Figure 12.23. SIGNIFICANCE TESTING: Meta-Analysis

EM9112: The Globe and Mail, November 2, 1991, page D10

META-ANALYSIS

Lies, damned lies and statistics

Getting a tad miffed by all those contradictory medical studies? A brand-new method of sifting the chaff from the oat bran is bringing analytical rigour to research

BY STEPHEN STRAUSS
SCIENCE REPORTER
TORONTO

MEDICAL scientists have taken to calling it "the oat-bran syndrome." Ten times a year, a small experiment produces an extremely promising result – for example, an oat-bran diet is found to lower blood-cholesterol levels – and the finding is immediately trumpeted across the media.

Then, lo and behold, some months later another study appears that says the first one was all wet. Either there is nothing to the first study – or far more confusing, the new research suggests that there is actually some deleterious effect from, say, oat bran.

What is the ordinary person to think? Did one study cancel out the other? If a favourable third study weighs in, does oat bran win two-to-one? And if not, at what score *does* a finding triumph over its doubters?

If its proponents are right, the way out of the oat-bran syndrome lies along a bright, shiny statistical roadway called meta-analysis, a rigorous means of studying studies.

Scientists have been attempting overviews for 100 years or more, but it is only recently that they have grown sensitive to how intrinsically subjective they were. "In the past, you generally just asked experts to write review articles and tell you what he or she thought," says Dr. Stephen Thacker, author of a review of meta-analysis in the *Journal of the American Medical Association*.

How and why a person arrived at these conclusions was often not spelled out, says Dr. Thacker. "If you have room for judgement, then you have room for [unexpressed] differences of opinion. At least with meta-analysis you have the data laid out in front of you. You see why people disagree."

But meta-analysis promises more than just higher levels of objectivity, say its proponents. A painstaking statistical manipulation of a number of studies' data will reveal patterns that the most intuitive expert won't see.

The idea of a formal study of studies dates back to the work of English geneticist Ronald Fisher in the 1920s, though the term meta-analysis wasn't coined until 1976, by Gene Glass, a psychologist now at the University of Arizona. While applicable to results from all disciplines, meta-analysis took off in North America in the 1970s, as psychologists and other social scientists tried to make sense of the myriad studies that sought to evaluate the effect of social legislation. One classic example was an examination of how school integration affected performance levels of black students.

But in the last decade, and in particular in the past five years, the use of meta-analysis has exploded in medicine, with epidemiologists applying it to a wide variety of entanglements: Lead exposure and IQ scores, repeated caesarian sections, use of antibiotics in colon surgery, the link between birth-control pills and blood clots have all been subjected to the systematic study of studies.

SCIENTISTS begin a meta-analysis by gathering the available data, generally through a computer search for all relevant studies done in the area. They also try to locate the "fugitive literature," the unpublished data whose exclusion from an analysis may statistically skew results.

Data in hand, two statistical approaches are applied. The first, the "fixed-effect model," effectively lumps the splintered, small studies into one huge study. As an example of the power of the brute-lumping approach, Dr. Gordon Guyatt, an epidemiologist at McMaster University, in Hamilton, Ont., looked at 10 studies around the world that tried to determine if an exercise regimen after a heart attack lowered death rates. While individual studies pointed toward exercise being life-saving, only one had demonstrated a statistically significant decline in deaths.

However, under the hard-eyed gaze of a fixed-effect meta-analysis, exercise was seen to have decreased death rates by 25 per cent.

That was a truth only visible after lumping. "All the individual studies were just too small to see the effect," says Dr. Guyatt.

The other general approach in meta-analysis is known as the "random-effects model." This considers the variations between studies. For example: Is an abnormally large positive or negative effect in one or two studies skewing the whole meta-analysis? Are there significant sex or age differences? While there are a number of random-effect techniques used to approach this problem, one very powerful method involves shuffling and reshuffling the statistical deck.

If there are 20 studies in your sample, this approach may ask what happens if any three studies are left out. Do we still have the same effect? All combinations of 17 studies in and three out are tried. If an effect is still seen after a complete shuffling of all original studies, the laws of statistical probability shout that some real phenomenon has been found.

While meta-analyses are appearing regularly in major medical journals, the process is not without its critics.

John Bailar III, a McGill University biostatistician and epidemiologist, has been sharply dismissive. "One difficulty is the old problem of garbage in, garbage out" he says. "The problem is that different study groups are really different in substantially different ways."

Furthermore, he says, the reliance on statistics "sciencizes" the subjective qualifications made by old-style reviewers. Numbers mask doubts and consequently, "there is an intellectually unjustified air of precision to a meta-analysis."

But more basically, he doubts that meta-analysis has arrived at any new truths. "I have yet to see a clear demonstration that meta-amalysis has taught us something we would not have learned in the older way."

A painstaking statistical examination of number of studies' data will reveal patterns that the most intuitive expert won't see

Meta-analysis' defenders generally agree with Dr. Bailar that its uncritical acceptance opens the way for statistical hocus-pocus. "It's hard to do a meta-analysis and even harder to do a good one," says Ingram Olkin of Stanford University.

However, he and others believe that their methods do unveil truths that experts either missed or would only see much later on. The example often pointed to is the studies of heart attacks and aspirin spearheaded by meta-analysis apostle Richard Peto of Oxford University. By virtue of their sample size, Dr. Peto and his associates were able to show in 1988 what several smaller

(*continued overleaf*)

studies had failed to do: Aspirin can reduce the likelihood of heart attack.

To honour their championing of meta-analysis, Dr. Peto and his mentor, Oxford University epidemiologist Richard Doll, recently shared the $700,000 (U.S.) Helmut Horten Research Award, which was established by a German supermarket magnate to recognize advances in medicine.

While it is becoming a hot way to do science overviews, meta-analysis may also begin to have an effect on research in general. Researchers in Boston are suggesting that a computer data bank of studies in contentious areas be set up. Any new findings would be added to the list and quickly meta-analyzed.

When a paper is published, it could include with all its other data a statistical report card on itself. This would describe how close the new evidence brought the whole controversy to resolution – something perhaps as simple as: "Two more studies of this size and the case will be closed."

That may not totally eliminate the problems of conflicting evidence, say statisticians, but it should make it much more difficult for a scientist, or a journalist, to get up and cry: Oat bran.

1 The basic idea behind meta-analysis is *combining* the data of several or many studies on a topic for analysis (although there are a number of technical aspects of how this combining is done). What fundamental *statistical principle* is involved in the idea of combining? Explain briefly.

- What *proviso* should always be attached to this principle? Outline how this proviso bears on meta-analysis.
  - Where in the article EM9112 reprinted overleaf on page 12.55 and above is this proviso hinted at? Explain briefly.

2 What *statistical parallel* is there between a *run of nine* out-of-control signal from a control chart for averages and the discussion in the fourth and fifth paragraphs of the middle column overleaf on page 12.55 of Dr. Guyatt's meta-analysis of ten studies of the possible benefits of an exercise regimen after a heart attack. [Recall that the effect of exercise on death rates in *healthy* people is the subject of the article EM9023 reprinted in Figure 9.15 of the STAT 220 Course Materials.]

- What fundamental criterion for (scientific) accuracy of an Answer is involved in these two contects? Explain briefly.

3 Outline the *statistical* matter underlying the statement in the third paragraph of the middle column overleaf on page 12.55: *They also try to locate the "fugitive literature," the unpublished data whose exclusion from an analysis may statistically skew results.*

- What *non*-statistical matter is also involved? Explain briefly.
  - Can you think of an analogy with much of what is described in the *media* to this non-statistical matter?

4 Explain what you understand from a statistical perspective by the discussion in the third paragraph of the right-hand column of the article EM9112 reprinted overleaf on page 12.55, which finishes: *If an effect is still seen after a complete shuffling of all original studies, the laws of statistical probability shout that some real phenomenon has been found.*

5 Outline the *statistical* matters involved in the discussion in the fourth-last paragraph of the right-hand column overleaf on page 12.55: ..... *statistics* "sciencizes" *the subjective qualifications made by old-style reviewers. Numbers masks doubts and consequently, "there is an intellectually unjustified air of precision to meta-analysis."*

6 Suggest briefly how you might *defend* meta-analysis against Dr. Bailar's charge, given in the third-last paragraph of the right-hand column overleaf on page 12.55: *"I have yet to see a clear demonstration that meta-analysis has taught us something we would not have learned in the older way".*

7 Discuss critically the two sentences in the second-last paragraph above:

- *When a paper is published, it could include with all its other data a statistical report card on itself.*
- *This would describe how close the new evidence brought the whole controversy to resolution – something perhaps as simple as: "Two more studies of this size and the case will be closed."*

8 An essential aspect of the proper use of statistical tests of significance is the clear recognition of the distinction between the *statistical significance* and the *practical importance* of the Answer. Identify explicitly the place(s) in the article EM9112 reprinted overleaf on page 12.55 and above where this distinction is important; explain briefly in each case you identify.