Chapter 9: Factors Critical to the Design and Execution of Epidemiologic Studies and Description of an Innovative Technology to Follow the Progression From Normal to Cancer Tissue

Montserrat Garcia-Closas, Susan E. Hankinson, Shuk-mei Ho, Donald C. Malins, Nayak L. Polissar, Stefan N. Schaefer, Yingzhong Su, Mark A. Vinson

The results obtained from experimental studies of estrogen carcinogenesis need validation in epidemiologic studies. Such studies present additional challenges, however, because variations in human populations are much greater than those in experimental systems and in animal models. Because epidemiologic studies are often used to evaluate modest differences in risk factors, it is essential to minimize sources of errors and to maximize sensitivity, reproducibility, and specificity. In the first part of this chapter, critical factors in designing and executing epidemiologic studies, as well as the influence of sample collection, processing, and storage on data reliability, are discussed. One of the most important requirements is attaining sufficient statistical power to assess small genetic effects and to evaluate interactions between genetic and environmental factors. The second part of this chapter describes innovative technology, namely, Fourier transform-infrared (FT-IR) spectra of DNA that reveal major structural differences at various stages of the progression from normal to cancer tissue. The structural differences become evident from wavenumber-by-wavenumber statistical comparisons of the mean FT-IR spectra of DNA from normal to cancer tissues. This analysis has allowed distinguishing benign tissues from cancer and metastatic tissues in human breast, prostate, and ovarian cancers. This analysis, which requires less than 1 μg of DNA, is predicted to be used for detecting early cancer-related changes at the level of DNA, rather than at the cellular level. [J Natl Cancer Inst Monogr 2000;27:147–56]

In the study of estrogen carcinogenesis, it has become apparent that results obtained from experimental studies need validation in epidemiology/population studies. However, because variations in a human population are much greater than those existing in experimental systems and in animal models, population studies present additional challenges. In addition, because epidemiologic studies are frequently used to evaluate modest differences in risk factors and, therefore, in their design, it is essential to minimize sources of errors and technical variations and to choose methods with maximum sensitivity, reproducibility, and specificity. In these regards, this chapter focuses on two important topics: One deals with technical issues in study design and in statistical power requirements, and the other focuses on the development of a new technology to measure a surrogate cancer risk marker.

Several methodologic challenges and technical hurdles in designing and executing epidemiologic/population studies are discussed in this chapter. Important issues that are discussed include reproducibility of laboratory assays, limitations imposed by the small amount of plasma/serum collected, and the validity of using a single sample per subject. The chapter also discusses in detail the influences of sample collection, processing, and storage methods on data reliability. Finally, the importance of attaining adequate statistical power by reaching the required sample sizes is highlighted.

Several important lessons are enumerated. 1) Collection protocols need to minimize variations in factors that are not of etiologic interest by standardizing case and control subjects on these factors. 2) Sample collection, processing, and storage procedures must be subjected to stringent scrutiny to ensure that variations in these steps will not mask the modest differences expected to exist in the risk factors of interest. 3) Study design must take into consideration the limitations linked to within-person variation over time as well as the single sample per subject collection method and, therefore, whenever possible, repeated sampling should be considered. 4) Comparison of data collected from different laboratories may be difficult because large variations exist in different study populations and in laboratory methods. Introduction of a standardization or validation program should be considered for multisite analyses. 5) To attain statistical power in case–control studies, larger sample sizes are needed for studies that are assessing small genetic effects. Furthermore, if the goal of the study is to evaluate interaction among factors, sample sizes need to be increased accordingly.

The second half of this chapter focuses on breakthrough technology referred to as the Fourier transform-infrared/statistics model, which has been successfully adapted for analyses of DNA changes in cancer and precancerous tissues *(1–6)*. Infrared spectra generated by applying infrared beams to sample DNA produced a large number of spectra. Fourier transform spectral data analyses, coupled with statistical comparisons, yield a few

Affiliation of authors: M. Garcia-Closas, Environmental Epidemiology Branch, National Cancer Institute, Bethesda, MD; S. E. Hankinson, Department of Epidemiology, Harvard School of Public Health and Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; S. Ho, University of Massachusetts Medical School, Department of Surgery, Division of Urology, University Campus, Worcester; D. C. Malins, Molecular Epidemiology Program, Pacific Northwest Research Institute, Seattle, WA; N. L. Polissar, The Mountain-Whisper Light Statistical Consulting, Seattle, and Department of Biostatistics, University of Washington, Seattle; S. N. Schaefer, Y. Su, M. A. Vinson, Molecular Epidemiology Program, Pacific Northwest Research Institute, Seattle.

Correspondence to: Shuk-mei Ho, Ph.D., University of Massachusetts Medical School, Department of Surgery, Division of Urology, Office of Urologic and Translation Research, 55 Lake Ave., No., Worcester, MA (e-mail: Shuk-Mei.Ho@UMASSMED.EDU).

See "Notes" following "References."

© Oxford University Press

principal components that were shown to be sufficient for discriminating DNA alterations in precancerous and cancer DNA samples. The technology has recently been developed further to produce cancer probability-risk score models for breast and prostate cancers. With the use of a model developed for breast cancer, normal tissue, primary tumors, and metastasizing tumors were correctly discriminated at more than 80% probability. The method also has the added advantage of requiring only a small amount of DNA (<1.0 μg) and, therefore, is potentially suitable for analyses of needle biopsy samples. Should the technology live up to its promise, it will provide a highly sensitive and reliable diagnostic and risk-assessment method for clinical and population studies.

Study Design Considerations in the Assessment of Cancer Risk in Relation to Genetic Polymorphisms

Polymorphisms in genes coding for enzymes or receptors involved in the metabolism and intracellular transport of estrogens could influence the risk of developing breast cancer *(7)*. The case–control design is the most commonly used to evaluate associations between genetic polymorphisms and the risk of common diseases in the population, as well as interactions between genetic and environmental risk factors *(8)*. In this type of study, the odds ratio is used to measure the association between a particular genotype and the risk of disease. The number of case and control subjects that would need to be included in such studies to have an 80% power (two-sided test with 5% Type I error) to detect a genotype effect, as a function of the prevalence of the genotype and the magnitude of the effect, is illustrated in Fig. 1. According to Fig. 1, studies including 200 case patients and 200 control subjects or fewer will be able to detect moderate to large genetic effects (odds ratio [OR] ⩾ 2.0) for a wide range of genotype prevalences. However, a minimum of 400 case patients and 400 control subjects will be needed to detect small genetic effects (OR = 1.5).

It is likely that a gene coding for a particular metabolizing enzyme confers disease susceptibility in combination with genes coding for other enzymes involved in the same metabolic pathway or in combination with other determinants of substrate levels. Therefore, studies should be designed to be able to evaluate potential gene–gene and gene–environment interactions. The assessment of interactions requires large sample sizes to attain adequate statistical power, especially when the factors under study are either very rare or very common or when the magnitude of the interaction is modest *(9–12)*. Estrogen-related risk factors, such as reproductive characteristics or body size, have small to moderate effects on breast cancer risk, and susceptibility genotypes, such as those for metabolizing enzymes, are also likely to have small to moderate effects. Therefore, we expect to observe modest interactions, unless the genetic or environmental factors are very rare, and, in both situations, large samples are needed. The sample size requirements to have an 80% power to detect an example of a twofold multiplicative interaction (ratio of stratum-specific ORs of 2.0), as a function of the prevalence of the environmental and genetic factors, is presented in Fig. 2. In this example, it is assumed that the genotype effect in the absence of the exposure and the exposure effect in the absence of the genotype are both 2.0 and that the genetic and environmental factors occur independently in the population. This figure indicates that a minimum of about 550 case patients and 550 control subjects will be needed to study this type of interaction (genotype and exposure prevalence of 30%–50%), and a much larger sample size will be required to study less common or more common genetic or environmental factors, especially if prevalences are lower than 20%. Similarly, the sample size will increase as the genotype or exposure prevalence becomes larger than 30%–50%.

The sample size to study multiplicative interactions is mainly determined by the susceptibility genotype and exposure prevalence and by the magnitude of the interaction *(11)*. The magnitude of the effect of the genotype in the absence of the exposure and the effect of the exposure in the absence of the genotype

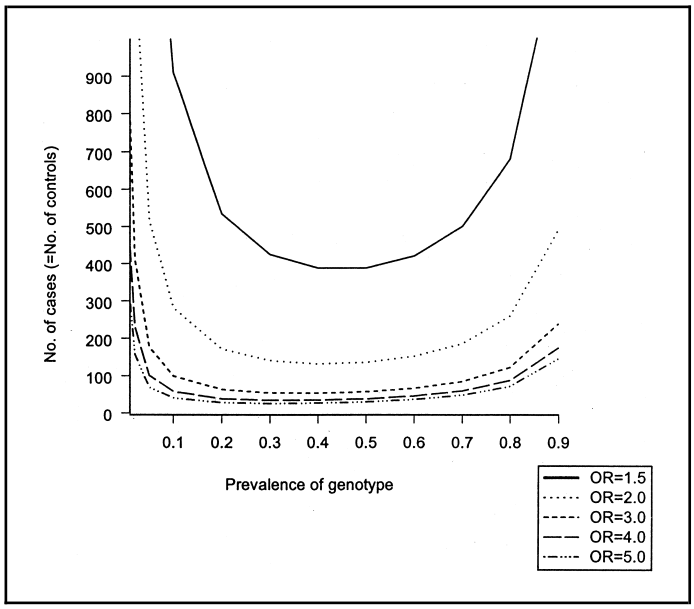

Fig. 1. Number of case patients required to have an 80% power to detect a genotype effect with the use of a two-sided test with 5% Type I error, as a function of genotype prevalence and magnitude of effect.

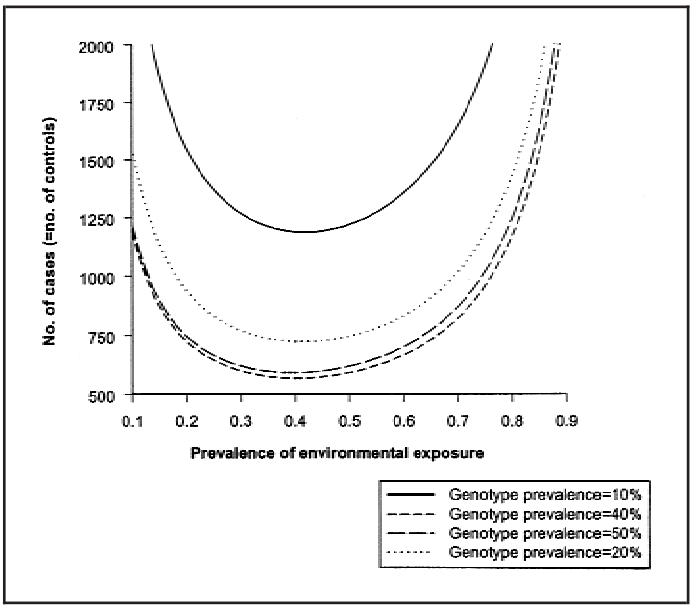

Fig. 2. Number of case patients required to have an 80% power to detect an example of a twofold gene–environment interaction with the use of a two-sided test with 5% Type I error, as a function of genotype and exposure prevalence.

affect the sample size to a lesser extent. Thus, the sample sizes illustrated in Fig. 2 will be similar for other examples of twofold gene–environment interactions in which the genotype or exposure effects are smaller than 2.0. Moreover, the sample size needed to study a 1.5-fold interaction, rather than a twofold interaction, will be two to three times higher.

Accurate information is essential for the study of environmental exposures and their potential interactions with genetic polymorphisms *(13,14)*. Misclassification of exposure, either differential or nondifferential with respect to disease status, will tend to underestimate a multiplicative interaction effect, provided that exposure misclassification is independent of the genotype and that the exposure status is independent of the genotype in the population *(15)*. As a consequence, the sample size required to detect the attenuated interaction with adequate statistical power will be increased.

The impact of misclassification on estimation and sample size is highly dependent on the exposure prevalence *(13,14)*. For instance, the use of an instrument with low sensitivity (proportion of exposed subjects correctly classified) can have a very strong impact when the exposure frequency is high, but it may not be so deleterious when the frequency is low. It should also be taken into account that even small errors in the genotype determination can have substantial impact on the sample size, especially when environmental exposure is also measured with error. The impact of exposure and genotype misclassification on sample size, when the prevalence of both factors is 50%, is illustrated in Table 1. The sample sizes correspond to the same example of gene–environment interaction as in Fig. 2. In the absence of genotype misclassification, an exposure sensitivity of 80% increases the sample size about 1.7 times (from 590 to 1600 subjects); in the absence of exposure misclassification, a genotype sensitivity of 95% increases the sample size 0.5 times (from 590 to 900 subjects). However, in the presence of both genotype and exposure misclassification, the sample size will be further increased to 2045 case patients and 2045 control subjects.

Alternative study designs can be more efficient to study interactions under certain circumstances and can address potential biases common in case–control studies *(10,16,17)*. For instance, the case-only design can be very efficient in detecting gene–gene and gene–environment interactions, provided that the genes and the environmental exposure are independent in the population; the case–parental control design addresses the problem of confounding of gene effects by ethnicity or population stratification; and cohort studies can address the problem of disease bias because biologic samples and questionnaire data are collected prior to the onset of disease. Mixed study designs or different sampling strategies, such as oversampling women with a positive family history of breast cancer to increase the frequency of potential susceptibility genotypes, may also help to address potential biases or increase the efficiency to study particular hypotheses.

Table 1. Impact of exposure and genotype misclassification on sample size to detect a twofold interaction*

Scenario	Exposure sensitivity, %	Genotype sensitivity, %	No. of subjects
No misclassification	100	100	590
Exposure misclassification	80	100	1600
Genotype misclassification	100	95	900
Exposure and genotype misclassification	80	95	2045

*Interaction model described in text; prevalence of both genotype and exposure = 50%; genotype and exposure specificity = 100%.

Sex Steroid Hormones: Technical Hurdles in Population Studies

Challenges in Epidemiologic Studies

The epidemiologic study of steroid hormones in relation to disease risk poses several methodologic challenges. Frequently, we wish to detect only modest (but etiologically important) differences in hormone levels between study subjects. To optimize the chance of detecting these differences, sources of error and technical variation in our results must be minimized. Protocols for the collection, processing, and storage of study urine or blood samples must be evaluated and optimized. Laboratory assay procedures must be reproducible, require a small sample volume, and, preferably, be consistent across studies. In addition, the ability of a single hormone measurement (as is available in most population studies) to reflect long-term hormone levels must be evaluated. Each of these issues is discussed below.

Blood or Urine Collection, Processing, and Storage

Sample Collection

It is important to establish a specific collection protocol that minimizes the sources of variation in hormone levels that are not of etiologic interest. These sources of variation include fasting status, time of day the sample is collected, and, for premenopausal women, phase of the menstrual cycle. Not standardizing case and control subjects on these factors (in either the design or analysis phase of a study) could substantially attenuate hormone/disease associations or, if the distribution of these factors varied by chance between case and control groups, an association may be detected that does not in truth exist. For example, several adrenal androgens [e.g., dehydroepiandrosterone (DHEA)] and prolactin have substantial diurnal variation *(18)*.

Collecting all blood samples at a single time in the day, matching control subjects to case patients on the time of sample collection, or controlling for time of day of collection in the statistical analysis will remove any noise associated with the circadian variation in hormone levels.

For premenopausal women, the effect of the menstrual cycle on hormone levels is important to consider. A number of hormones, particularly estrogens, fluctuate substantially over the menstrual cycle *(18)*. Thus, similarly to what was described above, to allow a valid comparison between case and control subjects, it is necessary to either collect all samples at approximately the same time in the cycle, match on cycle day, or carefully control for cycle day in the analysis. In general, timing the luteal sample from the first day of the next menstrual cycle is more accurate than counting from day 1 of the current cycle, as the luteal phase is more consistent in length than the follicular phase *(19)*. Accurate matching in the luteal phase requires knowing when the next menstrual cycle began (i.e., the cycle after sample collection). This requires recontacting study participants (or having the participants recontact study staff by mail, for example), thus adding an additional challenge to epidemiologic studies of premenopausal hormone levels. It is in part due to the complexity of collecting well-timed samples from premeno-

pausal women that few studies of premenopausal endogenous hormones and cancer risk have been conducted.

Sample Processing

Ideally, biologic samples would be processed and either analyzed or frozen immediately after sample collection. This action will minimize any deterioration in hormone levels over time. However, in some large population studies, particularly those with a geographically dispersed population, immediate processing and storage are not feasible. The effect of delayed processing and storage on any parameter of interest must be evaluated before study implementation.

Prior to a large blood collection effort in the Nurses' Health Study, the effect of a delay in blood processing on steroid hormone levels was evaluated (Table 2). The stability of endogenous hormones in plasma prepared from whole blood that had been stored for 24 or 48 hours in a sealed Styrofoam mailer cooled with a frozen gel pack was evaluated relative to samples that were immediately processed and frozen *(20).* Overall, the delay in sample processing resulted in little change in hormone levels. Estradiol, percentage of free estradiol, androstenedione, and prolactin all changed by less than 5% per day. The hormone with the greatest percentage of change per day was testosterone (9.5% per day). However, even with this degree of change, the true between-person variation substantially outweighed the error introduced by the delay in processing and the laboratory analysis, as evidenced by the high intraclass correlation coefficient [ICC = 0.86; i.e., between-person variation/(between-person variation + within-person variation)]. More recently, these blood collection methods have also been evaluated for their effect on plasma insulin-like growth factor-I (IGF-I) and insulin-like growth factor-binding protein 3 (IGFBP-3) levels *(21).* IGF-I and IGFBP-3 levels in samples that were processed and serum frozen immediately after venipuncture (the standard processing methods) were compared with samples that were stored in heparinized whole blood for 24–36 hours before processing (mimicking blood collection conditions used in certain studies). The mean IGF-I and IGFBP-3 values were almost identical, and the intraclass correlations between results of the two collection methods were 0.98 for IGF-I and 0.96 for IGFBP-3—again showing that the collection methods did not adversely affect sample integrity.

Sample Storage

Freezer alarm and back-up systems must be in place to prevent thawing or warming of study samples. Twenty-four-hour alarms should be in place, and manual checks of the freezer temperature should be conducted periodically. For added security, each individual's sample should be split between freezers, so that, in the event that a freezer thaws, only part of a sample from any one participant will be lost. To maintain the ability to identify stored samples, cryotubes should be labeled before freezing, using labels with adhesive specifically designed for low temperatures.

Several different freezing options are available: storage in mechanical freezers at either 20 °C or at −70 °C, or in the vapor phase (temperature range, −130 °C to −196 °C) or the liquid phase (constant at −196°C) of liquid nitrogen freezers. Some concern regarding the suitability of upright front-loading mechanical freezers for long-term sample storage was raised in a study conducted by Su et al. *(22).* They evaluated temperature variations in upright mechanical freezers and found that, for freezers set at −80 °C, the internal temperature ranged from −90 °C to −43.5 °C, with the warmer regions being the upper and front sections of the freezers. We are unaware of similar evaluations in chest freezers.

Several studies have used frozen specimens with little if any sign of degradation in the hormone level. Mean levels of plasma testosterone, estradiol, androstenedione, and percentage of free estradiol in samples stored at −70 °C for either 6 or 8 years were not significantly different, although estrone levels were slightly higher in samples stored for only 6 years *(23).* Plasma estradiol, sex hormone-binding globulin (SHBG)-bound estradiol, free estradiol, and prolactin levels remained stable after archiving at −70 °C for 6 months to 6 years *(24,25).* In other study *(26),* both estradiol and prolactin were observed to be stable at −70 °C for 3 years; although testosterone levels varied modestly over the 3 years, the rank correlation remained high (approximate $r = .9$). In contrast to other steroids, plasma progesterone levels were reported to decrease by 40% over a 3-year period, although, again, the rank correlation over time was high ($r = .98$). However, in a second study *(27)* in which plasma was stored at −20 °C, progesterone levels were reported to increase 2.8% per year of storage. Thus, the stability of plasma or serum progesterone levels over time is uncertain, and additional evaluations are needed. Stability of samples in −70 °C or colder over a period of more than 8 years has not yet been evaluated. Although DHEA sulfate has been reported stable when stored at −20 °C for 10–15 years *(28),* the percentage of free (versus bound) estradiol *(29)* and testosterone *(30)* has been reported to increase significantly with storage at this temperature; thus, freezing at −70 °C or colder is preferred.

Although the above studies suggest that hormone levels tend

Table 2. Plasma hormone levels from postmenopausal women according to delay in processing specimens after phlebotomy

Hormone	Mean hormone levels according to delay in processing after phlebotomy			% change per day*	CV_b/CV_w†
	0 h	24 h	48 h		
Androstenedione, ng/L	516	493	482	−3.4 (−6.9, 0.2)	3.0
Estradiol, ng/L	19	19	20	3.5 (−1.7, 8.7)	1.5
% free estradiol‡	1.8	1.9	1.8	−1.3 (−4.7, 2.1)	4.0
Prolactin, mg/L	8.8	8.9	8.8	−0.4 (−1.9, 2.7)	9.2
Sex hormone-binding globulin, nmol/L	31.3	30.1	32.1	1.3 (−2.4, 5.0)	4.4
Testosterone, ng/L	221	242	262	9.5 (1.7, 17.3)	3.5

*Numbers in parentheses = 95% confidence intervals associated with the percent change per day.

†Ratio of the between-person (CV_b) and within-person (CV_w) coefficients of variation.

‡n = 5; all others, n = 9 subjects.

to be stable for relatively long periods if stored at −70 °C or colder, it remains advisable to match case patients and control subjects on length of sample storage to minimize the effect of any modest changes on case/control comparisons. In addition, we are not aware of data that address the stability of hormone levels with repeated freezing and thawing of the biologic material; thus, it is recommended that freeze/thaw cycles be minimized.

Validity of Using a Single Sample per Subject

Although average long-term hormone levels are often of primary interest in epidemiologic studies of hormones and cancer, for both economic and logistical reasons, frequently only one blood sample is collected per study subject. The degree to which this one sample can represent an individual's long-term levels depends on the degree of within-person variation (relative to the between-person variation) over time. The more representative a single sample is of long-term levels, the greater the chance of detecting true differences between study subjects. This issue has been evaluated in several studies.

Correlations for plasma estrogens over approximately a 2-year period in postmenopausal women ranged from 0.36 *(31)* to 0.94 for the percentage of bioavailable estradiol (Table 3) *(25)*. How well a single postmenopausal hormone measure reflects levels over a 3-year period was evaluated and an ICC of 0.66–0.92 was found for the sex steroid hormones *(32)*. Prolactin had a somewhat lower ICC of 0.53. For IGF-I, among 24 adults who had two blood samples drawn on average 6 weeks apart, the ICC was 0.94 ($P = .001$) *(33)*.

In a study of 26 premenopausal women in which two luteal phase samples were collected about 1.5 years apart, the correlation for the repeated samples was 0.70 for androstenedione and 0.73 for testosterone (Table 3) *(34)*. In samples from 71 premenopausal women that were collected randomly throughout the menstrual cycle over a 1- to 2-year period, the ICC was 0.72 (95% confidence interval [CI] = 0.57–0.83) for percentage of free estradiol and 0.83 (95% CI = 0.73–0.90) for SHBG-bound estradiol ("nonbioavailable" estradiol) *(25)*. In the same study *(24)*, the ICC for prolactin was 0.48 (95% CI = 0.31–0.63). More recently, repeated luteal phase samples were collected 1 year apart from 60 premenopausal women; the ICCs were 0.85 for DHEA sulfate and 0.60 for total testosterone *(35)*. In a study of urinary estrogen metabolite levels, the ICC for the ratio of urinary 2-hydroxyestrone to 16-α-hydroxyestrone over a 6-month period was 0.67 *(36)*.

In the study by Muti et al. *(35)*, the ICC for estradiol in the luteal phase was reported to be just 0.06, suggesting extremely poor reproducibility over time. However, this result may be related to the investigators' inability to exclude women with anovulatory cycles or to pinpoint when in the menstrual cycle the sample was collected. (In this study, the samples were collected on days 20–24, counting forward from day 1 of the cycle, and the date of start of the next menstrual cycle was not available). In a more recent study *(37)*, we found much higher reproducibility (ICC = 0.62) over a 1-year period when we included only ovulatory women who provided their samples in the midluteal phase of their cycle. Thus, the reproducibility of estradiol levels (in the same phase of the cycle) in premenopausal women needs further evaluation. Intraclass correlations for plasma estrogens from the largest studies to address this issue are provided in Table 3.

Although the data are not entirely consistent, in general, this level of reproducibility (ICC of 0.5–0.8) is similar to that found for other biologic variables, such as blood pressure, pulse, and cholesterol measurements, all exposures that are considered to be reasonably well measured and are consistent predictors of disease in epidemiologic studies *(38)*. Of note, reproducibility data such as these (that measure within-person variation in levels over time) can also be used to explicitly correct for measurement error in studies of plasma hormones and disease risk *(39)*.

Table 3. Intraclass correlation coefficients (ICC) and 95% confidence intervals (95% CI) for plasma estrogens in samples collected over a 1- to 3-year period

Hormone	No. of women	ICC	95% CI
Among postmenopausal women*			
Estradiol	79	0.68	(0.59–0.80)
% free estradiol	79	0.80	(0.73–0.87)
% bioavailable estradiol	79	0.86	(0.82–0.92)
Estrone	79	0.74	(0.66–0.83)
Estrone sulfate	79	0.75	(0.67–0.84)
Among premenopausal women			
Estradiol (luteal phase)†	60	0.06	(0.00–§)
% free estradiol‡	71	0.72	(0.57–0.83)
% SHBG estradiol‡	71	0.83	(0.73–0.90)
Estradiol (luteal phase)‖	39	0.62	(0.43–0.78)
Estradiol (follicular phase)‖	85	0.53	(0.37–0.67)

*Hankinson et al. *(32)*; three measurements collected over a 2- to 3-year period.

†Muti et al. *(35)*; two measurements collected over a 2- to 3-year period.

‡Toniolo et al. *(25)*; two to three measurements collected over a 2- to 3-year period. SHBG = sex hormone-binding globulin.

§Not provided.

‖Two measurements over a 1-year period *(37)*. Luteal phase estradiol was calculated among women with progesterone level ⩾ 300 ng/dl who collected each sample 4–10 days before the start of the next cycle.

Assay and Laboratory Precision

Overview

In contrast to clinical needs, in epidemiologic studies we are generally interested in detecting modest differences within the normal range of hormone levels; laboratory error could easily result in true (and important) associations being missed. This issue has been particularly important in the measurement of plasma estrogens in postmenopausal women, as normal levels are in the picogram per milliliter range and between-person variation in levels is relatively small. Given the limited quantity of plasma collected in most population studies, being able to conduct the assay with a small plasma volume is also important and makes high reproducibility (and high sensitivity) even more difficult to achieve. Another issue is the varying sensitivity and specificity of hormone assays used by different laboratories and, thus, by different epidemiologic studies. These differences make the comparison of findings between published studies difficult.

Reproducibility and Validity of Hormone Assays

Several studies have been conducted to assess the ability of laboratories to reproducibly measure plasma steroid levels in postmenopausal women *(40–43)*. We sent replicate samples of plasma to each of four well-established endocrine laboratories in the United States on one or two separate occasions. All replicate

samples were handled identically during processing, storage, and retrieval, and they were labeled to preclude their identification by the receiving laboratory. The within-person coefficient of variation, a measure of laboratory error frequently reported by laboratories, was consistently low (<15%) for several hormones. For estrone and estradiol, however, hormones present at low levels in postmenopausal women, the laboratory error was often large (>25%), and the ratio of between-person variation to laboratory error was often less than 2.0. Several other studies have also reported variability in assay reproducibility of both plasma *(41,42)* and urinary *(43)* steroid hormones, although results depended on the laboratory conducting the assay, the specific hormone, and the menopausal status of the woman.

A number of factors may have influenced the variable reproducibility observed. First, differences in laboratory methods may be important. For example, in our study, although all laboratories used radioimmunoassay (RIA) to measure estradiol, two laboratories used celite column chromatography and one laboratory used LH20 Sephadex column chromatography prior to RIA, whereas the fourth laboratory did not use a separation step prior to RIA. The laboratory method could not have been the only source of error, however, because results also varied within a single laboratory (e.g., the CV for estradiol ranged from 8% to 59%). These substantial differences might relate to a change either in the laboratory personnel or in the reagents and equipment used in the assays or perhaps varying levels of performance by the same technician or piece of equipment over time. In addition, some laboratories may be set up primarily to assay clinical specimens. The level of error tolerable in a clinical setting, in which the distinction between normal and abnormal hormone levels is of primary interest, is substantially greater than that which can be tolerated in epidemiologic research, in which relatively small differences within the spectrum of normal hormone levels are the subject of investigation.

Another technical challenge in studies of hormones is that a number of different laboratory methods are used to measure the same hormone, and no standardization or validation programs exist. For example, in several studies in which plasma estradiol was measured in postmenopausal women, mean levels were 9 pg/mL *(44),* 13 pg/mL *(45),* and 28 pg/mL *(46).* To what degree these differences represent different study populations or simply differences in laboratory methods is unclear and complicates any comparison of results between the studies. The comparison of different laboratory methods against a "gold standard" would be helpful in resolving this issue; however, it is unclear which analytic method would be most appropriate as the gold standard.

Summary and General Recommendations

On the basis of our current knowledge, several recommendations can be made to epidemiologists wanting to use hormone measurements in their research. Close collaboration with laboratory experts should be obtained in the planning stages of a study and should continue through its conclusion. Any variation from the standard collection and processing procedures should be evaluated prior to their implementation. Before having any study blood samples analyzed, laboratory performance should be independently evaluated. After this initial assessment, a proportion of samples sent to the laboratory with each batch of study samples should be quality-control specimens that are indistinguishable from the case and control specimens. Matched case–control pairs should be handled identically and together, shipped in the same batch, and assayed in the same analytical run. All assays should be conducted without knowledge of the case/control status. Identical handling of all case and control specimens is critical to validity, as any possible deterioration related to collection, processing, or storage should affect case and control specimens equally and will not appreciably affect measures of association. Finally, collection of repeated blood or urine samples from a subset of study subjects should be considered; this collection will allow both the evaluation of within-person variability over time and the use of measurement error correction techniques in the calculation of relative risks.

Fourier Transform-Infrared/Statistics Models

Fourier transform-infrared (FT-IR) spectra of DNA have revealed major structural differences at various stages in the progression of morphologically normal estrogen-responsive tissues (ERT) to cancer *(1–5).* Reactions of the hydroxyl radical (·OH) with the base *(1–6,47–50)* and deoxyribose *(1–5)* structures have been implicated as major contributors to these modifications, although other factors, to include hypermethylation *(51)* and the formation of depurinating adducts *(52),* may modify DNA spectra. In ERT (e.g., the human breast), the ·OH is believed to arise from the metal-catalyzed decomposition of H_2O_2, which is produced from redox cycling of catechol estrogen metabolites *(48)* and certain xenobiotics (e.g., aromatic hydrocarbons) *(53).* The structural differences are evident from wavenumber-by-wavenumber statistical comparisons of the mean FT-IR spectra of DNA (extracted with pheonol) *(1)* from normal and transformed tissues (e.g., normal prostate versus prostate cancer) *(5).* Principal component analysis (PCA) *(4)* allows most of the information in each spectrum to be represented by a few principal components (PCs), the first three usually accounting for more than 80% of the total variance. Each PC score is a weighted sum of spectral absorbances. Plots can be constructed on the basis of the first two or three PCs. In these plots, a point represents a single spectrum, and groups (clusters) of points represent the DNA from a particular tissue type (e.g., prostate cancer). In the carcinogenic transformation of one tissue type to another (e.g., normal → cancer), the location of the cluster and its diversity in PC space are important measures of DNA change *(49).* When spectral differences exist between the DNA of tissue groups in a disease progression (e.g., normal tissue → cancer), discriminant analysis can be used to establish cancer prediction models, such as those reported for breast *(1,4)* and prostate *(4,5)* cancers.

Prototype prediction models, based on multivariate analysis of infrared spectral data, have been developed, and they have an ability to potentially differentiate between tissue groups that were not satisfactorily differentiated by simpler statistical models. These models can potentially distinguish nonmetastatic primary tumors from those with disseminated metastases. Examples of FT-IR/statistics models for predicting cancer-related changes in DNA prior to evidence for cellular transformations are presented, together with discussion of their clinical and etiologic implications.

Significant differences were found between the mean absorbances of DNA from the morphologically normal ovary (O_n) and ovarian adenocarcinoma (AC) over most of the spectral region (Fig. 3, A) *(54).* The *P* values are presented for each wavenumber (Fig. 3, B). Statistically significant differences (from about 1650 cm^{-1} to 1680 cm^{-1}, 1200 cm^{-1} to 1260 cm^{-1}, and 1000 cm^{-1} to 1150 cm^{-1}) are evident in spectral areas assigned to

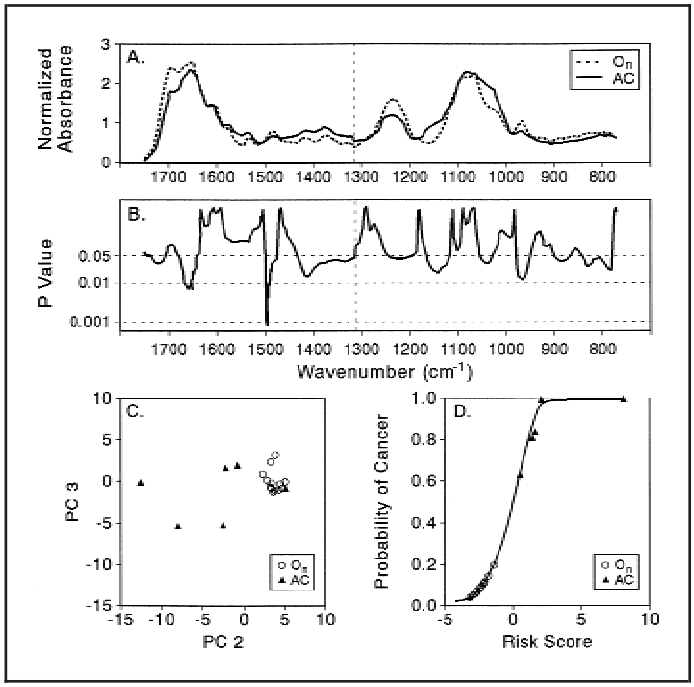

Fig. 3. Spectral comparisons of ovarian DNA *(53).* **A**) Grand mean DNA spectra of morphologically normal ovarian tissue (O_n; n = 13) and primary ovarian adenocarcinoma (AC; n = 6); **B**) *P* values for spectral comparison in **A** (unequal variance *t* test) *(53);* **C**) principal component (PC) plot comparing spectra of ovarian DNA from morphologically normal tissues (O_n) and primary adenocarcinoma (AC) *(53);* and **D**) plot of the probability of ovarian cancer with the risk score for the O_n and AC. The null hypothesis that the PC scores do not discriminate between the groups is rejected with $P<.001$. Reprinted with permission © 1998 from the National Academy of Sciences, U.S.A.

vibrations of the nucleotide bases, the PO_2^- group and deoxyribose, respectively *(54).* PCA of the spectral data provided two major PCs that were plotted against each other (Fig. 3, C). The plot revealed that the O_n formed a tight, ordered group of points, whereas the ACs were highly diverse and relatively disordered. The relationship between the probability of ovarian cancer and the risk score derived by discriminant analysis is shown in Fig 3, D. The ovarian cancer group is located primarily at the top of the sigmoid-like curve, and the noncancer group is located at the bottom. The predicted probability scores rise rapidly over a narrow range, which reflects a high degree of discrimination between the groups. The disorder reflected in the AC and the metastasized primary ovarian adenocarcinomas (AC_m) contrasts with the order in the O_n and distant ovarian metastases to the colon (AC_{dm}), as apparent from the mean spectral comparisons (Fig. 4, A) and the PC plot in which the points of each group substantially overlap (Fig. 4, B) *(54).* Despite the inability to discriminate between the two ordered DNA systems with the use of spectral comparisons and PCA, comparisons of standard deviations of absorbances at each wavenumber over the entire spectral range revealed increased spectral diversity in the AC_{dm} in regions assigned to base vibrations but not in those relating to the furanose ring. This finding is consistent with the presence of increased base mutations in the DNA of the distant metastases *(54).*

Comparisons of the mean spectra of DNA from morphologically normal breast tissues obtained from breast reduction surgery (reduction mammoplasty tissues, RMT) and invasive ductal carcinoma (IDC) tissues revealed characteristic differences in spectral regions assigned to the base and deoxyribose structures *(1–4).* A three-dimensional plot of the points from PCA is given in Fig. 5, A. The points representing the RMT are clustered primarily in the upper-left region of the plot. The IDC, comprising primary tumors with and without evidence for axillary metastases, are broadly dispersed, thus indicating considerable structural diversity. Discriminant analysis of the spectral data provided a relationship between the probability of cancer and the risk score (Fig. 5, B), having a sensitivity of 86% and a specificity of 81%. The DNA of "benign" (microscopically normal) tissues from near the breast tumors of 11 women (not included

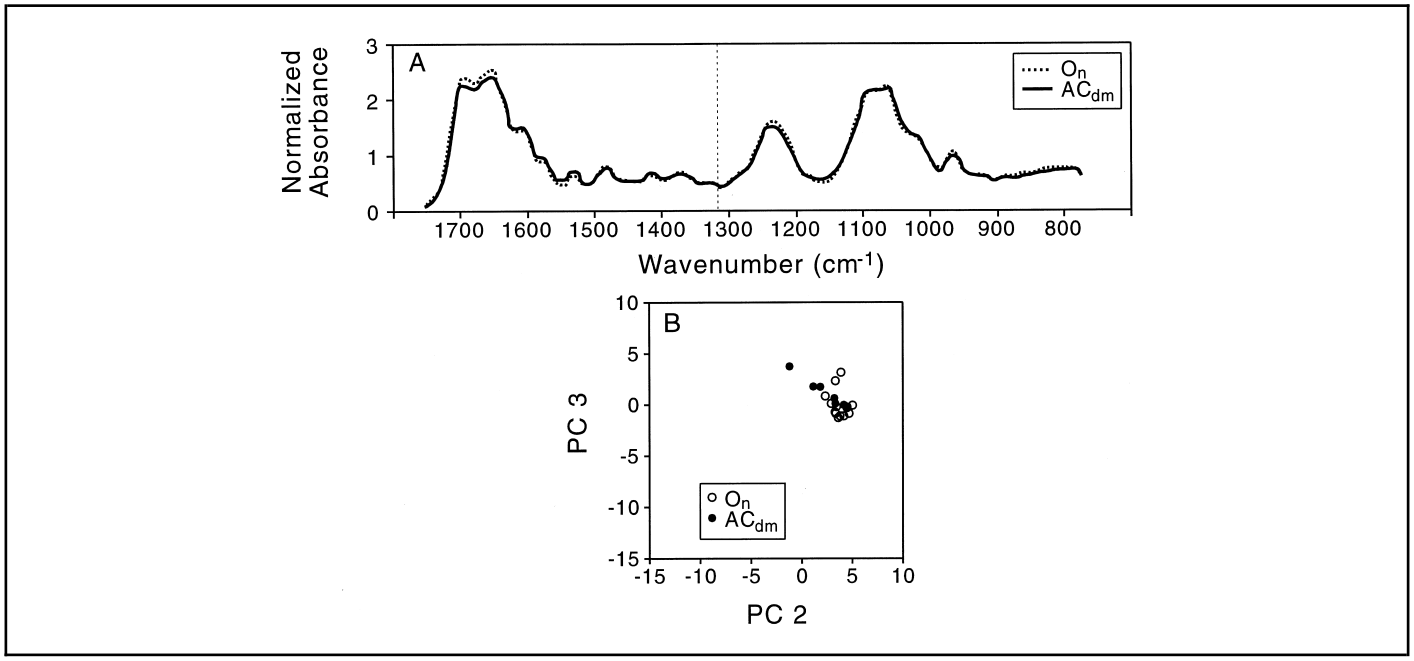

Fig. 4. Spectral comparisons of ovarian DNA *(53).* **A**) Mean spectra of DNA from the morphologically normal ovary (O_n; n = 13) compared with mean spectra of DNA from ovarian adenocarcinoma metastases to the colon (AC_{dm}; n = 7); **B**) principal components plot comparing the mean DNA spectra from each group shown in **A.** Reprinted with permission © 1998, National Academy of Sciences, U.S.A.

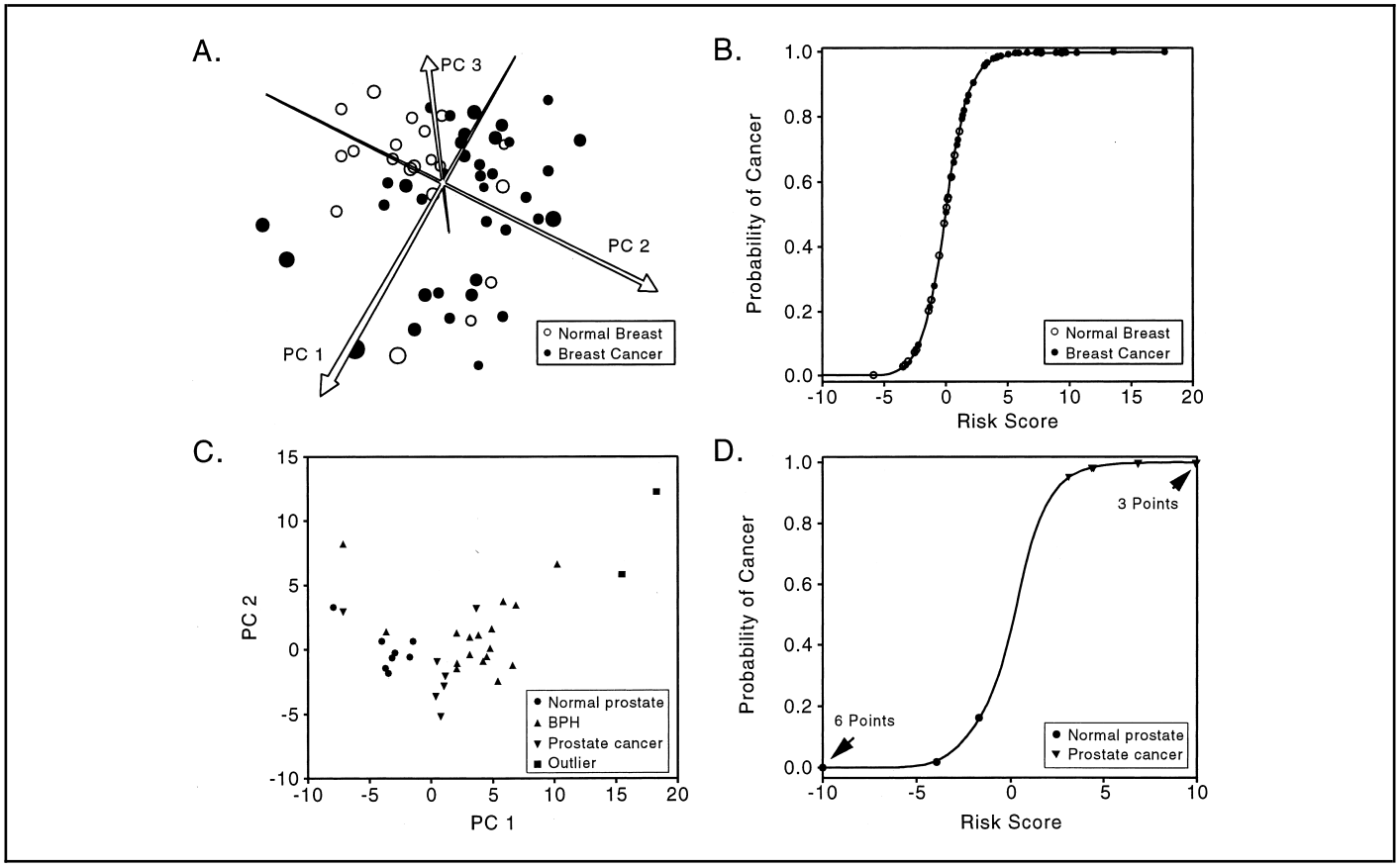

Fig. 5. Spectral comparisons of breast DNA *(4).* **A**) Three-dimensional plot of principal component (PC) scores of DNA spectra from normal breast (n = 21) and breast cancer (invasive ductal carcinoma [IDC]; n = 37) tissues showing distinct clustering of each group *(1,3);* **B**) plot of the probability of cancer with the risk score for the normal breast and breast cancer. The null hypothesis that the PC scores do not discriminate between the groups is rejected with $P<.0001$ *(1);* **C**) two-dimensional plot of PC scores of DNA spectra from normal prostate (n = 8), benign prostatic hyperplasia (BPH; n = 18) and prostate cancer (adenocarcinoma; n = 8) in which the clustering is distinct *(5);* **D**) plot of the probability of cancer versus the risk score for normal prostate and prostate cancer. The null hypothesis that the PC scores do not discriminate between the groups is rejected with $P = .009$ *(5).* Reprinted with permission by *Nature Medicine.* Portions A and B originally appear in Nat Med 1997;3:927–30.

in the predictive model) was analyzed *(4).* When the scores were used in the breast cancer probability-risk score model (Fig. 5, B), 10 of 11 had a predicted cancer probability of more than 75%. This finding is consistent with data showing that tissue near a breast tumor has a high risk for forming a second cancer lesion *(55).*

In studies of the human prostate, the mean spectral differences between the DNA of normal tissue and the DNA of prostatic adenocarcinoma were substantial. The PC plot revealed pronounced discrimination between DNA spectra of normal and cancer tissues *(4,5).* A similarly effective separation was obtained between the clusters of DNA points representing normal tissue, prostatic cancer, and benign prostatic hyperplasia (BPH) (Fig. 5, C). The discriminant analysis models that predict disease probability (normal prostate tissue versus prostate cancer; normal prostate tissue versus BPH) had sensitivities and specificities of 100% for both comparisons. These models are based on more PC scores than the two-dimensional PC plots. The relationship between the normal prostate DNA and the DNA of prostatic adenocarcinoma, expressed in terms of cancer probabilities, is shown in Fig. 5, D.

Prototype statistical models, based on FT-IR spectroscopy, are being tested in our laboratory. These models hold promise for distinguishing the DNA from primary tumors and metastasizing primary tumors (those that have given rise to dissemi-

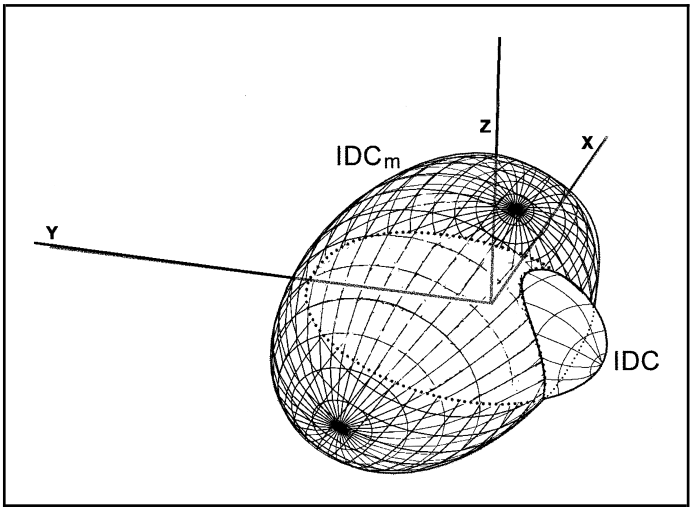

Fig. 6. Three-dimensional ellipsoids based on a multivariate model for principal component (PC) scores of DNA spectra of invasive ductal carcinoma (IDC) and metastasized IDC (those that give rise to disseminated metastases; IDC_m). The ellipsoids contain an expected 90% of the population of each group (e.g., IDC samples). (*See* text for details.)

nated metastases). The FT-IR/statistics models based on simple linear logistic regression, such as those shown in Fig. 5, B, did not effectively differentiate these groups. By use of models based on multivariate normal distributions of the first three PCs, a three-dimensional projection (Fig. 6) was constructed to contain a designated percentage (i.e., 90%) of the population of a group. In a model with 90% probability, such as that shown in Fig. 6, a randomly selected IDC_m spectrum would likely fall inside the appropriate three-dimensional figure (i.e., only an expected 10% of DNA spectra in the population of IDC_m spectra would fall outside the model). By use of this DNA model, normal breast tissue (RMT), primary breast tumors (IDC), and metastasizing primary breast tumors (IDC_m) were correctly classified as follows: 89% (16 of 18), 97% (31 of 32), and 82% (18 of 22), respectively. The discrimination between the IDC and the IDC_m is a potentially important basis for identifying metastasis in primary tumors, prior to evidence for malignant cells at distant sites. The prototype model (Fig. 6), which is presently based on a limited number of samples, can be applied to other systems having larger databases.

The FT-IR/statistics models have the ability to identify subtle changes in DNA in relation to the progression of normal tissues to diseased states. We are unaware of other techniques with the power to accomplish such a high degree of discrimination between DNA of natural systems. It is now possible to analyze less than 1.0 μg of DNA with the use of FT-IR spectral techniques recently developed in our laboratory. This will eventually allow the FT-IR/statistics technology to be applied to less than 1.0 mg of tissue, thus broadening the application to small biologic samples (e.g., fine-needle biopsy tissues). Future uses of the technology would be expected to encompass diverse areas of cancer research and clinical practice, as previously described *(4)*. For example, with the use of the FT-IR/statistics technology, the potential exists for detecting early cancer-related changes at the level of DNA, rather than at the cellular level, thereby affording a distinct advantage in patient treatment.

References

(1) Malins DC, Polissar NL, Nishikida K, Holmes EH, Gardner HS, Gunselman SJ. The etiology and prediction of breast cancer: Fourier transform-infrared spectroscopy reveals progressive alterations in breast DNA leading to a cancer-like phenotype in a high proportion of normal women. Cancer 1995;75:503–17.

(2) Malins DC, Polissar NL, Gunselman SJ. Tumor progression to the metastatic state involves structural modifications in DNA markedly different from those associated with primary tumor formation. Proc Natl Acad Sci U S A 1996;93:14047–52.

(3) Malins DC, Polissar NL, Gunselman SJ. Progression of human breast cancers to the metastatic state is linked to hydroxyl radical-induced DNA damage. Proc Natl Acad Sci U S A 1996;93:2557–63.

(4) Malins DC, Polissar NL, Su Y, Gardner HS, Gunselman SJ. A new structural analysis of DNA using statistical models of infrared spectra. Nat Med 1997;3:927–30.

(5) Malins DC, Polissar NL, Gunselman SJ. Models of DNA structure achieve almost perfect discrimination between normal prostate, benign prostatic hyperplasia (BPH), and adenocarcinoma and have a high potential for predicting BPH and prostate cancer. Proc Natl Acad Sci U S A 1997;94:259–64.

(6) Malins DC, Holmes EH, Polissar NL, Gunselman SJ. The etiology of breast cancer. Characteristic alterations in hydroxyl radical-induced DNA base lesions during oncogenesis with potential for evaluating incidence risk. Cancer 1993;71:3036–43.

(7) Feigelson HS, Henderson BE. Estrogens and breast cancer. Carcinogenesis 1996;17:2279–84.

(8) Khoury MJ, Beaty TH. Applications of the case–control method in genetic epidemiology. Epidemiol Rev 1994;16:134–50.

(9) Hwang SJ, Beaty T, Liang KY, Coresh J, Khoury MJ. Minimum sample size estimation to detect gene–environment interaction in case–control designs. Am J Epidemiol 1994;140:1029–37.

(10) Goldstein AM, Falk RT, Korczak JF, Lubin JH. Detecting gene–environment interactions using a case–control design. Genet Epidemiol 1997;14:1085–9.

(11) Foppa I, Spiegelman D. Power and sample size calculations for case–control studies of gene–environment interactions with a polytomous exposure variable. Am J Epidemiol 1997;146:596–604.

(12) Garcia-Closas M, Lubin JH. Power and sample size calculations in case–control studies of gene–environment interactions: comments on different approaches. Am J Epidemiol 1999;149:689–92.

(13) Rothman N, Garcia-Closas M, Stewart WT, Lubin JH. The impact of misclassification in case–control studies of gene–environment interactions. In: Vineis P, Malats N, Lang M, d'Errico N, Caporaso N, Cuzick J, et al., editors. Metabolic polymorphisms and susceptibility to cancer. Lyon (France): IARC; 1999. p. 89–96.

(14) Garcia-Closas M, Rothman N, Lubin J. Misclassification in case–control studies of gene–environment interactions: assessment of bias and sample size. Cancer Epidemiol Biomarkers Prev 1999;8:1043–50.

(15) Garcia-Closas M, Thompson WD, Robins JM. Differential misclassification and the assessment of gene–environment interactions in case–control studies. Am J Epidemiol 1998;147:426–33.

(16) Khoury MJ, Flanders WD. Nontraditional epidemiologic approaches in the analysis of gene–environment interaction: case–control studies with no controls! Am J Epidemiol 1996;144:207–13.

(17) Whittemore AS, Nelson LM. Study design considerations in genetic epidemiology: theoretical and practical considerations. J Natl Cancer Inst Monogr 1999;26:61–9.

(18) Yen SS, Jaffe RB, editors. Reproductive endocrinology: physiology, pathophysiology, and clinical management. Philadelphia (PA): WB Saunders; 1986.

(19) Lenton EA, Landgren BM, Sexton L, Harper R. Normal variation in the length of the follicular phase of the menstrual cycle: effect of chronological age. Br J Obstet Gynaecol 1984;91:681–4.

(20) Hankinson SE, London SJ, Chute CG, Barbieri RL, Jones LA, Kaplan LA, et al. Effect of transport conditions on the stability of biochemical markers in blood. Clin Chem 1989;35:2313–6.

(21) Hankinson SE, Willett WC, Colditz GA, Hunter DJ, Michaud DS, Deroo B, et al. Circulating concentrations of insulin-like growth factor-I and risk of breast cancer. Lancet 1998;351:1393–6.

(22) Su SC, Garbers S, Rieper TD, Toniolo P. Temperature variations in upright mechanical freezers. Cancer Epidemiol Biomarkers Prev 1996;5:139–40.

(23) Cauley JA, Gutai JP, Kuller LH, Dai WS. Usefulness of sex steroid hormone levels in predicting coronary artery disease in men. Am J Cardiol 1987;60:771–7.

(24) Koenig KL, Toniolo P, Bruning PF, Bonfrer JM, Shore RE, Pasternack BD. Reliability of serum prolactin measurements in women. Cancer Epidemiol Biomarkers Prev 1993;2:411–4.

(25) Toniolo P, Koenig KL, Pasternack BS, Banerjee S, Rosenberg C, Shore RE, et al. Reliability of measurements of total, protein-bound, and unbound estradiol in serum. Cancer Epidemiol Biomarkers Prev 1994;3:47–50.

(26) Bolelli G, Muti P, Micheli A, Sciajno R, Franceschetti F, Krogh V, et al. Validity for epidemiological studies of long-term cryoconservation of steroid and protein hormones in serum and plasma. Cancer Epidemiol Biomarkers Prev 1995;4:509–13.

(27) Thomas HV, Key TJ, Allen DS, Moore JW, Dowsett M, Fentiman IS, et al. A prospective study of endogenous serum hormone concentrations and breast cancer risk in premenopausal women on the island of Guernsey. Br J Cancer 1997;75:1075–9.

(28) Orentreich N, Brind JL, Rizer RL, Vogelman JH. Age changes and sex differences in serum dehydroepiandrosterone sulfate concentrations throughout adulthood. J Clin Endocrinol Metab 1984;59:551–5.

(29) Siiteri PK, Simberg N, Murai J. Estrogens and breast cancer. Ann N Y Acad Sci 1986;464:100–5.

(30) Langley MS, Hammond GL, Bardsley A, Sellwood RA, Anderson DC. Serum steroid binding proteins and the bioavailability of estradiol in relation to breast diseases. J Natl Cancer Inst 1985;75:823–9.

(31) Cauley JA, Gutai JP, Kuller LH, Powell JG. Reliability and interrelations

among serum sex hormones in postmenopausal women. Am J Epidemiol 1991;133:50–7.
(32) Hankinson SE, Manson JE, Spiegelman D, Willett WC, Longcope C, Speizer FE. Reproducibility of plasma hormone levels in postmenopausal women over a 2- to 3-year period. Cancer Epidemiol Biomarkers Prev 1995;4:649–54.
(33) Goodman-Gruen D, Barrett-Connor E. Epidemiology of insulin-like growth factor-I in elderly men and women. The Rancho Bernardo Study. Am J Epidemiol 1997;145:180–6.
(34) Micheli A, Muti P, Pisani P, Secreto G, Recchione C, Totis A, et al. Repeated serum and urinary androgen measurements in premenopausal and postmenopausal women. J Clin Epidemiol 1991;44:1055–61.
(35) Muti P, Trevisan M, Micheli A, Krogh V, Bolelli G, Sciajno R, et al. Reliability of serum hormones in premenopausal and postmenopausal women over a one-year period. Cancer Epidemiol Biomarkers Prev 1996; 5:917–22
(36) Pasagian-Macaulay A, Meilahn EN, Bradlow HL, Sepkovic DW, Buhari AM, Simkin-Silverman L, et al. Urinary markers of estrogen metabolism 2- and 16α-hydroxylation in premenopausal women. Steroids 1996;61:461–7.
(37) Michaud DS, Manson JE, Spiegelman D, Barbieri RL, Sepkovic DW, Bradlow HL, et al. Reproducibility of plasma and urinary sex hormone levels in premenopausal women over a one-year period. Cancer Epidemiol Biomarkers Prev 1999;8:1059–64.
(38) Willett WC. Nutritional epidemiology. New York (NY): Oxford University Press; 1990.
(39) Rosner B, Spiegelman D, Willett WC. Correction of logistic regression relative risk estimates and confidence intervals for random within-person measurement error. Am J Epidemiol 1992;136:1400–13.
(40) Hankinson SE, Manson JE, London SJ, Willett WC, Speizer FE. Laboratory reproducibility of endogenous hormone levels in postmenopausal women. Cancer Epidemiol Biomarkers Prev 1994;3:51–6.
(41) Potischman N, Falk RT, Laiming VA, Siiteri PK, Hoover RN. Reproducibility of laboratory assays for steroid hormones and sex hormone-binding globulin. Cancer Res 1994;54:5363–7.
(42) Gail MH, Fears TR, Hoover RN, Chandler DW, Donaldson JL, Hyer MB, et al. Reproducibility studies and interlaboratory concordance for assays of serum hormone levels: estrone, estradiol, estrone sulfate and progesterone. Cancer Epidemiol Biomarkers Prev 1996;5:835–44.
(43) Ziegler RG, Rossi SC, Fears TR, Bradlow HL, Adlercreutz H, Sepkovic D, et al. Quantifying estrogen metabolism: an evaluation of the reproducibility and validity of enzyme immunoassays for 2-hydroxyestrone and 16α-hydroxyestrone in urine. Environ Health Perspect 1997;105(suppl 3): 607–14.
(44) Hankinson SE, Willett WC, Manson JE, Colditz GA, Hunter DJ, Spiegelman D, et al. Plasma sex steroid hormone levels and risk of breast cancer in postmenopausal women. J Natl Cancer Inst 1998;90:1292–9.
(45) Dorgan JF, Longcope C, Stephenson HE Jr, Falk RT, Miller R, Franz C, et al. Relation of prediagnostic serum estrogen and androgen levels to breast cancer risk. Cancer Epidemiol Biomarkers Prev 1996;5:533–9.
(46) Toniolo PG, Levitz M, Zeleniuch-Jacquotte A, Banerjee S, Koenig KL, Shore RE, et al. A prospective study of endogenous estrogens and breast cancer in postmenopausal women. J Natl Cancer Inst 1995;87:190–7.
(47) Von Sonntag C, Hagen U, Schon-Bopp A, Schulte-Frohlinde D. Radiation-induced strand breaks in DNA: chemical and enzymatic analysis of end groups and mechanistic aspects. Adv Radiat Biol 1981;9:109–42.
(48) Liehr JG. Hormone-associated cancer: mechanistic similarities between human breast cancer and estrogen-induced kidney carcinogenesis in hamsters. Environ Health Perspect 1997;105(suppl 3):565–9.
(49) Yager JG, Liehr JG. Molecular mechanisms of estrogen carcinogenesis. Annu Rev Pharmacol Toxicol 1996;36:203–32.
(50) Liehr JG. Dual role of oestrogens as hormones and pro-carcinogens: tumour initiation by metabolic activation of oestrogens. Eur J Cancer Prev 1997;6:3–10.
(51) Weitzman SA, Turk PW, Milkowski DH, Kozlowski K. Free radical adducts induce alterations in DNA cytosine methylation. Proc Natl Acad Sci U S A 1994;91:1261–4.
(52) Cavalieri EL, Stack DE, Devanesan PD, Todorovic R, Dwivedy I, Higginbotham S, et al. Molecular origin of cancer: catechol estrogen-3,4-quinones as endogenous tumor initiators. Proc Natl Acad Sci U S A 1997;94: 10937–42.
(53) Frenkel K, Wei L, Wei H. 7,12-dimethylbenz[*a*]anthracene induces oxidative DNA modification in vivo. Free Radic Biol Med 1995;19:373–80.
(54) Malins DC, Polissar NL, Schaefer S, Su Y, Vinson M. A unified theory of carcinogenesis based on order–disorder transitions in DNA structure as studied in the human ovary and breast. Proc Natl Acad Sci U S A 1998; 95:7637–42.
(55) Henderson CI. Risk factors for breast cancer development. Cancer 1993; 71:2129s-40s.

Notes

Supported by Public Health Service grants CA49449 and CA67262 from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services (S. E. Hankinson); and by U.S. Army Medical Research and Materiel Command Contract DAMD17-95-1-5062 (D. C. Malins).

S. E. Hankinson thanks Stacey Missmer for her input into this manuscript and Sandra Melanson for excellent technical assistance. D. C. Malins thanks the National Cancer Institute Cooperative Human Tissue Network for tissues and pathology data.