

Review

Design Options for Molecular Epidemiology Research within Cohort Studies

Andrew G. Rundle,¹ Paolo Vineis,^{2,3} and Habibul Ahsan¹

¹Department of Epidemiology, Mailman School of Public Health and Herbert Irving Comprehensive Cancer Center, Columbia University, New York, New York; ²Imperial College London, London, United Kingdom; and ³University of Torino, Turin, Italy

Abstract

Past discussions of the relative strengths of nested case-control and case-cohort designs have not fully considered cohorts with stored biological samples in which biomarker analyses are planned. Issues related to biomarker analyses can affect an investigator's choice of design and the conduct of these two designs. The key issues identified are effects of analytic batch, long-term storage, and freeze-thaw cycles on biomarkers. In comparison with the nested case-control design, the case-cohort design is less able to handle these challenges. Problems arise because most implementations of the case-cohort design do not allow for simultaneous evaluation of biomarkers in cases and reference group members, and there is no matching. By design, the nested case-control study controls for storage duration and the batching of biological samples from cases

and controls is logistically simple. The allowance for matching also means that subjects can be matched on the number of freeze-thaw cycles experienced by the biological sample. However, the matching generates complex data sets that can be more difficult to analyze, and the costly biomarker data generated from the controls has few uses outside of testing the specific hypotheses of the study. In addition, because the same subject can serve as a control and a case, or multiple times as a control, biomarker analyses and sample batching can be more complex than initially anticipated. However, in total, of the two designs, the nested case-control study is better suited for studying biomarkers that can be influenced by analytic batch, long-term storage, and freeze-thaw cycles. (Cancer Epidemiol Biomarkers Prev 2005;14(8):1899–907)

Introduction

Although population- and hospital-based case-control studies (1-3) are commonly used for molecular epidemiologic studies, a growing number of large cohort studies has been established over the past decade in which biological samples have been collected and stored for future molecular epidemiology studies (4, 5). By avoiding the problems of questionable temporal relationships between biomarkers and disease risk that hamper traditional case-control studies, cohort studies can examine the relationships in a methodologically superior manner. The stored biological samples collected from the cohort members at baseline are precious and represent a finite resource that must be used as efficiently as possible. Two related, yet methodologically distinct study designs—the case-cohort and the nested case-control studies—offer logistic efficiency over full cohort analyses and are typically used for molecular epidemiologic studies within prospective cohort studies. Prior literature has described the differences between these two designs and the trade-offs in selecting one design over another (6-11). However, the use of biomarkers adds another level of complexity in making a choice between the two designs and in the conduct of the studies. These issues

have been briefly discussed in a previous publication (12). In this paper, we describe in depth how biomarker analyses alter the calculus of deciding which design to use and some of the ways in which the study conduct is altered.

In the case-cohort design, which was first proposed as the “case-base” design by Miettinen (13) and subsequently refined by Prentice (14), a random subset (known as “subcohort”) of the total baseline cohort is selected as a “control” or reference group. In some instances, the subcohort is chosen as a stratified random sample, with stratification on important confounding factors (15, 16). To estimate effects, the exposure distribution among this reference group is compared with that in a case group of interest arising from this cohort. Apart from random errors, the distributions of exposure (i.e., biomarker) or any other factors in this random subset should represent that of the total cohort at the beginning of the observation period. Because the ratio of exposed to unexposed individuals (with respect to the biomarker of interest) in the reference group is same as in the total cohort at baseline, the estimated odds ratio using a case-cohort study represents the risk ratio for the same biomarker-disease relationship had a full cohort analysis involving all cohort members been done (17-19). Alternatively, the follow-up data for the subcohort (including those who develop disease) may be treated as representative of the person-time experience of the total cohort, and case-cohort studies can be analyzed with appropriate weighting to estimate the rate ratio for the biomarker-disease relationship (9, 14). If a baseline cohort member later develops the disease of interest during the observation period, this member is also treated as a case and thus also contributes to the exposure distribution among the cases (9, 14). The major advantage of this design is that the same reference group, who are a random subset of the baseline cohort, can be used for multiple different case groups arising from the cohort (6). This advantage relies

Received 11/21/04; revised 5/17/05; accepted 5/26/05.

Grant support: NIH grants KO7-CA92348-01A1, 1R01CA102484, and 1P42ES10349; European Community (5th Framework Programme) grant QLK4-CT-1999-00927 (P. Vineis); and a grant of the Compagnia di San Paolo to the Institute for Scientific Interchange Foundation. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Requests for reprints: Andrew Rundle, Department of Epidemiology, Mailman School of Public Health and Herbert Irving Comprehensive Cancer Center, Columbia University, 722 West 168th, Room 730, New York, NY 10032. Phone: 212-305-7619; Fax: 212-305-9413. E-mail: agr3@columbia.edu

Copyright © 2005 American Association for Cancer Research.

doi:10.1158/1055-9965.EPI-04-0860

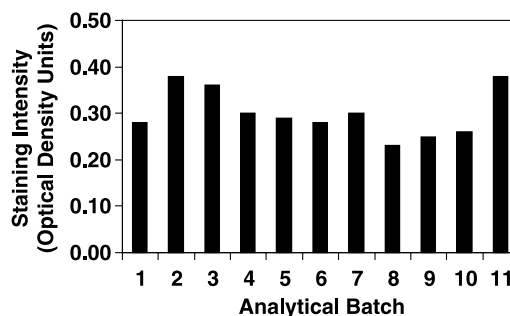
on the supposition that the quality of exposure measurement in the subcohort is the same as for the case series. This is likely to be true for stored questionnaire data (20), the modeling of occupational exposures based on industrial hygiene records and job histories (21), or the modeling of ambient exposures using data-based residential histories and air monitoring records (22). As will be discussed, this assumption may not hold for biomarker analyses. A major limitation of this design is that subcohort members cannot be matched to cases on any potential confounders because subcohort members are selected randomly from the total cohort.

In the nested case-control design, controls are selected concurrently as cases develop (i.e., they are selected from the cohort members who are still disease-free at the time a case occurs). This sampling procedure is called incidence density sampling. By matching on the time at risk (i.e., duration between enrollment/biospecimen collection and disease onset), the ratio of exposed to unexposed number of controls represents, on average, the ratio of exposed to unexposed person-time for the total cohort. This allows the odds ratio derived from the nested case-control study to estimate the incidence rate ratio for the biomarker-disease relationship had a full cohort analysis involving all cohort members been done. A cohort member selected as a control at a given time because he is at risk for disease at the time a case occurs may be again selected as a control for a subsequent case as long as he remains unaffected by the disease of interest. The major advantage of the nested case-control design is its ability to match controls with cases on follow-up duration between biological sample collection and disease development. In addition to matching on follow-up duration, it is common for controls to be matched to cases on other characteristics, such as age and gender. However, additional layers of matching make the control series less representative of the person-time experience of the total cohort. In fact, the primary complaint with nested case-control studies is that the intricate matching used in the design means that the control series is highly selected and cannot be used for other studies and do not provide representative data on the entire cohort (11).

The Use of Biomarkers Brings Additional Layers of Complexity to Nested Case-Control and Case-Cohort Studies

As noted above, there are several papers in the literature that describe the relative strengths and weaknesses of the nested case-control and case-cohort designs and provide guidance in selecting a study design. These papers have discussed both practical or logistic concerns and comparisons of statistical efficiency and power (6, 7, 10). However, the use of biomarkers adds additional complexities to comparing these two designs (12). The pertinent issues that must be considered with the addition of biomarkers to these study designs are analytic batch effects, storage effects, and freeze-thaw cycles. Below, we describe these issues in general terms without regard to a particular study design. In later sections of this paper, we describe how these issues impact the case-cohort and nested case-controls designs.

Batch Effects. Analytic batches can be defined as a group of biological samples that are analyzed together under a particular set of conditions. In studies in which samples are analyzed in multiple batches, the laboratory strives to duplicate the relevant conditions across batches. The goal is to assure that analytic results derived from any given sample are not influenced by the particular analytic batch in which the sample was analyzed. However, our experience and published articles have shown that this goal is not often met for many laboratory assays (23-27). Figure 1 shows the mean polycyclic aromatic hydrocarbon-DNA adduct in prostate tissue samples



1 P<0.001 by ANOVA

Figure 1. Polycyclic aromatic hydrocarbon-DNA adduct by immunohistochemistry in prostate tissue by batch ($P < 0.001$, ANOVA).

by batch measured by immunohistochemistry (28). The prostate samples were assigned to batches using a random number generator. The difference between the mean of the highest and lowest batch is 0.15 absorbance units or 39% of the highest batch. Each batch contained two laboratory control slides cut from the same tissue block. Thus, it is possible to use the control slides to partially normalize the data shown here (28). However, because of random error in the assay within batch, there is variability in the control slides and normalization is not completely efficient; therefore, batch effects remain. Even commercial ELISA kits from the same manufacturer have shown substantial variation among different lots of kits (24).

Batch effects can create random noise or bias (29, 30). The effect of batch is expected to have the same consequences as misclassification of exposure in epidemiologic studies. Such misclassification usually implies an underestimation of the association between analyte and disease if the measurement error is evenly distributed among cases and controls. However, when the errors are unevenly distributed, bias arises. Hence, it is important to have the same proportion of cases and controls in different batches to avoid such bias. Batch effects can occur with and without time effects. In the example from the work on prostate tissue samples, the effect of batch does not seem to be related to time; all of the samples were analyzed within a short period of time and there does not seem to be a strong trend across batches. However, there can also be time-related effects where laboratory results seem to drift with time, either producing higher or lower results over time (30).

Batches are created through a number of constraints that prevent all of the samples from being analyzed together. These constraints may be technological, such as the number of wells on an ELISA or PCR plate or the number of samples that can be loaded into robotic device. The constraints might be administrative and driven by the need to complete particular milestones for funding agencies or for publications or by the availability of reagents. They might also be logistic and determined by the rate of subject recruitment and transport and shipping limitations. Finally, batches might be defined by the limits of endurance of the laboratory technicians or other staffing issues.

It should also be noted that there are often batches and meta-batches (i.e., batches nested within batches). For instance, a coordinating center might send a laboratory 200 samples in one batch. However, the assay technology may only be able to analyze 40 samples per run, and so the shipped batch is analyzed in five batches over the course of 5 weeks. At some later date, perhaps ≥ 6 months later, a second meta-batch of samples would be shipped from the coordinating center to the laboratory, and again these samples would be analyzed in multiple laboratory runs. This is essentially the case with the prostate tissue samples shown in Fig. 1. These 11 batches of samples represent the first meta-batch sent to

the laboratory from the coordinating center. After several more months of recruitment, a second large meta-batch of samples will be sent to the laboratory where they will be split into smaller analytic batches.

Storage Effects. The second issue is the concern that for some biomarkers the levels of the biomarker may be influenced by the duration of sample storage, or that storage duration will reduce the accuracy of the assay. This is a concern often discussed during study planning and in reviewers' comments to grant applications but the published data demonstrating storage effects on biomarker level is sparse. Antioxidant micronutrients in plasma (31, 32), and cotinine and creatinine in urine, have been reported to be stable in samples stored for many years (33). Benzo(a)pyrene-DNA adducts have been reported to be stable in samples stored for 10 months (34). However, levels of other markers, such as lipoprotein A, total and high-density lipoprotein cholesterol, triglycerides (35-37), HIV p24 antigen (38), free prostate-specific antigen (39), progesterone (40), estradiol (41), hepatitis virus C RNA concentrations (42), and salivary IgA (43), have been shown to change during sample storage. Additionally, for paraffin-embedded tissue sections, the reliability of immunohistochemical and fluorescence *in situ* hybridization assays for HER2 decline with storage time (44), as does antigenicity for tissue microarrays (45). Typically, investigators worry about biomarker levels decaying, with measured levels of biomarkers declining over time; however, there are examples of measured biomarker levels increasing over time (40). Storage-induced changes in biomarker levels can be substantial. Evans et al. (35) have reported a 19% decrease in lipoprotein A in serum samples stored for 3 years at -70°C . Additionally, Bolelli et al. (40) reported a 30% increase in serum-free testosterone and a 40% decrease in progesterone in samples stored at -80°C for 3 years. In part, the sparse data on this issue may reflect the logistic difficulties in conducting a study to test whether long-term storage impacts biomarker levels.

Another concern is that for mechanical freezers, the internal temperature can vary by location and time within a freezer and temperature can vary across freezers (46, 47). After long-term storage in such freezers, it may be appropriate to match samples on storage location in the freezer (47). Storing samples in the liquid phase of liquid nitrogen, as is being done in the European Prospective Investigation into Cancer and Nutrition (EPIC), probably provides the most consistent storage conditions. However, in large cohort studies such as EPIC, samples are stored centrally and then shipped to the investigator's laboratories for analyses, where they may be stored in mechanical freezers for substantial time periods. It may even be advisable to physically arrange the samples in these mechanical freezers in a configuration appropriate for the study design, with matched sets of cases and controls stored in close proximity for nested case-control studies or in a random order for case-cohort studies.

The impact of storage effects on study validity depends on whether it reduces assay precision or alters the level of the biomarker. If assay precision is altered by storage effects, random error may be increased for biomarkers stored longer. This will reduce statistical power and, depending on how the study is implemented, may cause bias. If storage duration alters the biomarker level, the effect on study validity depends on whether biomarker levels change in a half-life-like fashion at an absolute linear rate or a relative rate. Depending on how the study is implemented and the mathematical function describing how biomarker levels change, bias may be away or toward the null.

It should also be understood that for large multicenter cohorts, it may take many years to assemble the entire cohort and thus storage time may be affected by logistic considerations. If recruitment begins at different centers at different times, storage duration becomes associated with site and the

demographic and exposure characteristics of the subjects recruited at that site. In EPIC, the Oxford site was one of the first to begin recruiting and subjects recruited at this site tend to have a higher socioeconomic status. A large proportion of the subjects were vegetarians. Thus, sample storage duration is not randomly distributed across lifestyle risk factors. A related issue is that sample handling and processing may vary by site and may influence biomarker levels. Thus, whereas there is less empirical data demonstrating that duration of storage may alter biomarker analyses, it is an issue of concern.

Freeze-Thaw Cycles. It is typical for whole-blood samples drawn from study subjects to be separated into constituent parts, such as WBC, serum, and RBC, and then frozen in multiple small aliquots. However, many laboratory assays require sample volumes that are smaller than the stored aliquots, so during a particular study a sample aliquot may be thawed, sampled from, and then the remaining biological sample refrozen for future studies. However, it is possible that thawing and refreezing may alter components of the biological sample. Thus, when the sample is thawed a second time to be assayed for a new biomarker study, the level of the biomarker may differ from the level that would have been measured if the sample was being thawed for the first time. There are several physical and chemical mechanisms through which the freezing process may alter the chemical properties of a biological sample (48). During freezing, there is a concentration of solutes in the residual liquid phase of the sample and ionic strength increases that may cause protein precipitation and denaturation (48, 49). Furthermore, as ice crystals precipitate out of solution, the pH of the remaining liquid phase can change by several pH points, which may cause lipid degradation and protein denaturation (48-50).

It has been shown that levels of serum or plasma cholesterol (51), cotinine (52), micronutrients (51), reproductive hormones (23, 51, 53), insulin-like growth factors (54), saliva progesterone (55), C-reactive protein (56) and antibodies to some infectious diseases (57), and PCR-based detection of hepatitis C (58) are stable despite several cycles of freezing and thawing. However, cycles of freezing and thawing have been shown to impact measures of lipoprotein A (59), anticardiolipin antibodies (48), endogenous antioxidants (superoxide dismutase, reduced glutathione; ref. 60), hepatitis C viral RNA quantification (61), saliva cortisol (55), epidermal growth factor receptor (62), and tissue inhibitor of metalloproteinase-1 (63); freeze-thaw cycles have also been shown to degrade DNA (64, 65). The impact of freeze-thaw cycles on biomarker levels can be substantial. Sgoutas et al. (59) reported a 27% decrease in lipoprotein A levels as measured by ELISA after two freeze-thaw cycles to -20°C , and a 23% decrease after four cycles to -70°C . Belleste et al. (65) reported that after three freeze-thaw cycles, the content of DNA stored in sterile distilled water was reduced by more than half. In these experiments, the copy number of the target DNA sequence and the volume of distilled water influenced DNA decay (65).

Thus, biomarker levels and study results could be influenced by the sequence in which hypotheses are tested in a given bank of samples. Thus, for a given aliquot of biological material in which an investigator plans to measure several biomarkers, the sequence of assays and required freeze-thaws must be carefully planned. For instance, in a saliva sample, cortisol should be analyzed within the first or second round of freeze-thaws, whereas progesterone can be analyzed in a later round (55).

Design Considerations: Batch Effects, Sample Decay, and Freeze-Thaw Cycles

For biomarkers that do not suffer from batch effects, do not change during storage, and are not influenced by freeze-thaw

cycles, the calculus in deciding whether to use a nested case-control study or a case-cohort study is the same as for studies using other stable measures, such as questionnaire data or exposure-job history matrices. For example, with vigilant laboratory quality control, genotyping assays are thought to have little batch effect and the DNA is thought to be stable (44, 66). For biomarkers that suffer from decay and/or batch effects, the calculus in choosing a study design begins to change.

In the following text, we will describe the issues for case-cohort studies and nested case-control studies. In this discussion, we refer to grant funding periods, which reflect the U.S. and European scientists' experience of the work of conducting a nested case-control or case-cohort study (case identification, control or subcohort selection, and laboratory analyses) being done within a maximum of 3- to 5-year period set by funding agencies. Due to differences in the way research funds are distributed internationally, scientists in other countries may not experience such sharp temporal delineations. However, a key issue in thinking about the conduct of nested case-control or case-cohort studies is how the work required for the study relates temporarily to the life course of the underlying cohort.

Case-Cohort Studies

The Initial Case-Cohort Study. For the initial case-cohort study being launched within a cohort, there are two scenarios that need to be considered. The first scenario is if an investigator were to launch a new cohort with an *a priori* plan for conducting case-cohort analyses at a future date. This might be termed a prospective case-cohort study. The second is if a cohort has been in existence for some time and, after a period of follow-up, a case-cohort study is conducted within the cohort. This might be termed a retrospective case-cohort study. In the prospective scenario, the subcohort is selected as soon as the cohort is fully assembled and the cases are identified later during the course of cohort follow-up. In this prospective scenario, there are incentives to begin laboratory analyses of samples from subcohort members as soon as the subcohort is selected. First, this allows the bulk of the laboratory work to be completed quickly. The ability to quickly begin laboratory work is similar to the logistic efficiency advantage that has been previously cited for case-cohort studies when extensive field work is required to gather data from study subjects (6). Furthermore, early completion of the bulk of the laboratory work allows for cross-sectional studies of correlates of biomarkers to begin right away. Whereas it should not drive design decisions, one of the advantages of the case-cohort design is that, because the subcohort is a random sample of the cohort, cross-sectional analyses of biomarkers in the subcohort provide valid information on the overall cohort. In addition, prevalent cases identified at baseline as the cohort is assembled can be compared with the subcohort.

For example, in the Health Effects of Arsenic Longitudinal Study—a prospective cohort study of arsenic exposure and premalignant, malignant, and nonmalignant outcomes—researchers identified substantial numbers of premalignant skin lesions during the baseline assessment of the cohort (67). Incident cases of skin lesions and other outcomes are being ascertained during the follow-up of this cohort for prospective case-cohort analyses of the effects of arsenic, genotype, and other biomarkers on these lesions. In addition, the researchers exploited the baseline data to investigate the same associations in a cross-sectional fashion by comparing individuals with prevalent skin lesion ascertained during cohort assembly with the subcohort group. Some of the biomarker data from the subcohort samples generated as part of the cross-sectional studies, e.g., urinary arsenic species and genetic polymorphisms, will also be used in future case-cohort analyses involving incident cases in this cohort.

In such a prospective case-cohort study, as incident cases accrue during follow-up, the investigators might begin analyzing biological samples from cases. However, if subcohort samples were analyzed at the beginning of the grant funding period, this approach would very closely align case-cohort status to duration of sample storage and analytic batch, setting the stage for bias. Biological samples from cases that arise from the subjects selected to be part of the subcohort will already have been analyzed, and thus there will be no difference in the circumstances of analyses for these cases and the subcohort. In general, however, the majority of cases are expected to arise from the vast bulk of study subjects not selected as part of the subcohort. Because samples from these cases were not analyzed with the subcohort, the circumstances of analyses for these samples will likely differ from those of the subcohort.

If batch effect is the main concern and storage time is unlikely to affect the biomarker, then a valid approach is to wait until the end of follow-up to begin laboratory analyses when all cases of interest are identified. Then, samples could be randomly assigned to analytic batches with proportional numbers of case and subcohort samples in each batch. If multiple case types are of interest, then batch assignment can also take multiple case types into consideration. On the other hand, if both batch effect and storage time are of concern, then an alternative approach would be to analyze subcohort and case samples as cases accrue by creating balanced batches with proportional numbers of case and subcohort samples in each batch. In this approach, the laboratory work is spread out over the duration of the grant period, with the pace of laboratory work dictated by case accrual. This approach requires good estimates of the likely incidence rate for the cases so that the number of subcohort members to be assigned to each batch can be estimated. Careful estimation of the number of subcohort samples to be assayed in each batch over the time course of follow-up is most critical if the batch effect includes a time component. With this approach, case and subcohort samples analyzed together would not be matched on other characteristics; subcohort members would still be randomly chosen. Theoretically, with very careful planning, it is possible to include multiple case groups in analytic batches in this approach if the various case series are ascertained concurrently. However, it may not be feasible if different types of cases have different induction periods, causing largely nonoverlapping occurrence of different case types during the follow-up period (see below). However, because differences in sample storage duration may still exist between cases and subcohort members, it may be appropriate to control for sample storage duration in the statistical analyses whenever effects of storage duration are of concern.

The retrospective scenario is more typical in which a case-cohort study would be implemented in an existing cohort that has been followed-up for a number of years and accrued enough cases for analyses. In this scenario, an investigator would retrospectively select a subcohort from the baseline members of the cohort and compare them to the cases. Samples from the subcohort and the case series would be analyzed contemporaneously, and samples should be randomized to batches to avoid bias due to batch effects. This is similar to first option in the prospective scenario when storage time is not of concern. If the samples vary in the amount of time they have been frozen and sample decay is a concern, then control for storage time in the statistical analyses is the only option, as the opportunity for simultaneous batch assignment as cases accrue is no longer available. If storage time varies by case-subcohort status, statistical control for storage time should remove any bias due to storage-related changes in the biological sample. Additionally, in the more complex situation where storage time is associated with potential risk factors for disease, potential bias can be removed by statistical control for storage time.

For the first case-cohort study launched in a cohort, whether implemented prospectively or retrospectively, it does not seem that the effects of freeze-thaw cycles will inherently bias the analyses. Unless previous case-only studies required case samples or previous cross-sectional studies using subcohort and prevalent case samples required subcohort members to undergo freeze-thaws, it is unlikely that the prevalence of samples that have been thawed will differ between cases and the subcohort members. Even if previous biomarker studies had been conducted within a cohort, it is likely that there will be multiple aliquots of each sample fraction available for analyses, so if one has previously been thawed, another could be used. Should the extent of freeze-thaw cycles differ between the case series and the subcohort, then statistical control for the freeze-thaw cycles could be implemented.

Issues Related to Analyses of the Subsequent Case Series and the Original Subcohort. One of the often-cited strengths of case-cohort studies is that the subcohort can be used as a reference group for a variety of different case groups (6). One could imagine a case-cohort study implemented early in cohort follow-up for a common cancer and then after much further follow-up the subcohort would be used as a reference group for a series of rarer or long-latency cancers. Here, tremendous logistic efficiency is gained because data from the reference group already exists in study databases. However, in studies using biomarkers susceptible to batch or storage effects, this advantage may essentially be outweighed by the potential for bias.

It may be of interest to analyze the same biomarkers in subsequent case series that were previously analyzed in the subcohort and a prior case series. However, the samples from successive case series will be analyzed in new analytic batches, separate from the subcohort samples previously analyzed, essentially guaranteeing batch effects. If the batch effect is not related to time (i.e., batch effect represents a random effect), it is theoretically possible that, overall, the batches of case samples will reflect the same random noise as the subcohort and bias will be minimal. However, this possibility cannot be verified in the data, calling into question the results. If the batch effect includes a time-related component, as seen in laboratory drift, bias is guaranteed. Another possibility is that technology may have progressed in the time between analyses of the subcohort and analyses of the new case series. If samples from the new case series are to be analyzed by an assay that has been incrementally improved or with a completely new technology, assay results from the new case series may differ substantially from assay results previously generated for the subcohort. When batch effect or changing technology is the main concern and the effects of storage time and freeze-thawing are less of a concern, then one may wish to analyze a new aliquot of samples from the previously analyzed subcohort samples. Analyses can proceed, taking case-control status into consideration during batches or by randomizing subjects to batches. However, having to reanalyze samples from the subcohort nullifies the efficiency advantage of the case-cohort design.

If effects of storage time are of concern and the same biomarkers are analyzed in the new case series that were previously analyzed in the subcohort, the situation becomes less manageable and control for storage time becomes increasingly difficult. As the duration of time between the laboratory analyses conducted in the subcohort and the analyses conducted for the new case series grows, the difference in the length of time the samples were stored before being analyzed for case series and subcohort members increases. This makes statistical control for duration of storage more difficult for longer latency cancers due to sparse data at the tails of the distribution of storage times.

Alternatively, the hypotheses to be tested in subsequent case series may involve biomarkers that were not previously

analyzed in the subcohort. In this scenario, biological samples from the new case series and the previously identified subcohort could be analyzed for the new biomarker, and samples could be randomized to batches to combat bias due to batch effects. However, depending on the biological material to be analyzed, the sample aliquots from the subcohort may have already been thawed and refrozen in prior rounds of laboratory analyses. For instance, if a serum biomarker was analyzed in the original case-cohort analyses for one case type and a different serum biomarker is of interest in the subsequent case-cohort study for a different case type, serum aliquots from the subcohort may have already gone through a round of thawing and refreezing. The samples from the new case series are less likely to have been previously thawed because the samples were not of interest until the new case series was identified. Thus, when new biomarkers are to be analyzed in subsequent case series and the subcohort, storage conditions may impact study validity.

Issues in the Statistical Analysis of Biomarker Data. An advantage of the case-cohort design is that the time scale for the follow-up analyses is not fixed by the design of the study (6). In a nested case-control study, controls are typically matched to cases on duration of follow-up and/or age and thus all analyses must use this as the time scale. In a case-cohort study, an investigator may select the time scale best suited for the particular analyses (6). A common time scale for cohort and case-cohort analyses is duration of follow-up; however, age and year can both independently influence disease risk and age maybe the more appropriate time scale (6, 68).

The selection of the underlying time scale may alter which subcohort members are included in risk set with a particular case. If duration of follow-up is used as the time scale, then for a case occurring after T_X duration of follow-up only those subcohort members who contribute T_X duration or more of follow-up and are still at risk of disease will be included in the risk set with the case (9). Subcohort members who are the same age as the case but who have been not been followed for T_X duration, perhaps because they were enrolled into the cohort later during the enrollment period, will not be included in the risk set with this case. The situation is different if age is used as the time scale. For a case who develops disease at age 60 who has been followed for T_X duration, members of the subcohort who are age 60 or older and who are still at risk of disease are included in the risk set, whether or not they have been followed for T_X duration.

Depending on the context of the study and the biomarker in question, different time scales might be more or less appropriate. For a genetic polymorphism for which the effect is thought to occur over the duration of the life course, age might be the most appropriate time scale rather than duration of follow-up. For a biomarker that might be influenced by a strong generational cohort effect, such as blood lead levels (69), calendar year might be more appropriate. For a biomarker that reflects a new onset of exposure, such as a new occupational exposure, time since exposure onset might be most appropriate. In general, the case-cohort design can accommodate separate analyses of each of these biomarkers using different time scales for each analysis. In a nested case-control study, each of the biomarkers could be analyzed but they would all use the same time scale set by the matching criteria. However, in some implementations of the case-cohort design, the use of biomarkers may limit the investigator's flexibility in choosing time scales. Previously, it was suggested that in the prospective case-cohort design, subcohort members could be proportionately assigned to laboratory batches based on the incidence of case accrual. However, the suggestion of proportionate assignment of subcohort members assumes that duration of follow-up is

the appropriate underlying time scale. In some instances, it may be more appropriate to prospectively assign subcohort members to laboratory batches based on the expected age distribution of cases.

Another possible consideration in the statistical analyses is the appropriate weighting of cases that arise from outside of the subcohort (9). Approaches described by Prentice (14) and Barlow (70) use a weighting of 1 in the denominator of the pseudo-likelihood for cases that arise outside the subcohort, whereas an approach by Self and Prentice (71) uses a weighting of 0. Furthermore, the approach of Barlow weights the cases in the subcohort before failure and the subcohort controls by the inverse of the sampling fraction. At this time, the optimal weighting scheme for case-cohort analyses is unclear (9). However, the extent to which biomarker data from cases arising outside the subcohort have been differently affected by batch effects and storage conditions may impact the performance of the weighting schemes. Different weighting schemes may be more or less appropriate for different biomarker analysis scenarios.

Nested Case-Control Studies. In nested case-control studies, controls are matched to cases typically using a quite intricate criteria. At a minimum, controls are matched to cases on length of follow-up since entry into the cohort study. In a case-control study nested in the Physicians Health Study, controls were additionally matched to cases on age, smoking status (never, ex, current), cigarettes per day in current smokers (1-19, 20-39, and 40+ cigarettes per day), and analytic batch (4). In Genair/EPIC, a case-control study of noncurrent smokers nested in EPIC, controls were matched to cases on age, gender, country, smoking status (never, ex-smoker), and analytic batch (72). A result of this matching is that controls are not a representative sample of the cohort at baseline, and thus cannot be used as controls for future case groups as can be done with subcohort members in a case-cohort study. Studies of future case series in EPIC will not be able to use the Genair/EPIC control series and its biomarker data and will require a new reference group and new biomarker analyses. Major advantages of nested case-control design are that matching on length of follow-up, as is common, effectively matches on sample storage time. Also, individual matching allows for easy matching on analytic batch, removing batch effects. However, depending on how the study is conducted, there are several nuances that need to be considered.

In a nested case-control study, a control matched to a case at time A may later develop the disease and becomes a case at time B. In this scenario, the study subject is treated as both a control (at time A) and a case (at time B). For cancer outcomes, this might be expected to be a rare occurrence, but when one considers matching controls to cases on age, smoking status, and cigarettes smoked per day (73), the probability of a control later becoming a case is not trivial. Additionally, because of the amount of matching, there may be relatively few subjects available to be selected as controls within certain strata of matching variables. Thus, a subject may be selected as a control for a case at time A and again as a control at time B. For questionnaire data, the same data point, e.g., cigarettes smoked per day, is used for the subject at time A when she is a control and at time B when she is a case or if selected again as a control. However, things are more complex when biomarker data is used from a subject who contributes information more than once in a study.

Biomarker Analyses When a Subject Appears Multiple Times in a Nested Case-Control Study. The possibility of a subject occurring multiple times within a nested case-control study has implications for sample batching. If a nested case-control study is implemented at the end of cohort follow-up, batching can be arranged such that a single biological sample from an individual that occurs in the data set multiple times can be placed in one batch, along with all the subjects to which that individual is

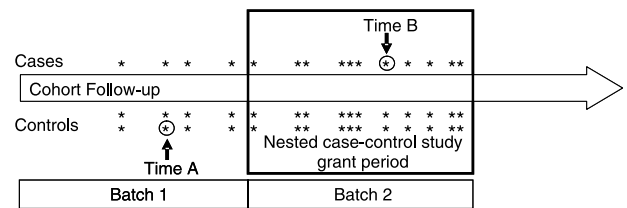


Figure 2. Laboratory batching in a nested case-control study in which a control later becomes a case. This figure illustrates a nested case-control study in which two controls are matched to each case. The circled study subject is a control at time A and becomes a case at time B. Samples from all of the cases and controls accrued before the grant period are analyzed in batch 1 at the beginning of the grant period. Cases and controls that accrue during the grant period are analyzed in batch 2 at the end of the grant period. A sample from the circled subject should be analyzed in both batch 1 and batch 2.

matched. Another scenario is when a nested case-control study is implemented such that samples from cases and controls accrued in the past are analyzed, whereas follow-up continues prospectively for another couple of years. Here, it is possible for a subject selected to be a control for a past case (time A) to become a case during the continued follow-up and be matched to a set of controls at time B. In this scenario, a biological sample from the subject may have already been analyzed previously when the subject acted as a control at time A (see Fig. 2). However, it would be improper for this data point from time A to also be used as the data point when the subject is a case at time B. This is because, if time A biomarker data, when the subject was a control, is used when the subject is a case at time B, biomarker data from the controls at time B will come from a different laboratory batch than the biomarker data from the case at time B. Furthermore, storage duration for cases and controls will differ at time B if the data point from time A is used for the case at time B. It would be most appropriate if a new aliquot of biological material from the case at time B were analyzed with the controls matched to the case at time B. Likewise, if a subject serves as a control for multiple cases, multiple samples from that control should be analyzed, one for each instance where the subject is represented in the data set. This is analogous to making sure the quality of exposure data is the same for cases and controls. Clearly, this issue is a larger problem for more common diseases and may not be impossible to address if the laboratory assay consumes large volumes of biological sample. An additional problem occurs if the biomarker is influenced by freeze-thaw cycles and the same sample aliquot used at time A must be thawed out again for analyses at time B. In this circumstance, one must weigh the measurement concerns arising from batch effects versus the measurement concerns associated with free-thaw cycles. One option would be to use aliquots from the subjects matched to the individual in question, which have also been previously thawed.

Cross-Sectional Analyses among Controls. An issue with nested case-control studies is the validity of cross-sectional analyses conducted in the controls. For instance, it might be of interest to assess whether a biomarker is associated with particular exposures in the controls. Analyses have been conducted in controls to determine whether carcinogen-adduct levels are associated with various genetic polymorphisms and demographic and lifestyle characteristics (73, 74). As pointed out above, such analyses in the subcohort of a case-cohort study would be valid and results could be extrapolated to the entire cohort. We emphasize that the choice between conducting a nested case-control or case-cohort study should be driven by the need to most validly assess exposure-disease hypotheses. However, these types of secondary analyses have been published and there are concerns regarding their validity.

In a nested case-control study, it is typical that controls are matched to cases in a 2:1 or 3:1 ratio. Due to the individual matching of controls to cases, the control series are not independent observations; in fact, each set of controls constitutes a highly intercorrelated cluster (i.e., many variables in each control set will be correlated within control sets because each control was tightly matched to a case, and thus is correlated with the case). For instance, among controls matched to lung cancer cases in Genair/EPIC, body mass index, which was not a matching variable, is significantly correlated within pairs of matched controls ($r = 0.17$, $P = 0.01$). Such correlations can also occur for biomarker levels due to matching on batch within case-control sets. Thus, biomarkers levels and potential determinants of biomarkers may be intercorrelated within matched control sets creating very clustered data structures. This is illustrated in Fig. 3. This clustering must be taken into account in the cross-sectional analyses. β coefficients from standard least-square linear regression, logistic regression analyses, or t tests will typically have artificially small SE values because of this clustering and robust variance estimators will be necessary. An appropriate approach is to use generalized estimating equations or mixed models and to use the match ID variable as the cluster variable.

A separate issue is that the controls are typically not a representative sample of the cohort at baseline, and caution is required when extrapolating results of cross-sectional analyses in the controls to the overall cohort. It is possible that cross-sectional results in controls will produce a quite biased representation of associations in the overall cohort at baseline. The issues of bias and precision are illustrated in Table 1. The association between body mass index and gender and age was analyzed in the controls from Genair/EPIC and the corresponding portion of the EPIC cohort in which Genair/EPIC was nested. Analyses of the controls from Genair/EPIC were conducted using linear regression analyses and then with generalized estimating equations. Comparisons of cross-sectional analyses show that linear regression β coefficients calculated from the controls are quite different from those calculated from the corresponding cohort at baseline. Additionally, the SE values of the β coefficients are substantially smaller in the linear regression analyses of the controls compared with generalized estimating equation analyses of the controls.

There are several possible reasons for the discrepancies in the β coefficients. One possibility is that the relationship between age and body mass index may be nonlinear and plateau at older

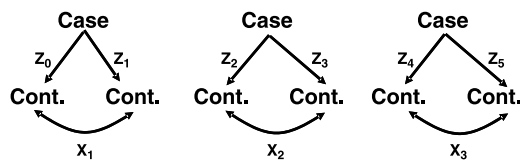


Figure 3. Intercorrelations among control sets due to matching to cases. This figure illustrates three matched case-control sets from a larger nested case-control study population and shows the clustering of control subjects. Instead of treating controls as a series of independent observations, investigators must acknowledge that pairs of controls may represent highly correlated clusters. In this illustration, instead of treating the control series as six independent subjects, the control series should be visualized as three clusters of two subjects. X s and Z s represent correlation coefficients for variables in the case-control sets. If controls were matched to cases on age and gender, then the individual controls matched to the case will be highly correlated with each other for these characteristics. Because biological samples from the controls were likely to have been assayed in the same laboratory batch, the biomarker data is likely to be correlated within matched pairs as well.

Table 1. Analyses of body mass index, gender, and age in controls from Genair/EPIC and in the EPIC cohort

	EPIC*† (β)	Controls from Genair/EPIC* (β , SE)	Controls from Genair/EPIC‡ (β , SE)
Gender	-1.06	-0.46, 0.150	-0.46, 0.164
Age	0.08	0.03, 0.008	0.03, 0.010

*Analyses by linear regression.

†Restricted to the portion of EPIC that Genair/EPIC was nested in; nonsmokers and ex-smokers who quit smoking >10 years before enrollment into EPIC.

‡Analyses by generalized estimating equations.

ages. The selected controls are likely to be older than the overall cohort due to matching to the cases. Thus, analyses in controls from Genair/EPIC may occur in a flatter portion of the overall age-body mass index curve and generate a lower β coefficient. Another possibility is that the distribution of important uncontrolled confounders differs between the Genair/EPIC controls and the overall cohort, producing different confounding effects in the analyses. A final possibility is that one or more of the matching factors are effect modifiers of the effects of age and gender and the analyses in the Genair/EPIC controls really represent stratum-specific analyses within a larger context of interaction. In any case, the problem is that because the controls are not generally representative of the overall cohort, care must be taken in extrapolating cross-sectional results from the controls to the overall population.

Discussion

Substantial resources have been invested in launching large cohort studies that collect biological samples. It is prohibitively inefficient to analyze biological samples from each of the study subjects and then conduct standard cohort analyses. As such, nested case-control and case-cohort analyses serve as the primary tools for conducting molecular epidemiologic studies within cohorts. A recent review shows that the nested case-control approach is more commonly used than the case-cohort, probably because case-cohort methods have been developed more recently (9). However, there is dissatisfaction with the nested case-control approach because the matching usually creates very complex data sets that are not intuitively understood and have limited uses beyond the immediate case-control analyses.

A body of literature already exists that compares and contrasts the strengths of these two study designs; thus far, there has been little consideration of how the addition of biomarker analyses alters the balance (12). The concerns raised by biomarker analyses regard assuring that the quality of the biomarker data is comparable in the case series and the reference group (controls or subcohort). We have identified three key areas of concern: batch effects, storage effects on biological samples, and freeze-thaw effects on biological samples. Table 2 summarizes the strengths and weaknesses of the two designs in addressing these issues. In our own experience, batch effects present the most consistent problem, yet of the areas of concern, batch effects seem to be the least well described in the literature. Batch effects can generate random noise, bias, or may cause clustering in the data. The manner in which these effects can impact a study depends on the temporal relationship between the time period in which the work of a nested case-control or case-cohort study is done and the life course of the underlying cohort. The effects also depend on when in the time period the laboratory begins the analytic runs of the samples. These timing issues influence batching decisions, differences in storage duration between cases and subcohort members, and the likelihood that sample aliquots will have endured freeze-thaw cycles.

Table 2. Comparison of case-cohort and nested case-control designs with respect to biomarker-related issues

Sensitivity of analyte to:	Case-cohort design			Nested case-control
	Prospective, initial case series	Retrospective, initial case series	Subsequent case series	
Batch effect, not related to time	Controllable with proportional batching, or by doing laboratory work at the end of follow-up	Samples can be randomized to batches to remove batch effects	There may be no bias, but presence of bias is not detectable	Matching on batch will control for effects and is logistically simple
Batch effect, related to time	Controllable with proportional batching, but is more difficult, or by doing laboratory work at the end of follow-up	Samples can be randomized to batches to remove batch effects	Bias will occur	Matching on batch will control for effects and is logistically simple
Storage length	Can be addressed through statistical control for storage length	Can be addressed through statistical control for storage length	Statistical control becomes less efficient as follow-up gets longer	The design matches on duration of follow-up and thus usually on storage time
Thawing/freezing cycles	Can be addressed through statistical control	Can be addressed through statistical control	Likely to be a problem; statistical control may be possible	Possible to match on number of thaws
None of the above	Case-cohort has clear advantages	Case-cohort has clear advantages	Case-cohort has clear advantages	Nested case-control has few advantages

In terms of batching samples to particular analytic runs, the nested case-control design provides logistically the simplest approach. Samples can be identified for batching as the controls are matched to the cases. The only potentially awkward circumstance occurs if a study subject is represented multiple times in the study. In this circumstance, an investigator would ideally analyze separate aliquots of that subject's biological sample at each point in which the subject appears in the study. In the case-cohort design, the ideal situation occurs if all of the cases for analyses and the subcohort have been identified at the beginning of the time period in which the work of the case-cohort study is done. In this circumstance, biological samples can be randomly assigned to batches. If case selection is ongoing and biological samples are being analyzed before the entire study sample is identified, creating laboratory batches that are balanced with respect to case-subcohort status is challenging.

On balance, for biomarkers that suffer from batch effects and/or storage decay, we feel the nested case-control study provides a better approach to dealing with these issues. The primary complaint with the nested case-control study is that the matching is not intuitive and that investment in generating biomarker data from the controls cannot be used as reference data for other case series or compared with other populations. Thus, there is a tremendous investment in generating data that has few secondary uses. Typically, the case-cohort design is thought to provide an answer to these issues, but analytic batch effects and storage effects on samples can largely nullify this advantage. For biomarkers that do not suffer from storage effects, batch effects, and freeze-thaw cycles, the issues in choosing between study designs are essentially the same as confronted with questionnaire data. In this case, we feel that the investment in generating reference data should be leveraged as much as possible and that the case-cohort design is the preferred approach.

References

- Gammon MD, Santella RM, Neugut AI, et al. Environmental toxins and breast cancer on Long Island. I. Polycyclic aromatic hydrocarbon DNA adducts [see comment]. *Cancer Epidemiol Biomarkers Prev* 2002;11:677–85.
- Rundle A, Tang D, Hibshoosh H, et al. The relationship between genetic damage from polycyclic aromatic hydrocarbons in breast tissue and breast cancer. *Carcinogenesis* 2000;21:1281–9.
- Ambrosone CB, Freudenheim JL, Graham S, et al. Cigarette smoking, N-acetyltransferase 2 genetic polymorphisms, and breast cancer risk [see comment]. *JAMA* 1996;276:1494–501.
- Tang D, Phillips D, Stampfer M, et al. Association between carcinogen-DNA adducts in white blood cells and lung cancer risk in the physicians health study. *Cancer Res* 2001;61:6708–12.
- Riboli E, Kaaks R. The EPIC Project: rationale and study design. *European Prospective Investigation into Cancer and Nutrition. Int J Epidemiol* 1997;26:s6–14.
- Wacholder S. Practical considerations in choosing between the case-cohort and nested case-control designs. *Epidemiology* 1991;2:155–8.
- Wacholder S, Gail M, Pee D. Selecting an efficient design for assessing exposure-disease relationships in an assembled cohort. *Biometrics* 1991;47:63–76.
- Wacholder S, Silverman DT, McLaughlin JK, Mandel JS. Selection of controls in case-control studies. III. Design options. *Am J Epidemiol* 1992;135:1042–50.
- Barlow WE, Ichikawa L, Rosner D, Izumi S. Analysis of case-cohort designs. *J Clin Epidemiol* 1999;52:1165–72.
- Langholz B, Thomas DC. Nested case-control and case-cohort methods of sampling from a cohort: a critical comparison. *Am J Epidemiol* 1990;131:169–76.
- Suissa S, Edwards MD, Boivin JF. External comparisons from nested case-control designs. *Epidemiology* 1998;9:72–8.
- Hunter D. Methodological issues in the use of biological markers in cancer epidemiology: cohort studies. In: Toniolo P, Boffetta P, Shuker D, Rothman N, Hulka B, Pearce N, editors. *Application of biomarkers in cancer epidemiology* 142. Lyon: IARC; 1997. p. 39–42.
- Miettinen O. Design options in epidemiologic research: an update. *Scand J Work Environ Health* 1982;8:7–14.
- Prentice R. A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* 1986;73:1–11.
- Salomaa V, Matei C, Aleksic N, et al. Soluble thrombomodulin as a predictor of incident coronary heart disease and symptomless carotid artery atherosclerosis in the Atherosclerosis Risk in Communities (ARIC) Study: a case-cohort study [see comment]. *Lancet* 1999;353:1729–34.
- Ballantyne CM, Hoogeveen RC, Bang H, et al. Lipoprotein-associated phospholipase A2, high-sensitivity C-reactive protein, and risk for incident coronary heart disease in middle-aged men and women in the Atherosclerosis Risk in Communities (ARIC) study [see comment]. *Circulation* 2004;109:837–42.
- Sato T. Risk ratio estimation in case-cohort studies. *Environ Health Perspect* 1994;102 Suppl 8:53–6.
- Schouten EG, Dekker JM, Kok FJ, et al. Risk ratio and rate ratio estimation in case-cohort designs: hypertension and cardiovascular mortality [see comment]. *Stat Med* 1993;12:1733–45.
- Rothman K. *Modern Epidemiology*. Boston (MA): Little, Brown and Company; 1986.
- Dirx MJ, Voorrips LE, Goldbohm RA, van den Brandt PA. Baseline recreational physical activity, history of sports participation, and postmenopausal breast carcinoma risk in the Netherlands Cohort Study. *Cancer* 2001;92:1638–49.
- Zeka A, Eisen EA, Kriebel D, Gore R, Wegman DH. Risk of upper aerodigestive tract cancers in a case-cohort study of autoworkers exposed to metalworking fluids. *Occup Environ Med* 2004;61:426–31.
- Abrahamowicz M, Schopflocher T, Leffondre K, du Berger R, Krewski D. Flexible modeling of exposure-response relationship between long-term average levels of particulate air pollution and mortality in the American Cancer Society study. *J Toxicol Environ Health A* 2003;66:1625–54.

23. Falk RT, Rossi SC, Fears TR, et al. A new ELISA kit for measuring urinary 2-hydroxyestrone, 16 α -hydroxyestrone, and their ratio: reproducibility, validity, and assay performance after freeze-thaw cycling and preservation by boric acid. *Cancer Epidemiol Biomarkers Prev* 2000;9:81–7.
24. Aziz N, Nishanian P, Mitsuyasu R, Detels R, Fahey JL. Variables that affect assays for plasma cytokines and soluble activation markers. *Clin Diagn Lab Immunol* 1999;6:89–95.
25. Tworoger SS, Yasui Y, Chang L, Stanczyk FZ, McTiernan A. Specimen allocation in longitudinal biomarker studies: controlling subject-specific effects by design. *Cancer Epidemiol Biomarkers Prev* 2004;13:1257–60.
26. Taioli E, Kinney P, Zhitkovich A, et al. Application of reliability models to studies of biomarker validation. *Environ Health Perspect* 1994;102:306–9.
27. Falk RT, Gail MH, Fears TR, et al. Reproducibility and validity of radioimmunoassays for urinary hormones and metabolites in pre- and postmenopausal women. *Cancer Epidemiol Biomarkers Prev* 1999;8:567–77.
28. Rybicki BA, Rundle A, Savera AT, Sankey SS, Tang D. Polycyclic aromatic hydrocarbon-DNA adducts in prostate cancer. *Cancer Res* 2004;64:8854–9.
29. Rundle A. A molecular epidemiologic case-control study of breast cancer. *Environmental Health Sciences*. New York: Mailman School of Public Health; 2000. p. 152.
30. Schulte P, Perera F. *Molecular epidemiology: principles and practices*. San Diego (CA): Academic Press; 1993.
31. Comstock GW, Norkus EP, Hoffman SC, Xu MW, Helzlsouer KJ. Stability of ascorbic acid, carotenoids, retinol, and tocopherols in plasma stored at -70°C for 4 years. *Cancer Epidemiol Biomarkers Prev* 1995;4:505–7.
32. Comstock GW, Alberg AJ, Helzlsouer KJ. Reported effects of long-term freezer storage on concentrations of retinol, β -carotene, and α -tocopherol in serum or plasma summarized. *Clin Chem* 1993;39:1075–8.
33. Riboli E, Haley NJ, De Waard F, Saracci R. Validity of urinary biomarkers of exposure to tobacco smoke following prolonged storage. *Int J Epidemiol* 1995;24:354–8.
34. Binkova B, Hubalek F, Sram R. Stability of benzo(a)pyrene DNA adducts in rat tissues during their long-term storage at -20°C or -80°C . *Mutat Res* 1996;371:229–35.
35. Evans RW, Sankey SS, Hauth BA, Sutton-Tyrrell K, Kamboh MI, Kuller LH. Effect of sample storage on quantitation of lipoprotein(a) by an enzyme-linked immunosorbent assay. *Lipids* 1996;31:1197–203.
36. Kronenberg F, Lobentanz EM, Konig P, Utermann G, Dieplinger H. Effect of sample storage on the measurement of lipoprotein[a], apolipoproteins B and A-IV, total and high density lipoprotein cholesterol and triglycerides. *J Lipid Res* 1994;35:1318–28.
37. Shih WJ, Bachorik PS, Haga JA, Myers GL, Stein EA. Estimating the long-term effects of storage at -70°C on cholesterol, triglyceride, and HDL-cholesterol measurements in stored sera. *Clin Chem* 2000;46:351–64.
38. Lathey JL, Marschner IC, Kabat B, Spector SA. Deterioration of detectable human immunodeficiency virus serum p24 antigen in samples stored for batch testing. *J Clin Microbiol* 1997;35:631–5.
39. Woodrum D, French C, Shamel LB. Stability of free prostate-specific antigen in serum samples under a variety of sample collection and sample storage conditions. *Urology* 1996;48:33–9.
40. Bolelli G, Muti P, Micheli A, 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.
41. Phillips GB, Yano K, Stemmermann GN. Serum sex hormone levels and myocardial infarction in the Honolulu Heart Program. Pitfalls in prospective studies on sex hormones. *J Clin Epidemiol* 1988;41:1151–6.
42. Lagging LM, Garcia CE, Westin J, et al. Comparison of serum hepatitis C virus RNA and core antigen concentrations and determination of whether levels are associated with liver histology or affected by specimen storage time. *J Clin Microbiol* 2002;40:4224–9.
43. Ng V, Koh D, Fu Q, Chia SE. Effects of storage time on stability of salivary immunoglobulin A and lysozyme. *Clin Chim Acta* 2003;338:131–4.
44. Risio M, De Rosa G, Sarotto I, et al. HER2 testing in gastric cancer: molecular morphology and storage time-related changes in archival samples. *Int J Oncol* 2003;23:1381–7.
45. Fergenbaum JH, Garcia-Closas M, Hewitt SM, Lissowska J, Sakoda LC, Sherman ME. Loss of antigenicity in stored sections of breast cancer tissue microarrays. *Cancer Epidemiol Biomarkers Prev* 2004;13:667–72.
46. Su SC, Garbers S, Rieper TD, Toniolo P. Temperature variations in upright mechanical freezers. *Cancer Epidemiol Biomarkers Prev* 1996;5:139–40.
47. Helsing KJ, Hoffman SC, Comstock GW. Temperature variations in chest-type mechanical freezers. *Clin Chem* 2000;46:1861.
48. Brey RL, Cote SA, McGlasson DL, Triplett DA, Barna LK. Effects of repeated freeze-thaw cycles on anticardiolipin antibody immunoreactivity. *Am J Clin Pathol* 1994;102:586–8.
49. van den Berg L, Rose D. Effect of freezing on the pH and composition of sodium and potassium phosphate solutions: the reciprocal system $\text{KH}_2\text{PO}_4\text{-Na}_2\text{HPO}_4\text{-H}_2\text{O}$. *Arch Biochem Biophys* 1959;81:319–29.
50. Williams-Smith DL, Bray RC, Barber MJ, Tsopanakis AD, Vincent SP. Changes in apparent pH on freezing aqueous buffer solutions and their relevance to biochemical electron-paramagnetic-resonance spectroscopy. *Biochem J* 1977;167:593–600.
51. Comstock GW, Burke AE, Norkus EP, Gordon GB, Hoffman SC, Helzlsouer KJ. Effects of repeated freeze-thaw cycles on concentrations of cholesterol, micronutrients, and hormones in human plasma and serum. *Clin Chem* 2001;47:139–42.
52. Niedbala RS, Haley N, Kardos S, Kardos K. Automated homogeneous immunoassay analysis of cotinine in urine. *J Anal Toxicol* 2002;26:166–70.
53. Reyna R, Traynor KD, Hines G, Boots LR, Azziz R. Repeated freezing and thawing does not generally alter assay results for several commonly studied reproductive hormones. *Fertil Steril* 2001;76:823–5.
54. Yu H, Mistry J, Nicari MJ, et al. Insulin-like growth factors (IGF-I, free IGF-I and IGF-II) and insulin-like growth factor binding proteins (IGFBP-2, IGFBP-3, IGFBP-6, and ALS) in blood circulation. *J Clin Lab Anal* 1999;13:166–72.
55. Groschl M, Wagner R, Rauh M, Dorr HG. Stability of salivary steroids: the influences of storage, food and dental care. *Steroids* 2001;66:737–41.
56. Aziz N, Fahey JL, Detels R, Butch AW. Analytical performance of a highly sensitive C-reactive protein-based immunoassay and the effects of laboratory variables on levels of protein in blood. *Clin Diagn Lab Immunol* 2003;10:652–7.
57. Pinsky NA, Huddleston JM, Jacobson RM, Wollan PC, Poland GA. Effect of multiple freeze-thaw cycles on detection of measles, mumps, and rubella virus antibodies. *Clin Diagn Lab Immunol* 2003;10:19–21.
58. Quan CM, Krajden M, Zhao J, Chan AW. High-performance liquid chromatography to assess the effect of serum storage conditions on the detection of hepatitis C virus by the polymerase chain reaction. *J Virol Methods* 1993;43:299–307.
59. Sgoutas DS, Tuten T. Effect of freezing and thawing of serum on the immunoassay of lipoprotein(a). *Clin Chem* 1992;38:1873–7.
60. Bilodeau JF, Chatterjee S, Sirard MA, Gagnon C. Levels of antioxidant defenses are decreased in bovine spermatozoa after a cycle of freezing and thawing. *Mol Reprod Dev* 2000;55:282–8.
61. Halfon P, Khiri H, Gerolami V, et al. Impact of various handling and storage conditions on quantitative detection of hepatitis C virus RNA. *J Hepatol* 1996;25:307–11.
62. McLeay WR, Horsfall DJ, Seshadri R, Morrison DA, Saccone GT. Epidermal growth factor receptor in breast cancer: storage conditions affecting measurement, and relationship to steroid receptors. *Breast Cancer Res Treat* 1992;22:141–51.
63. Holten-Andersen MN, Schroll AS, Brunner N, Nielsen HJ, Hogdall CK, Hogdall EV. Evaluation of sample handling in relation to levels of tissue inhibitor of metalloproteinases-1 measured in blood by immunoassay. *Int J Biol Markers* 2003;18:170–6.
64. Lahiri DK, Schnabel B. DNA isolation by a rapid method from human blood samples: effects of MgCl_2 , EDTA, storage time, and temperature on DNA yield and quality. *Biochem Genet* 1993;31:321–8.
65. Bellele B, Flori P, Hafid J, Raberin H, Tran Manh Sung R. Influence of the quantity of nonspecific DNA and repeated freezing and thawing of samples on the quantification of DNA by the Light Cycler. *J Microbiol Methods* 2003;55:213–9.
66. Nederhand RJ, Droog S, Klufft C, Simoons ML, de Maat MP; Investigators of the EUROPA trial. Logistics and quality control for DNA sampling in large multicenter studies. *J Thromb Haemost* 2003;1:987–91.
67. Ahsan H, Chen Y, Parvez M, et al. Health Effects of Arsenic Exposure Longitudinal Study (HEALS)—design and description of a multidisciplinary collaborative investigation. *J Expo Anal Environ Epidemiol*. In press 2005.
68. Breslow N, Lubin J, Marek P, Langholz B. Multiplicative models and cohort analysis. *J Am Stat Assoc* 1983;78:1–12.
69. Pirkle JL, Brody DJ, Gunter EW, et al. The decline in blood lead levels in the United States. The National Health and Nutrition Examination Surveys (NHANES) [see comment]. *JAMA* 1994;272:284–91.
70. Barlow W. Robust variance estimation for the case-cohort studies. *Biometrics* 1994;50:1064–72.
71. Self S, Prentice R. Asymptotic distribution theory and efficiency results for case-cohort studies. *Ann Stat* 1988;16:64–81.
72. Vineis P, Airoldi L, Veglia F, et al. Environmental tobacco smoke and the risk of respiratory cancer and COPD in ex-smokers and never smokers in the EPIC prospective study. *Br Med J* 2005;330:277–80.
73. Perera FP, Mooney LA, Stampfer M, et al. Physicians' Health Cohort Study. Associations between carcinogen-DNA damage, glutathione S-transferase genotypes, and risk of lung cancer in the prospective Physicians' Health Cohort Study. *Carcinogenesis* 2002;23:1641–6.
74. Sun CA, Wu DM, Wang LY, Chen CJ, You SL, Santella RM. Determinants of formation of aflatoxin-albumin adducts: a seven-township study in Taiwan. *Br J Cancer* 2002;87:966–70.