Surrogate end points in women's health research: science, protoscience, and pseudoscience

David A. Grimes, M.D., Kenneth F. Schulz, Ph.D., M.B.A., and Elizabeth G. Raymond, M.D., M.P.H.

Family Health International, Research Triangle Park; and Department of Obstetrics and Gynecology, University of North Carolina School of Medicine, Chapel Hill, North Carolina

A surrogate end point (e.g., a laboratory test or image) serves as a proxy for a clinical end point of importance (e.g., fracture, thrombosis, or death). Adoption and use of surrogate end points lacking validation, especially in cardiovascular medicine, have caused thousands of patients' deaths, a serious violation of the ethical principle of beneficence. (Fertil Steril® 2010;93:1731–4. ©2010 by American Society for Reproductive Medicine.)

"People studying biomarkers don't understand the concept of validation." (1)

Uncritical use of unproven surrogate end points is medically dangerous (2). Basing clinical practice on such biomarkers has caused the needless deaths of thousands of patients (3–5). Stemming from widespread naïveté about surrogate end points, this threat is real and ongoing, analogous to uncritical adoption of unproven technologies (6).

This problem is acute in the field of hormonal contraception. Based on surrogate end points lacking validation, a recent report inferred risks of venous thromboembolism (VTE) in women using depo-medroxyprogesterone acetate (DMPA) for contraception (Table 1) (7). The Food and Drug Administration required addition of a "black box warning" to the label of DMPA restricting use because of concerns about possible fracture risk; this was inferred from data showing that the drug is associated with bone mineral density changes (8–9). Another report warned about VTE risk with levonorgestrel 1.5 mg as an emergency contraceptive after noting effects on hemostatic factors (10). Based on only 12 women, the authors advised that, "For individuals or patients with a genetic predisposition or a transiently disturbed haemostatic balance even the small changes like the ones seen in the present study might be harmful" (10). The same first author proposed that sex hormone-binding globulin might be a "risk marker" for VTE, also without credible validation (11).

Others have advised against prescribing specific combined oral contraceptives based on a single laboratory test (12). After observing that pills containing drospirenone increased the normalized activated protein C sensitivity ratios, researchers concluded that these pills should not be used for new oral contraceptive clients because they confer an increased risk of thrombosis (12). A subsequent prospective cohort study of almost 60,000 women failed to confirm this conclusion (13). Other researchers have similarly inferred an elevated risk of thromboembolism from the contraceptive patch based on its effects on various blood tests without first proving the clinical significance of those effects (14). Such unscientific inferences can damage women's health by depriving them of effective contraceptive options.

Received November 12, 2009; revised December 21, 2009; accepted December 22, 2009; published online February 12, 2010.
D.G. has nothing to disclose. K.S. has nothing to disclose. E.R. has nothing to disclose.
Supported in part by Family Health International (FHI) with funds from the U.S. Agency for International Development and the National Institutes of Health. Views expressed in this article do not necessarily reflect those of FHI or the funding agencies.
Reprint requests: Dr. Grimes, Family Health International, P. O. Box 13950, Research Triangle Park, NC 27709 (FAX: 919-544-7261; E-mail: dgrimes@fhi.org).

A HIERARCHY OF SCIENCES

Types of science may be viewed in three levels, in decreasing order of credibility.

Level One: Science

Science is the branch of knowledge that produces theoretic explanations for natural phenomena based on experiments and observations (15). Other definitions state that science is systematic knowledge of the world gained through observation and experimentation (16). The key principle is empiricism: science is testable and its theories refutable. When its hypotheses are tested then confirmed or rejected, it becomes science.

Level Two: Protoscience

Protoscience denotes a new area of knowledge still in the process of becoming established as legitimate (17). Like science, protoscience follows scientific principles, including the willingness to be refuted by evidence or replaced by more credible theories. As such, protoscience can be considered nascent science. Its theories and predictions are consistent with known evidence. However, they have yet to be tested empirically.

Protoscience can ascend to legitimate science. For example, Wegener's theory of continental drift was promoted to science after the mechanism of plate tectonics was documented. The protosciences of alchemy and astrology arose before the scientific method but later spawned the sciences of chemistry and astronomy. Most surrogate end points used today, such as coagulation tests to infer VTE risks with hormonal contraception, reflect protoscience. They correlate

TABLE 1

Examples of unproven and invalid surrogate end points in women's health.

Intervention	Surrogate end point	Clinical end point of interest	Validation	Study
Unproven				
Depo-medroxyprogesterone acetate for contraception	Coagulation tests	Venous thromboembolism	Not done	Randomized controlled trial, 14 participants, all given same drug, no clinical end points (7)
Depo-medroxyprogesterone acetate for contraception	Bone mineral density	Fracture	Not done	Black box warning from FDA (8)
Levonorgestrel for emergency contraception	Coagulation tests	Venous thromboembolism	Not done	Randomized controlled trial, 12 participants, no clinical end points (10)
Combined oral contraceptives	Sex hormone–binding globulin	Venous thromboembolism	Not done	Randomized trial, 35 participants, no clinical end points (11)
Combined oral contraceptives containing drospirenone	Coagulation test	Venous thromboembolism	Not done	Before-after study, 156 participants, no clinical end points (12)
Transdermal hormonal contraceptive	Coagulation tests	Venous thromboembolism	Not done	Randomized controlled trial, 24 participants, no clinical end points (14)
Invalid clinical correlate				
Menopausal estrogen-progestin	Lipoprotein levels	Myocardial infarction	Lipoprotein profile improved, but more heart disease	Randomized controlled trial (20)
Fluoride	Bone mineral density	Fracture	No effect or harmful effect on fracture	Randomized controlled trials (25, 26)

Grimes. Surrogate end points. Fertil Steril 2010.

with clinical end points, seem logical in light of current knowledge, but have yet to be validated (Table 1) (18).

Level Three: Pseudoscience

Pseudoscience, at the bottom of the hierarchy, purports to be science but does not follow scientific principles (16, 19). It commonly makes claims lacking evidence or claims that conflict with evidence. It may fail to provide the opportunity for testing and refutation. The passion of its advocates tends to be inversely related to the objective evidence available.

Protoscience can also descend to pseudoscience. The contemporary belief that zodiac signs govern human events reflects pseudoscience. Although alchemy spawned chemistry, to suggest that base metals can be transformed into gold constitutes pseudoscience. Using lipoprotein profiles to predict cardiovascular disease in women taking menopausal hormone therapy constitutes pseudoscience, because these biomarkers have been invalidated as surrogate end points (Table 1) (20).

RECOMMENDED NOMENCLATURE

Because of semantic confusion in the literature, the National Institutes of Health (NIH) sponsored a workshop to propose uniform terminology for biomarkers and surrogate end points (21, 22). Outcomes in clinical research can be viewed as belonging to three strata of decreasing clinical relevance (Table 2).

First Tier: Clinical End Points

The top tier includes clinical end points, such as illness, pregnancy, or death. Venous thromboembolism and fracture risk were clinical outcomes of interest in several of the studies cited above.

Second Tier: Valid Surrogate End Points

The middle tier includes valid surrogate end points that substitute for clinical outcomes of importance. A surrogate end point (a subset of biomarkers) is a proxy for an outcome that is rare (VTE) or that takes a long time to develop (cancer). To be validated, a surrogate end point must both correlate with the true outcome of interest and fully capture the effect of the treatment on the true outcome. This means that the surrogate accurately predicts the effect of the treatment on the clinical outcome of interest. Though most proposed surrogate end points meet the first criterion, few fulfill the second (23). Correlation with the true outcome is necessary but insufficient. The NIH conference nearly a decade ago advised against the term "surrogate marker," because it implies a substitute for a substitute rather than a substitute for a clinical end point of interest (21).