



Dr. Wulf H. Utian has served as Editor-in-Chief of *Menopause Management* since its inception in 1988. Arthur H. Bill Professor Emeritus of Reproductive Biology and Obstetrics and Gynecology at Case Western Reserve University, he is President of Rapid Medical Research, headquartered in Cleveland, and is Consultant in Women's Health to the Cleveland Clinic Foundation. He is a Fellow of both the Royal and American Colleges of Obstetricians and Gynecologists, a Fellow of the International College of Surgeons, and a board-certified reproductive endocrinologist.

A pioneer in menopause research, Dr. Utian founded the world's first menopause clinic in Cape Town, South Africa, in 1966 and established the Cleveland Menopause Clinic in 1983.

Recipient of many research grants and awards, he is the author of more than 150 scientific publications and five books. He is the Honorary Past-President of the International Menopause Society and Honorary Founding President and Executive Director of The North American Menopause Society. He is also Chairman of the Council of Affiliated Menopause Societies (CAMS) of the International Menopause Society.

Does “Statistically Significant” Always Equate to “Clinically Relevant?”

There is a widening chasm between the points of view of epidemiologists/research investigators and practicing clinicians regarding the appropriate interpretation of major randomized trials and observational studies. The demand to practice “evidence-based medicine” in a medical world largely deficient of the necessary evidence that would pertain to all specific medical situations is difficult; and, at the very least, it certainly is frustrating to individuals on both sides of the desk.

Every clinical research study essentially is reported as three major components. First, there are the research hypothesis and protocol; second, the data (ie, the results or numerical outcome derived from conducting the research protocol); and third, the discussion or interpretation of the results by the investigators. Let me expand on these components.

The research hypothesis usually questions the efficacy and safety of a therapy, drug, behavior modification, or whatever on a given population. Studies may be observational or prospective and randomized in design. Each has advantages and disadvantages. An immediate problem relates to the selection of study population. To enroll adequate numbers of subjects to enable statistical power, the prevalence of the problem being addressed needs to be known. Often a study population is selected that may not necessarily be reflective of the clinical population that usually is seen in routine practice.

The data are the data. That is, results are presented with statistical analysis. Often investigators will utilize multiple statistical instruments in an attempt to demonstrate a significant difference between the study and the control group. Sometimes a result may be pursued with excessive zeal, with subanalyses of subanalyses. To quote Dr. John Bailar, “investigators must be especially wary of any tendency to misinterpret or over interpret study results in which they have a professional or personal stake.”

There is the rub. The discussion section of the paper may expose authors' bias towards justifying their hypothesis, irrespective of how weak the data may be. Thus, it is in the discussion and conclusion sections where truth most often becomes opinion. However, these opinions and not the more complicated data are most often picked up by the media and eventually presented to the public at large.

So, returning to the original question: When does “statistically significant” equate to “clinically relevant?” The medical literature does not make the answer simple for the practicing clinician. An excellent example comes from two reports out of the same investigation, namely, the Women's Health Initiative. When comparing placebo to conjugated equine estrogens and medroxyprogesterone acetate on breast

cancer outcome in postmenopausal women, the investigators report in the results section that the difference reaches “almost nominal statistical significance” (ie, not statistically significant). But, in the discussion section, they emphasize “the substantial risks for cardiovascular disease and breast cancer...”

In another example from the Women’s Health Initiative, looking for differences in parameters of “quality of life,” the investigators state in the results section that “there were small but statistically significant positive effects of estrogen plus progestin on physical functioning ($P<0.001$), bodily pain ($P<0.001$), and sleep disturbance ($P<0.001$).” However, in their conclusion they state that “in this trial estrogen plus progestin did not have a clinically meaningful effect on health-related quality of life.”³ In this instance, it would appear that there is a bias towards confirming adverse effects of EPT. That is, in the first example the breast cancer findings are reported as statistically insignificant but are regarded as clinically relevant. In the other instance, the findings are reported as statistically significant but, in the opinion of the investigators, clinically irrelevant.

Statistically significant indeed may be clinically irrelevant under many circumstances. A study reporting a few additional cases of a severe disease per 10,000 women per year in an older population may be clinically insignificant for a younger population in which the prevalence of that disease is extremely low. Yet, the reverse may not hold. That is, a marginally significant increased complication rate of a drug in a younger population may be very relevant clinically in an older population with an already existing higher prevalence of that disease.

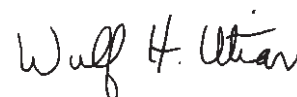
Thus, a statistically significant finding may indicate either real benefit or risk, depending on specific circumstances. Consequently, the application of these findings to decision-making in clinical practice may be more challenging. Each study really represents one piece of a giant jigsaw puzzle. Moreover, it should be apparent that there is no such thing as a “definitive study” that will provide answers for all populations.

All of this exposes a huge flaw in the brave new world of “evidence-based medicine.” A large proportion of clinical practice is based on decades of experience, but not necessarily on confirmative research studies. In some instances, the situation is so obvious that it would be absurd to demand further evidence. For example, surgery for acute appendicitis without a published evidence-based foundation is standard practice; and to demand evidence, for instance, on the basis of a randomized trial against antibiotic therapy would appear both dangerous and ridiculous. On the other hand,

particularly in areas of preventive medicine, it may not be possible to directly extrapolate excellent research studies to a different population that varies by age or other risk factors; and a decision as to future action requires difficult clinical judgments to determine potential risk or benefit.

Clinical practitioners should not despair. Nor should they grant excess credence to any one research report, meta-analysis, or focused review. Instead, they should consider all facts relevant to each individual patient. The unique profile inevitably requires a unique clinical opinion. Provided that a patient is given an understandable summary of the balance of risk and benefit, that discussion is documented, and that appropriate follow-up planned, you are practicing good medicine. Ultimately, whether “statistically significant” is “clinically relevant” still has to be decided between the woman and her health provider, and that remains the challenge of the “practice of medicine!”

In this issue of *Menopause Management* we publish the 2003 NAMS Position Statement on the use of hormone therapy in peri- and postmenopausal women. An expert panel wrestled with the dilemma discussed here, and the final report serves as an excellent example of how all studies need to be considered when developing appropriate clinical approaches to situations not yet fully resolved, but for which there are multiple—and often conflicting—scientific publications.



Wulf H. Utian, MD, PhD
*Executive Director and
Honorary Founding President
The North American Menopause Society*

References

1. Bailar J. Hormone replacement therapy and cardiovascular disease. *N Engl J Med* 2003;349:521-2.
2. Writing Group for the Women’s Health Initiative Investigators. Risks and benefits of estrogen plus progestin in healthy postmenopausal women. *JAMA* 2002;288:321-33.
3. Hays J, Ockene JK, Brunner RT, et al, for the Women’s Health Initiative Investigators. Effects of estrogen plus progestin on health-related quality of life. *N Engl J Med* 2003;348:1839-54.