Molecular Epidemiology of Physical Activity and Cancer

Andrew Rundle

Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, New York

Abstract

As in other areas of epidemiology, researchers studying physical activity and cancer have begun to include laboratory analyses of biological specimens in their studies. The incorporation of these “biomarkers” into epidemiology has been termed molecular epidemiology and is an approach primarily developed to study chemical carcinogens. Thus far, there has been no discussion in the field on how the established molecular epidemiologic framework might be adapted for research into physical activity, what methodologic needs exist, what the goals of such an approach might be, and what limitations exist. This article relates the literature on molecular epidemiology to the needs of physical activity research and tries to set research priorities for the field as it moves in this new direction. Although this approach will be very useful for investigating the mechanisms through which physical activity exerts effects, there are several challenges for physical activity epidemiologists in adapting molecular epidemiologic approaches. Primarily, there are currently no available biomarkers that might be considered measures of exposure or biologically effective dose. In addition, most available biomarkers of intermediate effects have been tested in training studies at activity levels much higher than those seen in population-based epidemiologic studies. Thus, it is not clear whether these biomarkers are valid at lower activity levels. Furthermore, the nature of the relationship between activity and many available biomarkers depends very much on the context of the activity. Addressing these issues should be a priority if we are to develop a molecular epidemiologic paradigm for studying physical activity. (Cancer Epidemiol Biomarkers Prev 2005;14(1):227–36)

Introduction

Although there is consistent epidemiologic evidence showing a protective effect of physical activity for some cancers, surprisingly little is known about the mechanisms through which activity exerts its effects. It has been suggested that biomarker studies would be very useful in evaluating the role of physical activity in cancer prevention (1-4). Biomarkers refer to the measurement of biological parameters that reflect events along the causal chain between exposure and disease (5). Although these calls to action have been met with general approval, there has not yet been a discussion of what types of biomarkers and what strategies would be most useful. Such studies would fall under the rubric of molecular epidemiology, an analytic paradigm originally developed for the investigation of chemical carcinogens (5). Thus far, there has been no discussion of how this framework might be adapted for research into physical activity and cancer. This article will (a) review the molecular epidemiologic framework developed for chemical carcinogens, (b) review the rationale for conducting biomarker studies of physical activity, (c) discuss the role such studies would play in the investigation of various cancers, and (d) explore how such studies would relate to currently developed molecular epidemiologic paradigms.

The Traditional Molecular Epidemiologic Paradigm

A premise of molecular epidemiology is that examination of physical activity and cancer have begun to include laboratory analyses of biological specimens in their studies. The incorporation of these “biomarkers” into epidemiology has been termed molecular epidemiology and is an approach primarily developed to study chemical carcinogens. Thus far, there has been no discussion in the field on how the established molecular epidemiologic framework might be adapted for research into physical activity, what methodologic needs exist, what the goals of such an approach might be, and what limitations exist. This article relates the literature on molecular epidemiology to the needs of physical activity research and tries to set research priorities for the field as it moves in this new direction. Although this approach will be very useful for investigating the mechanisms through which physical activity exerts effects, there are several challenges for physical activity epidemiologists in adapting molecular epidemiologic approaches. Primarily, there are currently no available biomarkers that might be considered measures of exposure or biologically effective dose. In addition, most available biomarkers of intermediate effects have been tested in training studies at activity levels much higher than those seen in population-based epidemiologic studies. Thus, it is not clear whether these biomarkers are valid at lower activity levels. Furthermore, the nature of the relationship between activity and many available biomarkers depends very much on the context of the activity. Addressing these issues should be a priority if we are to develop a molecular epidemiologic paradigm for studying physical activity. (Cancer Epidemiol Biomarkers Prev 2005;14(1):227–36)
epidemiologic studies of physical activity, it must be decided how physical activity and related biomarkers fit into this paradigm.

The Rationale for Biomarker Studies of Physical Activity

Molecular epidemiology was originally developed to crack the “black box” between exposures to xenobiotics and cancer. If anything, the relationship between physical activity and cancer represents a more opaque black box than the relationship between chemical exposures and cancer. Long before the advent of molecular epidemiology, it was understood that chemical carcinogens could form DNA adducts and cause mutations through various means (21-23). There is far less information on how physical activity exerts its effects and what type of intermediate events exist. Furthermore, physical activity patterns are likely to have complex associations with other determinants of cancer risk, such as diet, occupation, and smoking. The lack of data on intermediate events and the likely presence of confounding effects emphasize the need for biomarker studies to generate mechanistic data to inform causal inference and guide future intervention studies.

For cancers for which traditional epidemiology has not provided consistent data on the role of physical activity, biomarker studies could be used to test mechanistic hypotheses and generate biological data useful in the process of causal inference. Furthermore, biomarker studies could be used to test and refine causal hypotheses, such as investigating associations with biomarker-defined subsets of tumors or by providing information on what components of activity may be important. For cancers for which traditional epidemiology has already provided sufficient data to designate causality, biomarker studies could be used to elucidate the mechanisms through which activity exerts its protective effects. This would be useful in identifying new targets for interventions. In addition, biomarkers would be useful as surrogate outcomes in intervention trials to test various physical activity–based prevention programs.

Instances Where Traditional Epidemiologic Data Are Inconclusive. For several cancers epidemiologic data are not consistent or the observed effect is suspect because of the presence of an overwhelming confounder. For example, the hypothesis that physical activity protects against prostate cancer has been investigated in over 30 studies. Fourteen studies show a protective effect, 13 studies show no particular effect, and 4 found an increased risk with activity (24-26). Clearly, the data assessing potential relationships between prostate cancer and physical activity are inconclusive.

Of 13 cohort analyses published on physical activity and lung cancer, 8 have found significant protective effects of physical activity (27-34), 4 observed protective effects that were not statistically significant (35-37), and 1 observed a nonsignificant increased risk of lung cancer with activity (35). In addition, two of four case-control studies on physical activity and lung cancer have found protective effects due to activity (38-41). In the two case-control studies that did not observe protective effects, only occupational activity was studied, which was assessed by job title only (38, 39). However, the primary concern with reports on lung cancer and physical activity is the overwhelming effect of cigarette smoking and the likelihood that activity levels are associated with smoking behavior. Most of the studies have had limited data on smoking and activity levels, and so even after control for smoking, residual confounding with smoking could still explain the observed protective effect of activity. Thus, the data suggesting a protective effect of activity on lung cancer are considered inconclusive.

In instances in which traditional epidemiology is inconclusive or suspect, biomarkers downstream from physical activity can be used to generate data relevant to causal inference. Mechanistic studies using biomarkers can examine whether physical activity acts as an antecedent to a known risk factor or as an effect modifier of a known disease process. An example of physical activity acting as an antecedent to a known risk factor is the hypothesis that activity levels influence estrogen levels, which in turn influence breast cancer risk (42, 43). As an effect modifier, physical activity would modify the risk associated with a known risk factor; that is, an exposure would be hypothesized to be less harmful among active compared with sedentary individuals. An example is the hypothesis that physically active smokers would have a lower risk of lung cancer than sedentary smokers because activity induces protective antioxidant enzymes (43-45). Data demonstrating or refuting a hypothesized biological basis for a protective effect would be useful in developing a causal understanding and addressing concerns regarding confounding. Likewise, such studies can determine whether activity is likely to be important at specific times in life or is only important for specific subtypes of the tumor in question. Such data could be then used to generate new or more refined hypotheses that could be tested in population-based epidemiologic studies.

Instances Where Associations with Physical Activity Are Generally Accepted. Sufficient data have been generated through traditional epidemiologic approaches such that a protective effect of physical activity on breast and colon cancer is generally accepted (43, 46-48). For breast and colon cancer, epidemiologic studies that incorporate biomarkers would be useful for determining the mechanisms through which physical activity exerts its effects. These studies could use biomarkers as the outcome or could be etiologic studies that test whether a biomarker is a causal intermediate that mediates the effect of activity on disease risk. Such studies would help answer remaining questions related to the type and extent of activity needed to elicit a protective effect, whether there are critical time periods in which activity is important, and...
whether the effect of activity is modified by other factors. This information would be very important for designing effective intervention trials and creating population-based prevention programs. Furthermore, the identification of the mechanistic pathway may provide us with new targets for intervention that can be effectively modified by means other than activity or in ways that increase the effect of activity. In much the same way that the identification of estrogen-related risk factors for breast cancer helped provide the rationale for tamoxifen chemoprevention, the biological pathways that respond to physical activity may be amenable to pharmaceutical or other interventions. From a public health perspective, the promotion of physical activity is preferable to developing a pill that replaces physical activity. However, the development of pharmaceuticals that target activity responsive pathways may be useful for certain segments of the population, for instance, individuals with conditions that prevent them from engaging in the required level of activity.

The use of biomarkers as surrogate end points in intervention trials would also greatly aid physical activity research (3, 49). The use of validated intermediate biomarkers rather than clinical end points as outcomes would expedite intervention trials allowing for the efficient testing of hypotheses (49). Furthermore, studies that incorporate repeated biomarker measures and cross-over designs could be used to determine the time course over which exercise exerts its effects. In addition, there are questions that cannot easily be answered by observational designs and are better addressed in intervention trials. For instance, an important question is, if active individuals have a lower risk of cancer, will sedentary individuals who become active gain the same protective effect (3)? Because most people do not commonly change their activity patterns, observational studies are not a powerful tool for observing the effects of change (50). Another important role for intervention trials is to determine whether efficacious prevention programs are palatable and sustainable in the community and thus likely to be adopted. Changes in cancer-related biomarkers could be used to determine whether those who have adopted an exercise program actually engage in enough activity to affect cancer risk.

Currently the best example of an exercise intervention trial using intermediate end points as the outcome is the Physical Activity for Total Health study being conducted by McTiernan and colleagues (51). In this study, 173 overweight, sedentary, postmenopausal women were randomized to receive an activity intervention or weekly stretching session for a year. The intervention arm engaged in 45 minutes a day of moderate exercise, five times a week. During the first 3 months intervention arm participants attended three supervised sessions per week at a study facility and exercised 2 days a week at home. For the rest of the study period the intervention arm participants attended one supervised session per week and exercised at home or a facility four times a week. The control arm attended one supervised stretching session a week. The intervention was shown to reduce body weight, body fat, serum estrone, estradiol, and free estradiol (42, 52). Overall, the effects on serum hormones were statistically significant after 3 months of intervention but not after 12 months (42). The effects on serum hormones were primarily observed among women who lost body fat during the intervention, and in this subgroup the effects on estradiol and free estradiol were significant after both 3 and 12 months of intervention (42).

Biomarkers of Exposure and Physical Fitness

In principle, physical activity can be conceptualized as an exposure existing outside of the black box. In the same way that data can be gathered on occupational, dietary, or tobacco-related exposures, data on an individual’s activity levels can be gathered. However, it is difficult to identify or even conceptualize appropriate biomarkers of exposure or biologically effective dose for physical activity. In the realm of physical activity, there is no currently known biomarker akin to aflatoxin B1-albumin adducts, which serve as a biomarker of biologically effective dose in studies of aflatoxin exposure (53, 54). Within the traditional molecular epidemiologic framework, biomarkers of exposure and biologically effective dose measure, respectively, the amount of an external exposure that has entered the body and the amount of the exposure that interacted with a macromolecular target (5, 6). The use of these biomarkers is thought to increase the validity of exposure assessment (6). These concepts do not translate well to the realm of physical activity. Physical activity does not represent an ambient external exposure, like a chemical component of air pollution, in which only a portion of the external dose is absorbed into the body. For physical activity there is no conceptual discrepancy between the external ambient concentrations and the dose that enters the body. Thus, this rationale for using biomarkers of exposure or biologically effective dose does not apply to studies of physical activity. However, exposure-related biomarkers of physical activity would be useful in addressing other difficulties in measuring physical activity.

In instances in which exposures are ambient and unnoticed, or exposure assessment relies on a subject’s memory, the use of biomarkers of exposure or biologically effective dose is expected to improve exposure assessment (6). In many ways, the measurement of physical activity fits this description. For large-scale epidemiologic studies, the assessment of physical activity relies on questionnaires and thus on a study subject’s memory and subjective recall of the duration and intensity of his or her activity. In addition, although regular patterns of training or gym usage may be readily recalled, there are unnoticed and ambient types of activity that may have important effects on health. For many people, activities of daily living, such as cleaning the house, transportation, or taking care of children, may compose a large portion of their activity and differentiate them from truly sedentary individuals. Thus, it is generally acknowledged that physical activity is often poorly measured (43, 55). The current lack of suitable exposure-related biomarkers means that one of the often cited advantages of molecular epidemiology, improved measurement of exposure, will not be realized by adapting this approach to the study of physical activity.

Although not biomarkers in the traditional sense, it is worth considering the use of measures of physical fitness, such as aerobic capacity or resting pulse, as markers of activity. Although they certainly complement measures of activity, using measures of physical fitness to improve measurement of physical activity suffers from several drawbacks. Fitness and physical activity represent different constructs that are not interchangeable and one is not a more refined measure of the other. Fitness is a performance characteristic defined as “the ability to carry out daily tasks with vigor and alertness, without undue fatigue, and with ample energy to enjoy leisure-time pursuits and to meet unforeseen emergencies” (56). Fitness includes health-related aspects, such as cardiorespiratory endurance and body composition, and skill-related aspects, such as balance and speed (56). Physical activity is a behavior defined as “bodily movement produced by skeletal muscles that results in energy expenditure” (56). Thus, the use of fitness as a biomarker of activity may negatively affect the construct validity of studies seeking to understand the role of physical activity. Activity among nonfit subjects can have important health benefits potentially relevant to cancer. For example, in relation to obesity and type II diabetes, overweight/obese individuals who increase physical activity can reduce insulin resistance even in the absence of concurrent weight loss (57-60).
Similarly among the nonobese, physical activity done at levels insufficient to influence body mass or maximal oxygen uptake can still improve insulin action (61). It has been hypothesized that insulin resistance plays a role in the development of pancreatic cancer and underlies observed associations between physical activity, obesity, and pancreatic cancer (62-64). A study of pancreatic cancer that solely uses measures of fitness as a marker of activity may yield inappropriate conclusions about activity.

Clearly, the current lack of biomarkers that can be conceptualized as measures of exposure or the biologically effective dose of physical activity means that the use of molecular epidemiologic approaches will not improve the measurement of physical activity. If the benefits of bringing molecular epidemiologic techniques to the study of physical activity are to be fully realized, the development of such biomarkers should be a research priority.

### Biomarkers of Effect, Altered Function, and Susceptibility

Due to the lack of biomarkers of exposure to physical activity, most of the biomarkers suggested for use in molecular epidemiologic studies of physical activity are best conceptualized as biomarkers of effect, altered function, or susceptibility. Biomarkers of effect and altered function are defined as "processes that are intermediate between exposure and disease" or as "early biological or biochemical changes in the target tissue that result from the action of the carcinogen and are thought to be either a step in the carcinogenic process or correlate closely with that process" (5, 6). Biomarkers of susceptibility represent processes that are thought to modify the effects of an exposure that causes cancer. Table 1 shows examples from the literature of proposed mechanisms through which physical activity might exert its effects and relevant biomarkers that could be incorporated into studies. Figure 2 shows how these proposed mechanisms are thought to relate to the stages of carcinogenesis. For studies of physical activity, these biomarkers either represent a hypothesized risk factor for disease or factors that modify the effects of other risk factors or exposures. That is, physical activity may represent an antecedent exposure that influences a previously identified risk factor or represents an antecedent to a susceptibility factor that interacts with a disease pathway. The correct specification of a biomarker influenced by physical activity as a risk factor on the disease path or as an effect modifier of the disease process is critically important. In the molecular epidemiologic literature on chemical carcinogens, confusion regarding the relationship of antecedent and effect modifying biomarkers has led to the misinterpretation of several important data sets (20).

Hormone levels are an example of a biomarker of effect that are thought to be influenced by physical activity and are themselves a hypothesized causal factor of disease (43). It is thought that high androgen levels cause prostate cancer and high estrogen levels cause breast cancer (1, 43). Physical activity has been hypothesized to protect against these two cancers by influencing the levels of these hormones (1, 43). In a cohort study that has stored blood samples, one could assess the extent to which lower hormone levels mediate any association between physical activity and breast cancer. Physical activity can thus be integrated into molecular epidemiologic models for these cancers, as shown in Fig. 3.

If indeed causal, the observed protective effect of physical activity on lung cancer is likely to occur because physical activity in some manner reduces the effect of cigarette smoke carcinogens. In this case, physical activity would be conceptualized as a susceptibility factor that acts as an effect modifier. Causal models representing effect modification could be diagrammed as shown in Fig. 4. For instance, it has been hypothesized that physical activity increases the body’s defenses against oxidative stress and carcinogens (43, 44, 82, 83) and thus may protect the lung against reactive oxygen species and carcinogens in cigarette smoke (see Fig. 5). Such a causal model could be tested by assessing the impact of physical activity on glutathione peroxidase or catalase activity (biomarkers of endogenous oxidative defenses) and levels of 8-oxodeoxyguanosine, a biomarker of oxidative DNA damage.

<table>
<thead>
<tr>
<th>Table 1. Physical activity responsive pathways and related biomarkers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Potential physical activity responsive pathways</td>
</tr>
<tr>
<td>Immune function</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Growth factors and growth factor binding proteins</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Sex hormones and binding proteins</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Endogenous antioxidant enzyme systems</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Oxidative stress</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>DNA repair</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Phase II xenobiotic enzyme Systems</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
In the case of colon cancer, physical activity has been hypothesized to decrease the transit time of food through the colon, reducing mucosal exposure to dietary carcinogens (43). This mechanism could be investigated by testing whether physical activity is associated with lower levels of heterocyclic amine adducts in individuals with diets high in broiled meats (84). Again, in this example, physical activity is conceptualized as a susceptibility factor that induces a process that lowers the risk associated with a presumed causal exposure. In epidemiologic terms, physical activity is hypothesized to be an effect modifier that reduces the main effects of an exposure.

The relationship between many of these proposed biomarkers and activity can be complex. Several of these markers seem to have U- or J-shaped dose-response curves in relation to activity (85, 86). Furthermore, differential effects have been noted that seem to depend on whether the activity is acute or chronic, is done by trained or untrained individuals, and is of moderate or exhaustive intensity (2, 85, 87, 88). For instance, single bouts of intense exercise lead to the generation of reactive oxygen species and DNA damage (79, 89, 90). However, regular exercise seems to cause an adaptive response that induces endogenous antioxidant enzyme systems and lowers oxidative stress (70, 77-81). Miyazaki and colleagues showed that bouts of exhaustive exercise caused increased lipid peroxidation, indicative of increased oxidative stress (70). However, training increased glutathione peroxidase and superoxide dismutase activity and reduced the levels of lipid peroxidation caused by bouts of exhaustive exercise (70). In regard to immune function, strenuous bouts of activity have been associated with immune suppression as measured by increased upper respiratory infections and by reductions in natural killer cell counts and activity (86, 91). However, moderate exercise training seems to reduce the risk of upper respiratory tract infections (66, 86), and this effect may be due to increases in salivary IgA levels seen with training (92). Likewise several studies have found increases in natural killer cell count and activity associated with programs of exercise training (65, 66, 93-95). Similarly cross-sectional analyses of
sedentary elderly adults and regularly exercising elderly adults have found higher natural killer cell counts among exercisers (96). A firm understanding of the nature of the dose-response curve and its modification by the context of exercise is important for modeling the effect of activity on intermediate biomarkers and on disease outcome.

As with other areas of research in carcinogenesis, physical activity researchers are starting to study gene and protein expression with microarray technology, and such data may be useful for identifying biomarkers of effect (97-99). At this time, the literature on human studies is quite limited. In a study of muscle aging, Roth and colleagues used muscle biopsies to assess the expression of ~4,000 genes in response to strength training (98). In total, 69 genes showed altered expression in response to a 9-week strength-training program, and the majority of changes represented down-regulation. Connolly and colleagues have studied circulating peripheral blood mononuclear cells and shown alterations in expression for 311 genes after a single bout of heavy exercise (99). Up-regulated genes included those related to stress, inflammation, growth and repair (99). At this time, however, it is unclear how an individual's global patterns of gene expression might be incorporated into traditional epidemiologic designs. But work such as that of Roth and colleagues and Connolly and colleagues could identify new candidate genes whose expression might serve as useful biomarkers of effect. As with other biomarkers, the dose response between activity and gene expression is unknown, and it is unclear how other methodologic issues such as multiple comparisons, individual variability, and measurement error affect microarray results (99, 100). Still, gene and protein expression studies are likely to be useful for understanding the effects of physical activity.

**Strategies for the Use of Biomarkers in Studies of Physical Activity**

The molecular epidemiologic literature regarding chemical carcinogens has made the distinction between etiologic investigations and transitional studies (101). This distinction is also useful for developing a molecular epidemiologic approach to the investigation of physical activity.

**Transitional Studies.** Transitional studies are defined as studies that “bridge the gap between laboratory experimentation and population-based epidemiology” (101). Such studies have one of several goals: to validate biomarkers so that they can be used in population-based studies; to determine sources of intersubject variability, thus identifying potential confounders, antecedents, or effect modifiers; to evaluate the feasibility of using the marker in the field; to investigate proposed mechanistic pathways; and to define exposure-effect relations (101). In general, transitional studies treat the biomarker of interest as the outcome. The majority of molecular epidemiologic studies of chemical carcinogens that utilize intermediate biomarkers, such as carcinogen-DNA adducts, are transitional in nature. There are, in contrast, relatively few studies that...
have assessed the association between intermediate biomarkers and disease (15, 17, 102, 103).

There is a particular need for traditional studies of physical activity and biomarkers. Most studies linking physical activity to a biomarker, usually of biological effect, have occurred within the realm of sports physiology or clinical experiments and have been in the context of high activity levels. For instance, Evelo and colleagues trained 23 men and 18 women to a biomarker, usually of biological effect, have occurred within the realm of sports physiology or clinical experiments and have been in the context of high activity levels. For instance, Evelo and colleagues trained 23 men and 18 women to a biomarker, usually of biological effect, have occurred within the realm of sports physiology or clinical experiments and have been in the context of high activity levels (2). Likewise, increases in immune function have been noted in comparisons between sedentary and actively training individuals (104).

It is important to recognize that epidemiologic studies linking physical activity and reduced cancer risk have noted effects at activity levels far lower than typically used in exercise physiology experiments. It is not clear whether biomarker responses seen at high activity levels are relevant to associations seen in the epidemiologic literature. Furthermore, the small sample sizes common in these studies prevent consideration of potential confounders and effect modifiers that may be important to consider in large epidemiologic studies. An example of the discrepancies that can arise between small clinical studies conducted at high training levels and large population-based studies can be seen in the cardiovascular literature. Several small, clinical training studies have observed associations between training and blood lipid profiles (105-108). However, a large population-based study only observed strong associations between activity levels and blood lipid levels when an interaction with the ApoE4 genotype was taken into consideration (109).

A greater focus on large-scale cross-sectional studies of physical activity and biomarkers of effect are warranted to confirm that biological effects seen in training studies are relevant to the lower activity levels typical of population-based epidemiologic studies. Such studies will be useful in identifying potential confounders and effect modifiers that should be considered in etiologic studies. For instance, some of the effects of training on the antioxidant glutathione system seen in small clinical studies have been verified in larger scale studies of the general population (110-112). These larger studies have also identified other important correlates of improved antioxidant glutathione status, such as cigarette smoking, body mass index, and gender, factors likely to be associated with activity (110-112). Also of interest are multilitered studies in which multiple markers along a proposed causal chain are assessed. Taking the example of physical activity and antioxidant enzymes, for instance, it would be of interest to determine whether physical activity levels and increased antioxidant enzyme activity were associated with lower levels of oxidative DNA damage. Statistical analyses could determine whether variation in antioxidant enzyme activity mediated any inverse association between physical activity levels and oxidative DNA damage.

One perceived weakness of larger scale studies of the general population is that activity levels are likely to be assessed using questionnaires (55), whereas more objective measures, such as accelerometers or pedometers, may only be feasible for smaller studies (113, 114). Generally, questionnaires are hampered by measurement error, and, furthermore, older questionnaires lacked content validity, focusing only on selected recreational activities and not considering occupational or home activities (43, 55). Poor questionnaire reliability reduces the ability of a study to identify relationships (55, 115). However, despite these issues, questionnaire-based studies have revealed important relationships between activity and certain cancers (2, 30, 43). Because the relationship between questionnaire data and a causal intermediate is likely to be stronger than the relationship between questionnaire data and disease, questionnaire data are likely to provide sufficient information for larger scale translational studies that use biomarkers as the outcome (20).

**Etiologic studies.** Etiologic studies include measures of disease outcome, and optimally would assess measures of activity and activity-related biomarkers of effect representing hypothesized mechanisms through which physical activity exerts its effects. In light of the large number of proposed mechanisms through which physical activity might exert its effects, it is important that these biomarkers have previously been validated and tested in translational studies. The optimal design would be a case-cohort or nested case-control study conducted within a cohort study that collected baseline blood samples. Population- or hospital-based case-control studies are problematic because in these designs blood samples are collected from cases at the time of or after diagnosis. Disease status may influence biomarker levels and/or may influence recent activity patterns that, in turn, could influence biomarker levels (116).

The European Prospective Investigation of Cancer and Nutrition (EPIC) is an excellent example of a study in which molecular epidemiologic studies of physical activity can be nested (117). Baseline blood samples are available for biomarker analyses, as are data on physical activity and important potential confounding factors, such as dietary practices and smoking (117, 118). An appropriate strategy would be to conduct a nested case-control study of a particular cancer within the EPIC cohort and to analyze stored samples for a biomarker of interest. Given the lack of exposure-related biomarkers, biomarkers of effect would be the most likely candidates for such a study. The biomarker should be well characterized as being activity responsive and should represent a mechanism through which activity is hypothesized to exert its effect. The data would then be analyzed to determine whether the biomarker mediates the association between activity and disease. In this example, sedentary behavior is thought of as the primary “exposure” of interest and as an antecedent to a biological mechanism that causes cancer. Alternatively, physical activity could be hypothesized to be a susceptibility factor that modifies the effect of another exposure. Under this hypothesis, an appropriate strategy would be to test for an interaction between the activity-related biomarker and the primary exposure of interest. For instance, one could test for an interaction between smoking history and an immune function marker that is thought to be positively influenced by physical activity.

Biomarkers would also be useful in etiologic studies to define more etiologically homogeneous subgroups of cancers that may be more strongly associated with physical activity (119). That is, specific molecular characteristics of tumors, such as p53 mutation or estrogen receptor expression, may define a subset of cancers linked to a specific exposure by virtue of the exposure having caused the mutation or expression of that marker (119-121). Alternatively, tumors not expressing the marker in question may have been caused by other exposures. If indeed an exposure is associated only with a particular subset of a cancer diagnosis, for instance, estrogen receptor–positive breast cancer, studies that include all breast cancers would underestimate the effect of the exposure or even fail to find the association. Biomarkers measured in tumors can be used to test hypotheses regarding etiologic heterogeneity in either the case-only study design or in a case-control or cohort study using a polytomous outcome variable (119).

Enger and colleagues have used this strategy to study physical activity, body size, and breast cancer (121). Both of these risk factors have been hypothesized to act through a hormone-related pathway. Thus, it was hypothesized that obesity and a lack of physical activity would be associated with increased risk of estrogen receptor (ER) and progesterone receptor (PR) positive tumors, but not ER−/PR− tumors (121).
For postmenopausal breast cancer they found increasing body mass index to be associated with case-control status only when ER+/PR+ cases were compared with controls. However, they found the protective effect of physical activity did not vary by ER/PR status. Thus, ER/PR status defined etiologic heterogeneity with respect to body mass index, but not with respect to physical activity, suggesting a sex hormone–related causal pathway for body mass index only. Similarly, Gammon and colleagues have used polyclonal logistic regression to assess whether the protective effect of physical activity differs by p53 status in tumor tissue (120). Again in this example the biomarker did not define tumor subgroups that were etiologically homogeneous with regard to activity levels.

**Conclusion**

Molecular epidemiology was developed to address the challenges of linking xenobiotic exposures to cancer development. There is a great hope that adapting these approaches to the study of physical activity and cancer will aid in elucidating new associations, identifying mechanistic pathways, and validating prevention programs. A primary utility of biomarker studies will be to illuminate the mechanisms through which activity exerts its effects. The elucidation of these mechanisms will hopefully provide new targets for lifestyle or pharmaceutical interventions that may more efficiently modulate these pathways, producing a greater preventive effect than physical activity. This is analogous to how the identification of reproductive risk factors for breast cancer led to investigations of the role of reproductive hormones, which then helped provide a rationale for the use of chemopreventive agents that target reproductive hormone–related pathways. Another utility will be the use of biomarkers to refine causal hypotheses and potentially uncover relationships between physical activity and cancer that have thus far been hidden. Of particular interest is the possibility that sedentary life-styles might only be associated with particular subtypes of cancers and that these subtypes may be identifiable by biomarkers measured in tumor samples. Biomarkers may also be very useful as intermediate outcomes in intervention trials, increasing the rate at which intervention strategies can be tested.

However, there are several challenges to applying molecular epidemiologic approaches to the study of physical activity and cancer that must be surmounted before these hopes can be fulfilled. A primary challenge is the current lack of biomarkers of exposure or biologically effective dose for physical activity. Without major developments in this area, the application of molecular epidemiologic approaches will not solve the difficult measurement issues inherent in physical activity research. An additional challenge is defining the relationship between physical activity and biomarkers of effect, altered function, and susceptibility. Many of the dose-response curves do not appear to be linear, and for many biomarkers, the effects of acute bouts of exhaustive activity differ from the effects of training. Linear, and for many biomarkers, the effects of acute bouts of intense exercise physiology studies. Addressing these issues should be a prominent component of the physical activity and cancer research agenda for the coming years.

**Acknowledgments**

I thank Drs. Neugut, Vineis, Ahsan, and Halim and Ms. Campbell for their thoughtful comments on the manuscript.

**References**
