

## Epidemiologic Theory Counts

### *Harm, Knowledge, Action, and the People's Health*

In this final chapter, I address in one more way the question undergirding this book: why care about epidemiologic theories of disease distribution—past, present, or future? After all, theoretical debates—including over even what constitutes “evidence” and how to think about it—are commonplace in scientific disciplines, and in fact core to the critical conduct of science (Cohen, 1985; Ziman, 2000; Mjøset, 2002; Sober, 2008). Vying for the most intellectually and empirically sound, robust, coherent, and comprehensive account of the phenomena of interest has long been standard scientific practice—and there is no reason the field of epidemiology should be any exception.

But in epidemiology, a distinction does matter. Theoretical debates about substantive explanations for disease occurrence are not simply about solving a puzzle correctly. Nor are they merely ivory-tower “academic” disputes. The difference has to do with our domain: “population distributions of disease, disability, death, and health and their determinants and deterrents, across time and space” (Krieger, 2001a; *see* **Chapter 1**). Ethically and intellectually, we are bound, in the time-honored phrase of the Hippocratic Oath, at the very least “to do no harm” (The Oath, 1983, p. 67; Nutton, 2004, pp. 66–69). Making this responsibility amply clear is the necessity of epidemiologic research to conform to the 1964 Declaration of Helsinki regarding “Ethical Principles for Medical Research Involving Human Subjects” (World Medical Association, 1964 [2008]), along with whatever are our country-specific regulations regarding human research.

However, in the case of epidemiology, I would extend the current argument to state that our accountability reaches beyond the injunction not to harm individuals who participate in epidemiologic studies. Also at issue are the ways epidemiologic theories and the research they animate contribute—potentially helpfully, potentially harmfully—to efforts to understand and address who and what is responsible for observed burdens and distributions of health, disease, and well-being.

Hence, to conclude the case that epidemiologic theory counts, in this final chapter I consider several examples in which choice of epidemiologic theory has made all the difference in whether people's health is harmed or helped.

## First Do No Harm—and Second, Do Some Good: Some Examples of Why Theory Matters

Spanning a variety of disease outcomes, and involving a range of epidemiologic theories, I consider below four examples that illustrate the limitations of relying solely on either a biomedical approach stripped of societal context or a social approach focused only on social exposures—and why instead we are better served by theories attentive to society and biology, in context.

**Example 1: Hormone Therapy, Cardiovascular Disease, and Breast Cancer—Biomedical Disregard for Social Determinants of Health Leads to Iatrogenic Disease.** Widespread use of hormone therapy (HT) in the United States and other affluent countries took off in the 1960s (Seaman & Seaman, 1977; McCrae, 1983; Lock, 1993; Seaman, 2003; Krieger et al., 2005; Stefanick, 2005; Houck, 2006; Watkins, 2007). The promise was to alleviate menopausal symptoms and to keep women “*Feminine Forever*,” as proclaimed in 1966 by the title of one widely influential U.S. book (Wilson, 1966), whose publication and distribution was later revealed to be sponsored by Ayerst, a pharmaceutical company (Seaman & Seaman, 1977) (see **Textbox 8–1**). During this time, the idea that hormonal “deficiencies” led to menopause morphed into the biomedical view that menopause was itself a hormonal “deficiency disease” that could and should be treated by administration of “female” sex hormones (e.g., estrogen) (U.S. Federal Security Agency, 1950; Ayerst Labs, 1960; Rhoades, 1965; Wilson, 1966; Castallo, 1967; Seaman & Seaman, 1977; McCrea, 1983; Seaman, 2003; Krieger et al., 2005; Houck, 2006; Watkins, 2007; Foxcroft, 2009).

As background to this practice, starting in the early 1900s physicians had begun prescribing tablets made from cow ovaries to women undergoing menopause (Seaman, 2003; Stefanick, 2005; Houck, 2006; Watkins, 2007). Not surprisingly, results were mixed, given no control over the myriad substances present in cow ovaries, let alone in what quantities. By the 1930s, new laboratory techniques enabled pharmaceutical companies to start producing and marketing more carefully formulated versions of what was then termed “female sex hormone therapy” (Oudshorn, 1994; Houck, 2006; Watkins, 2007). They did so, ironically (in hindsight), at the same time laboratories were first discovering estrogen’s carcinogenicity (Krieger et al., 2005; Gaudillière, 2006). Opening the door to even wider use, in 1942 the U.S. Food and Drug Administration (FDA) approved Premarin (an estrogen-only HT product) for treating menopause (Seaman, 2003; Stefanick, 2005; Watkins, 2007). Thereafter, the pace of the mass production and marketing of what became termed *hormone replacement therapy* picked up, as it became increasingly acceptable for physicians to prescribe and women to take hormonal pills to regulate their sexual and reproductive health, as also sanctioned by the FDA’s approval in 1960 of the oral contraceptive pill (Gordon, 1976; Seaman & Seaman, 1977; Oudshorn, 1994; Seaman, 2003; Houck, 2006; Watkins, 2007). Although some researchers at the time did raise questions about the soundness of the science favoring HT and expressed concerns about HT-associated risks, their work had little impact (Krieger et al., 2005; Houck, 2006; Watkins, 2007).

As use of HT increased in the 1960s, so did hypotheses as to its benefits, leading to its reframing as not simply a “restorative” but actually “preventive” medicine. Fueling this thinking was recognition of differences for women versus men in the epidemiology of cardiovascular disease, an outcome whose rising mortality rates had led it to become, by the early 1960s, a—if not *the*—leading cause of death in most affluent countries (Marmot & Elliott, 2005). Specifically, age of onset was earlier among men compared to women,

**Textbox 8-1. A Half-Century's Changing and Contested Perspectives and Data on—and Recurrent Questions About—Menopause, its Medicalization, and Hormone Therapy: Biomedical and Social Accounts, 1950–2005.**

**United States, Federal Security Agency, Public Health Service. *Menopause*. (Health Information Series No. 15). Washington, DC: US Govt Printing Office, 1950.**

p. 1: "Menopause means the end of menstrual periods, and therefore, the end of childbearing years. It is also called the *climacteric* or '*change of life*' (italics in original). It is nature's plan for protecting women against childbearing beyond their years of greatest physical energy."

p. 2: "Many women have very few mild symptoms; some have none at all; with a few, the discomfort is very severe.

The symptoms are caused by the disappearance of the female sex hormone which the ovaries produce. The same ones occur when the ovaries are removed surgically because of disease (surgical menopause). After a period of months or a year or two, the body adjusts itself and the symptoms disappear. While this adjustment is taking place, hot flushes, etc., can appear.

Modern medical treatment is very successful in relieving symptoms of menopause. The doctor gives his patient medicine containing the ovarian hormones (or chemicals which act like it). In other words, he puts back in her body what nature is no longer producing. The treatment is continued until her body adjusts itself and hot flashes and other symptoms disappear. Treatment is necessary for a period of several months or for a year or two. *Medical care at this time also helps to correct the causes of nervousness and low spirits that often go along with the menopause* (italics in original)..."

p. 3: "Don't discuss your emotional or physical worries with relatives and neighbors during menopause; they haven't the medical knowledge to help you. Too, they might pass on some ancient 'change of life' superstition that could worry you. Rely upon your doctor or hospital clinic for information, advice, and medical treatment..."

Remember that the menopause is not a complete change of life. The normal sex urges remain and women retain their usual reaction to sex long after the menopause. *There is nothing abnormal about the change of life and nothing unusual about the continuation of happy married relations afterward.*" (italics in original)

p. 3: "Each woman should see her doctor or consult a hospital clinic as soon as she notices she is starting her 'change of life'... *The modern woman relies upon her doctor for advice and goes through menopause with a minimum of physical and emotional discomfort.*" (italics in original)

**Wilson RA. *Feminine Forever*. New York: M. Evans (distributed in association with Lippincott), 1966.**

p. 18: "... menopause—far from being an act of fate or a state of mind—is in fact a deficiency disease. By way of rough analogy, you might think of menopause as a condition similar to diabetes. Both are caused by a lack of a certain substance in the body chemistry. To cure diabetes, we supply the lacking substance in the form of insulin. A similar logic can be applied to menopause—the missing hormones can be replaced."

p. 25: "In short, menopause must at last be recognized as a major medical problem in modern society. Women, after all, have the right to remain women. They should not have to live as sexual neuters for half their lives. The treatment and cure of menopause thus becomes a social and moral obligation."

pp. 43–44: "... while women during their fertile years are virtually immune to coronary disease and high blood pressure, the menopausal woman—lacking female hormones—soon loses this advantage and becomes as prone to heart trouble and strokes as a man of similar age. These are the secondary effects of her castration.

As for the primary effects, they are quite simple. Deprived of its natural fluids, the entire genital system dries up. The breasts become flabby and shrink, and the vagina becomes stiff and unyielding. The brittleness often causes chronic inflammation and skin cracks that become infected and make sexual intercourse impossible.

Additional physical consequences of castration [include]... nervousness, irritability, anxiety, apprehension, hot flushes, night sweats, joint pains, melancholia, palpitations, crying spells, weakness, dizziness, severe headache, poor concentration, loss of memory, chronic indigestion, insomnia, frequent urination, itching of the skin, dryness of eye, nose, and mouth, and a backache...

The effects of menopausal castration, as is evident from this list of symptoms, are by no means confined to the sexual organs. Because the chemical balance of the entire organism is disrupted, menopausal castration amounts to a mutilation of the whole body. I have known cases where the resulting physical and mental anguish was so unbearable that the patient committed suicide.

While not all women are affected by menopause to this extreme degree, no woman can be sure of escaping the horror of this living decay. Every woman faces the threat of extreme suffering and incapacity."

pp. 62–63: "... estrogen may be termed the hormone of feminine attraction and well-being. As it courses through the body in the bloodstream, its effects are indeed varied and manifold. Aside from keeping a woman sexually attractive and potent, it preserves the strengths of her bones, the glow of her skin, the gloss of her hair. It prevents the development of high blood pressure, heart disease, and strokes. It tends to prevent diabetes and diseases of the urinary bladder, it keeps the kidneys from wasting or losing salt in the urine—a vital matter in the regulation of tissue fluids throughout the body...

Through an ingenious mechanism by which estrogen acts on the pituitary gland at the base of the brain, it has a direct effect on a woman's emotional state. To a woman, estrogen acts as the carrier of that mysterious life force that motivates work, study, ambition, and that marvelous urge toward excellence that inspires the best of human beings.

Granted, the achievement of any woman cannot be attributed to estrogen alone—estrogen cannot command the fickle element of plain luck. Even with an abundance of estrogen, she still has to cope with the vicissitudes of existence—heredity, accident, environment, childhood training, financial status, and so forth. But at least estrogen puts her in a position where she can take advantage of whatever lucky breaks come her way. No matter what her particular sphere of activity may be—in the home, in business, in the arts or professions—a woman cannot live up to her opportunities unless she has her full quota of estrogen."

p. 158: "Estrogen therapy, far from causing cancer, tends to prevent it."

pp. 163–164: "These, then, are the facts to keep in mind when you hear estrogen discussed with suspicion: the inept and illogical mouse experiment that gave rise to the estrogen-cancer myth, and the inverse relationship of estrogen to breast and genital cancer (high estrogen in youth: low cancer incidence... low estrogen in age: high cancer incidence).

Keeping a woman rich in her ovarian hormones, especially estrogen and progesterone, lessens the incidence of malignant lesions, including breast and genital cancer."

**Boston Women's Health Book Collective. *Our Bodies, Ourselves*. Revised and expanded. New York: Simon & Schuster, 1976.**

p. 327: "Chapter 17. Menopause. Even though menopause has been a neutral or positive experience for many women, the physical and emotional changes associated with it are often misunderstood and mystifying... Even when physical symptoms are minimal or under control, menopause is often a more negative experience than it needs to be because of our society's attitude toward us during that time. The popular stereotype of the menopausal woman has been primarily negative: she is exhausted, irritable, unsexy, hard to live with, irrationally depressed, unwillingly suffering a 'change' that marks the end of her active (re)productive life."

p. 328: "In a society which equates our sexuality with our ability to have children, menopause is wrongly thought to mean the end of our sexuality—the end of our sexual pleasure, or even the total end of our sex life."

p. 330: "About one out of every five women will have no (or just a few) menopausal symptoms. Although most women do experience some bothersome symptoms, many of these will not actually require treatment. We should seek help whenever symptoms interfere with our normal activities, particularly because continuous and unrelieved distress may result in depression."

pp. 330–332: "Estrogen replacement therapy (ERT)... Those doctors who view menopause as a deficiency disease strongly argue the usefulness of widespread estrogen therapy. Others recommend a much more cautious approach."

ERT involves serious risks. Most notably, ERT has been linked with endometrial (uterine) cancer, especially among women taking estrogen for longer than one year... As with the pill, taking estrogen increases the risk of blood clots and hypertension...

Much more research is needed in the area of ERT, particularly its effect on osteoporosis, heart disease, cancer, and mental depression. Because of present uncertainties, many women are very cautious about ERT, choosing it only when symptoms are severe and when no contraindications are present."

p. 333: "Those of us looking ahead to menopause or just beginning to experience it can find little material that explains what most women go through during menopause. Most research has been done on 'clinical samples'—that is, on the minority of women who have chosen or been forced to seek medical care because of the severity of their symptoms. Consequently, we know very little about what menopause is like for all the women who never seek medical help."

**Petitti DB, Perlman JA, Sidney S. Postmenopausal estrogen use and heart disease. *N Engl J Med* 1986; 315:131–132.**

p. 131: "Findings from our recently completed study of mortality and hospitalization for cardiovascular disease in postmenopausal estrogen users, based on information from the Walnut Creek Contraceptive Drug Study, may shed light on the reason for the contradictory results of other studies on the relation between cardiovascular disease and postmenopausal estrogen use."

The Walnut Creek Contraceptive Drug Study has been described in detail elsewhere. Briefly, 16,638 women between the ages of 18 and 54 were recruited to the study in the late 1960s and early 1970s and provided information on their use of oral contraceptives and other sex-steroid hormones at recruitment and approximately yearly through 1978. The current analyses uses data from follow-up for mortality through December 31, 1983..."

p. 132: "After adjustment for smoking, alcohol consumption, and history of hypertension, the [age-adjusted] relative risk of death from cardiovascular disease was 0.5 in women with postmenopausal estrogen use... The [age-adjusted] relative risk of

death from accidents, homicide, and suicide was also substantially lower in postmenopausal estrogen users than in women who had never used sex-steroid hormones, and this lower risk persisted after adjustment for smoking, alcohol consumption, Quetelet index, and history of hypertension.

There is no biologically plausible reason for a protective effect of postmenopausal estrogen use on mortality from accidents, homicide, and suicide. We believe our results are best explained by the assumption that postmenopausal estrogen users in this cohort are healthier than those who had no postmenopausal estrogen use, in ways that have not been quantified and cannot be adjusted for. The selection of healthier women for estrogen use in this population is not necessarily a characteristic shared by other populations. If there are differences in the degree of selection for postmenopausal estrogen use between populations, then the disparate results of studies published to date are understandable. In the face of such selection, the question of the effect of postmenopausal estrogen use may be answerable with validity only through a randomized clinical trial."

**Rosenberg L. Hormone replacement therapy: the need for reconsideration.**  
*Am J Public Health* 1993; 83:1670-1673.

p. 1670: "Millions of menopausal women are taking hormone supplements. Observational studies suggest that unopposed estrogens reduce the risk of cardiovascular disease and fractures and increase the risk of endometrial cancer and, possibly, breast cancer. In the absence of information from randomized trials, how much of the apparent beneficial effect on heart disease is due to the tendency of healthier women to use these drugs is unknown. The effect on the cardiovascular system of estrogen taken with a progestin is unknown, and this regimen may increase the risk of breast cancer. An approach to health and illness that focuses on a single cause or preventive and on single organ systems is severely limited. Alternative ways to improve cardiovascular health and skeletal health that do not increase the risk of cancer are available. A reconsideration of the appropriate use of hormone supplements is needed."

p. 1671: One reason for the use of hormone supplements is the view held by some that menopause is a deficiency disease that requires treatment, as exemplified by the use of the term "estrogen-deficient" to characterize menopausal women. Menopausal women are indeed deficient in endogenous estrogens, relative to premenopausal women, if the focus is solely on the role of endogenous estrogens as preventives of fractures and heart disease. If the focus is shifted to the role of endogenous estrogens in the etiology of, for example, breast cancer (which occurs more frequently in women who have an early menarche or a late menopause), then premenopausal women are "hyperestrogenic" and postmenopausal women have a more desirable level. The same could be said for other diseases as well, including endometrial cancer, ovarian cancer, and uterine fibroids..."

**Writing Group for the Women's Health Initiative Investigators. Risk and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial.** *JAMA* 2002; 288:321-333.

p. 321 (abstract):

CONTEXT: Despite decades of accumulated observational evidence, the balance of risks and benefits for hormone use in healthy postmenopausal women remains uncertain.

OBJECTIVE: To assess the major health benefits and risks of the most commonly used combined hormone preparation in the United States.

**DESIGN:** Estrogen plus progestin component of the Women's Health Initiative, a randomized controlled primary prevention trial (planned duration, 8.5 years) in which 16608 postmenopausal women aged 50–79 years with an intact uterus at baseline were recruited by 40 U.S. clinical centers in 1993–1998.

**INTERVENTIONS:** Participants received conjugated equine estrogens, 0.625 mg/d, plus medroxyprogesterone acetate, 2.5 mg/d, in 1 tablet ( $n = 8506$ ) or placebo ( $n = 8102$ ).

**MAIN OUTCOMES MEASURES:** The primary outcome was coronary heart disease (CHD) (nonfatal myocardial infarction and CHD death), with invasive breast cancer as the primary adverse outcome. A global index summarizing the balance of risks and benefits included the 2 primary outcomes plus stroke, pulmonary embolism (PE), endometrial cancer, colorectal cancer, hip fracture, and death due to other causes.

**RESULTS:** On May 31, 2002, after a mean of 5.2 years of follow-up, the data and safety monitoring board recommended stopping the trial of estrogen plus progestin vs placebo because the test statistic for invasive breast cancer exceeded the stopping boundary for this adverse effect and the global index statistic supported risks exceeding benefits. This report includes data on the major clinical outcomes through April 30, 2002. Estimated hazard ratios (HRs) (nominal 95% confidence intervals [CIs]) were as follows: CHD, 1.29 (1.02–1.63) with 286 cases; breast cancer, 1.26 (1.00–1.59) with 290 cases; stroke, 1.41 (1.07–1.85) with 212 cases; PE, 2.13 (1.39–3.25) with 101 cases; colorectal cancer, 0.63 (0.43–0.92) with 112 cases; endometrial cancer, 0.83 (0.47–1.47) with 47 cases; hip fracture, 0.66 (0.45–0.98) with 106 cases; and death due to other causes, 0.92 (0.74–1.14) with 331 cases. Corresponding HRs (nominal 95% CIs) for composite outcomes were 1.22 (1.09–1.36) for total cardiovascular disease (arterial and venous disease), 1.03 (0.90–1.17) for total cancer, 0.76 (0.69–0.85) for combined fractures, 0.98 (0.82–1.18) for total mortality, and 1.15 (1.03–1.28) for the global index. Absolute excess risks per 10,000 person-years attributable to estrogen plus progestin were 7 more CHD events, 8 more strokes, 8 more PEs, and 8 more invasive breast cancers, while absolute risk reductions per 10,000 person-years were 6 fewer colorectal cancers and 5 fewer hip fractures. The absolute excess risk of events included in the global index was 19 per 10,000 person-years.

**CONCLUSIONS:** Overall health risks exceeded benefits from use of combined estrogen plus progestin for an average 5.2-year follow-up among healthy postmenopausal U.S. women. All-cause mortality was not affected during the trial. The risk-benefit profile found in this trial is not consistent with the requirements for a viable intervention for primary prevention of chronic diseases, and the results indicate that this regimen should not be initiated or continued for primary prevention of CHD.

**NIH State-of-the-Science Panel. National Institutes of Health  
State-of-the-Science Conference Statement: Management of  
Menopause-Related Symptoms. *Ann Int Med* 2005; 142:1005–1013.**

p. 1005: "Menopause is a natural process that occurs in women's lives as part of normal aging. Many women go through the menopausal transition with few or no symptoms... The focus of this report is to identify menopausal symptoms and assess treatments for them on the basis of existing scientific evidence..."

p. 1005: "Estrogen, either by itself or with progestins, has been the therapy of choice for decades for relieving menopause-related symptoms. Epidemiologic studies in the 1980s and 1990s suggested that estrogen-containing therapy might protect women from heart disease and other serious medical problems. The Women's Health Initiative (WHI) was a large clinical trial of postmenopausal women (age range, 50 to



79 years [mean, 63.2 years]) that was designed to see whether estrogen with or without progestin therapy could prevent chronic conditions, such as heart disease and dementia. The estrogen with progestin portion of the trial ended early because of increased incidence of breast cancer. There were increases in blood clots, stroke, and heart disease among women who received this treatment as well. These findings raised serious questions about the safety of estrogen to treat symptoms of menopause. Many women stopped hormone replacement therapy, and some searched for alternative therapies. To reflect a shift of focus from 'replacement' to use of hormones for relief of symptoms, we will use the term *menopausal hormone therapy*, which includes a range of doses and preparations of estrogens and progestins."

p. 1010: " 1. Menopause is the permanent cessation of menstrual periods that occurs naturally in women, usually in their early 50s. Many women have few or no symptoms; these women are not in need of medical treatments... 6.... much more research is needed to clearly define the natural history of menopause, associated symptoms, and effectiveness and safety of treatment for bothersome symptoms. Natural histories are important for both science and policy. Knowing how many women transit menopause with few or no symptoms, and how many manage menopause largely on their own, can lead to public health information that empowers women and increases their self-reliance... "

pp. 1010-1011: "Menopause is 'medicalized' in contemporary U.S. society. There is great need to develop and disseminate information that emphasizes menopause as a normal, healthy phase of women's lives and promotes its demedicalization. Medical case and future clinical trials are best focused on women with the most severe and prolonged symptoms. Barriers to professional care for these women should be removed."

and their and overall rates of were mortality higher (Bush, 1990; Epstein, 2005; Stamler, 2005). Based on the biomedical view that sex hormones were chiefly responsible for myriad male/female differences, including not only for reproductive health but also for nonreproductive health outcomes and other behavioral and cognitive characteristics and capacities (Dreifus, 1977; Ruzek, 1978; Hubbard et al., 1982; Oudshorn, 1994; Fee & Krieger, 1994), this framework led to positing that "female sex hormones" were protective against cardiovascular disease.

In the mid-1960s the "Coronary Drug Project" became the first clinical trial of estrogen to prevent cardiovascular disease—and was conducted solely with men (Coronary Drug Project, 1970; Coronary Drug Project, 1973). Its results defied expectation: the estrogen arm of the study was abruptly halted in the mid-1970s because the men given Premarin had unexpectedly high cardiovascular mortality (resulting from thromboembolic events and myocardial infarction) (Coronary Drug Project, 1970; Coronary Drug Project, 1973; Petitti, 2004; Stefanick, 2005). Further crimping enthusiasm for HT, shortly thereafter, epidemiologic research implicated HT (still estrogen-only) in rising rates of endometrial cancer (Seaman & Seaman, 1977; McCrae, 1983; Seaman, 2003; Stefanick, 2005; Houck, 2006; Watkins, 2007).

The turnabout occurred in the 1980s with the introduction of estrogen-progestin HT formulations, whereby progestin was added to prevent estrogen-induced endometrial cancer (Seaman, 2003; Stefanick, 2005; Houck, 2006; Watkins, 2007). During the 1980s,



a raft of observational studies reported that HT use was associated with reduced risk of osteoporosis, cardiovascular disease risk markers, and cardiovascular disease among women (Seaman, 2003; Stefanick, 2005; Houck, 2006; Watkins, 2007). By the early 1990s, major epidemiologic reviews (including meta-analyses) concluded that HT warranted use as preventive medication (Stampfer & Colditz, 1991; Grady et al., 1992). Even so, a number of articles in the epidemiologic literature warned about the dangers of causal inference based on observational data and also raised concerns about risk of breast cancer (Petitti et al., 1986; Barrett-Connor, 1992; Brinton & Schairer, 1993; Rosenberg, 1993), albeit with little impact. Consequently, in 1991 and 1992, considering the extant human observational evidence as well as animal and other laboratory studies, the FDA and American College of Physicians respectively recommended use of HT to prevent cardiovascular disease (Seaman, 2003; Stefanik, 2005). The argument was that any small risk for breast cancer was outweighed by the larger danger of future cardiovascular disease.

In 2002, however, routine acceptance of HT was shattered by results from the Women's Health Initiative (WHI; Writing Group/WHI, 2002), the largest randomized clinical trial of HT, conducted in a population of mainly healthy U.S. women, which like the smaller 1998 Heart and Estrogen/progestin study (Hulley et al., 1998) and its 2002 follow-up (Grady et al., 2002), found that contrary to expectations, HT did not decrease—and may, in fact, have increased risk of cardiovascular disease—while also confirming that long-term use of the combined estrogen plus progestin HT increased risk of breast cancer. Publication of the results triggered a dramatic global drop in HT prescriptions and marketing, with the U.S. decline in HT use by 2004 exceeding 50% compared to just before the WHI results (Ettinger et al., 2003; Hersh et al., 2004; Buist et al., 2004; Majumdar et al., 2004; Haas et al., 2004; Stefanick, 2005; Kelly et al., 2005; Kim et al., 2005; Wei et al., 2005; Hing & Brett, 2006; Guay et al., 2007).

The findings also led to fierce debates in the epidemiologic literature over reasons for the discrepancies between the observational and clinical trial results (Humphrey et al., 2002; Grodstein et al., 2003; Stampfer, 2004; Vandenbroucke, 2004; Barret-Connor, 2004; Kuller, 2004; Petitti, 2004; Lawlor et al., 2004a; Prentice et al., 2005a; Barrett-Connor et al., 2005; Willett et al., 2006; Manson & Bassuk, 2007; Barrett-Connor, 2007; Banks & Canfell, 2009). Diametrically opposed hypotheses posited: (1) confounding, selection bias, and measurement error biased the observational studies, versus (2) the clinical trial results were misleading because the women in the WHI were typically older and more likely to have atherosclerosis (and, hence, have plaque vulnerable to HT-caused disruption) than women in the observational studies (this latter hypothesis, also known as the “timing” hypothesis, has not been supported, at the time of preparing this chapter, by the most recent re-analysis of WHI data; Prentice et al., 2009). Further spurring the debates, new legal evidence has revealed that between 1998 and 2005, 26 peer-reviewed scientific papers backing use of HT and downplaying its risks, published in 18 different medical journals (including such respectable publications as the *American Journal of Obstetrics and Gynecology*), were secretly written by medical ghostwriters paid by Wyeth, the pharmaceutical company that manufactures Premarin; the listed academic authors, who did little or no writing, were solicited solely to provide credibility (Singer, 2009).

Where does epidemiologic theory fit into this saga? Centrally: by illuminating who and what is studied—as well as who and what is ignored. In this case, the contending perspectives pit the biomedical model against a social epidemiologic perspective (Krieger, 2005a).

Notably, starting in the latter part of the 1980s, some epidemiologists began to question whether the supposed protective effect of long-term use of HT on risk of cardiovascular

disease largely reflected a combination of selection bias and confounding by social class (Petitti et al., 1986; Rosenberg, 1993). Behind this concern lay an awareness—amply confirmed by subsequent research (Humphrey et al., 2002; Nelson et al., 2002; Lawlor et al., 2004b; McPherson & Hemminki, 2004)—that women from more affluent backgrounds (childhood as well as current), with better health (including no contraindications against exogenous hormone use), were the women most likely to be prescribed (and who could afford) HT—and least likely to be stricken by cardiovascular disease (Krieger, 2003a; Petitti, 2004; Krieger et al., 2005; Barrett-Connor et al., 2005; Rossouw, 2006). This alternative hypothesis, grounded in a social epidemiologic perspective, was nevertheless discounted by biomedically-oriented researchers, in part because they believed controlling for one or two socioeconomic variables was sufficient to deal with confounding by social class (Stampfer & Colditz, 1991; Grodstein et al., 2003; Stampfer, 2004). It received serious attention only after the WHI published its results in 2002. Indeed, the rapidity of the turn-around from being controversial to argue that socioeconomic position could be an important confounder (Grodstein et al., 2003; Stampfer, 2004) to this being a “common sense” proposition (Petitti & Freedman, 2005; Barrett-Connor et al., 2005; Watkins, 2007; Banks & Canfell, 2009) is nothing short of remarkable.

Reckoned in human terms, the cost of biomedical disregard for social epidemiologic critiques of the HT hypothesis can be seen in breast cancer incidence rates. Prompted by the WHI results, new epidemiologic analyses published between 2002 and 2005, using data from the United States, Europe, and Australia, have estimated that HT might account for 10% to 25% of observed breast cancer cases (Beral et al., 2003; Bakken et al., 2004; Coombs et al., 2005a; Coombs et al., 2005b). Moreover, as per **Figure 1–2 in Chapter 1** (Krieger, 2008a), between 2006 and 2009, 14 population-based studies—eight from the United States (Clark et al., 2006; Ravdin et al., 2007a; Ravdin et al., 2007b; Clark et al., 2007; Hausauer et al., 2007; Stewart et al., 2007; Jemal et al., 2007; Glass et al., 2007), five European (Bouchardy et al., 2006; Katalinic & Rawal, 2008; Kumle, 2008; Verkooijen et al., 2008; Parkin, 2009), and one Australian (Canfell et al., 2008), have documented notable and otherwise unexpected annual declines of breast cancer incidence, especially among women age 50 and older with estrogen receptor-positive (ER+) tumors. All have attributed these trends to the dramatic reduction in use of HT following the July 2002 publication of the WHI results.

Whether these declines can be causally linked to reduced HT use is now, not surprisingly, an active area of research. Suggesting the hypothesis has merit, one line of evidence indicates that observed trends are unlikely to result from declines in breast cancer detection or changes in other major risk factors (Smigal et al., 2006; Hausauer et al., 2007; Jemal et al., 2007; Glass et al., 2007; Chlebowski et al., 2009; Roberts, 2009). Specificity of the decline is also shown by new results indicating that at least with the United States, the declines in breast cancer incidence rates occurred solely among the women most likely to use HT—that is, White and living in affluent counties with ER+ tumors, and with no decline evident among the remaining women, both of color and/or living in poorer counties (Krieger et al., 2010). Further supporting the hypothesis is a second line of evidence regarding the biological plausibility of a short lag time between cessation of HT exposure and a decline in risk of developing a detectable incident breast cancer, considering how steroids, including HT, can act as breast cancer tumor promoters (Bradlow & Sepkovic, 2004; Dietel et al., 2005; Yager & Davidson, 2006; Cordera & Jordan, 2006; Ravdin et al., 2007b).

At one level, the recent disturbing findings linking HT use to iatrogenic breast cancer point to the dangers of inadequately understood pharmacologic manipulation of complex hormonal systems (Seaman, 2003; Krieger et al., 2005; Krieger, 2008)—a caution that

ought be remembered (Krieger et al., 2010) when considering past and present proposals to prevent breast cancer by administering regimens of powerful hormones to healthy young women (Hendersen et al., 1993; Pike & Spicer, 2000; de Waard & Thijssen, 2005; Medina, 2005; Tsubura et al., 2008). At another level, these findings underscore how a narrow biomedical conceptualization of menopause as a “disease” ignored alternative possibilities that, from an evolutionary standpoint, menopause could be a beneficial or neutral trait—and that its consequences for health, regardless, must be addressed in relation to many, not just single, health outcomes (Leidy, 1999, NIH, 2005).

Also exposed is the problematic longstanding reductionist framing of sex hormones (construed solely as *sex hormones*) as fundamental to explaining women's and men's behavior and biology (Oudshorn, 1994; Fee & Krieger, 1994; Doyal, 1995; Krieger et al., 2005; Payne, 2006). This theoretical perspective has been importantly challenged by social epidemiologic research attending to critical distinctions between—and the health consequences of—gender and sex-linked biology (Fee & Krieger, 1994; Doyal, 1995; Krieger, 2003b; Payne, 2006), as discussed in **Chapter 7** (see **Table 7-1**). Additional examples illustrating the importance of considering whether one, the other, both, or neither are germane are two figures appearing in **Chapter 1**. The first, **Figure 1-2**, also related to the breast cancer story, depicts secular changes in U.S. twentieth century age-specific breast cancer incidence rates among White women, revealing an especially steep rise—then fall—only among women age 55 years and older (Krieger, 2008a). The second, **Figure 1-3**, shows disease-specific changes in mortality rate ratios for women compared to men for mid-nineteenth to mid-twentieth century England and Wales (Morris, 1955, Morris, 1957, pp. 1–2). In both cases, the observed temporal changes in women's rates of disease and male/female risk differences cannot plausibly be explained simply by intrinsic “fixed” sex-linked traits; alterations in exogenous exposures must also be at play.

A question thus worth asking is: Had “sex hormones” been conceptualized as one particular variety of hormones that affect cell proliferation, rather than as specialized molecules preoccupied with sex, might it have been possible to avoid pharmacologic change of women's hormone levels being portrayed benignly as *hormone replacement therapy* and instead have this more appropriately seen as *hormone manipulation*, with attendant implications for cell proliferation, including increased risk of cancer (Krieger et al., 2005)? Also relevant is the role of the “invisible industrialist” in promoting use of HT (Krieger et al., 2005), a hidden factor strikingly revealed by a political economy of health perspective. The larger implication is that biomedical disregard for social epidemiologic theoretical frameworks can literally harm people's health—along with the corollary that debate over findings as framed by contending theories can lead to new insights and more valid and beneficial results.

**Example 2: Peptic Ulcers, Stress, *Helicobacter pylori*, and Allergies—From Psychosocial to Biomedical Extremes, Leaving Questions About Susceptibility and Unintended Consequences of Treatment Still Open.** The second example offers a contrary case: when singular attention to social determinants can lead to neglect of relevant biophysical exposures—even as wholesale reversion to a decontextualized biomedical approach can both miss etiologic clues and promote therapeutic interventions with unanticipated problems. The example concerns the now well-rehearsed story of peptic ulcers, psychosocial stress, and *H. pylori* (Thagard 1999, pp. 39–97, 364–366; Atherton, 2006), albeit with a twist.

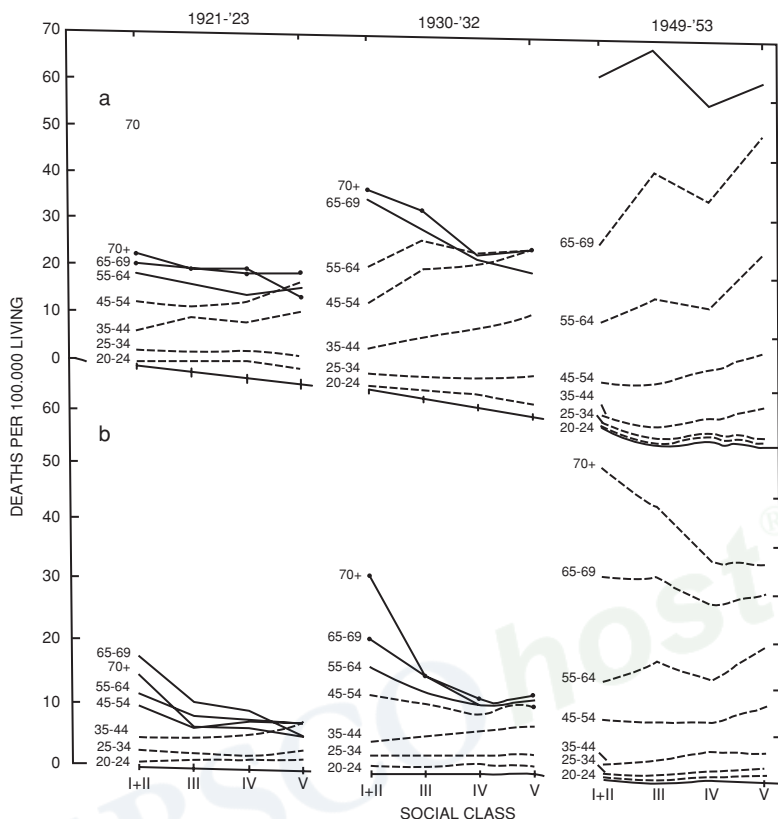
Briefly stated, during much of the twentieth century, the cause of peptic ulcers, following the tenets of psychosocial theories, was attributed to “stress.” The hypothesis was that “stress” increased stomach acids, leading to perforation of the stomach wall and hence

peptic ulcers (Richardson, 1985; Levenstein, 2000; Grob, 2003). With peptic ulcer conceptualized as a “disease of civilization,” the early twentieth century rise and then subsequent fall of peptic ulcer mortality in industrialized countries—shown to reflect changing risks by birth cohort in pathbreaking analyses published in 1962 by Mervyn Susser (b. 1921) and Zena Stein (b. 1921) (Susser & Stein, 1962; Susser & Stein, 2002; **Figure 8–1**)—was attributed to the initial strain of and then adaptation to the quicker pace of life in the twentieth century. Treatments included not only surgery and modified diets but also stress-reduction therapies (Grob, 2003).

Starting in 1984, however, when Barry Marshall (b. 1951) and Robin Warren (b. 1937) discovered an unidentified bacteria (later identified as *H. pylori*) in the stomach of patients with gastritis and peptic ulceration (Marshall & Warren, 1984), the etiologic orientation began to shift, such that by 1994, an NIH consensus document deemed *H. pylori* to be the leading cause of peptic ulcer (NIH, 1994). “Civilization” became re-interpreted as improved hygiene, with better sanitation reducing risk of early life fecal–oral transmission while simultaneously increasing age at infection in childhood (as prevalence was waning); later age at infection, moreover, was posited to increase risk of developing peptic ulcer, leading to its rise in early twentieth century birth cohorts in industrialized countries (Sonnenberg et al., 2002; Leung, 2006; Atherton, 2006). The successful treatment of peptic ulcers by antibiotics vindicated the biomedical approach to both etiology and intervention and discredited the psychosocial accounts, whose etiologic understandings had promoted psychological and dietary treatments incapable of eradicating the harmful infection (NIH, 1994; Thagard, 1999; Danesh, 1999; Marshall, 2002; Leung, 2006; Kandulski et al., 2008).

Yet, suggesting that the story may not be quite so simple as “biomedical = correct” and “psychosocial = wrong” are two sets of considerations—one involving etiology, the other medical interventions (both curative and preventive). Starting with etiology, one relevant finding is that only 15% to 20% of persons with *H. pylori* infection develop ulcers (Levenstein, 2000; Atherton, 2006; Choung & Talley, 2008; Kandulski et al., 2008). Moreover, some countries, such as India, have very high infection rates but show marked regional variation in rates of peptic ulcer (Akhter et al., 2007; Leong, 2009). These well-known phenomena, whereby not everyone infected by a microorganism becomes ill, and population variation in disease rates exists despite virtually ubiquitous infection, in turn raises the equally well-known question: What additional conditions are required for infection to become disease?

One hypothesis proposed in the *H. pylori* literature focuses on genetic susceptibility, whereby a “match” is needed between host polymorphisms and whichever strain of *H. pylori* (a bacteria with very high genetic variability [Dykhuizen & Kalia, 2008]) has colonized the stomach (Leung, 2006; Atherton, 2006; Kandulski et al., 2008; Snaith et al., 2008). Still other evidence, however, has led to a re-emergence and refinement of the “stress” hypothesis (Levenstein, 2000; Levenstein, 2002; Choung & Talley, 2008). In contrast to the prior version, where “stress” was thought to induce ulcer formation directly, the two updated “stress”-related pathways posit that adverse stress can alter behavior and physiology in a way that increases susceptibility to infection by *H. pylori*, resulting in disease. One route is by stress increasing alcohol and aspirin consumption, both of which have been shown to increase risk of developing an ulcer; the other is by stress impairing immune responses relevant to wound healing (Levenstein, 2000; Levenstein, 2002; Choung & Talley, 2008). Thus, although adverse psychosocial exposures clearly are not obligate for development of peptic ulcers arising from *H. pylori* infection, and by themselves cannot explain the changing incidence rates over time, that does not mean they are completely irrelevant to explaining disease occurrence. Both the genetic “match” and the psychosocial



**Figure 8-1.** Changing class gradients in mortality because of peptic ulcer by birth cohort: analysis of 1921–1923 through 1949–1953 mortality data from the United Kingdom by Susser and Stein. (Susser & Stein, 1962, p. 118). (©1962 Elsevier, Inc. Reprinted with permission.)

Classes: Class I + II = “Professional, Managerial, and Technical Occupations”; Class III = “Skilled Occupations”; Class IV = “Partly-Skilled Occupations”; Class V = “Unskilled Occupations”

**Description and interpretation of the changing class gradients (in effect arguing against a monotonic gradient invariant to time, in this case, historical generation):**

p. 116: *Social Class*.

Analysis of changes in mortality by social class over the last three censuses again shows upward age-shifts, which suggest that each generation is carrying forward its own particular risk. At each successive census, a more or less regular pattern of mortality recurred in age-groups, which were older by the interval, which had elapsed between the censuses.

Statistics relate only to males, and are first available for the period of the 1921 census (Morris and Titmuss, 1944). From 1921 to 1923 death-rates from gastric ulcer showed a social class gradient increasing from the higher to the lower classes up to the age of 55; this gradient flattened between 55 and 70, and was reversed over the age of 70. A decade later, in the period 1930–32, the gradient increasing from higher to lower classes was apparent up to age 65 and then reversed—i.e., ten years older than at the previous census. In the period 1949–52 the gradient increasing from higher to lower classes persisted up to the age of 70 before it flattened... Although death rates in old age must be interpreted with caution, the trend fits the expectation.

hypotheses can be correct (i.e., be complementary, not antagonistic, each explaining a different part of the pattern).

Second, raising questions about appropriate medical interventions to cure or prevent *H. pylori* infection, several new investigations (albeit not all) have reported an increased risk of asthma among persons lacking childhood infection by *H. pylori* (Jarvis et al., 2004; Chen & Blaser, 2007; Blaser et al., 2008; Kandulski et al., 2008). A new interpretation, based on long-term historical and evolutionary thinking, including recognition that *H. pylori* has ubiquitously colonized the human gut since Paleolithic times (Atherton, 2006; Blaser et al., 2008; Dykhuizen & Kalia, 2008) is that early life exposure to relevant microbes—in this case, possibly *H. pylori* or else another microbe that tracks with *H. pylori* exposure—may be important for healthy immune system development and function (a.k.a. the “hygiene hypothesis”) (Jarvis et al., 2004; Atherton, 2006; Chen & Blaser, 2007; Blaser et al., 2008). Additionally, some—but not all—studies have suggested that adults treated by antibiotics for *H. pylori* infection may be at increased risk of gastro-esophageal reflux disease, Barrett’s esophagus, and adenocarcinoma of the esophagus, a consequence of *H. pylori* eradication affecting stomach acidity (Ahmed & Sechi, 2005; Atherton, 2006; Kandulski et al., 2008). Consequently, a biomedical emphasis on eradicating *H. pylori* infection by antibiotics, necessary for treating adults with ulcers, and also on preventing infection in children via vaccines could have consequences other than averting *H. pylori*-related illness (Chen & Blaser, 2007; Blaser et al., 2008). As with HT, biomedical “magic bullets” can well hit targets others than those anticipated by reductionist analysis, a possibility more readily considered by non-reductionist frameworks.

Thus, in this second example, as with the first, critical social, biological, evolutionary, and historical thinking are necessary for robust hypotheses and apt interventions. The omission of any one can lead to incomplete understanding and potentially harmful consequences.

**Example 3: Diabetes and Indigenous Health—From “Thrifty Genes,” Fictional History, and Racialized Disease to Reckoning With the Transgenerational Biological Embodiment of Social and Ecological Injustice.** Epidemiologic analysis of the mid-twentieth century sudden emergence of high and rising rates of diabetes among some—but not all—Indigenous peoples, especially those living in countries of the global North (Joe & Young, 1993; Young, 1994; Kunitz, 1994; Young, 2000; Gohdes & Action, 2000; Ferriera & Lang, 2006; Gracey & King, 2009; King et al., 2009; Cunningham, 2009), constitutes the third example of why epidemiologic theory matters for etiologic analysis. At issue is the harm caused by problematic causal inferences of epidemiologic theories of disease distribution that selectively—or erroneously—invoke history and neglect addressing jointly the societal and biophysical (including ecological) determinants of current and changing population patterns of health and health inequities.

In 1962, the geneticist James V. Neel (1915–2000) (Lindee, 2001), attempting to unravel the puzzle of rising rates of diabetes among people who, in his words “have come to enjoy the blessings of civilization” (Neel, 1962, p. 357), proposed his then novel hypothesis of the “thrifty genotype,” a vivid phrase he coined precisely because it was “a somewhat colloquial but expressive term” (Neel, 1962, p. 354). Neel’s thesis, foreshadowing present-day biomedical/lifestyle versions of “evolutionary medicine” and “gene–environment interaction” frameworks (as discussed in **Chapter 5**), was premised on the assumption (asserted without any supporting references) that humanity, for most of its existence as hunter-gatherers, experienced cycles of “feast-and famine” (see **Textbox 8–2**).

Positing a survival advantage for individuals predisposed to metabolize energy frugally (because “during a period of starvation” they would have “an extra pound of adipose reserve” (Neel, 1962, p. 355)), Neel theorized that this trait would become harmful as food



**Textbox 8-2. "Thrifty Gene" Hypothesis: Fictional Pasts Versus Historical Realities, 1962-2009.**

**"Thrifty Gene" hypothesis: Neel's initial (1962) and modified (1999) propositions**

**1962: original statement of hypothesis by James V. Neel (1915-2000), in: Neel JV. Diabetes mellitus: a "thrifty" genotype rendered detrimental by "progress"? *Am J Human Genet* 1962; 14:353-362.**

p. 353: "For the population geneticist, diabetes mellitus has long presented an enigma. Here is a relatively frequent disease, often interfering with reproduction by virtue of an onset during the reproductive or even pre-reproductive years, with a well-defined genetic basis, perhaps as simple in many families as a single recessive or incompletely recessive gene. If the considerable frequency of the disease is of relatively long duration in the history of our species, how can this be accounted for in the face of the obvious and strong genetic selection against the condition? If, on the other hand, the frequency is a relatively recent phenomenon, what changes in the environment are responsible for the increase?"

p. 354: "... in the early years of life the diabetic genotype is, to employ a colloquial but expressive term, a 'thrifty' genotype, in the sense of being exceptionally efficient in the intake and/or utilization of food... The precise physiologic basis for this 'thriftness' remains unclear. There are obvious possibilities. Thus, if after stimulation of the islets of Langerhans they continued to function longer in the predisposed than the normal, this could depress the blood sugar level unduly, resulting in hunger and increased food intake..."

p. 355: "A second possible mechanism to be considered involves a pancreas more rapidly responsive to increases in the level of blood glucose. In this connection it must be remembered that during the first 99 per cent or more of man's life on earth, while he existed as a hunter and gatherer, it was often feast or famine. Periods of gorging alternated with periods of greatly reduced food intake. The individual whose pancreatic responses minimized post-prandial glycosuria might have, during a period of starvation, an extra pound of adipose reserve..."

p. 357: "We now come to the problem posed by the relatively high frequency of the disease diabetes... three lines of thought suggest there has been a true increase in the frequency of the disease as more and more people have come to enjoy the blessings of civilization. Firstly, obesity appears to be, on the whole, a rarer phenomenon in primitive cultures than our own. There is less opportunity to indulge a hypertrophied appetite, and/or the lower mean caloric intake and greater physical activity of these primitive groups provide less of a stimulus to insulin production; this in turn means less stimulation of the antagonist-producing mechanism. Secondly, the action of the adrenal steroids in bringing to light the subclinical diabetic, in addition to the well-recognized effect of these compounds on



gluconeogenesis, may also be through stimulation of antagonist production. Since the response of the adrenal cortex to alarm situations is now less often followed by motor activity than in the past, one may postulate a disturbance in the physiologic balance established in the course of human evolution. Thirdly, the well-known glucose-mobilizing effects of adrenalin release are now not followed to the same extent by physical activity as seems to have been the case for primitive societies; since this calls for a greater insulin production than would otherwise be the case, here again it is an opportunity for increased stimulation of the insulin antagonist mechanism. These latter two thoughts would tend to place diabetes in the poorly defined areas of the 'stress diseases'—and indeed, the physiologic evidence to this effect is right now at least as convincing as it is for regarding peptic ulcer and coronary hypertension as stress disease."

pp. 359–360: "Should the foregoing thoughts prove correct, then diabetes underlines one of the ethical dilemmas of modern medicine. If the dietary and cultural conditions which elicit the relatively high frequency of diabetes in the Western World are destined to spread and persist over the entire globe, then, to the extent that modern medicine makes it possible for diabetics to propagate, it interferes with genetic evolution. But if, on the other hand, the mounting pressure of population numbers means an eventual decline in the standard of living with, in many parts of the world, a persistence or return to seasonal fluctuations in the availability of food, then efforts to preserve the diabetic genotype through this transient period of plenty are in the interests of mankind. Here is a striking illustration of the need for caution in approaching what at first glance seem to be 'obvious' eugenic considerations!" NB: (a) Neel provided no evidence or documentation that hunter-gatherers routinely experienced "feast or famine" nor did the article ever discuss diabetes among Indigenous populations, yet (b) most of the application of the "thrifty gene" hypothesis has been in relation to selection effects of "feast vs famine" as translated to presumed differences in genotype of Indigenous vs non-Indigenous populations, even though two of Neel's three lines of evidence focused on diabetes as a "modern" "stress disease." (See also example 2, above, on peptic ulcer as a "stress" disease).

**1998–1999: Neel's qualification—yet upholding—the hypothesis, in:**

**a) Neel JV. The "thrifty genotype" in 1998. *Nutr Rev* 1999; 57:S2–S9.**

p. S3: "A more telling observation concerning the role of lifestyle in the emergence of NIDDM [non-insulin dependent diabetes mellitus, i.e., Type 2 diabetes] in Amerindians involves the Pima Indians of southern

Arizon and a closely related group, the Pima Indians of the Sierra Madre Mountains of northern Mexico, two groups estimated to have separated some 700–1000 years ago. The results of studies done in those groups give no support to the notion that the high frequency of NIDDM in reservation Amerindians might be due simply to an ethnic predisposition—rather, it must predominantly reflect lifestyle changes.”

p. 54: “Despite all these advances in our understanding of NIDDM, the nature of the environmentally precipitated genetic maladjustments that result in the disease remain obscure. Given the intensity of the current effort to localize and characterize the genes, the functioning of which seems to be compromised in NIDDM, speculation at this time concerning their nature seems of little value. However, the concept of a ‘thrifty genotype’ remains as viable as when first advanced...”

**b) Neel JV, Weder AB, Julius S. Type II diabetes, essential hypertension, and obesity as “syndromes of impaired genetic homeostasis”: the “thrifty genotype” hypothesis enters the 21st century. *Perspect Biol Med* 1998; 42:44–74.**

pp. 45–46: “The Arizona Pima are the textbook material for the emergence of NIDDM and obesity among Amerindians following acculturation. Forced on a government dole by the rapacious diversion by the early settlers of the West of the water essential to their irrigation-style agriculture, the Pima of necessity adapted a high-fat, highly refined, government dispensed diet at the same time that they were forced to abandon any pretense of the agricultural work ethic. Currently among adults, the BMI (kg wt/ht in m<sup>2</sup>) averages  $33.4 \pm 7.5$  and 37 percent of men and 54 percent of women meet the criteria for NIDDM. The Mexican Pima, by contrast, still pursue a much more traditional lifestyle. The average BMI was  $24.9 \pm 4.0$ , and only 2 of 19 women ( $36 \pm 13$  years of age) and 1 of 16 men ( $48 \pm 14$  years of age) were diabetic. The diet of the Mexican Pima was largely beans, corn, and potatoes, grown by traditional and physically very demanding techniques.”

p. 49: “... it is now clear that the original thrifty genotype hypothesis, with its emphasis on feast or famine, presented a grossly overly simplistic view of the physiological adjustments involved in the transition from the lifestyle of our ancestors to life in the high-tech fast lane.”

pp. 60–61: “The various recent developments regarding NIDDM, essential hypertension and obesity suggest both a modification and a broadening of the original concept of a thrifty genotype. It now seems preferable to conceptualize these diseases as resulting from previously adaptive

multifactorial genotypes, the integrated functioning of whose the complexly altered environment in which these genotypes now find themselves. Some terminological problems must be dealt with. The term 'thrifty genotype' has served its purpose, overtaken by the growing complexity of modern genetic medicine... The ultimate genetic complexity of each of these diseases qualifies it for the term *syndrome* [italics in original]. Perhaps collectively we can speak of 'syndromes of impaired genetic homeostasis' or, more colloquially, the 'civilization syndromes,' or the 'altered life style syndromes,' to which other diseases may be added.

This viewpoint creates some terminological issues. As the genes associated with these three diseases are identified, let us not repeat the mistakes in genetic terminology that have and still are occurring in the field of cancer genetics. Although it is convenient as genes associated with the various malignancies are identified to speak of proto-oncogenes and, following the demonstration of the appropriate mutations, of oncogenes, these genes are for the most part cell cycle or housekeeping genes, playing important roles in normal cell activities; the sooner these roles are elucidated, the better for clear thinking. The same considerations will apply as the nature of the genes associated with the failed homeostatic syndromes are elucidated. They are not "NIDDM genes" or "hypertension genes," and in the end, developing the appropriate perspective will be facilitated by avoiding this shortcut terminology and developing a nomenclature more appropriate to the function of the gene."

**Rebuttals of the premises of and evidence for the "thrifty gene" hypothesis**

**Benyshek DC, Watson JT. Exploring the thrifty genotype's food-shortage assumptions: a cross-cultural comparison of ethnographic accounts of food security among foraging and agricultural societies. *Am J Physiol Anthropol* 2006; 131:120-126.**

p. 120: "The 'thrifty genotype hypothesis' has become firmly entrenched as one of the orienting concepts in biomedical anthropology, since first being proposed by Neel ([1962] *Am. J. Hum. Genet.* 14:353-362) over 40 years ago. Its influence on inquiries into the evolutionary origins of diabetes, lactose intolerance, and other metabolic disorders can hardly be underestimated, as evidenced by its continued citation in many top scientific and medical journals. However, its fundamental assumption, that foragers are more likely to experience regular and severe food shortages than sedentary agriculturalists, remains largely untested. The present report tests this assumption by making a cross-cultural statistical comparison of the quantity of available food and the frequency and extent of food shortages among 94 foraging and agricultural

societies as reported in the ethnographic record. Our results indicate that there is no statistical difference ( $P < 0.05$ ) in the quantity of available food, or the frequency or extent of food shortages in these reports between preindustrial foragers, recent foragers, and agriculturalists. The findings presented here add to a growing literature that calls into question assumptions about forager food insecurity and nutritional status in general, and ultimately, the very foundation of the thrifty genotype hypothesis: the presumed food shortages that selected for a 'thrifty' metabolism in past foraging populations."

**Paradies YC, Montoya MJ, Fullerton SM. Racialized genetics and the study of complex diseases: the thrifty genotype revisited. *Perspectives Biol Med* 2007; 50: 203–227.**

pp. 212–213: "Although it has been argued that famines are incapable of producing the differential mortality required to select for a thrifty genotype, it is, nonetheless, plausible that metabolic efficiency, as well as energy balance and storage, would have been a selective advantage in most human populations during the last few million years. Hence, under the racialized THG [thrifty genome hypothesis], there is a need for an intense environmental impetus to reverse this selection pressure for Europeans first (or only), in order to explain why this ethnoracial group does not suffer from T2DM [Type 2 diabetes mellitus] at the same rate as other ethnoracial groups. It has been suggested that agriculture, which became a dominant mode of subsistence 10,000 years ago for some human societies, could be this impetus.

Underlying the proposal of agriculture as this environmental impetus is the assumption that its advent was associated with an abundant food supply. There is little evidence, however, that agricultural societies were free of feast-and-famine cycles.... Furthermore, some hunter-gather societies were free of feast-and-famine cycles altogether. For instance, Nauruans and other Pacific populations, who now suffer from a very high prevalence of T2DM, formerly lived in thinly populated tropical islands that boasted a generous food supply all year round.

To complicate this simple evolutionary story further, a number of disadvantaged ethnoracial groups, who are thought to now have the putative thrifty genotype, have been using agriculture for many thousands of years. For example, the Pima Indians, the leitmotif of the TGH, have been farming intensively for over 2,000 years."

Ferreira ML, Sanchez T, Nix B. Touch the heart of the people: cultural rebuilding and the Potawatow Health Village in California. In: Ferreira ML, Lang CG (eds). *Indigenous Peoples and Diabetes: Community Empowerment and Wellness*. Durham, NC: Carolina Academic Press, 2006; pp. 459–492.

pp. 467–469: “As a Yurok outreach worker at United Indian Health Services recently put it, ‘to say I am an American Indian means I am or will be diabetic.’ A recent statistical analysis of the Yurok database, however, reveals that there is a weak, negative correlation between diabetes and quantum of Indian blood... The Pearson correlation between diabetes and blood quantum is  $-0.218$  ( $p < 0.001$ ), which indicates that as quantum increases towards full bloods, the frequency of diabetes decreases...”

In fact, few medical studies actually correlate individual genetic admixture with the occurrence of diabetes, when stating that Indian blood quantum structures risk for diabetes. Most studies take ‘Indian heritage’ for granted, and use it as a major variable to establish genetic causative links for the diabetes epidemic. The problem here is, of course, one of social identity, because it is closely related to the question ‘Who is an Indian?’—which appears as the most frequently asked question in the Bureau of Indian Affairs’ website. While there are nations, such as the Pima of the Gila River Indian Community, which require a minimum of  $\frac{1}{4}$  ‘Indian blood’ for tribal membership, in other groups who have ‘open enrollment,’ such as the Cherokee Tribe of Oklahoma, there are members today who are  $1/2048$  Cherokee (the fractions are always multiples of 2). Genetically speaking, ‘Indian heritage’ obviously has different meanings for enrolled members of the Pima and Cherokee nations.”

But the hypothesis  
lives on...

**National Institute of Diabetes and Digestive and Kidney Diseases. The Pima Indians: obesity and diabetes.** Available at: <http://diabetes.niddk.nih.gov/DM/pubs/pima/obesity/obesity.htm> (Accessed: July 22, 2009).

NIDDK research conducted on the Pima Indians for the past 30 years has helped scientists prove that obesity is a major risk factor in the development of diabetes. One-half of adult Pima Indians have diabetes and 95% of those with diabetes are overweight.

These studies, carried out with the help of the Pima Indians, have shown that before gaining weight, overweight people have a slower metabolic rate compared to people of the same weight. This slower metabolic rate, combined with a high fat diet and a genetic tendency to retain fat may cause the epidemic overweight seen in the Pima Indians, scientists believe.

Along with genetic make-up, diet is a key factor to healthy lifestyle. The influence of traditional desert crops on the

metabolism of the Pima Indians is being studied to determine how to prevent the onset of diabetes and obesity. Scientists use the "thrifty gene" theory proposed in 1962 by geneticist James Neel to help explain why many Pima Indians are overweight. Neel's theory is based on the fact that for thousands of years populations who relied on farming, hunting, and fishing for food, such as the Pima Indians, experienced alternating periods of feast and famine. Neel said that to adapt to these extreme changes in caloric needs, these people developed a thrifty gene that allowed them to store fat during times of plenty so that they would not starve during times of famine.

Dr. Eric Ravussin, a visiting scientist at the Phoenix Epidemiology and Clinical Research Branch at NIDDK, has studied obesity in the Pima Indians since 1984. He believes the thrifty gene theory applies to the Pimas.

The Pima Indians maintained much of their traditional way of life and economy until the late 19th century, when their water supply was diverted by American farmers settling upstream, according to Ravussin. At that time, their 2,000-year-old tradition of irrigation and agriculture was disrupted, causing poverty, malnutrition and even starvation. The Pima community had to fall back on the lard, sugar and white flour the U.S. government gave them to survive, says Ravussin.

However, World War II brought great social and economic change for American Indians. Those who entered military service joined Caucasian units. Many other American Indians migrated from reservations to cities for factory employment and their estimated cash income more than doubled from 1940 to 1944.

When the war and the economic boom ended, most Native Americans returned to the reservations, but contact with the larger society had profoundly affected the Pimas' way of life. Ravussin says it is no surprise that the increase in unhealthy weight among the Pima Indians occurred in those born post-World War II.

During this century people world-wide experienced more prosperity and leisure time, and less physical work. Since the 1920s, all Americans have consumed more fat and sugar and less starch and fiber. The greatest changes have occurred in consumption of fat. In the 1890s, the traditional Pima Indian diet consisted of only about 15 percent fat and was high in starch and fiber, but currently almost 40 percent of the calories in the Pima diet is derived from fat. As the typical American diet became more available on the reservation after the war, people became more overweight.

"The only way to correct obesity is to eat less fat and exercise regularly," Ravussin says.

Recently, Ravussin visited a Pima community living as their ancestors did in a remote area of the Sierra Madre mountains of Mexico. These Mexican Pimas are genetically the same as the Pima Indians of Arizona. Out of

35 Mexican Pimas studied, only three had diabetes and the population as a whole was not overweight, according to Ravussin.

"We've learned from this study of the Mexican Pimas that if the Pima Indians of Arizona could return to some of their traditions, including a high degree of physical activity and a diet with less fat and more starch, we might be able to reduce the rate, and surely the severity, of unhealthy weight in most of the population," Ravussin says.

"However, this is not as easy as it sounds because of factors such as genetic influences that are difficult to change. Our research focuses on determining the most effective way to bring about permanent weight loss in light of these factors," Ravussin adds.

**Candib LM. Obesity and diabetes in vulnerable populations: reflections on proximal and distal causes. *Ann Fam Med* 2007; 5:547-556.**

p. 550: "**Thrifty genotype.** This hypothesis proposes that the ability to conserve calories by laying down abdominal fat offers a genetic advantage for selection of this genotype during periods of food scarcity in human history. All humans are likely to have genetically determined mechanisms to survive caloric restriction, but some people may have thrifter genotypes than others, resulting in a higher likelihood of obesity when they are faced with higher caloric loads. With increases in caloric intake and decreases in physical activity brought about by economic and social changes, the genetically driven tendency toward central fat deposition becomes a health hazard. Such visceral fat is directly related to the development of insulin resistance and eventually diabetes. This hypothesis has been criticized for its potential to confuse genes with race and its application of biological determinism in the explanation of diabetes without addressing the social determinants of health, however, unique ethnic patterns of abdominal fat deposition resulting from a variety of genetic mutations are demonstrable. Compared with white Europeans, ethnic groups from developing countries—as well as native Americans, native Canadians, Maori, Asian Pacific islanders, and many other native populations—are more vulnerable to the development of diabetes, possibly because of genetic selection for those who could withstand starvation. Why the rate of diabetes among Europeans is lower despite famines in European history is subject to speculation, perhaps because of the differing patterns of famine, differing ways regions addressed food scarcity, or out-migration of those most at risk for diabetes."



Vogel F, Motulsky AG. Human and medical genetics. In: Detels R, McEwen J, Beaglehole R, Tanaka H (eds). *Oxford Textbook of Public Health*. Oxford: Oxford University Press, 2004. Available at: <http://www.r2library.com> (Accessed: July 25, 2009).

"Some decades ago Neel (1962) proposed the hypothesis that type 2 diabetes might be caused by a 'thrifty genotype.' He suggested that the gene or genes underlying diabetes might be an adaptation to long-lasting conditions of food shortage and starvation. Genes that increased mobilization of carbohydrate may have enabled their carriers to survive and reproduce. There is circumstantial evidence in favor of this hypothesis. India is a country in which the majority of the population have suffered for a long time from food shortages. Indians who have emigrated and are living under affluent conditions have a higher frequency of type 2 diabetes. Among some Amerindian tribes, diabetes and obesity have also become very common under the conditions of the current Western American diet. Type 2 diabetes, despite its genetic determination, can often be prevented by avoiding over-nutrition."

availability increased—a hypothesis aptly captured by his article's title: "Diabetes mellitus: a 'thrifty' genotype rendered detrimental by 'progress'?" (Neel, 1962, p. 353). Aggravating this tendency toward increased obesity and hence elevated diabetes risk, Neel argued, would be the physiological consequences of civilization's "new types of stresses" (Neel, 1962, p. 359) and increased propensity for sedentary living, with people in modern as compared to what he termed *primitive* societies less likely to respond with physical activity (whether "fight" or "flight") to "alarm situations" that "increased stimulation of the insulin antagonist mechanism" (Neel, 1962, p. 357) (see **Textbox 8-2**).

Comparing selection for the putative "thrifty genotype" to selection for sickle cell trait in populations living in regions "hyperendemic for falciparum malaria" (another genetic condition he had studied [Lindee, 2001]), Neel ended his article with a section titled "some eugenic considerations" (Neel, 1962, p. 359). Speculating that "the mounting pressure of population numbers" would lead to global food shortages, he argued—contrary to what he termed "'obvious' eugenic considerations!"—that under such conditions, "modern medicine" would be justified in its efforts to "preserve" (as opposed to eliminate) the "diabetic phenotype" because the "thrifty gene" would once again confer a survival advantage (Neel, 1962, pp. 369–360; see **Textbox 8-2**).

In his original 1962 article, Neel made no mention of any Indigenous populations; he would do so later, in a highly influential article published in 1982 (Neel, 1982). Instead, Neel's initial intent was to explain rising rates of diabetes in the "Western World" (Neel, 1962, p. 359). Nevertheless, the "thrifty gene" hypothesis was swiftly deployed (starting in the mid-1960s) to account for fast rising rates of diabetes in the twentieth century among American Indians and Aboriginal Australians, populations in which diabetes had been previously extremely rare or non-existent (Knowler et al., 1983; Knowler et al., 1990; Knowler et al., 1993; Ravussin et al., 1994; Young, 1994; Swinburn, 1996; McDermott, 1998; Bennett, 1999; Paradies et al., 2007). It soon became "one of the orienting concepts

in biomedical anthropology” (Benyshek & Watson, 2006, p. 120) (*see Textbox 8-2*) and has achieved “textbook” status both for explaining diabetes epidemiology among Indigenous peoples worldwide *and* as an exemplar of “gene–environment interaction,” as elaborated in numerous contemporary reference texts (see, for example: Williams, 2003; Vogel & Motulsky, 2004; Hetzel et al., 2004; Inzucchi & Sherwin, 2007; Leonard, 2008) (*see Textbox 8-2*). As such, the hypothesis features prominently in the website for the longest-running joint NIH and Indian Health Service studies on diabetes among American Indians, established in 1965 among the Pima Indians (NIDDK, 2009) (*see Textbox 8-2*).

Yet, despite the “thrifty genotype” hypothesis’ popularity, substantial evidence contradicts its core assumptions, overall and as applied to Indigenous peoples (*see Textbox 8-2*), as Neel himself came to recognize (Neel et al., 1998; Neel, 1999). Three major flaws may be summarized as follows:

- Faulty assumption 1: Risk for diabetes is determined by specific “diabetes genes,” and these genes are more prevalent in Indigenous populations.* First, Neel’s initial presumption that diabetes has “a well-defined genetic basis, perhaps as simple in many families as a single recessive or incompletely recessive gene” (Neel, 1962, p. 353) has been refuted by contemporary research. Findings instead implicate myriad candidate genes expressed in diverse biological pathways involving energy intake and expenditure (Vogel & Motulsky, 2004; Hetzel et al., 2004; Prentice et al., 2005b; Paradies et al., 2007; Inzucchi & Sherwin, 2007; Prentice et al., 2008). The argument is not that genes are irrelevant but that it is erroneous to posit there exists a specific “thrifty genotype.” Second, despite 40-plus years of intensive investigation, to date no unique set of “Indigenous genes” has been identified that singularly predict diabetes occurrence among diverse Indigenous populations (McDermott, 1998; Ferreira & Lang, 2006; Paradies et al., 2007), even as claims have been made for entirely different candidate genes in disparate groups (e.g., in Canada [Hegele et al., 2003] and Australia [Busfield et al., 2002]). Although a lack of evidence is not equivalent to negative evidence, nevertheless the implication (as with other types of health inequities) is that gene expression, not gene frequency, is what matters (Krieger, 2005a).
- Faulty assumption 2: Risk of famine is greatest in “primitive” societies.* Neel’s unrefereed—and oft-repeated—assertion that the “primitive” hunter-gatherer societies accounting for “the first 99 percent or more of man’s [*sic*] life on earth” endemically experienced cycles of “feast-and-famine” (Neel, 1962, p. 355), whereas subsequent agriculturalist societies did not, is belied by both historical and contemporary evidence. The evidence refuting his claim was clearly available at the time Neel proposed his hypothesis and has been mounting since (Sigerist, 1951 [1979]; Crosby, 1986; Swinburn, 1996; Ströhle & Wolters, 2004; Prentice et al., 2005; Speakman, 2006; Benyshek & Watson, 2006; Paradies et al., 2007; Ó Gráda, 2009). Empirical research, for example, has found no difference exists “in the quantity of available food, or the frequency or extent of food shortages... between preindustrial foragers, recent foragers, and agriculturalists” (Benyshek & Watson, 2006, p. 120). Moreover, not only were ancient agricultural civilizations, such as Egypt, Mesopotamia, and China, plagued by famines (Sigerist, 1951 [1979]; Prentice et al., 2005b; Ó Gráda, 2009), but as summarized by Mark Nathan Cohen, an anthropologist with expertise in population and agriculture (Cohen, 1989, p. 97):

“The recent histories of India, Russia, China, France, and most of the rest of Europe at least until the nineteenth century display a record of frequent and severe famine that is not exceeded or

even matched in simpler societies, much of the famine being attributed not to climate but to the failure of—and even burden imposed by—central administrative mechanisms”.

Although debates continue over the importance of recent famine regarding its potential selective effects on fertility and metabolism versus mortality (Speakman, 2006; Prentice et al., 2008; Speakman, 2008), Neel's initial supposition about “primitive” societies and cycles of “feast-and-famine” stands as refuted—even as it is still repeated uncritically in contemporary literature (see, for example: Lindsay & Bennett, 2001; Zimmet & Thomas, 2003; Chakravarthy & Booth, 2004; Candib, 2007; Kuzawa et al., 2008; Leonard, 2008; Servio et al., 2009).

—*Faulty assumption 3: Incorporation of Indigenous peoples into modern states reduced their risk of famine and increased their caloric intake.* The assumptions that Indigenous peoples prior to colonization were solely hunter-gatherers and/or experienced famine only before—but not after—colonization are likewise demonstrably false (Crosby, 1986; Weatherford, 1988; Viola & Margolis, 1991; Nabokov, 1991; Kunitz, 1994; Jackson, 1993; Mann, 2005; Carson et al., 2007). Within the United States, for example, the Pima Indians of southern Arizona, who since the 1970s have been repeatedly stated to have the highest rates of diabetes among U.S. Indian nations, if not the world (Bennett et al., 1971; Knowler et al., 1983; Knowler et al., 1990; Smith et al., 1993; Bennett, 1999), were, along with the Hohokam, their likely predecessors, agriculturalists for centuries before the Spaniards arrived in 1535 (Castetter & Underhill, 1935; Smith et al., 1993; Sheridan, 2006). Instead, their experiences of famine, like those of other American Indian nations, date chiefly to the late nineteenth and early twentieth centuries, a consequence of U.S. policies involving forced relocation into reservations, followed by forced reliance on U.S. government-supplied “ration” and “commodity” foods that, like the diets of children forced to attend Indian boarding schools, were inadequate both in quantity and quality (i.e., primarily starch and meat, with few or no vegetables, fresh fruit, eggs, or milk) (Nabokov, 1991; Smith et al., 1993; Jackson, 1993; Sheridan, 2006; Omura, 2006; Lang, 2006; Martinez et al., 2009).

The subsequent post-World War II reliance on cheap energy-dense nutrition-poor food and lack of access to affordable high-quality food by predominantly impoverished American Indians—both on and off reservation (Joe & Young, 1993; Snip, 2000; Ferriera & Lang, 2006; Candib, 2007), moreover, is not unique to Indigenous peoples but instead common to impoverished populations in many wealthy and middle-income countries, with attendant increased risks of obesity and diabetes (Tanumihardjo et al., 2007; Larson et al., 2009; Gracey & King, 2009). What differs is the intensity and compressed generational timeframe of these events: from traditional foods to famine to the poverty version of the so-called “Western diet,” and from a traditional to cash economy. Similarly, Indigenous peoples in the Pacific whose rates of diabetes (especially among impoverished urban populations) have suddenly risen in the twentieth century (Hales & Barker, 1992; Kunitz, 1994; Gracey & King, 2009; Cunningham, 2009) were unlikely to have experienced famine prior to colonization. Illustrating this point is the example of the Naurans, who, given their low population density and the availability of tropical vegetation and fish throughout the year (Baschetti, 1998; Paradies et al., 2007), had sufficient food pre-contact—but under Japanese rule during World War II, experienced “forced labour, exile, and starvation,” followed by “post-war prosperity from phosphate exports... accompanied by sedentariness, obesity, and diabetes” (McDermott, 1998).

Thus, although the diversity of experiences of Indigenous peoples and their contingent health profiles, both before and after their forced incorporation into modern states, defies simple generalizations (Young, 1994; Kunitz, 1994; Ferriera & Lang, 2006; Carson et al., 2007; Gracey & King, 2009; King et al., 2009; Cunningham, 2009), the assumption of a long period of “feast-and-famine” sufficient to select for a “thrifty genotype” uniquely or uniformly among Indigenous populations is untenable. So, too, is the assumption that their experiences of famine occurred only pre-colonization and was followed simply by “plenty.”

Recognizing threats to the tenets of the “thrifty gene” hypothesis, in 1998 and 1999, shortly before his death, Neel published a set of articles—one lengthy (Neel et al., 1998) and one abridged (Neel, 1999)—in which he both tempered and defended his hypothesis (see **Textbox 8–2**). Together, these articles superseded Neel’s initial 1982 updating of his hypothesis, in which he included “alternative physiological mechanisms, not understood 20 years ago” (Neel, 1982, p. 284) and also tellingly admitted that “although I have talked rather glibly about a ‘thrifty genotype,’ I have been quite discrete as to the precise genetic basis of the genotype” (Neel, 1982, p. 290). Accepting, 16 years later, that both the terminology of *thrifty genotype* and the emphasis on “feast-and famine” was “overly simplistic” (Neel et al., 1998, p. 49), Neel further concluded that there was “no evidence for a strong ethnic predisposition” (Neel et al., 1998, p. 45) to Type 2 diabetes in “Amerindians.” He instead attributed the rapid emergence of diabetes among the Pima to forced “acculturation,” including the destruction of their farming culture through settlers’ diversion of their water (Neel et al., 1998, p. 46) (see **Textbox 8–2**).

Arguing that “NIDDM [non-insulin dependent diabetes, i.e., Type 2 diabetes] is an etiologically (genetically) heterogeneous entity” (Neel et al., 1998, p. 46) with “complex genetics,” Neel accordingly cautioned against referring simplistically to “NIDDM genes.” Going further, he stated it was important to avoid the errors made in “cancer genetics,” whereby “clear thinking” was hindered by labeling particular genes “proto-oncogenes” and “oncogenes,” as if they were exclusively relevant to cancer, when instead such “genes are for the most part cell cycle or housekeeping genes, playing important roles in normal cell activities” (Neel et al., 1998, p. 61). Nevertheless, despite discussing the etiologic importance of the “environment,” specifically in relation to diet and physical activity, Neel continued to argue that diabetes and such kindred conditions as obesity and hypertension should be conceptualized as “‘syndromes of impaired genetic homeostasis,’ or, more colloquially, the ‘civilization syndromes,’ or the ‘altered life styles syndromes,’ to which other diseases may be added” (Neel et al., 1998, p. 61; see **Textbox 8–2**).

One current alternative to the *thrifty genotype* hypothesis is now the *thrifty phenotype* hypothesis, first proposed in 1992 by C. Nicholas Hales (1935–2005) and David JP Barker (b. 1938) (Hales & Barker, 1992) and elaborated and debated since (Swinburn, 1996; Hales et al., 1997; Bennett, 1999; Hales & Barker, 2001; Lindsay & Bennett, 2001; Prentice et al., 2005b). Premised on a life-course framework (see **Chapter 6**), this alternative hypothesis posits that poor nutrition in early life (pre- and post-natal) produces “permanent changes in glucose-insulin metabolism... which, combined with the effects of obesity, ageing and physical inactivity, are the most important factors in determining type 2 diabetes” (Hales & Barker, 2001, p. 5). A corollary is that there can be nongenetically determined familial transgenerational transmission of risk (e.g., because of shared poor nutrition across generations or the physiologic consequences of being a fetus born to an overweight, obese, or diabetic mother) (Hillier et al., 2007; Smith et al., 2009).

Thus, in contrast to the *thrifty genotype* hypothesis, which holds that malnutrition acted as “a selection pressure over many generations to alter the genetic make up of the population” (Lindsay & Bennett, 2001, p. 24), the *thrifty phenotype* hypothesis focuses on

present-day early life exposures and proposes physiological mechanisms involving expression of common genes. In effect “de-racializing” both the genotype and disease, the hypothesis has no need to maintain the claim, unsupported by contemporary data (McDermott, 1998; Ferriera & Lang, 2006; Paradies et al., 2007), that there exists a specific genotype predisposed to diabetes that is uniquely shared by the highly heterogeneous populations comprising Indigenous peoples. Instead calling into question the continued use of resources to search for unique Indigenous *thrifty genes*, the *thrifty phenotype* hypothesis has instead shifted attention to health-damaging effects of recent (i.e., late nineteenth to twenty-first centuries) adverse conditions (Hales & Barker, 1992; McDermott, 1998; Speakman, 2006; Paradies et al., 2007).

What has been the harm caused by a nearly half-century prioritizing of “thrifty genes,” premised on inaccurate history, to explain twentieth and twenty-first century spatiotemporal patterns of varying degrees of excess and rising rates of diabetes among diverse Indigenous peoples? First and foremost, the entrenched focus by scientists and health professionals on an assumed genetically-determined Indigenous vulnerability to common (a.k.a. “normal”) “Western” foods effectively relegated primary prevention to a secondary consideration (McDermott, 1998; Ferriera & Lang, 2006). Efforts instead prioritized clinical care and diabetes self-management—albeit within the funding constraints typically undercutting provision of appropriate health services to Indigenous populations (Knowler et al., 1990; Knowler et al., 1993; Gohdes & Acton, 2000; Roubideaux et al., 2000; Wilson et al., 2005; Warner, 2006; Pavkov et al., 2008). Although providing medical care to those in need is, of course, imperative, the point is that the racialized *thrifty genotype* perspective deemed this was all that could be done, ignoring possibilities for primary prevention.

Second and related, a prevalent corollary of the *thrifty gene* hypothesis is the fatalistic view that “being Indian means to be diabetic” (Ferriera & Lang, 2006, p. 15; see also McDermott, 1998; Unnatural Causes, 2009). As stated in one of the handful of books on diabetes and Indigenous Peoples providing Indigenous perspectives, “[p]atients and their families reason that if diabetes is genetic or in their blood, ‘there is nothing that can be done about it’” (Ferriera et al., 2006, p. 470). Commenting in 1994 on a three-decade long federally funded study of diabetes among the U.S. Pima Indians, James W. Justice, who did similar research among the nearby related Tohono O’odham Indians, sadly observed (Justice 1993, p. 73):

“It is sometimes disheartening to consider that with all of our abilities to detect diabetes and begin early intervention, we (i.e., IHS (Indian Health Services) and NIH (National Institutes of Health)) failed to prevent the disaster that has overtaken the Tohono O’odham people and other American Indian Tribes in the United States.”

The neglect of context is further underscored by the existence, to date, of only *one* age-period-cohort analysis of diabetes mortality among American Indians. Based on New Mexican mortality records, its limited findings suggest the existence of a birth cohort effect beginning in 1912 (not explained) and a strong period effect starting in 1960 (attributed to not only rising prevalence but also to the increase in health centers and improved accuracy of death certificates) (Gilliand et al., 1997).

Etiologic analyses and interventions premised on the alternative assumptions of sociopolitical, psychosocial, and ecologically and historically oriented epidemiologic theories of disease distribution offer a different vantage (Joe & Young, 1993; McDermott, 1998; Ferriera & Lang, 2006; Carson et al., 2007; Bartlett et al., 2007; King et al., 2009; Cunningham, 2009; Martinez et al., 2009). Reinterpreting diabetes as not only a bodily

disease but also “representative and symptomatic of the disruption in the physiology *and* cultures of Indigenous Peoples around the world” (italics in original) (Raymer, 2006, p. 313), these alternatives detail how actual—including recent—histories and contemporary realities of political, economic, territorial, cultural, and ecological expropriation, combined with global changes in the political economy and ecology of food (production, commerce, consumption, and ultimately food sovereignty [La Via Campesina, 2003; IPC, 2009; Rosset, 2009; McMichael, 2009]), have, since the mid-twentieth century, jointly increased of risk diabetes. At issue are their impact on: (1) local economies of food, work, and land ownership; (2) resources for and participation in Indigenous traditions; and (3) access to and resources for appropriate health services (Joe & Young, 1993; McDermott, 1998; Warner, 2006; Ferriera & Lang, 2006; Carson et al., 2007; Bartlett et al., 2007; King et al., 2009; Cunningham, 2009; Martinez et al., 2009).

Together, these sociopolitical phenomena have become embodied, affecting both birth outcomes and birth cohorts, via both transgenerational and concurrent influences on social and biophysical conditions across the life-course (Joe & Young, 1993; McDermott, 1998; Ferriera & Lang, 2006; Carson et al., 2007; Bartlett et al., 2007; King et al., 2009). Particularly salient is the research noted earlier showing the long reach of past damage, whereby children born to mothers affected by obesity, diabetes, and even just high (although not clinically abnormal) insulin levels are themselves at greater risk for these outcomes because of metabolic alterations, independent of heredity *per se* (Hillier et al., 2007; Smith et al., 2009).

Rejecting ill-founded genetic determinism, new Indigenous-led initiatives, many informed by what are referred to as “decolonizing methodologies” (Smith, 1999; Bartlett et al., 2007), are thus developing innovative approaches to prevent diabetes and improve its prognosis. They are doing so by reclaiming history, reclaiming traditional practices and foods, reorienting health interventions to address societal determinants of health, redesigning health systems to be culturally safe and inclusive, and, in the case of the Tohono O’odham, Pima, and Maricopa Indians, literally reclaiming the waters of the Gila River to enable them to farm once again, having finally won an 80-plus year battle contesting the diversion of the river’s waters to non-reservation non-Indigenous farmlands, suburbs, and cities, including Phoenix (Castetter & Underhill, 1935 (1978); Kraker, 2004; Applied Research Center, 2005; Sheridan, 2006; Archibold, 2008; TOCA, 2009; Unnatural Causes, 2009; Martinez et al., 2009) (see **Textbox 8–3**). Because these initiatives are new, it is too soon to tell whether they will succeed where prior efforts have not. Nevertheless, what stands out is the stark contrast between the analytic approaches—and practical public health implications—of epidemiologic theories of disease distribution that do versus do not engage with actual societal, historical, geographic, and ecologic contexts and relevant spatiotemporal scales and levels.

**Example 4: Short-Sighted Analyses—The Impact of Curtailing and Depoliticizing Relevant Time-Frames on Analyses of Temporal Trends in Health Inequities.** The fourth and final case concerns a different sort of harm—at the population and policy level, with implications for individuals but without harm directed at individuals *per se*. The example is a recent debate involving different theoretical trends within social epidemiology, with the empirical dispute focused on the question: As population health improves, do relative and absolute social inequalities in health widen or shrink? (Shaw et al., 1999; Phelan & Link, 2005; Mechanic, 2005; Kunitz & Pesis-Katz, 2005; Cutler et al., 2006; Kunitz, 2006; Siddiqi & Hertzman, 2007; Krieger et al., 2008a; Beckfield & Krieger, 2009).

An increasingly common view, typically drawing on recent data from the United States, is that relative, if not also absolute, health disparities are bound to increase as mortality rates decline, largely because groups with the most education and most resources are most



**Textbox 8–3. Contemporary Alternative Analysis of Diabetes Epidemiology by Tohono O’odham Indians (Arizona, United States) and Corresponding Prevention Strategies**

**Diabetes prevalence: from nonexistent prior to 1950 to among the world’s highest**

**Pre-to-mid-1950s:**  
“Earlier government reports do not mention diabetes as a health problem among the Tohono O’odham nor is diabetes mentioned by the Tohono O’odham tribal economic plans for 1949, although infant mortality, malnutrition and other problems were listed... The first extensive health status survey of the Tohono O’odham was conducted by Kraus and Jones during 1952-1953... They reported a diabetes prevalence rate of 5 per 1,000 population for Tohono O’odham and 9 per 1000 for the Pimas.” (Justice 1994, p. 74)

**1960s–1980s**  
Estimated diabetes prevalence rates among persons > 25 years old (per 1000 population) (Justice, 1994, pp. 77, 79):

	1965	1985
Men	153.9	257.8
Women	144.7	373.0

**21st century:**  
‘... the Tohono O’odham Nation has the highest diabetes rate in the world; over 50 percent of adults have adult-onset diabetes. (Applied Research Center 2005; p. 41)

**Ecological and economic context**

**Before encounter with Europeans in late 1600s:**  
“When Jesuit missionary Eusebio Francisco Kino criss-crossed the Pimería Alta in the early 1690s, he and his frequent traveling companion, Juan Mateo Manje, distinguished among different groups of O’odham, including the Sobas along the Río de la Concepción and the Papabotas (Papagos [Tohono O’odham])...” (Sheridan, 2006, p 26)  
“Before Kino and his companions rode down the Santa Cruz, O’odham... living along the few rivers and streams of the Sonoran Desert pursued three complementary subsistence strategies, all of which depended upon an intricate knowledge of plants, animals, and climate. During the spring and summer, they farmed floodplain fields or arroyo deltas. Throughout the year, they harvested wild plants and hunted wild game. When rains were abundant, washes ran and rivers flowed, watering their desert cultigens. When heat and drought withered crops, the O’odham relied entirely upon the seeds, fruits, roots, and caudices produced by the desert itself. Their agriculture was based upon three thousand years of accumulated knowledge about plant physiology and microclimates in the Sonoran Desert. O’odham in the Santa Cruz, San Pedro, and Magdalena-Altar-Concepción watersheds practiced irrigation agriculture, constructing



brush weirs that diverted water into earthen canals that led to their fields. O'odham along the Gila River... may not have needed to expend the energy to build weirs and canals. Instead, they simply may have planted their seeds in swales and islands after the Gila's seasonal floodwaters receded." (Sheridan, 2006, p. 34)

*"Historical Tohono O'odham Food System.* The traditional food system supported a local economy, maintained the people's physical well-being, and provided the material foundation for Tohono O'odham culture. For many centuries, the Tohono O'odham and their ancestors combined a series of well-adapted strategies of producing food in the arid lands of the Sonoran Desert. The three parts of this traditional Tohono O'odham food system were:

**Ak Chin Farming**—Using the flood waters that accompany the summer monsoons, thousands of acres were planted with crops that are nutritious and well-adapted to the short, hot growing season. Many of these foods were eaten fresh and preserved for use throughout the rest of the year.

**Harvesting Wild Foods**—Throughout the year, the desert provides a wide variety of wild foods that were collected and eaten. These wild foods included cholla buds, the fruit of different cacti, mesquite bean pods and acorns. Many of these foods were preserved for use throughout the year.

**Hunting**—The animals of the desert also provided an important source of nutrition. The hunting of rabbits, deer, havalina and other desert dwellers was a significant supplement to the foods grown in O'odham fields and collected in the desert. The combination of flood-based farming during the summer rains, collection of wild foods, and hunting provided the O'odham with a rich and varied diet.

In addition to providing healthy foods, all of these activities (and their cultural supports such as traditional dancing) promoted high levels of physical activity and fitness." (TOCA, 2009)

#### **20th century CE:**

*"Tohono O'odham Community Action (TOCA)* is based in Sells, Arizona, on the 4,600-square-mile Tohono O'odham Reservation, in the heart of the Sonora Desert. The tribe now has around 24,000 members. Until the mid 1900s, the O'odham used traditional agricultural practices they had developed over a thousand years. But a series of government policies seriously undermined their ability to continue these practices. Federal food programs introduced processed foods, displacing traditional nutrition. O'odham were encouraged to take jobs as field laborers for large irrigated cotton farms that surrounded O'odham land, resulting in many families leaving for six to eight months a year and being unable to maintain their own fields. Nearby development lowered the flood table and, as a result of governmental flood control

projects, water became scarce, and flood waters were eliminated from important lands. On top of these devastating changes, large numbers of children were forcibly sent to boarding schools, where they were severely punished for speaking their language and participating in their culture. All of these factors resulted in a break in the transfer of knowledge and traditions.

These changes wreaked havoc on O'odham agriculture. In the 1920s, over 20,000 acres of flood plain were cultivated using flash-flood irrigation conducive to the area's pattern of frequent summer monsoons. But by 1949 only 2,500 acres were cultivated, and by 2000 only a few acres were cultivated. There are other major challenges as well. The reservation is extremely rural and has the lowest per capita income of all U.S. reservations. The Tohono O'odham Nation has the highest diabetes rate in the world; over 50 percent of adults have adult-onset diabetes. The major changes in diet and community have certainly played a role in the diabetes epidemic." (Applied Research Center, 2005; p. 41)

"The steady change from a farming, hunting, and gathering society to one in which most individuals became wage earners occurred very slowly until the 1930s. During the 1930s, cotton farming off reservation needed manual labor, and by 1939, one third of all dollars earned by reservation residents came from this work... By 1960, most of the dollars earned by reservation Papagos [Tohono O'odham] came from off reservation work or from allotments paid by outside agencies, including the Bureau of Indian Affairs Welfare Program... Once wage earning became the primary source for food purchasing, the six trading posts on or near the reservation change their inventories. As late as 1949, most stores carried only beans, syrup, sugar, flour, coffee, lard, and powdered or canned milk. By the early 1960s, these same on-reservation stores started to carry high caloric pre-packed sweets, such as carbonated beverages (i.e., 'soda pop'), candy, potato chips, and cakes. None of these foods were mentioned by Ross in her 1941 survey of dietary habits nor by Van Cleft, who in 1954 calculated an in-depth survey of the buying habits of 16 families at 'the trading posts'... Finally in 1959, large quantities of refined flour, sugar, and canned fruits high in sugar became available from the U.S. Department of Agriculture surplus commodity food program, and by 1965 were being distributed widely." (Justice, 1994, pp. 115-117)

"More than a hundred years ago, the Gila River, siphoned off by farmers upstream, all but dried up here in the parched flats south of Phoenix, plunging an Indian community that had depended on it for centuries of farming into starvation and poverty... Most of the water was diverted in the late 19th century, slowing the Gila River to a trickle. It was a startling turn of events for a tribe

whose ancestors had thrived on the river for generations through an elaborate system of ditches and laterals, some of them still visible today. The construction of the Coolidge Dam, completed in 1928, by the federal government was intended to restore some of the lost water, but the reservation never received enough to bring back farming in any big way. Later diversions also depleted the Salt River, which runs north of the reservation and helped support farming as well. As the water disappeared and the Pima switched to government rations as their staple, obesity, alcoholism and diabetes exploded..." (Archibold, 2008)

#### **21st century CE:**

"Per capita income on the Tohono O'odham Nation is \$6998 (compared with \$21,994 nationally), the lowest of all U.S. reservations. Median family income is \$21,223 (compared with \$50,046 nationally). 41.7% of all households and 50.6% of households with children are below the poverty level (compared to the U.S. averages of 9.2% and 13.6%, respectively). Only 31.3% of the adult population is currently employed." (TOCA, 2009)

"Fewer than half of the Tohono O'odham community's adults have completed high school, the lowest rate of all U.S. Native American tribes." (TOCA, 2009)

#### **Etiologic analysis**

"The primary cause of diabetes within the community is the change from a diet consisting primarily of traditional food and the destruction of a sustainable Tohono O'odham food system." (TOCA, 2009)

#### **Interventions**

##### **TOCA programs:**

"TOCA's goal is to develop a food system and then encourage people to make healthy choices. TOCA's food system project focuses on three incentives: health, culture, and economy. In addition to their health benefits, traditional foods and crops are closely related to O'odham cultural identity. Many of TOCA's programs work to encourage the continuity of these linked traditions. TOCA is also working to encourage production and supply of traditional foods." (Applied Research Center, 2005, p. 44)

"TOCA's Tohono O'odham Food System and wellness initiatives combat the highest rates of diabetes in the world while simultaneously creating economic opportunity. By reintroducing traditional food production to the community, TOCA is stimulating improved community health, cultural revitalization and economic opportunity... This program has established a working farm to grow food once part of the daily O'odham diet, is actively documenting song, stories, harvesting, cultivation and processing methods, and is providing traditional foods for sale in the community and to hospitals, schools, and elderly lunch programs." (TOCA, 2009)

**Reclamation of river waters for farming:**

"... the long lush fields of the Pimas and Maricopas began to wither in the late 1800s, when Mormon farmers upstream began diverting huge amounts of water from the Gila...

The Gila River Indian Community first went to court in 1925, and has spent millions of dollars in an effort to quantify its water rights. Originally, the tribes made a claim on the Gila and Salt rivers, which form the boundaries of the reservation. In 1974, they sued for nearly the entire annual flow of the Gila, almost 2 million acre-feet. If the tribes had pursued that claim, it would have posed a major threat to the water supply of fast-growing Phoenix, according to City Water Manager Tom Buschatzke. Phoenix took the threat so seriously that, in 1988, it created an adjudication section within its law department, staffed by an attorney, a paralegal, a historian and a hydrologist, to try to work out a settlement with the Gila River Community and other tribes...

It took eight years, but the city, working with the Arizona Department of Water Resources, the Central Arizona Water Conservation District, the Bureau of Indian Affairs, the Bureau of Reclamation, irrigation districts and other metropolitan cities, finally came to an agreement with the Gila River Community..." (Kraker, 2004)

"...[the settlement] provides the reservation 653,500 acre-feet of water a year (an acre-foot is equivalent to about one family's water use annually) coming from a mix of sources, with the Central Arizona Project tapping the Colorado River providing the biggest share. It also includes the \$680 million to rebuild the irrigation system and to provide drainage, water monitoring, and other benefits... The reservation has discussed farming some 150,000 acres, 40 percent of its 372,000 acres... And it will take much effort to reverse the legacy of poor health, though programs abound, intended for the young and old, to combat diabetes... 'When we lost that water, we lost generations of farming,' said Janet Haskie, a community gardener. 'Then people had the attitude like, 'They owe us. I'm going to take these rations.' So now we have to start over again, a little at a time.'" (Archibold, 2008)

able to take advantage of new knowledge and technology (Phelan & Link, 2005; Mechanic, 2005; Cutler et al., 2006). By contrast, others, reviewing contemporary data from other countries (such as Canada), have posited that as mortality rates drop, health inequities tend to "flatten up" (Siddiqi & Hertzman, 2007), largely because improvements in population health are driven by "'pulling up' the health of the lower groups" (Siddiqi & Hertzman, 2007, p. 592). Still others, examining both the North American and European data over longer spans of time, hold that no one general pattern can be expected; instead, the circumstances leading to improvements in population health and affecting the magnitude of health

inequities are historically contingent and depend on the societal context and its public health, political, and economic priorities (Shaw et al., 1999; Kunitz & Pesis-Katz, 2005; Kunitz, 2006; Krieger et al., 2008a; Beckfield & Krieger, 2009). At a theoretical level, the contrast is between hypotheses premised on a depoliticized and ahistorical population health perspective versus those that are explicit about political and historical context, as discussed in **Chapters 6 and 7**.

Such a debate has profound policy implications. If, for example, increased health inequities inevitably accompany improvements in population health, it would suggest that the focus on health inequities be secondary to concerns about overall secular trends (Cutler et al., 2006)—in effect, a “trickle-down” approach to rectifying health inequities. Conversely, if improvements in overall population rates chiefly result from larger gains among those faring worst, then it would suggest that as long as population rates improve, health inequities should decline (Siddiqi & Hertzman, 2007)—that is, these trends necessarily move together. If, however, the relationship between population health and the magnitude of health inequities is more variable, it would imply resources are needed to tackle both concerns (Krieger et al., 2008a; Beckfield & Krieger, 2009).

Providing insight into this debate are **Figures 1–1a and 1–1b** in **Chapter 1**, showing the past half-century of trends in U.S. premature mortality (death before age 65 years) and infant death rates by county income level and race/ethnicity (Krieger et al., 2008a). These figures address the partial picture provided by most prior U.S. analysis, which, in part because of data limitations, typically have focused on post-1980 trends, mainly regarding racial/ethnic disparities in mortality, but with a few also including socioeconomic data and with a handful also extending back to 1968 and two using data from 1960 in conjunction with post-1969 data (Pappas et al., 1993; Singh & Yu, 1995; Singh & Yu, 1996a; Singh & Yu, 1996b; Schalick et al., 2000; Levine et al., 2001; Hillemeier et al., 2001; Kington & Nickens, 2001; Williams, 2001; Singh & Siapush, 2002; Singh, 2003; Ronzio, 2003; Satcher et al., 2005; Murray et al., 2006; Singh & Kogan, 2007; Ezzati et al., 2008). Together, this prior work on racial/ethnic and socioeconomic inequities in U.S. mortality has tended to support the hypothesis that as overall deaths rates have declined, social inequities in mortality have increased.

The yearly mortality data presented in the figures in **Chapter 1**, by contrast, span from 1960 to 2002, an interval that precedes as well as encompasses the period of the mid-1960s, a time when new U.S. federal policies were enacted with the intent of reducing socioeconomic and racial/ethnic inequalities, overall and also in relation to medical care (Davis & Schoen, 1978; O'Connor, 2001; Fairclough, 2001; Conley & Springer, 2001; Turncock & Atchison, 2002; Quadagno & McDonald, 2003; Navarro & Muntaner, 2004; Duncan & Chase-Lansdale, 2004; Kunitz & Pesis-Katz, 2005; Krieger et al., 2008a). Examples include the various federal policies constituting the “War on Poverty,” the 1964 U.S. Civil Rights Act, and the establishment of Medicare, Medicaid, and community health centers (Davis & Schoen, 1978; Cooper et al., 1981; O'Connor, 2001; Fairclough, 2001; Turncock & Atchison, 2002; Quadagno & McDonald, 2003; Duncan & Chase-Lansdale, 2004; Almond et al., 2006; Navarro & Muntaner, 2004; Smith, 2005; Lefkowitz, 2007). The selected timeframe likewise encompasses subsequent periods of active debate and change regarding government policies and spending on anti-poverty and civil rights initiatives, including post-1980 policies to “roll back” the welfare state (O'Connor, 2001; Fairclough, 2001; Turncock & Atchison, 2002; Henwood, 2003; Duncan & Chase-Lansdale, 2004; Navarro & Muntaner, 2004; Auerbach et al., 2006; Beckfield & Krieger, 2009).

The study's *a priori* prediction, borne out by the results shown in **Figures 1–1a and 1–1b**, was that the societal changes during the study time period (1960–2002) would be embodied and manifested in reductions in socioeconomic and racial/ethnic health inequities that preceded the documented post-1980 widening of health disparities noted above.

As revealed by more detailed analyses (Krieger et al., 2008a) between 1966 and 1980, the relative and absolute socioeconomic disparities in premature mortality shrank, overall and especially among U.S. populations of color; and thereafter, starting in 1981, the relative and absolute socioeconomic gaps for premature mortality widened; similar trends occurred for infant deaths. These patterns, unlikely to be explained simply by changes in health behaviors or medical treatment (Krieger et al., 2008a), refute the view that improvements in population health by default entail growing or shrinking health disparities, whether absolute or relative.

The net implication is that the societal patterning of socioeconomic inequities in mortality within and across racial/ethnic groups is historically contingent: context matters (Shaw et al., 1999; Kunitz & Pesis-Katz, 2005; Kunitz, 2006; Krieger et al., 2008a; Beckfield & Krieger, 2009). Comparing the results for 1966 to 1980 versus 1981 to 2002, the early trends give grounds for hope; the latter augur poorly for the *Healthy People 2010* objective of eliminating U.S. socioeconomic and racial/ethnic health disparities (U.S. DHHS, 2000)—a target clearly not met. Enabling these patterns to be seen is research motivated by epidemiologic theories of disease distribution attuned to history and sociopolitical context. The harmful consequences, including for policy formulation, of overlooking relevant historical time periods is apparent. Getting it right matters: Death is inevitable. Premature mortality—and widening inequities in premature mortality—are not.

**Summary of Selected Cases: Illustrative, Not Isolated—Hence, Error, Harm, and the Vital Importance of Epidemiologic Theory.** If the above four examples were isolated problems, their impact would be “restricted” to the specific harm done to the diverse populations affected. But they are not.

Instead, and as the previous chapters recount, epidemiologists since the inception of the field have long reckoned with the knowledge that erroneous and inadequate explanations can cause damage—either literally harming particular individuals or else not averting preventable suffering, illness, and death. At one level, the awareness of the potential of epidemiologic research to do harm, not just good, is well-recognized in the obligate discussion of “Type I” and “Type II” errors that routinely appears in epidemiologic textbooks (*see*, for example, those reviewed in **Chapter 1**). Conceptualized in relation to the empirical statistical testing of hypotheses, these two types of error are defined as (Porta 2008, p.85):

- **Error, Type I** (Syn: alpha error). The error of wrongly rejecting a null hypothesis—that is, declaring that a difference exists when it does not.
- **Error, Type II** (Syn: beta error). The error of failing to reject a false null hypothesis—that is, declaring that a difference does not exist when, in fact, it does.

Commonly considered causes of these types of errors include various types of systematic error, different kinds of confounding, and inadequate sample size.

Bringing the discussion of error to literally another level, however, since the late 1990s, a third type of error has been added to the epidemiologic roster, defined as (Porta 2008, p. 85):

- **Error, Type III** Wrongly assessing the causes of interindividual variation within a population when the research question requires an analysis of causes of differences between population or time periods... Risk differences between individuals within a particular population may not have the same causes as differences in the average risk between two different populations.



This definition, introduced in 1999 by Sharon Schwartz and Kenneth Carpenter (Schwartz & Carpenter, 1999), recognizes that Type I and Type II errors do not exhaust the number of ways it is possible to get a right or wrong answer. Precursors to this definition of Type III error include the 1948 proposal, by the eminent biostatistician Frederick Mosteller (1916–2006), of “a third kind of error”—that of “correctly rejecting the null hypothesis for the wrong reason” (as “it is possible for the null hypothesis to be false”) (Mosteller, 1948, p. 61). More akin to the Schwartz and Carpenter approach is the 1957 definition provided by Allyn W. Kimbal: “the error committed by giving the right answer to the wrong problem” (Kimball, 1957, p. 134). Indeed, as the epidemiologist Major Greenwood (1880–1949) cogently observed in 1935: nature “always answers truthfully the question you ask her, not the question you *meant* to ask her but the one you *did* ask” (italics in original) (Greenwood, 1935, p. 67).

In offering their expanded definition of Type III error, Schwartz and Carpenter built on Geoffrey Rose’s social epidemiologic insight, discussed in **Chapter 6**, that causes of incidence are not necessarily the same as causes of cases (Rose, 1985). Also germane was the parallel recognition that “the sort of evidence gathered on the benefits of interventions aimed at individuals may not help in guiding policies directed towards reducing health inequalities” (Davey Smith et al., 2001). Motivated by then ongoing debates about individual-level versus societal determinants of population rates of homeless, obesity, and infant mortality, Schwartz and Carpenter’s concern was that studies designed to analyze individual-level differences in risk within a specified population were often wrongly interpreted as being causally informative about why rates change over time or differ between populations (Schwartz & Carpenter, 1999). For example, although inter-individual genetic variability could potentially contribute to explaining inter-individual variation in body mass index in a given population at a given point in time, such individual-level genetic variability could not by itself account for why obesity rates within a population were quickly rising, as “genetic variation does not change that rapidly”; instead, to explain the rising rates, data are needed on “whatever other changes have occurred between time periods (e.g., an increase in the pervasiveness of advertisements enticing people to eat, the number of fast food restaurants per square mile, or exposure to sedentary leisure activities)” (Schwartz & Carpenter, 1999, p. 1178). Schwartz and Carpenter consequently argued that valid epidemiologic research “requires consideration of the full range of risk factors at all levels organization” and that “to examine such exposures requires their overt consideration and different sampling, measurement, and conceptual frameworks” (Schwartz & Carpenter, 1999, p. 1179).

Stated another way, avoiding the kind of Type III error described by Schwartz and Carpenter requires explicitly engaging with epidemiologic theory. At issue is the role of theory in guiding the hypotheses that are tested in the first place, before even attempting to ascertain whether any particular study provides a valid test of the hypothesis under consideration. Methodological precision alone cannot suffice because choice of methods follows choice of question (Morris, 1957, p. 14; Krieger, 2007). Paying heed to epidemiologic theory, and theorizing deeply across a diverse array of determinants of distributions of disease in real societies at real points in time, is thus not a matter of “politically correct” science, as some conservative commentators have charged (Satel, 2000). It is, instead, a matter of doing *correct* science (Krieger, 2005a) and answering well the questions that epidemiology is best suited to answer.

Hence, as should by now be clear, epidemiologic theory counts—for good and for bad. It matters not just because of the potential to cause harm but also because it can lead to valuable knowledge that spurs possibilities for beneficial change. Otherwise, to paraphrase Morris, what’s the use of epidemiology? (Morris, 1957)



## Looking Ahead: Epidemiologic Theory, the People's Health, and the Stories that Bodies Tell

In the half-century since publication of Morris' *Uses of Epidemiology* (Morris, 1957) and MacMahon et al.'s *Epidemiologic Methods* (MacMahon et al., 1960), the field of epidemiology has grown enormously (Boslaugh, 2008; Susser & Stein, 2009). Epidemiologists have, in keeping with the "seven uses" of epidemiology that Morris outlined in 1957 (see **Chapter 6, Textbox 6-2**) (Morris, 1957, p.96), constructively contributed to the public's health. The roster of accomplishments includes: producing valuable knowledge about current and changing population distributions of health and disease at the community, national, and global levels; conducting etiologic research to elucidate the underlying causes of these patterns; quantifying the need for health services and evaluating the effectiveness of services provided; and helping identify both healthy "ways of living" and obstacles preventing individuals and communities from living healthy lives.

Like any other academic discipline, epidemiology has also been embroiled in myriad debates about the substance of its scholarship. Potentially calling the credibility of the field into question, the past decade alone has witnessed major disputes over why, in a number of high profile cases, discordant results about likely benefits versus harm have been produced by observational studies versus randomized clinical trials, including the case of HT, discussed above, and also use of vitamin supplements and other micronutrients (Davey Smith & Ebrahim, 2001; Lawlor et al., 2004a; Lawlor et al., 2004b; Lawlor et al., 2004c; von Elm & Egger, 2004; Ebrahim & Clarke, 2007). Added to this list of concerns is, most recently, why results of so many genetic association studies are, like the observational studies, often inconsistent, if not nonreproducible (Mayes et al., 1989; Davey Smith & Ebrahim, 2001; von Elm & Egger, 2004; Pocock et al., 2004; Ebrahim & Davey Smith, 2008; Little et al., 2009a; STROBE, 2009; STREGA, 2009). Typically, in the case of the observational studies, blame for discrepancies has been attributed to their inability to control adequately for confounding, above and beyond problems of reverse causation, measurement error, and selection bias (Davey Smith & Ebrahim, 2002; Ebrahim & Clarke, 2007; STROBE, 2009); the genetic studies, in turn, are postulated to suffer from problems of small sample size, inadequate characterization of study populations, biased analyses (STREGA, 2009), and perhaps even faulty hypotheses (Dickson et al., 2010). Reacting to the controversies and mixed evidence, some high-profile external critics, such as Gary Taubes, have gone so far as to argue that only epidemiologic evidence from randomized trials be given credence (Taubes, 1995; Taubes, 2007).

Within the field, however, more realistic and nuanced arguments have recognized the need for diverse study designs, each with their own limitations and strengths and each attuned to answering some types of questions but not others (Davey Smith et al., 2001; Barreto, 2004; Vandenbroucke, 2008). For example, just as randomized trials are ill-suited to investigate causes of disease and are incapable of addressing such critical epidemiologic questions as to whether (and, if so, why) age-period-cohort effects exist or rates of disease are changing over time, observational studies are poorly suited to evaluate whether intended therapies have their intended effects. Both types of studies, moreover, face the challenge of delineating the likely causal relationships between the phenomena of interest, so as to guide which variables should be included as exposures, outcomes, confounders, mediators, or effect modifiers and also how they should be measured. Both also must reckon with the profound problem of selection bias (Porta, 2008, pp. 225–226), which, by affecting who is and is not part of the study population, can skew the range of observed exposures and outcomes and hence the magnitude of observed exposure–outcome associations. Such bias can profoundly compromise not only the extent to which results "may apply, be relevant,

or be generalized to populations or groups that did not participate in the study” (Porta, 2008, p. 252), but also, the internal validity of the results themselves. Stated more bluntly, the argument that internal validity is paramount and comes before concerns about generalizability only makes sense if the study populations included provide a valid test of the hypothesis in question—and theory is vital to making this determination.

Taking stock of the field, recent articles have tallied up both epidemiology’s successes and failures (for examples, *see* **Textbox 8–4**) (Davey Smith & Ebrahim, 2001; Ness, 2009), with virtually all calling attention to the importance of the multiplicity of study populations, study designs, measures (including of relevant covariates), and modeling approaches needed to test hypotheses well and get the answers right. Representing one constructive step toward addressing recent criticisms of epidemiology are several new initiatives, including the STROBE (“STrengthening the Reporting of Observational Studies in Epidemiology”) and STREGA (“STrengthening the REporting of Genetic Association Studies”) guidelines, both of which have been published in major epidemiology and biomedical journals (Egger et al., 2007; von Elm et al., 2007a–2007g; Vandenbroucke et al., 2007a–2007c; von Elm et al., 2008; STROBE, 2009; Little et al., 2009a–2009g; STREGA, 2009). Intended to improve the rigor and transparency of how epidemiologic findings are presented, their recommendations set high standards for how studies are done.

Tellingly, however, none of these contemporary discussions about the strengths and limitations of epidemiologic research, nor the new STROBE and STREGA guidelines, include any explicit discussion about the relevance of epidemiologic theories of disease distribution for study hypotheses, methods, or interpretation of findings. Similarly, new efforts to improve causal analysis and inference—such as those using directed acyclic graphs (DAGs), which graphically encode relationships between variables (Greenland et al., 1999; Robins, 2001; Hérnan, 2002; Glymour, 2006; Fleisher & Diez Roux, 2008; Richiardi et al., 2008)—likewise remains mum about what determines which variables are even considered, let alone why and how they may be causally linked or contingently entangled, even as they do acknowledge that background knowledge is needed to understand the causal processes at play. Nor do contemporary epidemiologic textbooks offer much guidance, as discussed in **Chapter 1**.

Instead, epidemiologists are once again left to their own devices, free to populate more carefully structured causal webs with undertheorized assorted “variables,” as if hypotheses were independent of theoretical frameworks, and as if approaches to causal theorizing about determinants of population distributions and risk of disease are either self-evident, requiring no analysis, or else simply a matter of idiosyncratic inspiration (or ideological proclivities). If, however, transparency of assumptions is vital for valid scientific research, then explicit attention to epidemiologic theories of disease distributions is essential for the field—and as the above examples make clear, the cost of ignoring these theories can be high, whether measured in wasted effort or, more profoundly, in people’s lives.

But there is room for hope—at multiple levels. Within the past two decades, a renewed interest in epidemiologic theories of disease distribution has become apparent (*see* **Chapters 6 and 7**), largely prompted by the revitalization of social epidemiology and its focus on developing frameworks, concepts, models, and methods to explain—and inform efforts to alter—current and changing societal patterning of health, disease, and health inequities (Krieger, 1994; Krieger, 2001b). Although some of these approaches have been dismissive of research focusing on specific “risk factors” (including health behaviors) and biological aspects of pathogenesis, others have decidedly engaged with the biophysical and behavioral processes involved in disease etiology, albeit rejecting the decontextualized analytic mode of the biomedical and lifestyle perspectives. Conceptualizing disease processes and rates in relation to multiple levels of societally-shaped exposures, susceptibility,

**Textbox 8-4. Twenty-First Century Appraisals of Epidemiology's Triumphs and Failures Since 1950**

Triumphs	Failures
<p><b>Davey Smith G, Ebrahim S.</b> <b>Epidemiology—is it time to call it a day?</b> <i>Int J Epidemiol</i> 2001; 30:1–11</p> <ul style="list-style-type: none"><li>- cigarette smoking as a cause of lung cancer (and other diseases as well) (p. 2)</li><li>- evidence on the fetal origins of adult disease and the “rebirth of social physiology” (pp. 2–3)</li><li>- applied epidemiology in “poorer parts of the world”: eradication of smallpox; expanded programs of immunization; sanitary improvements (e.g., “slit latrines and deep tube wells”) (p. 3)</li></ul>	<p><b>Davey Smith G, Ebrahim S.</b> <b>Epidemiology—is it time to call it a day?</b> <i>Int J Epidemiol</i> 2001; 30:1–11</p> <ul style="list-style-type: none"><li>i. despite extensive epidemiologic research on peptic ulcer in relation to temporal trends, birth cohort effects, and possible risk factors, these investigations played no role in identification of <i>H. pylori</i> as a key causal agent (pp. 4–5)</li><li>ii. refutations by intervention studies (e.g., randomized clinical trials (RCTs)) of posited protective effects based on observational epidemiologic evidence for: (pp. 5–6)</li></ul>

**Topics included in 2009 special issue of Annals of Epidemiology on the “triumphs of epidemiology” (Ness 2009):**

1. folic-acid fortification to reduce risk of folic-acid preventable spina bifida and anencephaly (Oakley, 2009)
2. use of hepatitis B vaccine to prevent hepatocellular carcinoma (Palmer-Beasley, 2009); health impact of low-level lead exposure (Needleman, 2009);
3. elucidation of which types of human papilloma virus (HPV) cause which types of cancer, relevant to development of the new HPV vaccine (Koutsky, 2009);
4. prone sleeping position as a preventable cause of sudden infant death syndrome (Dwyer & Ponsoby, 2009);
5. identification of individuals uninfected by HIV despite high exposure and possible immunologic and genetic traits conferring “natural” protection (Detels, 2009);
6. health benefits of physical activity, including for cardiovascular and other chronic diseases (Blair & Morris, 2009);
7. health effects of particulate air pollution, including elevated risk of mortality (Dockery, 2009).

- a. hormone replacement therapy and risk of cardiovascular disease: observational = protective; RCTs = increased risk
- b. beta-carotene and risk of cardiovascular disease: observational = protective; RCTs = increased risk
- c. vitamin E and risk of cardiovascular disease: observational = protective; RCTs = increased risk
- d. vitamin C and risk of cardiovascular disease: observational = protective; RCTs = no protection
- e. fiber intake and risk of colon cancer: observational = protective; RCTs = no protection

and resistance across the life-course and transgenerationally, these latter frameworks essentially posit there is more to biology than biomedicine and more to behavior than lifestyle. Offering fresh approaches to analyzing disease distribution as the embodiment of societal, ecologic, and historical context, this new work additionally makes clear that alternatives to current inequities are possible—and that it is theory that enables us to see this, in eminently practical terms.

After all, as Raymond Williams (1921–1988) observed in his historical etymologic explication of the many meanings of “theory” (see **Chapter 1, Textbox 1–1**), the act of theorizing is “always in active relation to *practice* [italics in original]: an interaction between things done, things observed and (systematic) explanation of these” (Williams, 1983a, p. 317). To Williams, theory’s power to be transformative occurs when it when it makes “hope practical, rather than despair convincing” (Williams, 1983b, p. 240). It can do so by providing new means to think about and address unsolved problems (whether conceptual, methodological, or practical), by revealing connections between previously unlinked ideas or phenomena, and by sparking awareness that, to use the language of the day, “another world is possible” (World Social Forum, 2001). Williams’ conclusion that “if there are no easy answers there are still available and discoverable hard answers, and it is these that we can now learn to make and share” (Williams, 1983b, pp. 268–269), far from being disheartening, is one that encourages creative and cooperative efforts to develop theory, apply it, and learn from the experience—advice as apt for the work of epidemiologists and other scientists as it is for those who seek to translate the knowledge gained from scientific research into practical improvements in people’s ability to live meaningful, healthy, and dignified lives.

I accordingly offer three final examples of the difference, conceptually and empirically, that explicit attention to epidemiologic theory can make—for description, for etiologic analysis, and for action. The examples are:

**Textbox 8–5:** Data on social class, race/ethnicity and gender in public health surveillance systems

**Textbox 8–6:** Discrimination as a determinant of health inequities (see also **Figure 8–2**)

**Textbox 8–7:** New national policies and global recommendations for health equity, in ecologic context

In each case, I show how application of diverse social epidemiologic frameworks has revealed serious gaps in knowledge and expanded the range of questions, data, understanding, and evidence available for improving population health and promoting health equity.

In closing, epidemiology is complex science, one with a profound capacity to illuminate—or obscure—the stories that our bodies tell. Deeply engaged with so many facets of human existence and life on this planet, epidemiology as a discipline is a remarkable intellectual project, one constantly grappling with how lived experience translates into society’s epidemiologic profiles. As its past history and present state readily reveal, what knowledge the field offers depends on theories of disease distribution deployed, not just the methods used. Created by real people, in real historical, political, and ecologic contexts, epidemiologic theories of disease distribution necessarily draw on metaphors and mechanisms that reflect the contending worldviews and technological level of the societies in which these theories are developed, employed, and contested. Recognition of the critical role of theory in shaping epidemiologic inquiry in turn raises issues of accountability and agency, in relation to not only who and what is responsible for observed disease distributions but also how they are monitored, analyzed, and addressed.

**Textbox 8-5. The Difference Epidemiologic Theory Can Make: Description—  
“Painting a Truer Picture” of the People’s Health Using Social  
Epidemiologic Frameworks to Conceptualize Public Health  
Surveillance Data, Identify Gaps, and Fill Them, Using the Four  
Cases Examples of Class, Race/Ethnicity, Gender, and Global  
Health Inequities.**

**Public health surveillance: objective**      Public health surveillance systems (vital statistics, disease registries, health surveys) should, as Edgar Sydenstricker (1881–1936) observed when he was designing the first U.S. population-based morbidity studies in the 1920s (Sydenstricker, 1930), **“give glimpses of what the sanitarian has long wanted to see—a picture of the public-health situation as a whole, drawn in proper perspective and painted in true colors”** (bold added) (Sydenstricker, 1952, p. 280)

**Conceptual statement of problem**      **“The making of public health data: paradigms, politics, and policy”** (Krieger, 1992)  
p. 412: “If you don’t ask, you don’t know, and if you don’t know, you can’t act. These basic precepts lie at the core of current and long-standing controversies about the nature and politics of public health data within the United States. At issue is the routine omission of social class data from most data sources, such as national vital statistics, disease registries, hospital records, and even individual studies, along with the persistent treatment of ‘race’ and ‘sex’ as essentially biological variables, their consistent conflation with ethnicity and gender, and the pervasive silence about the social realities of class inequalities, racism, and sexism.  
Although concerns about measures used in public health data may seem removed from the tumult of everyday life and everyday struggles for dignity and health, the ways these data are collected and reported can profoundly affect how public health professionals and the public at large perceive public health problems and their support for or opposition to diverse public health problems. Label infant mortality a problem of ‘minorities’ and present data only on racial/ethnic differences in rates, and the white poor disappear from view; label it a ‘poverty’ issue and proffer data stratified only by income, and the impact of racism on people of color at each income level is hidden from sight; define the ‘race’ or socioeconomic position of the infant solely in terms of the mother’s characteristics, and the contribution of the father’s traits and household class position to patterns of infant mortality will likewise be observed. Any particular approach necessarily affects our ability to understand and alter social inequalities in health.  
Despite being fraught with political considerations, the making of public health data—that is, the processes whereby health professionals and institutions decide whether to obtain and present data in particular forms—usually is cast in apolitical terms. On rare occasions, the politics are patently transparent, as exemplified by the recent federal cancellation of two national sex surveys. Usually, however, conscious and unconscious decisions about what types of data to include and exclude are based on prevailing theories of disease causation, and the links between these theories and concurrent political concerns are often obscured by the claim that scientific knowledge is ‘objective’ and ‘neutral.’”

p. 422: "Science is at once objective and partisan. There is no escaping this fact. It is critical to recognize that the allegedly neutral stance of leaving out class and focusing on race, and treating race and sex as chiefly biological characteristics without reference to ethnicity and gender, is as thoroughly political as overt efforts to augment public health data bases without variables pertaining to social class and the everyday realities of both racism and sexism. The seemingly 'apolitical' stance is also fundamentally invalid, for by failing to address the full range of variation in population patterns of disease, it blocks our scientific efforts and professional duties to understand and improve the public's health."

**WHO Commission on Social Determinants of Health. *Closing the gap in a generation: health equity through action on the social determinants of health. Final report of the Commission on Social Determinants of Health.* (WHO CSDH, 2008)**

p. 20: "No data often means no recognition of the problem. Good evidence on levels of health and its distribution, and on the social determinants of health, is essential for understanding the scale of the problem, assessing the affects of action, and monitoring progress."

p. 181: "A minimum health equity surveillance system provides basic data on mortality and morbidity by socioeconomic and regional groups within countries... In addition to population averages, data on health outcomes should be provided in a stratified manner including stratification by: sex; at least two social markers (e.g., education, income/wealth, occupational class, ethnicity/race); at least one regional marker (e.g., rural/urban, province); include at least one summary measure of absolute health inequities between social groups, and one summary measure of relative health inequities between social groups. Good-quality data on the health of Indigenous Peoples should be available, where applicable."

#### **Case 1. U.S. Public Health Surveillance Systems: Limitations Imposed by a Biomedical Approach—the Problem of Missing Socioeconomic Data and an Ecosocial Solution**

##### **Empirical evidence of problem**

"Can we monitor socioeconomic inequalities in health? A survey of U.S. Health Departments' data collection and reporting practices" (Krieger et al., 1997); study abstract (p. 481)

*"Objective.* To evaluate the potential for and obstacles to routine monitoring of socioeconomic inequalities in health using U.S. vital statistics and disease registry data, the authors surveyed current data collection and reporting practices for specific socioeconomic variables.

*Methods.* In 1996 the authors mailed a self-administered survey to all of the 55 health department vital statistics offices reporting to the National Center for Health Statistics (NCHS) to determine what kinds of socioeconomic data they collected on birth and death certificates and in cancer, AIDS, and tuberculosis (TB) registries and what kinds of socioeconomic data were routinely reported in health department publications.

*Results.* Health departments routinely obtained data on occupation on death certificates in most cancer registries. they collected data on educational level for both birth and death certificates. None of the databases collected information on income, and few obtained



**Project designed to address the problem, informed by ecosocial theory**

data on employment status, health insurance carrier, or receipt of public assistance. When socioeconomic data were collected, they were usually not included in published reports (except for mother's educational level in birth certificate data). Obstacles cited to collecting and reporting socioeconomic data included lack of resources and concerns about the confidentiality and accuracy of the data. All databases, however, included residential address, suggesting records could be geocoded and linked to Census-based socioeconomic data.

*Conclusions.* U.S. state and Federal vital statistics and disease registries should routinely collect and publish socioeconomic data to improve efforts to monitor trends in and reduce social inequalities in health."

**Executive summary: *The US Public Health Disparities Geocoding Project* (Krieger et al., 2004)**

<b>The problem</b>	A lack of socioeconomic data in most U.S. public health surveillance systems.
<b>Why is this a problem?</b>	Absent these data, we cannot: (a) monitor socioeconomic inequalities in US health; (b) ascertain their contribution to racial/ethnic and gender inequalities in health; and (c) galvanize public concern, debate, and action concerning how we, as a nation, can achieve the vital goal of eliminating social disparities in health (Healthy People 2010 overarching objective #2)
<b>Possible solution</b>	Geocoding public health surveillance data and using census-derived area-based socioeconomic measures (ABSMs) to characterize both the cases and population in the catchment area, thereby enabling computation of rates stratified by the area-based measure of socioeconomic position.
<b>Knowledge gaps</b>	Unknown which ABSMs, at which level of geography, would be most apt for monitoring U.S. socioeconomic inequalities in health, overall, and within diverse racial/ethnic-gender groups.
<b>Methodologic study: The Public Health Disparities Geocoding Project</b>	We accordingly launched the Public Health Disparities Geocoding Project to ascertain which ABSMs, at which geographic level (census block group [BG], census tract [CT], or ZIP Code [ZC]), would be suitable for monitoring U.S. socioeconomic inequalities

### Key findings

in the health. Drawing on 1990 census data and public health surveillance systems of 2 New England states, Massachusetts and Rhode Island, we analyzed data for: (a) 7 types of outcomes: mortality (all cause and cause-specific), cancer incidence (all-sites and site-specific), low birth weight, childhood lead poisoning, sexually transmitted infections, tuberculosis, and nonfatal weapons-related injuries, and (b) 18 different ABSMs. We conducted these analyses for both the total population and diverse racial/ethnic-gender groups, at all 3 geographic levels.

Our key methodologic finding was that the ABSM most apt for monitoring socioeconomic inequalities in health was the census tract (CT) poverty level, since it: (a) consistently detected expected socioeconomic gradients in health across a wide range of health outcomes, among both the total population and diverse racial/ethnic-gender groups, (b) yielded maximal geocoding and linkage to area-based socioeconomic data (compared to BG and ZC data), and (c) was readily interpretable to and could feasibly be used by state health department staff. Using this measure, we were able to provide evidence of powerful socioeconomic gradients for virtually all the outcomes studied, using a common metric, and further demonstrated that: (a) adjusting solely for this measure substantially reduced excess risk observed in the Black and Hispanic compared to the White population, and (b) for half the outcomes, over 50% of cases overall would have been averted if everyone's risk equaled that of persons in the least impoverished CT, the only group that consistently achieved Healthy People 2000 goals a decade ahead of time.

### Recommendation

U.S. public health surveillance data should be geocoded and routinely analyzed using the CT-level measure "percent of persons below poverty," thereby enhancing efforts to track—and improve accountability for addressing—social disparities in health.

**State Health  
Departments that  
have issued reports  
using the  
methodology of  
the Public Health  
Disparities  
Geocoding Project**

"The Health of Washington State Supplement: a statewide assessment addressing health disparities by race, ethnic group, poverty and education." September 2004. <http://www.doh.wa.gov/HWS>

The Virginia Department of Health Epidemiology Profile 2007. <http://www.vdh.virginia.gov/epidemiology/DiseasePrevention/Profile2007.htm>

The 2008 Virginia Health Equity Report. <http://www.vdh.state.va.us/healthpolicy/2008report.htm>

Massachusetts Deaths 2008 (in press) <http://www.mass.gov/dph/rep>

**Public Health  
Disparities  
Geocoding Project  
Publications**

Krieger et al., 2001; Krieger et al., 2002a; Krieger et al., 2002b; Krieger et al., 2003a; Krieger et al., 2003b; Krieger et al., 2003c; Krieger et al., 2003d; Krieger et al., 2005b; Subramanian et al., 2005; Krieger, 2006; Subramanian et al., 2006a; Subramanian et al., 2006b; Rehkopf et al., 2006; Chen et al., 2006; Krieger et al., 2007; Krieger, 2009

**Case 2. Racial/Ethnic Data in Public Health Surveillance Systems—Unmasking the Politics of Data by Using Social Epidemiologic Approaches: PAHO, Brazil, and the United States (Krieger, 2000; Nobels, 2000; PAHO, 2002; Krieger, 2004b; Paixão, 2004; Travassos & Williams, 2004; Romera & da Cunha, 2006)**

**PAHO:** Division of Health and Human Development, Program on Public Policy and Health. Final Report Experts Workshop: Cultural diversity and disaggregation of statistical health information. Quito, Ecuador, 4-June 5, 2002. Washington, DC: Pan American Health Organization, 2002. (PAHO, 2002)

**rationale for collecting and reporting ethnic origin data—and problem with missing data**

p. 2: "The collection and dissemination of data by ethnic origin are essential in order to identify, monitor, and progressively eliminate inequities in health status and access to health services. This information is critical to the effort to ensure that prevention, promotion, and treatment programs are effective and to establish binding norms that will make it possible to achieve equity. The lack of high-quality, congruent data and analysis based on ethnic origin is a problem in the majority of the countries. The existence of information systems is vital for evidence-based decision-making, to achieve the proper allocation of limited resources, and to evaluate the effectiveness of interventions."

p. 3: "The goals of this project are to: 1) improve data collection and analytic capacity; 2) generate new information that expands our collective knowledge of the health situation of indigenous populations and populations of African descent."

**Brazil:  
problem of  
inadequate or  
missing racial  
data—and also  
economic  
data**

**Conceptual statement of problem:**

**Paixão M.** Waiting for the Sun: An Account of the (Precarious) Social Situation of the African Descendant Population in Contemporary Brazil. *J Black Studies* 2004; 34:743–765. (Paixão, 2004)

"Analysis of the statistical data on the economic and social situation of African descendants in Brazil demonstrates severe racial inequalities, which traditionally have been denied according to the conventional wisdom of the myth of "racial democracy." The absence of reliable statistical data has fueled the force of this myth; statistics produced recently are the result of pressure by the black movement. The demographic presence of African descendants in the Brazilian population is examined and social indices are broken down by color/race, income levels, educational standards, and health conditions. Specific Human Development Index analysis, disaggregating data for African descendants and whites in Brazil, demonstrates the severity of racial inequality in comparison with other countries of the world and Africa. The black population in Brazil is still characterized by the absence of collective social rights and by the wide gap separating its living standards from those of the Brazilian European descendant population."

**Empirical manifestation:**

**Romera DE, da Cunha CB.** Quality of socioeconomic and demographic data in relation to infant mortality in the Brazilian Mortality Information System (1996/2001). *Cad Saúde Pública* 2006; 22:673–681. (Romera & da Cunha, 2006)

"This study aimed to evaluate the quality of socioeconomic and demographic data in the Brazilian Mortality Information System (SIM), in relation to infant mortality. The article assesses the system's potential for monitoring inequalities in infant mortality in various States in the country. Accessibility, timeliness, methodological clarity, incompleteness, and consistency were explored as quality indicators. Selected variables were: race, birth weight, gestational age, medical care, parity, and maternal schooling, age, and occupation. The study also reviewed the system's working documentation and the scientific literature on infant mortality. Proportions of data incompleteness were calculated by region and State, identifying factors that might influence (in)completeness using logistic regression. Despite the database's accessibility and the relevance of most of its variables, the system has serious quality problems: confusing instructions in the information manual concerning missing data, misclassification of maternal occupation, lack of data on the informant's race/ethnicity, and high proportions of incomplete information. The system does not appear to be a reliable source for monitoring, evaluating, and planning measures to minimize infant health inequalities."

**United States: argument against excluding racial/ethnic data and the necessity of analyzing it critically**

**Krieger N. Data, "race," and politics: a commentary on the epidemiologic significance of California's Proposition 54. *J Epidemiol Community Health* 2004; 58: 632-633. (Krieger, 2004b)**

"Data for social justice and public health is akin to the proverbial two-edged sword. To the extent we base any of our claims about social injustice in evidence, we must use data—whether of the quantitative or qualitative sort. But data do not simply exist. Contrary to the literal definition of 'data' as 'that which is given,' data instead are duly conceived and collected, via the ideas and labor of those who would obtain the requisite evidence. In the case of epidemiology, moreover, we must often use population data appearing in categories that are far from ideal—precisely because the assumptions of those with the power to shape and accrue the data often differ from those who seek to use these data to illuminate and oppose social inequalities in health.

Instructively highlighting these tensions are issues that recently arose in relation to the California ballot initiative Proposition 54. Officially designated as the 'Classification by Race, Ethnicity, Color, or National Origin Initiative'—but called the "Racial Privacy Initiative" by its supporters (who previously sponsored the successful anti-affirmative action Proposition 209)—Proposition 54 sought to ban collection or use of racial/ethnic data by government agencies. Under the slogan "Think outside the box," the initiative's proponents claimed Proposition 54 would "end government's preferential treatment based on race, and junk a 17th-century racial classification system that has no place in 21st-century America."

Despite its seemingly "progressive" approach to discounting outdated modes of classifying "race," Proposition 54 nevertheless was soundly defeated (64% opposed) by a coalition lead in large part by public health advocates and researchers, who exposed how the absence of these data would translate to public harm, especially in relation to public health.

... Tellingly, both proponents and opponents of Proposition 54 condemned racism and unscientific beliefs about "race" as an "innate" characteristic. But, whereas proponents argued that racial/ethnic data should not be collected because "race" is not "real" (i.e., not "biological"), opponents countered that this stance patently ignored the social realities of "race," i.e., as a socially constructed category reflecting societal and individual histories of racial discrimination and dispossession.

The contradiction is therefore sharp—and unavoidable—and affects all research employing categories that bear the mark of social inequality. Data on social disparities in health has long been disparately interpreted as evidence of: (a) "innate" inferiority, (b) "cultural" inferiority, or (c) embodied consequences of social inequity. There is no "thinking outside of the box" devoid of context. In the case of racial/ethnic inequalities in health, when "color" is no longer a signal for denial of human dignity and human rights, we will live in—and the data will show—a multi-hued society with equality for all. Only by bringing into the open the issues of power and injustice that lie behind the "that which is given" of public health data can we work honestly with the data to promote social justice and human rights, which together comprise the foundation of public health."

### Case 3. Engendering Global Health Statistics: Enormous Gaps in Data Continue to Exist (Hedman et al., 1996; DESA, 2006; Lin et al., 2007; Jara, 2007).

Why gender statistics matter, including for health

Hedman B, Perucci F, Sundström P. *Engendering Statistics: A Tool for Change*. Stockholm, Sweden: Statistics Sweden, 1996. (Hedman et al., 1996)

p. 9: "Statistics and indicators on the situation of women and men in all spheres of society are an important tool in promoting equality. Gender statistics have an essential role in the elimination of stereotypes, in the formulation of policies and in monitoring progress towards full equality. The production of adequate gender statistics concerns the entire official statistical system as well as different statistical sources and fields. It also implies the development and improvement of concepts, definitions, classifications and methods." (italics in original)

Smith MK. Enhancing gender equity in health programmes: monitoring and evaluation. *Gender Development 2001*; 9:95–105. (Smith, 2001)

p. 95: "Over the past ten years there has been increasing international recognition of the vital role to be played by investment in health care in the poverty-reduction strategies supported by governments and international donors. Parallel to this, there has been a growing debate at national level on the need for gender analysis in mainstream health programming and policy. Previously, concern for women's and gender issues was confined to a narrow focus on women's reproductive role, and hence on mother-and-child services, rather than taking account of women's needs, caring roles, and access and utilisation of health services. Gender-sensitive monitoring and evaluation is an essential component of this new agenda. It is a key principle in gender work to question any assumptions that a particular project or programme reaches all members of a community and has a similar impact on all of them."

Extent of missing data: 1996 and 2006...

Hedman B, Perucci F, Sundström P. *Engendering Statistics: A Tool for Change*. Stockholm, Sweden: Statistics Sweden, 1996. (Hedman et al., 1996)

p. 77: "The following is a list of topics where data [stratified by gender] are particularly scarce." (NB: health outcomes are put in **bold**; the rest are all social determinants of health.)

- |  |                              |                                   |  |
|--|------------------------------|-----------------------------------|--|
| • Male fertility                       | • School dropout rates       | • Subsistence agriculture         | • <b>Violence against women/ domestic violence</b> |
| • Household composition and structure  | • Educational achievement    | • Unpaid work                     | • Economic decision making                         |
| • <b>Diseases and causes of death</b>  | • Fields of higher education | • Time use                        | • Decision making at the local level               |
| • Internal and international migration | • Access to credit           | • Individual and household income | • Decision making in the household                 |
|  | • Access to land             | • Income control                  | • Resources allocation within the household        |
|  | • Informal sector            | • Poverty                         |  |



and still missing data 10 years later...

**Department of Economic and Social Affairs, Statistics Division.**  
*The World's Women 2005: Progress in Statistics.* New York: United Nations, 2006. (DESA, 2006)

pp. 21–22: 'In the period 1995–2003, even basic statistical data such as the number of deaths of women and men and girls and boys are not being reported for many countries or areas. More than a third of the 204 countries or areas examined did not report the number of deaths by sex even once for the period 1995 to 2003. About half did not report deaths by cause, sex and age at least once in the same period... The region with the lowest proportion of countries or areas reporting deaths by sex is Africa. Only 18 out of 55 countries or areas, comprising 35 per cent of the region's population, reported national data on deaths by sex at least once in the period 1995–2003. In Asia, 33 countries or areas, representing 55 per cent of the region's population, and in Oceania 7 countries or areas, representing 76 per cent of the region's population, reported deaths by sex. Deaths by sex and age are reported by most countries or areas in North America, South America and Europe.'

pp. 22–23: 'According to the Beijing Platform for Action, son preference is one factor that contributes to differential mortality by sex. As a result, in some countries it is estimated that men outnumber women by 5 in every 100. A preference for sons remains deeply rooted in many societies and girls may have less access to nutrition, preventive care (such as immunization) and health care. Data on infant deaths by sex are needed to see where excess mortality among girls exists so that it can be addressed and eliminated. While total infant deaths were reported by 143 countries or areas in the period 1995–2003, fewer—114, representing 40 per cent of the world population—reported infant deaths by sex. The pattern of low reporting in Africa and Asia and high reporting in the other geographic regions, as seen with reporting deaths by sex, also prevails for infant deaths.'

p. 27: 'In general, countries fall into one of two groups: either they have a strong statistical capacity and have been able to report mortality data almost every year by sex, age and cause; or their reporting capacity is very limited and has not improved since 1975. Moreover, there is a clear association between the national reporting of mortality data by sex and age and the level of development. This is, at least partially, a consequence of the lack of well-functioning civil registration systems that record births and deaths in the less developed regions. However, there have been some notable improvements. There has been better reporting of deaths caused by HIV/AIDS. In addition, the implementation of international programmes such as Multiple Indicator Cluster Surveys and Demographic and Health Surveys have contributed to a wider availability of national data on some aspects of mortality, morbidity and disability.'

**Lin V, Gruszyn S, Ellickson C, Glover J, Silburn K, Wilson G, Poljski C.** Comparative evaluation of indicators for gender equity and health. *Int J Public Health* 2007; 52:S19–S26. (Lin et al., 2007)

p. S19: 'Methods: A comprehensive health information framework was developed on a generic framework by the ISO (2001) to use for the analysis of gender equity within mainstream health systems. A sample of 1 095 indicators used by key international organizations were mapped to this framework and assessed for technical quality and gender sensitivity.'

p. S21–22: '—most routine indicators, including basic health indicators such as infant mortality, were not reported disaggregated by sex or age, nor with a comparator or over time; although sex disaggregations were reported in indicators on life expectancy, education, workforce, and democracy;

— there were few age-disaggregated indicators and none that were disaggregated for ethnicity or socioeconomic groups;

— most sex-specific indicators described women, and were age-limited to women of reproductive age, or described reproductive outcomes (e. g. deliveries, births), while indicators on the health problems of females out of reproductive age (e. g., older women, girls) or pertaining to non-reproductive states (e. g., mental health) were largely missing.

— indicators with comparators compared females to males (i.e. used the male as the norm), and most were found in only six topics (including literacy, education, employment); and

— few indicators included a time element that would allow the assessment of change over time.

**Case  
examples:  
burn statistics**

**Smith MK. Enhancing gender equity in health programmes: monitoring and evaluation. Gender Development 2001; 9: 95–105. (Smith, 2001)**

pp. 98–99: "A gender-sensitive monitoring exercise focusing on the Assuit Burns Centre, a health project in Upper Egypt, provided an in-depth understanding of the gender dimensions of burns in a poor rural area of Upper Egypt, revealing women's particular vulnerability to burns, and their lack of access to health services... A gender analysis of burns in Upper Egypt tells us that it is usually women who are affected by burns, while they are performing their domestic tasks of cooking and baking... Although the physical complications of burns, such as disfigurement and disability, are the same for women and men, the social complications are different. Disfigured women have to face many prejudices which affect their social lives. For example, unmarried women who are burned may never get married, and may live almost as outcasts from society, which demands conformity to gendered expectations of female beauty and capacity for physical work."

**Case 4. New Approaches for Visualizing Between and Within Country Health Inequities and Their Social Determinants (Worldmapper, 2008; Gapminder, 2008)**

**Maps:**

**Chapter 1, Figures 1–4a through 1–4h: Maps from the "Worldmapper" project (Worldmapper, 2008).**

Using available sociodemographic and health data in a novel way, whereby each territory is sized in relation to the variable depicted, these maps from the "Worldmapper" project, premised on a social determinants of health analysis, are intended to give new insights

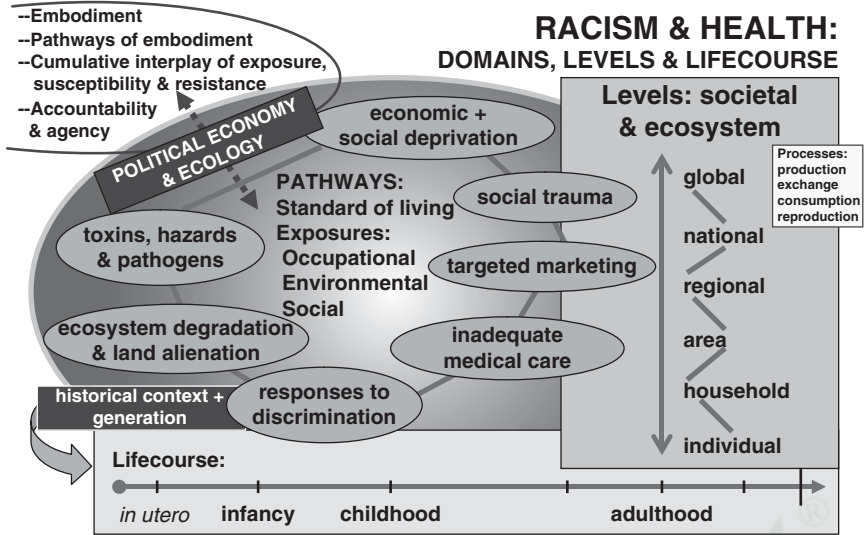
into—and spur action to address—global inequalities in health. Although these maps cannot, by themselves, answer the question “why” the observed inequities exist, the point here is the theoretical perspective used to create and juxtapose these maps in the first place—precisely so as to galvanize this very question of “why.” As stated by Danny Dorling (b. 1967), one of the team members who produced these maps (Dorling, 2007):

“Drawing maps is one way to engage more of our imagination to help understand the extent and arrangement of world inequalities in health... what I think matters most are the new ways of thinking that we foster as we redraw the images of the human anatomy of our planet in these ways. What do we need to be able to see—so that we can act?”

**Graphs: Figures 1–5 through 1–6 in Chapter 1: Graphs from the “Gapminder” project (Gapminder, 2008)**

Likewise animated by a social epidemiologic orientation, these graphs from the “Gapminder” project, developed by Hans Rosling (b. 1948), offer still another new way to present data on global health inequities. Intended to promote “**sustainable global development and achievement of the United Nations Millennium Development Goals**” (Gapminder, 2008), the figures’ stark depictions of the variable magnitudes of health inequities, within and between countries, at similar and different per capita income levels, and also over time, underscores both the variability in the magnitude of health inequities—and nothing inevitable about their size. The “why” questions prompted by data display reflect a theoretical orientation engaged with the social determinants of health and attuned to both historical and geographical contingency. Or, as stated by Rosling (Barone, 2007):

“Most people know only two types of countries, Western and third world, whereas I know 200 types of countries. I know each country’s gross national product, educational level, child mortality, main export products, and so on. We have a continuum of life conditions in the world... We want to know: How can we better measure and communicate the conditions of the poorest 1 to 2 billion people in the world?... There is a tsunami every month that could be cured by penicillin, for which there are no images and no reporting...”



**Figure 8–2.** Pathways by which racial discrimination can harm health: an ecosocial perspective. (Krieger, 1999/2000; Krieger, 2010a)

**Textbox 8–6. The Difference Epidemiologic Theory Can Make: Etiology—Racial Discrimination as a Societal Determinant of Health Inequities, Identified Using Social Epidemiologic Theories but Ignored by Biomedical and Lifestyle Approaches**

**A Personal Reflection on the Rapid Growth of Epidemiologic Research on Discrimination and Health Since the 1980s—and the Work that Lies Ahead:**

When I first began conducting epidemiologic research in the latter part of the 1980s and early 1990s on discrimination as a determinant of health inequities (Krieger, 1990; Krieger et al., 1993; Krieger & Sidney, 1996; Krieger, 2003) there were virtually no epidemiologic studies on this topic; what little research existed chiefly was in the psychology literature and usually not population-based. Enabling me to identify and start to address this important gap in the epidemiologic literature was the social production of disease framework (see **Chapter 6**); I soon found, however, that it gave insufficient guidance for thinking through how social inequalities become biologically expressed needed, a realization that spurred my work on developing ecosocial theory (Krieger, 1994)(see **Chapter 7**).

In 1999, I published the first major epidemiologic review article on discrimination and health (Krieger, 1999), in which I used ecosocial theory to delineate five major pathways by which discrimination could harm health: (1) economic and social deprivation; (2) toxic substances and hazardous conditions; (3) socially inflicted trauma; (4) targeted marketing of harmful products; and (5) inadequate or degrading medical

care—to which I have now added: (6) degradation of ecosystems, including as linked to systematic alienation of Indigenous populations from their lands and corresponding traditional economies (*see below*, and also as discussed in **Chapter 7**) (Krieger, 1999; Krieger, 2003c; Krieger, 2009; Krieger, 2010a). For this review, I was able to identify 21 population-based studies that explicitly tested hypotheses about associations between self-reported experiences of discrimination and health. Among these 21 investigations (3 of which addressed multiple types of discrimination), 15 focused on racial discrimination (the first published in 1984), 3 on gender discrimination (the first published in 1990), 4 on discrimination based on sexual orientation (the first published in 1994), and 1 on discrimination based on disability status (published in 1998); the range of health outcomes investigated spanned from somatic health to mental health to health behaviors.

Reflecting, in part, the growing influence of social epidemiologic frameworks in the field since the early 1990s, the pace of epidemiologic research on the health impacts of discrimination has quickened enormously. To date, much of the work has remained focused on racial discrimination and health, with the number of articles quickly rising, as evidenced by three recently published epidemiologic review articles: one published in 2003 included 53 studies, of which more than half (24) were published during or after 2000 (Williams et al., 2003), one published in 2006 included 138 studies published up through 2004 (Paradies, 2006), and a third published in 2009 included 115 articles that were indexed in PubMed between 2005 and 2007 alone (Williams & Mohammed, 2009). Initially, studies examining links between racial/ethnic discrimination and health were conducted almost exclusively in the United States; the research now, however, is starting to become global in scope, with studies now also conducted in Latin American and European countries, and also South Africa, New Zealand, and Australia (Krieger, 1999; Williams et al., 2003; Paradies, 2006; Williams & Mohammed, 2009).

Increasing interest in epidemiologic research on racial discrimination and health has also led to the development and refinement of methods to study racial discrimination at multiple levels (e.g., self-reported by individuals, in relation to residential racial segregation, in relation to health-care system factors, and at the government policy level), both across the life-course and transgenerationally. As discussed in the above four review articles (Krieger, 1999; Williams et al., 2003; Paradies, 2006; Williams & Mohammed, 2009), a 2004 National Research Council report focused on methods for measuring racial discrimination (National Research Council, 2004) and other recent review articles (Kressin et al., 2008; Pager & Sheperd, 2008; Pachter & Garcia, 2009), researchers currently employ a variety of methods and instruments (some validated, many not) to measure diverse aspects and levels of racial discrimination. Addressing current controversies about how epidemiologic research can best measure discrimination—at the individual, institutional, and societal levels—will require creative and critical empirical research; investigations tackling these challenges is clearly underway (Williams & Mohammed, 2009; Kressin et al., 2008; Carney et al., 2010; Sharif-Marco et al., 2009; Krieger et al., 2010).

Overall, epidemiologic research on discrimination and health, although perhaps no longer in its infancy (Krieger, 1999; Krieger, 2003c), is still very young, both conceptually and methodologically. The extant literature is thus not surprisingly full of inconsistent results and only just beginning to yield some robust findings that hold for specified outcomes across different studies. A safe prediction is that there will continue to be rapid and increasingly global growth in epidemiologic research engaged with the conceptual, methodologic, and substantive challenges of investigating the health consequences of diverse types of discrimination, singly and combined (Krieger, 2010a).

**Textbox 8-7. The Difference Epidemiologic Theory Can Make: Action—Contrasts Between Priorities Based on Research Guided by Individually Oriented Biomedical and Lifestyle Approaches Versus Social Epidemiologic Frameworks**

**(a) “10 Tips for Better Health”—a Satirical Contrast Between Recommendations Based on a Biomedical and Lifestyle Approach Versus a Social Determinants of Health Approach**

**Chief Medical Officer’s Top Ten Tips for Health—United Kingdom (1999) (Donaldson, 1999)**

- 1) Don’t smoke. If you can, stop. If you can’t, cut down.
- 2) Follow a balanced diet with plenty of fruit and vegetables.
- 3) Keep physically active.
- 4) Manage stress by, for example, talking things through and making time to relax.
- 5) If you drink alcohol, do so in moderation.
- 6) Cover up in the sun, and protect children from sunburn.
- 7) Practise safer sex.
- 8) Take up cancer screening opportunities.
- 9) Be safe on the roads: follow the Highway Code.
- 10) Learn the First Aid ABC—airways, breathing and circulation.

**Alternative Tips—Townsend Center for International Poverty Research (1999) (Gordon, 1999)**

- 1) Don’t be poor. If you are poor, try not to be poor for too long.
- 2) Don’t live in a deprived area. If you do, move.
- 3) Don’t be disabled or have a disabled child.
- 4) Don’t work in a stressful low-paid manual job.
- 5) Don’t live in damp, low quality housing or be homeless.
- 6) Be able to afford to pay for social activities and annual holidays.
- 7) Don’t be a lone parent.
- 8) Claim all benefits to which you are entitled.
- 9) Be able to afford to own a car.
- 10) Use education as an opportunity to improve your socio-economic position.

**(b) Experts “10 Steps to Health Inequalities” Developed for the United Kingdom’s 1998 *Independent Inquiry into Inequalities in Health* (the “Acheson Report”) (Independent Inquiry, 1998), Reflecting the Dominance of Individual-Level Biomedical and Lifestyle Studies—and a Social Epidemiologic Critique**

**Laurance J.** Experts’ 10 steps to health equality. *The Independent*, November 12, 1998. (Laurance, 1998)

“The 10 best ways of closing the health gap between the rich and the poor, highlighted as a priority by ministers, have been drawn up by an expert committee... The expert group that drew up the list included the editors of Britain’s two leading medical journals, the *Lancet* and the *British Medical Journal*, who assessed a range

**Davey Smith G, Ebrahim S, Frankel S.** How policy informs the evidence: “evidence based” thinking can lead to debased policy making. *BMJ* 2001; 322:184–185. (Davey Smith et al., 2001)

Who would not want health policy to be based on evidence? “Evidence based medicine” and “evidence based policy” have such reassuring and self-evidently desirable qualities that it may seem contrary to question their legitimacy in relation to reducing health inequalities. However, these terms are now so familiar that it is easy to forget the important



of studies on health inequalities and ranked the measures they recommended according to their chances of success... Dr Smith [editor of BMJ] said the group's recommendations could be criticized for concentrating on relatively small, medical interventions—rather than macro changes, such as adjustments to the tax and benefit system—but that was where the evidence was strongest."

- 1 Nicotine gum and patches free on the NHS. They double the chances of stopping smoking.
- 2 Pre-school education and child care. Strong evidence that it improves long-term prospects for children.
- 3 Fluoridation of drinking water. Cuts tooth decay.
- 4 Accident prevention—e.g., fit cars with soft bumpers. Accidents are the principal cause of deaths among young people.
- 5 Drugs education in schools. Prevents pupils becoming hooked.
- 6 Support around childbirth to promote breastfeeding and mental health. Good evidence of long-term benefits.
- 7 Improved access to NHS for ethnic minorities—e.g., by appointing link workers.
- 8 Adding folate to flour. Prevents spina bifida in babies; early evidence suggests it may prevent heart disease and Alzheimer's.
- 9 Free school milk.
- 10 Free smoke alarms. Good evidence they save lives.

**(c) New Twenty-First Century Examples of National Policies and Global Recommendations for Improving Population Health and Promoting Health Equity: Moving Beyond a Biomedical and Lifestyle Approach to One Informed by Social Epidemiologic Frameworks**

**Sweden's new Public Health Policy (2003)**

*Conceptual approach:*

**Swedish National Institute of Public Health. *Sweden's New Public Health Policy: National Public Health Objectives for Sweden*. Revised edition: 2003. (Swedish NIPH, 2003)**

question about what sort of data provide appropriate evidence for particular types of decisions. The sort of evidence gathered on the benefits of interventions aimed at individuals may not help in guiding policies directed towards reducing health inequalities... .

The task of the Acheson inquiry was to make recommendations that would reduce inequalities in health, not merely have a positive overall health benefit. For most of the evaluation group's suggested interventions there are no high quality controlled studies showing that they would reduce health inequalities... . Despite occasional rhetorical interest in wider determinants of health, evidence based assessments are largely restricted to individualized interventions. The *Cochrane Library* is unlikely ever to contain systematic reviews or trials of the effects of redistributive national fiscal policies, or of economic investment leading to reductions in unemployment, on health.

The insidious nature of this mismatch between evidence and policy is highlighted by the fact that the evaluation group is, as one would expect of such informed commentators, aware of the problem, while implicitly ignoring it. One of the evaluation group stated when launching the '10 steps to health equity,' 'Our recommendations are quite medical because are the sort that tend to have evidence behind them.'"

**WHO Commission on the Social Determinants of Health (2008)**

*Conceptual statement:*

**WHO Commission on Social Determinants of Health. *Closing the Gap in a Generation: Health Equity Through Action on the Social Determinants of Health. Final report of the Commission on Social Determinants of Health*. (WHO CSDH, 2008)**

pp. 5–6: “During the post-Second World War decades, the nature of health issues became increasingly more medical and professional. The discovery of new drugs and other irrefutable medical progress laid the foundations for a strong belief in the ability of doctors and the health service to solve all the major health problems. Health policy became increasingly synonymous with medical care policy, with the debates centering on how we should finance and recruit personnel to an ever-swelling hospital sector. Preventive health care tended increasingly to take a back seat.

Public health work began to regain a stronger position during the 1980s. The spread of AIDS dealt a deathblow to the belief in the health service’s ability to overcome major health problems and instead, many people began to question whether growing medical costs really did lead to an improvement in public health. The realization that there were large and growing class differences even in Swedish society also helped bring about a rethink in health policy...

... Whereas objectives had previously been based on diseases or health problems, health determinants were now chosen instead. Health determinants are factors in society or in our living conditions that contribute to good or bad health....

The benefit of using determinants as a basis is that the objectives will then be accessible for political decisions and can be influenced by certain types of societal measures. If we set objectives in terms of disease, e.g., to reduce the number of heart attacks, they do not provide any guidance as to what measures may be effective for achieving them... Using health determinants as the basis means the vast majority of public health work must take place outside the medical care service. Most of the factors that impact health are to be found outside the spheres of medical competence and knowledge.

p. viii: “...The poorest of the poor have high levels of illness and premature mortality. But poor health is not confined to those worst off. In countries at all levels of income, health and illness follow a social gradient: the lower the socioeconomic position, the worse the health.

It does not have to be this way and it is not right that it should be like this. Where systematic differences in health are judged to be avoidable by reasonable action they are, quite simply, unfair. It is this that we label health inequity. Putting right these inequities—the huge and remediable differences in health between and within countries—is a matter of social justice. Reducing health inequities is, for the Commission on Social Determinants of Health (hereafter, the Commission), an ethical imperative. Social injustice is killing people on a grand scale.”

p. 1: “The Commission takes a holistic view of social determinants of health. The poor health of the poor, the social gradient in health within countries, and the marked health inequities between countries are caused by the unequal distribution of power, income, goods, and services, globally and nationally, the consequent unfairness in the immediate, visible circumstances of peoples lives—their access to health care, schools, and education, their conditions of work and leisure, their homes, communities, towns, or cities—and their chance of leading a flourishing life. This unequal distribution of health-damaging experiences is not in any sense a ‘natural’ phenomenon but is the result of a toxic combination of poor social policies, and programmes, unfair economic arrangements, and bad politics. Together, the structural determinants and conditions of daily life constitute the social determinants of health and are responsible for a major part of health inequities between and within countries...

... Our core concerns with health equity must be part of the global community balancing the needs of social and economic development of the whole global population, health equity, and the urgency of dealing with climate change.”

When it comes to influencing unemployment figures, social security, housing segregation and alcohol habits, decisions taken in municipal assemblies and other democratic bodies play a much more important role than efforts made in the medical care sector...

... The overarching aim is to create the conditions for good health on equal terms for the entire population."

#### *Policy*

p. 6: "Eleven general objectives for public health work"

1. Participation and influence in society
2. Economic and social security
3. Secure and favorable conditions during childhood and adolescence
4. Healthier working life
5. Healthy and safe environments and products
6. Health and medical care that more actively promotes good health
7. Effective protection against communicable diseases
8. Safe sexuality and good reproductive health
9. Increased physical activity
10. Good eating habits and safe food
11. Reduced use of tobacco and alcohol, a society free from illicit drugs and doping and a reduction in the harmful effects of excessive gambling

p. 7: "The first six objectives relate to what are normally considered to be structural factors, i.e., conditions in society and our surroundings that can be influenced primarily by moulding public opinion and by taking political decisions on different levels. The last five objectives concern lifestyles which an individual can influence him/herself, but where the social environment normally plays a very important part."

#### *Recommendations*

p. 2: "The Commission's overarching recommendations"

1. Improve Daily Living Conditions.
2. Tackle the Inequitable Distribution of Power, Money, and Resources.
3. Measure and Understand the Problem and Assess the Impact of Action

Components:

pp. 3–9: Improve Daily Living Conditions

- "Equity from the start"
- "Healthy places healthy people"
- "Fair employment and decent work"
- "Social protection across the life-course"
- "Universal health care"

pp. 10–19: Tackle the Inequitable Distribution of Power, Money, and Resources

- "Health equity in all policies, systems, and programmes"
- "Fair financing"
- "Market responsibility"
- "Gender equity"
- "Political empowerment—inclusion and voice"
- "Good global governance"

pp. 20–21: "Measure and Understand the Problem and Assess the Impact of Action"

- "The social determinants of health: monitoring, research, and training"

*Perspectives that need to be incorporated:*

pp. 20–21: “a gender perspective on public health”

pp. 21–22: “a lifetime perspective on public health”

pp. 22–23: “Swedish public health in international perspective”

*Implications for the epidemiologic research agenda:*

p. 6: “It is important to clarify how a determinant impacts health. There is a relationship between greater economic inequality and poorer public health, but the mechanisms behind this relationship have not been particularly well clarified. This means in turn that the public health argument does not carry quite so much weight in the public debate as for example economic arguments do. Formulating public health objectives in terms of health determinants requires public health work to be very much knowledge-based.”

p. 20: “The majority of the research carried out in the health sector is basic medical research or research into disease, disease processes and their treatment. A vast amount of this research is financed by the pharmaceutical industry or by other economic interests associated with the medical care sector.

Research into preventive measures is performed to a substantially less extent and there is hardly any research at all into the social mechanisms of ill health. The latter constitutes just a small percentage of the total research performed.

Research policy reflects both an overconfidence in the medical care services’ ability to solve fundamental health problems and the strong economic interests that exist in the field of medical treatment. An individual

p. 22: Actors—“The role of governments through public sector action is fundamental to health equity. But the role is not government’s alone. Rather it is through the democratic processes of civil society participation and public policy-making, supported at the regional and global levels, backed by the research on what works for health equity, and with the collaboration of private actors, that real action for health equity is possible.”

*Implications for the epidemiologic research agenda:*

pp. 178–179: “The underlying causes of health inequity need to be understood, and evidence is needed on what types of interventions work best to reduce the problem. The evidence base needs strengthening in several respects. First, most health research (funding) remains overwhelming biomedically focused, whereas the largest health improvements arguably come from improvements in the social determinants of health. More interdisciplinary and multidisciplinary research on the social determinants of health is needed... Moreover, evidence on the social determinants of health can be context dependent. Responses to inequities will reflect a wide range of factors, including the culture and history of a society. Understanding the impact that context has on health inequities and the effectiveness of interventions requires a rich evidence base that includes both qualitative and quantitative data. Evidence needs to be judged on fitness for purpose—that is, does it convincingly answer the question asked—rather than on the basis of strict traditional hierarchies of evidence.”

p. 196: “Some of the overarching research needs that have emerged from the work of the Commission are:

1. The determinants of health in addition to the determinants of average population health:
  - understanding reasons for the relationship between social stratification and health outcomes;

and often deep-rooted biological approach dominates within the field of medicine, resulting in socially determined health discrepancies being studied relatively seldom or in many cases being ignored completely.

There is a substantial need for long-term competence building and research into the social causes of health and ill-health. Concerning basically all the social determinants of health, there is a need for research into the modes of action and the efficacy of various health policy strategies. Effective knowledge-based preventive measures need to be systematically developed.

In partnership with the Swedish Council for Working Life and Social Research, the Institute of Public Health has been instructed to analyse the Swedish public health research and propose improvements."

*For updates on the policy and additional background on its context, see:*

- (a) Swedish National Institute of Public Health. Public health policy—11 objectives. Updated March 12, 2009. (Swedish NHIP, 2009)
- (b) Sundin J, Willner W. Social Change and Health in Sweden: 250 Years of Politics and Practice. (2007) (Swedish NHIP, 2007)
- (c) Wall S, Persson G, Weinehall L. Public health in Sweden: facts, visions and lessons. (2003) (Wall et al., 2003)

- understanding the interaction between aspects of stratification (for example, gender, ethnicity, and income) and health inequities;

- quantifying the impact of supra-national political, economic and social systems on health and health inequities within and between countries.

- 2. Interventions, global to local, to address the social determinants of health and health equity...

- 3. Policy analysis...

- 4. Monitoring and measurement:

- developing new methodologies for measuring and monitoring health inequities, and for assessing the impact of population-level interventions."

*For updates and commentaries on the WHO CSDH final report, see:*

- (a) World Health Organization Executive Board Resolution: "Reducing health inequities through action on the social determinants of health" (EB124.46) (WHO, 2009)
- (b) The Wellcome Trust Centre for the History of Medicine at UCL: "The World Health Organization and the social determinants of health: assessing theory, policy and practice" (Wellcome Trust, 2008)

For epidemiologists—and others—to know the history of our field and its theories of disease distribution is vital. Such knowledge can help us avoid errors, spark new ideas, and enable us to be as critical conceptually as we are methodologically of the studies that comprise the substantive contributions of our field. At issue are not only the questions asked and interpretations offered but also those ignored. By making conscious use of epidemiologic theory and having informed debates over the different theoretical perspectives at play, we stand a better chance of producing epidemiologic knowledge truly useful for preventing disease, promoting health equity and improving the public's health. If epidemics are

that which befall the people, it is our professional obligation to do the best work we can, with the clearest thinking possible, to identify what will allow the people to stand once again. Our commitment to the people's health—and to explaining the people's health—demands no less. Embracing, extending, debating, and improving epidemiologic theory is one very good place to start.

EBSCOhost®