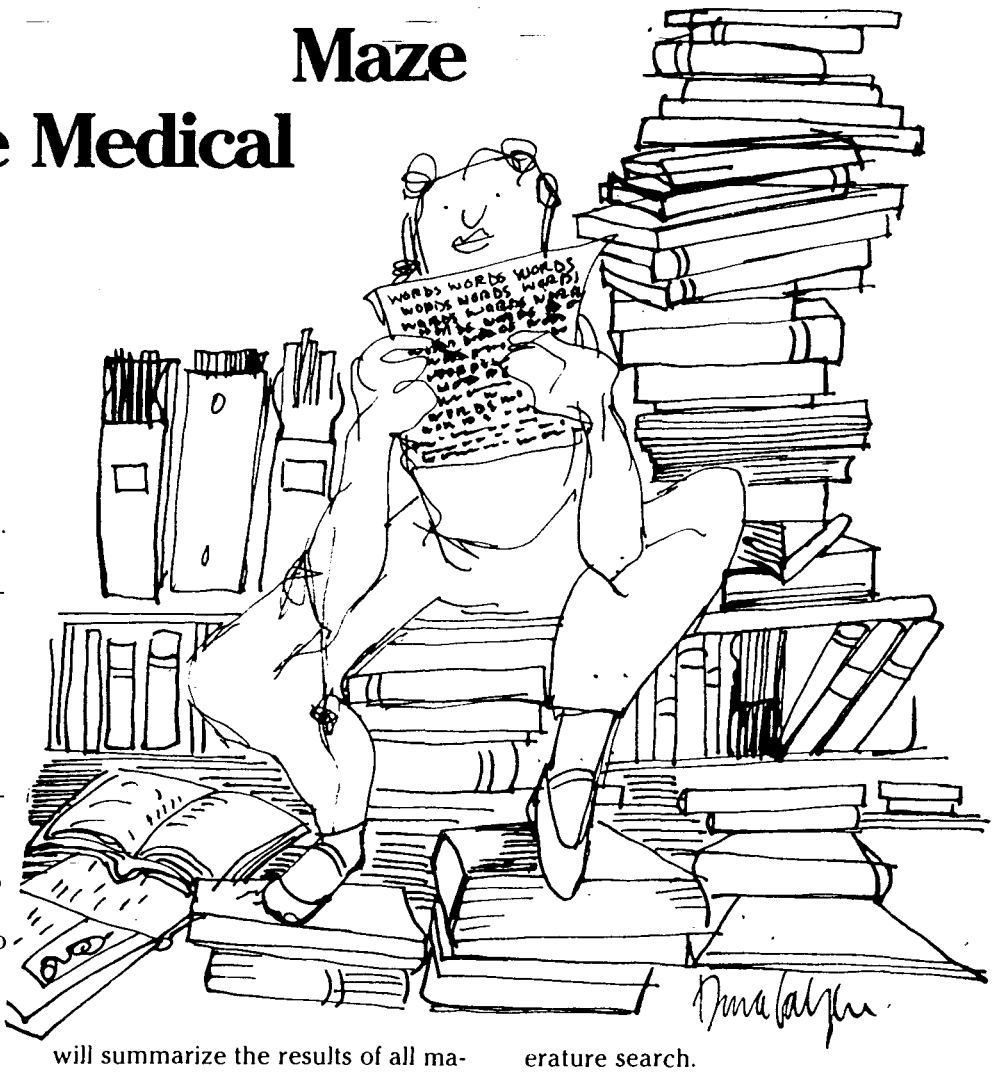
Barbara Mintzes

Through the Medical Maze



A woman who works in a community health clinic is invited to attend government hearings on contraception. Many of the presentations are on Depo Provera, a drug the federal government is considering for approval as a contraceptive. She hears references to many conflicting results of medical studies. Depo Provera appears either to be very risky for women to use, or completely safe. Medical studies are used to back both positions.

To go beyond her hunches on who stands to gain from an unsafe drug being approved, this woman needs to be able to read and judge medical studies for herself.

A woman may also want to read medical studies in order to make decisions about her own or her family's health care, or to decide what to do if she has received two different medical opinions on a health problem.

What we, as women and health consumers, need is information on how to read and judge medical studies. Many medical studies provide unreliable information due to problems in design. Unfortunately, some of this is intentional, as researchers sometimes receive funding for their research from foundations or corporations who have an interest in promoting certain "facts" and ignoring others. In order to examine the literature properly, we need to have a means of picking out which information to trust.

When considering where to begin, first choose what articles to read. A good approach is to look for recent reviews on a topic. A good review will summarize the results of all major studies on the topic and point out the main areas of disagreement among researchers.

Recent reviews can be found in *Indicus Medicus*, a listing of published medical studies available in university and some main branch city libraries. *Indicus Medicus* has separate volumes for each month of the current year and separate volumes by year for each preceding year. Studies are organized by subject. Review articles usually contain the word "review" in the title.

It is important not to stop at a literature review, since reviewers are often biased in the way they select, present, and interpret information. Therefore, you should check the original sources listed in the review and look for studies which the reviewer did not mention. This can be done either by using *Indicus Medicus* or by conducting a computerized literature search.

A computerized literature search is done for a fee, by a university or medical librarian who will also help you limit your topic to produce a useful list of articles. A properly conducted computerized literature search, including a list of titles and abstracts, should cost less than $50. You may also be able to get a search done for free. Doctors have access to a yearly allotment of free literature searches, and a friendly doctor may agree to arrange to have a search done for you.

Not all of the articles that you find will be useful to you. After reading the titles and abstracts, you can decide which are not relevant or obviously badly designed. If the study does seem worthwhile, read the whole study, since the abstract is so short that it omits a lot of information, and findings may be overdramatized or misrepresented.

Types of studies:

Randomized controlled or experimental trial This type of study usually uses volunteers to test the usefulness or safety of a drug or another type of treatment. The volunteers who receive the treatment form the study group and those who do not form the comparison or control group. In a well-designed study, there should be a random selection process such that everyone who enters the study has an equal chance of ending up in one group or the other, in the same way that a flipped coin has an equal chance of ending up as heads or tails. The study must also be "double-blinded," so that neither the volunteers nor the people who assess the effectiveness of the treatment know who is in the study or control group. Otherwise, people who know they are receiving a drug or other treatment may be more likely to get better because they know something is being done for them, and researchers who know that a person has received treatment may tend to judge that person as healthier than someone who has received no treatment.

A randomized controlled trial provides the only good evidence for whether a drug or other treatment is actually useful. Unfortunately, drugs are used for many purposes for which they have not been adequately tested or for which randomized controlled trials have not found them to be useful. For example, progesterone is used for the treatment of premenstrual syndrome in spite of randomized controlled trials which have shown it to be no more useful than a placebo. (In one trial, the placebo was more effective than the progesterone.) Unfortunately, many alternative treatments which may provide a safe and holistic way of treating health problems have not been tested for effectiveness in randomized controlled trials.

Cohort or prospective study A cohort or prospective study determines whether exposure to something causes a disease by following a group of exposed and unexposed people over a period of time. An example would be a study of two groups of women — women using the pill (the study group) and women using other forms of birth control — who are followed for 20 years in order to see if more women using the pill develop breast cancer.

Case-control or retrospective study This study looks back in time rather than forward. People with a disease or a condition are compared to people who do not have the disease. For example, in case-controlled studies of the pill and breast cancer women with breast cancer are interviewed to see whether and for how long they used the pill; the same questions are asked of a control group of women without breast cancer. A case-control study is much cheaper to do than a cohort study because study subjects don't need to be followed over a long period of time. Often, a problem is first identified with case-control studies and is later checked out in more depth using cohort studies.

Case-control studies are very useful for studying rare diseases. For example, strokes in young women are so rare that thousands of women on the pill could be followed in a cohort study without showing an increased risk of strokes in young women. However, a case-control study of young women who have had strokes can quickly show that women who have had strokes are more likely to have been on the pill.

Individual case study An individual case study looks at the development or treatment of disease in one individual and reports that something out of the ordinary happened. An example would be a case study of a man exposed to the drug DES before birth who developed testicular cancer.

Case series A case series reports on a group of people either with a similar disease or a similar exposure, without comparing them to a control group. A case series of a large number of people may look impressive, but it could provide misleading information. An example is the case series of 632 women who were given DES in the 1940s to help them with problem pregnancies. Based on the apparently positive results of this case series, DES was recommended for use in pregnancy to prevent miscarriage. As a result of these recommendations, the drug was prescribed to large numbers of pregnant women. However, a randomized controlled trial of DES in 1952 found it to be no more useful than a placebo in preventing miscarriage. In spite of this evidence, physicians continued to prescribe DES for this purpose until it was banned for use in pregnancy in 1971 because it was found to cause cancer in the daughters of women given the drug.

Glossary

Abstract: a short summary of the contents of an article.

Association: If two things happen together more often than would be expected by chance, they are said to be associated. This points to the possibility that one was caused by the other, but is not proof.

Blind study: A study in which people are unaware of whether they are in the study group or the control group. A study is double-blind if the people assessing the results also don't know.

Control group: A comparison group, consisting of people who are generally like the group being studied in all ways except the one characteristic being studied.

Matching: Choosing people for the control group with the same characteristics (such as age, sex, smoking) as people in the study group.

Odds ratio: This is an estimate of the relative risk (see below) which is calculated in case-control studies.

P value: The p value is the probability of the results occuring by chance or coincidence in the particular sample of people being studied. P=.05 translates to "the probability of this result occuring by chance is 5 per cent.

Placebo: (often referred to as a "sugar pill") A substance with no physical effects, but which looks like a medicine or another substance being tested.

Study design:

Sampling A group of people are usually studied in order to apply the study results to a larger group, such as people in general, women in gen-

Placebo effect: The proportion of people who will get better when taking a placebo.
Population: The larger group of people represented by a sample. It could be all women going through menopause, all Canadians, all pregnant women over 35, etc. depending on the study.
Random assignment: (randomization) Choosing who will be in the control or the study group in such a way that each person has an equal chance of being chosen for either group.

Relative risk: The probability of an exposed person developing a condition compared to the probablity of a non-exposed person developing the same condition.
Sample: A group of people who are part of a larger group, and are chosen to represent that larger group in a study.
Statistical significance: Study results are generally considered statistically significant if they have been calculated to have less than 5 per cent or a 1 in 20 chance of having occured by chance. Statistical significance says nothing about whether results are important or meaningful; it is a calculation of the likelihood that a random sample represents characteristics of a population.
Stratification: Dividing a group into subgroups (or "strata") on the basis of one or more characteristics such as age, sex, smoking, etc.

eral, etc. The group being studied is called a sample. The ability to generalize the results of the study to a larger group of people will depend on how the sample was selected. For example, the results of a study of women with breast cancer can only apply to all women with breast cancer if any women with breast cancer had an equal chance of being selected for the study. The results will not apply to younger women if only women over 50 are selected.

In randomized controlled trials, volunteers are frequently used. It is necessary to know what type of people volunteered for the study. For example, many drugs for general use are tested on young men, and the results may not apply to women or older people.

It is also important that roughly no more than 10 per cent of people are "lost to follow-up" in a study. The people who are "lost" may be very different from those who stay in a study, and their loss may bias the study results.

Sample size The larger the group studied, the more likely it is that the results of a study will be accurate, and not be the result of chance. For example, the results of a study of the sex ratio of babies would differ if they were based on the next 10 instead of the next thousand births. In the sample of 10 babies, the sex ratio will probably not be close to 50:50; in the sample of 1000 babies, it probably will.

Sometimes, the sample may appear to be larger than it is. A study may begin with a large sample, but the study results may have been calculated from a much smaller subgroup.

Study and control group Theoretically, the study and control groups are supposed to be identical in all respects except the one being studied. Obviously, this ideal can never be reached. However, researchers should make sure that the most important "risk factors" for the disease or health problem are equally distributed among the study and control group. This is usually done by stratifying the study and control groups, or dividing them into subgroups according to the presence of a risk factor. For example, age is a risk factor for breast cancer, since women are more likely to develop breast cancer as they grow older. A breast cancer study could stratify for age by limiting all comparisons to women of similar age groups. For example, women with breast cancer between the ages of 50 and 54 would be compared only with controls of the same age group.

After reading a few studies, you should become familiar with the most important known risk factors for a disease or condition. Age and sex are almost always major risk factors.

Follow-up period This is something to look for in studies of cancer or other diseases which are slow to develop. Twenty to 30 years may be needed for cancers to show up as a result of exposure to a cancer-causing substance.

Sometimes studies present follow-up time in terms of "person-years." Person-years is the number of people multiplied by the number of years each was followed. It can be a method of obscuring what really happened as 100 "women-years" may be 200 women followed for 6 months or 20 women followed for 5 years. Studies of effectiveness of birth control methods often use person-years of follow-up. These studies should also state actual numbers and length of time women were followed. as the effectiveness of some methods differs depending on how long a woman has used them.

Judging results:

Sometimes the results of a study will be useless because they are based on a questionable diagnostic test or on incomplete information. For example, a study on the safety of two types of IUDs in the April, 1986 issue of the *Canadian Medical Journal* may have missed many cases of PID. Women were asked whether they had ever had PID, and their medical records were only checked if they answered "yes". A woman who was not told by her doctor that she had a pelvic infection or PID, or who was misdiagnosed, would be classified as not having had PID. Many cases of PID may have been missed, resulting in an overestimate of the safety of the IUDs.

Information from interviews may be undependable if people's memories are affected by their expectations of what should have happened. For example, in a menstrual cycle study, women were asked to keep a daily diary of their mood for three months. At the end of the three months, they were asked to describe their moods during each part of their menstrual cycles during the previous three month period. Women remembered premenstrual time as mostly negative, although the same women had frequently described their mood as "elated" or in other ways very positive in their diaries at the time.

People's memories are also af-

fected by their opinion of what is good or socially acceptable behaviour. People usually underestimate the amount they drink or smoke and overestimate the amount they exercise. For example, the average alcohol consumption based on interviews is much lower than the average amount consumed based on the amount of alcohol sold in Canada.

Outcomes may be reported which misrepresent the situation. An example of this would be studies of *in vitro* fertilization in which success rates are measured in terms of total number of pregnancies. *In vitro* programs use pregnancy tests which are accurate 7 to 10 days after conception. At this early stage, about one third of the pregnancies end in miscarriage. For women considering in vitro fertilization, the important outcome is the number of babies born, not the number of pregnancies.

The results presented in the tables may not always back the claims in the discussion and/or abstract of the article. It is always important to carefully read through the tables, as they present the basic information, from which all conclusions are drawn. Sometimes the information in the tables is misrepresented in the text.

The results may not always point to a cause and effect relationship. Scientific studies tend to look at whether there is an association between two or more things, that is, whether they tend to happen together. Usually the real reason for the study is to provide evidence for a cause and effect relationship.

There are certain factors to be aware of when deciding whether one thing causes another. For example, is evidence available from randomized controlled trials, since this is considered the most dependable human evidence? Or is there a strong association of cause and effect in a number of studies? On any issue it will be possible to find studies which are contradictory, at which point you may be tempted to throw up your arms in despair. You need to assess each study individually and decide which ones are the most valid. Keep in mind that many studies are funded by groups, such as drug and tobacco companies or government agencies, which are interested in seeing specific results.

Another factor to consider in evaluating causal relationsips is did the cause come before the effect? This may seem obvious but is sometimes tricky to establish. For example, a study of stress and infertility would need to distinguish between stress which was caused by infertility and stress which occured before a person was aware of having fertility problems.

There are many other factors to consider. Is there a dose response? If a higher dose of X-rays causes more cancer than a lower dose, then the argument that X-rays cause cancer becomes stronger. Does the association make biological sense? An argument that AIDS can be spread on toilet seats does not make biological sense if it is known that the virus dies in the air and needs to be transmitted through body fluids. A bit of caution may be needed here, as sometimes accepted "biological sense" may be based on assumptions rather than something which is known. Generally, anyone with a new or outrageous theory should explain it in terms of how the human body and/or the disease work.

Finally, the usefulness of the results should be assessed. Will they make a difference to what treatments are available to women? Do they give insights into what causes a disease, and therefore how to avoid it? Do they open up new ways of looking at a situation? Are they a help or a hindrance in the struggle toward social equality for women?

Group research:

If you have not done health research or used medical or university libraries before, it can be very helpful to start researching a topic with one or two other women. Not only is it good to have someone to laugh with over outrageous comments from researchers, it is also helpful to have someone with which to check out hunches and directions for research.

At the Vancouver Women's Health Collective, we have started a "Journal Club." Women who are interested in research take turns bringing in an article for everyone to go through in nitpicking detail. It is amazing how much more we are able to pick out as a group. It also means that we build on one another's skills, so that we all are better able to pick out the useful from the useless information, and to bring our own perspective into reading medical literature in a way which values women's health.

Take this guide with you, and happy reading. Becoming familiar with reading and judging medical literature is not only helpful for making health care decisions. It can also take away some of the prestige of "privileged" information which only doctors and professionals seem to have access to. And it can give some insights into the role of science in our society, through a first-hand view of what "good" and "bad" science can be like.

Further Reading

Riegalman, Richard K. *Studying a Study and Testing a Test; How to Read the Medical Literature*, Little, Brown & Co., Boston, 1981.

Barbara Mintzes is a member of the Vancouver Women's Health Collective.