INTRODUCTION—WHY STUDIES OF CANCER PROGNOSIS AND OUTCOMES?

As a result of improved strategies for cancer detection at earlier stages, as well as improved treatment modalities, the number of cancer survivors in the population continues to increase. As of January 2004, it was estimated that there were 10.8 million cancer survivors in the United States, which is almost triple what it was in 1971 (1). Approximately 14% of the 10.8 million estimated survivors were diagnosed more than 20 years ago, and cancer survivors now represent approximately 3.7% of the population (2).

As the number of survivors increases, research on cancer survivorship has been identified as an area of major importance by the Institutes of Medicine (1), the President’s Cancer Panel (3), and the National Cancer Institute (4). For this large population of cancer survivors, there are many unanswered questions. Many cancer patients want to know what they can do to reduce symptoms during treatment, how they can protect themselves against recurring or secondary tumors, and how they can return to an active, healthy life (5). It is at this time after a cancer diagnosis that individuals are often most motivated to change their diet, exercise habits, and other health behaviors, and vast improvements in public health could be made among cancer survivors (6–12). Unfortunately, however, as stated by the American Institute of Cancer Research, “the painstaking process that yields science-based recommendations on diet and exercise for cancer survivors has not yet reached its conclusion” (5), and there are few guidelines or recommendations for cancer patients.

Because many cancer patients are in search of factors that may improve their health and reduce risk of recurrence, there is widespread use of nutritional supplements and complementary and alternative medicines (13). However, there is a paucity of data from...
Ulrich and Ambrosone

rigorously conducted studies that would support many behavior changes that may be adopted by cancer patients. Although it would be relatively harmless to find that many lifestyle factors may have no beneficial effects for cancer patients, there is the possibility that some factors could actually increase risk of adverse outcomes. For example, there are data indicating that antioxidant supplements may interfere with radiation therapy and many chemotherapeutic agents (14,15), that folate supplements may accelerate growth of malignant tumors (16,17), and that some herbal supplements, such as St. John’s wort, may directly affect the pharmacokinetics of cancer chemotherapeutic drugs (18). Thus, there is an urgent need for results from well-conducted studies to address and answer some of these questions.

Identifying predictors of cancer prognosis has, to date, been largely understudied by molecular epidemiologists, but it is becoming a more prominent research priority (19). In the absence of solid scientific data, it is frequently assumed that factors that reduce the risk of cancer must also have a positive influence on cancer survival. However, this assumption may be inaccurate, as noted above with regard to supplement use. Preliminary data also suggest that weight gain or weight reduction may have different effects on cancer prognosis than on etiology. Reasons for this dual response may be related to the transformation state of cells (e.g., a differential biological mechanism depending on the state of the cell, as detailed below for folate) or in the effects of health behaviors on quality of life and related immunological defense mechanisms. This suggests that epidemiologists should evaluate carefully the potential effects of exposures on prognosis, independent of their associations with etiology.

There is a wide range of research topics that can be addressed in molecular epidemiological studies of cancer prognosis, and such studies will be most fruitful if addressed with an interdisciplinary approach that includes strong biological knowledge. Comprehensive studies of cancer prognosis need to consider the role of molecular characteristics of the tumor in relation to treatment response, as well as the role of inherited genetic variability (polymorphisms) in drug metabolism pathways and response to treatment-related DNA damage in therapeutic toxicity and to treatment (pharmacogenetics) (20). As an additional benefit to studies of prognosis, they may inform studies of cancer etiology. Genetic polymorphisms are more likely to show measurable effects if a system is under stress, e.g., during chemo- or radiation therapy, which puts immense pressure on a cell’s DNA repair capacity and metabolism; identifying the most relevant genetic factors in this scenario can then inform epidemiological studies of gene-environment interaction where the environmental stressor may be comparably less strong (e.g., moderate smoking).

CANCER PROGNOSIS—A MULTIFACTORIAL OUTCOME

A large range of factors can contribute to cancer patients’ prognosis. Not surprisingly, studies of cancer treatment outcomes and prognosis are undertaken by a multitude of researchers who are interested in different research questions and outcomes. The most important outcomes include overall survival, disease-free survival, and, more recently, quality of life. In clinical trials investigating new treatment regimens, treatment-related toxicity is an important secondary outcome that is clearly associated with symptoms and quality of life. In addition, investigations may focus on surgical outcomes, comorbidities, a cancer patient’s ability and success in returning to work, economical outcomes, and many more.

Figure 1 illustrates the many influences and the complex interplay of factors that affect cancer prognosis. While this list is by no means comprehensive, it demonstrates the
A multitude of factors can influence cancer prognosis. Many of these factors are also interrelated or modify associations between other factors and cancer outcomes. Factors that are known to affect prognosis include tumor characteristics (e.g., stage and biological characteristics, such as microsatellite instability in colorectal cancers) (21), treatment modalities, surgical technique (which is of great prognostic significance for cancers that are difficult to resect, e.g., pancreatic cancer or rectal cancer), access to care, race/ethnicity, lifestyle factors (including smoking, body mass index, and physical activity), psychosocial factors, nutritional status, and also inherited genetic characteristics (polymorphisms). Several interrelations emerge with respect to race/ethnicity. Racial factors are indisputably linked to access to care and thus quality of the surgeon. In addition, racial factors are also correlated with genetic factors and tumor biology; for example, African-American women are more likely to be diagnosed with breast cancer with tumors that are high grade and negative for estrogen and progesterone receptors, associated with a poorer prognosis (22). Cancer prognosis itself, or how well a patient responds to an initial treatment regimen, will conversely affect the choice of future treatment modalities as well as quality of life, nutritional status, physical activity and BMI (summarized here as energy balance), and other lifestyle factors. A number of factors can modify the efficacy and toxicity of chemotherapeutic agents, such as genetic polymorphisms (20), gene expression and other tumor characteristics (23), and nutritional status of the patient. Finally, it is expected that genetic polymorphisms will modify associations among lifestyle factors, including nutrition or energy balance and cancer prognosis, similar to gene-environment interactions that have been observed in the context of studies of cancer etiology (24–26).

Traditionally, researchers have studied cancer prognostic factors in isolation. For example, they may have focused exclusively on tumor characteristics, such as gene expression. Alternatively, they may have studied primarily genetic polymorphisms without taking into account relevant epidemiological factors. Or, finally, they may have focused entirely on lifestyle factors without consideration of tumor characteristics or other biomarkers. While all of these approaches are valid and have yielded useful insights into the effects of specific factors on cancer prognosis, important gaps remain. A more comprehensive, integrated approach to studying cancer prognosis seems essential for understanding cancer outcomes, as illustrated below for folate status and prognosis.
limited. Researching the “larger picture” of cancer prognosis creates challenges and new opportunities for molecular epidemiologists.

Folate—an Example of Integrated Prognostic Research

The paradigm of integrated prognostic studies is illustrated here with the example of folate status and prognosis after colorectal cancer (Fig. 2). Within this thematic area, again, there are multiple interrelated components that may affect outcomes and may modify 5-fluorouracil (5-FU)-based treatment outcomes. 5-FU directly targets a key enzyme in folate metabolism, thymidylate synthase (TS), which converts uridine to thymidine. This inhibition results in a deficiency of thymidylate for DNA synthesis, which is compensated for by uracil misincorporation. Misincorporation of uracil into DNA causes repeated repair cycles with a greatly increased likelihood of single and double strand breaks. Thus, 5-FU functions as an antimetabolite. The use of folic acid (FA)-containing supplements before and after cancer diagnosis is thought to affect survival, possibly in opposite directions: while a higher folate status is associated with reduced cancer risk, presumably because of reduced mutation rates, there is increasing concern that the administration of folate once neoplastic or early neoplastic lesions are present can “feed the tumor,” i.e., foster growth of these lesions via a greater provision of nucleotides for DNA synthesis (16,17,27). Such a growth-enhancing effect is consistent with the upregulation of folate receptors and folate-related enzymes in many cancer types; most likely this upregulation reflects a greater need for folate for DNA synthesis to support rapidly growing tumors (28–30). Accordingly, folate-related tumor characteristics, such as the gene and protein expression of TS and enzymes in folate metabolism (31) are known to affect survival and modify 5-FU-based treatment. There is now evidence that gene expression in both the cancerous part of the tissue, as well as normal parts, play a role in the tumor’s folate status and cancer outcomes (32,33). It is not yet clear whether FA-containing supplement use prior to diagnosis affects these tumor characteristics, yet this is an important clinical question. Finally, inherited genetic polymorphisms in the folate pathway, such as 5,10-methylenetetrahydrofolate reductase (MTHFR)

Figure 2  The paradigm of integrated prognostic studies illustrated for folate status and colorectal cancer survival. Abbreviations: FA, folic acid containing supplements (see text for explanation); 5-FU, 5-fluorouracil.
rich and Ambrosones's challenges and new

with the example of this thematic area, outcomes and may directly targets a key converts uridylylate to e for DNA synthesis, oration of uracil into uridine, and The use of folic acid is thought to affect tus is associated with es, there is increasing neoplastic lesions are a greater provision of an effect is consistent in many cancer types; or DNA synthesis to indicate tumor characteristics in folate metabolism

There is now evidence as normal parts, play not yet clear whether or characteristics, yet ic polymorphisms in

reductase (MTHFR)

Molecular Epidemiological Designs for Prognosis

and functional TS variants can modify 5-FU toxicity and efficacy (20,34,35). Inherited genetic polymorphisms are, by default, also reflected in the tumor, unless loss of heterozygosity (i.e., loss of one allele in the cell due to chromosomal instability) has altered the tumor's genotype at a polymorphic locus.

STUDY DESIGNS FOR PROGNOSTIC STUDIES

Multiple study designs can be employed to investigate questions on cancer outcomes and prognosis. Table 1 summarizes and contrasts the general advantages and limitations of various approaches. These study designs should be considered complementary and chosen depending on the current scientific knowledge in a specific research area. For example, if it is not yet clear whether a factor, such as nutritional supplement use, may benefit or potentially harm patients, then observational studies constitute the first step to investigating this research question. Once observational evidence has accumulated and consistently supports a benefit, then intervention studies or randomized clinical trials (for secondary cancer prevention) are needed to confirm such an association with certainty.

Population-Based Cohort Studies

Molecular epidemiological studies of cancer prognosis may be based on follow-up of cases who participated in a case-control study of cancer risk or were identified in the context of a cohort study of cancer etiology. The establishment of a cohort specifically for an observational study of cancer outcomes may also be used to investigate factors contributing to cancer prognosis. Each of these approaches has strengths and weaknesses.

Follow-Up from Studies of Cancer Etiology: Case-Control and Cohort Studies

The conversion of a case-control study of cancer risk to one of cancer prognosis can be efficient and benefit from all of the efforts already put into ascertaining and enrolling cases, procuring blood specimens, and collecting extensive epidemiological data. Such a study may be conducted at a number of levels, from simply following up cases to ascertaining recurrence and survival and evaluating in relation to characteristics prior to and at diagnosis, to recontacting cases to assess behaviors after diagnosis and treatment, to an in-depth study involving medical record review to determine disease characteristics as well as treatments received, in addition to the collection of postdiagnostic epidemiological data. Each of these approaches entails determination of disease outcomes among the cases, which can be ascertained through a number of approaches. In general, follow-up is conducted by recontacting those who participated in the study to determine recurrence status; this approach requires foresight in the design of the original study, with permission in the study consent to recontact participants in the future and permission to obtain their future medical records. Deaths of cases can also be ascertained through checks on vital status through the state vital records and the National Death Index (NDI). Follow-up of cases from an etiological study that was not initially planned for conversion to a study of prognosis has a number of inherent weaknesses. Unless patients are recontacted at predetermined intervals to capture them at the same timepoint postdiagnosis, the questions that can be addressed are limited to behaviors prior to diagnosis in relation to treatment outcomes. Because there are gaps in understanding of potential lifestyle changes that patients can make to enhance their survival, a lack of data on the effects of postdiagnostic factors, such as diet, physical activity, weight gain, and supplements,
Table 1  Comparisons of Epidemiological Study Designs for Prognostic and Pharmacogenetic Studies Among Cancer Patients

<table>
<thead>
<tr>
<th>Population-based cohort study</th>
<th>Ancillary studies to clinical trials</th>
<th>Intervention studies or randomized controlled trials (secondary prevention)</th>
</tr>
</thead>
<tbody>
<tr>
<td>- General population</td>
<td>- Selected population</td>
<td>- Selected population</td>
</tr>
<tr>
<td>- &quot;Real-world&quot; treatment</td>
<td>- Selected treatment (often higher quality)</td>
<td>- Treatment of choice</td>
</tr>
<tr>
<td>- Heterogeneous treatment regimens</td>
<td>- Uniform treatment regimens</td>
<td>- Potentially heterogeneous treatment regimens</td>
</tr>
<tr>
<td>- Variable outcome assessment</td>
<td>- Excellent outcome assessment</td>
<td>- Uniform exposure (treatment) to prognostic variable that is tested</td>
</tr>
<tr>
<td>- Excellent assessment of health behaviors</td>
<td>- Limited assessment of health behaviors</td>
<td>- Excellent outcome assessment</td>
</tr>
<tr>
<td>- Single- or multicenter</td>
<td>- Multicenter</td>
<td>- Intervention or randomization on health behaviors</td>
</tr>
<tr>
<td>- Logistic challenges of multiple hospitals who participate in the prospective study, HIPAA regulations may differ</td>
<td>- Logistic challenges of many sites, cooperative group setting</td>
<td>- Single- or multicenter</td>
</tr>
<tr>
<td>- Prospective</td>
<td>- Advantage: already established collaborative setting</td>
<td>- Often small sample size</td>
</tr>
<tr>
<td>- Cannot establish causality</td>
<td>- Prospective or retrospective (for genetic testing)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>- Cannot establish causality</td>
<td></td>
</tr>
</tbody>
</table>

- Prospective
- Can establish causality (randomized controlled trials)
makes these types of studies less informative than those that obtain information from cases' postdiagnosis.

Of more use are etiological studies that are designed to conduct follow-up for recurrence and survival outcomes. With prognosis studies in mind, the cases can be consented at baseline for permission to recontact them, to obtain medical record information, and to retrieve their tissue blocks. It may be important to collect these data and biospecimens soon after enrollment into the study, because some Institutional Review Boards will not honor such consents for data and sample retrieval after a specified time period has elapsed. Additional queries regarding potential predictors of cancer prognosis, such as lifestyle factors, psychosocial factors, and complementary and alternative medicines, also should optimally be planned for implementation at specified timepoints after diagnosis. Of course, the obvious limitation for collection of these data in the context of an etiological study is funding. Although some aspects, such as permission to recontact, to review medical records, and to retrieve tissues, can easily be incorporated into an etiological study, the labor-intensive aspects of follow-up, recontact, and chart review can seldom be conducted within the context of a funded study of cancer risk. An additional complication in follow-up of cases ascertained in the context of a cohort study is the variable time from initial data collection among the healthy participants to cancer diagnosis, and then the variable times between diagnosis and follow-up assessments, unless there are resources to contact each case at the same specific timepoints postdiagnosis.

Prospective Observational Studies of Cancer Prognosis

Many of the limitations of conducting follow-up of cases in etiological studies can be overcome by the design of a prospective cohort study of cancer prognosis (36). In such a study, patients newly diagnosed with the incident, primary cancer of interest are ascertained and invited to participate. Ideally, cases will be enrolled, interviewed, and a blood specimen obtained prior to therapy for cancer. At enrollment, data can be collected on standard epidemiological factors prior to diagnosis and also on behaviors/characteristics at the time of diagnosis. Because the effects of some factors, such as folate or antioxidants, on treatment outcomes may be most dependent on their use during cancer therapy, the optimal study will collect data both at baseline and throughout cancer therapy, as well as at predetermined intervals throughout the follow-up period. In the context of a study specifically designed to evaluate the effects of behaviors and other factors during the postdiagnostic period, data can be rigorously collected that will likely provide important information for recommendations for cancer survivors to improve their prognosis. In an ongoing prospective observational study, data can also be collected on quality of life, psychosocial factors, and other variables that are not usually ascertained in epidemiological studies. With a prospective design, it will also be easier to collect extensive information on treatments received, including surgical procedures, chemotherapies, radiation therapies, and hormonal therapies for hormonally-related cancers. This type of study has the power to evaluate the effects of lifestyle factors on treatment outcomes, as well as gene-environment interactions, and also to examine the effects of epidemiological factors on outcomes in relation to specific cancer subtypes, determined through molecular characterization of the tumor. However, one limitation of this type of study is the heterogeneity of treatments received, which may be overcome through the implementation of a prospective follow-up study in the context of an ongoing clinical trial.
Studies Ancillary to Clinical Treatment Trials

There are many advantages to conducting prognostic follow-up studies in the context of a therapeutic clinical trial. For most studies, patients on the trial have more homogeneous disease characteristics, with eligibility criteria usually limited to subsets of disease characteristics, such as stage, grade, and nodal status. Because of the nature of the randomized clinical trial, initial chemotherapy regimens are consistent across each of the arms of the study, with all patients within an arm receiving the same drugs and dosages. Furthermore, endpoints are extremely reliable, with outcomes rigorously monitored for recurrence, disease-free progression, and survival, as well as toxicities experienced, usually using the NCI Common Toxicity Criteria or a similar standardized scale. All of these strengths reduce the number of sources of misclassification and minimize some sources of bias. As such, these studies may be quite advantageous for studies of the effects of pharmacogenetics on treatment outcomes, using DNA extracted from archived normal tissue, or for examining the role of tumor characteristics in cancer treatment outcomes. Currently, it is becoming more and more common for Cooperative Groups to collect and bank blood specimens in the context of clinical trials. These samples will provide an excellent source of DNA for pharmacogenetic studies, but the utility of serum may be somewhat limited due to the logistics associated with shipping blood samples from around the country. As pointed out in the chapter by Hankinson and Santella, variability in time to processing and differences in sample handling and shipping could introduce some systematic bias into subsequent studies.

One major limitation of conducting molecular epidemiological studies of cancer prognosis in the context of clinical trials is the lack of epidemiological and behavioral data on patients during and following treatment. However, this setting is ideal for the incorporation of questionnaires to assess diet, physical activity, supplement use, and other factors that may impact outcomes both during and following treatment. Recently, this has been initiated for specific studies in Cancer and Leukemia Group B (CALGB), resulting in findings of relationships between dietary patterns and colon cancer outcomes (37), and ongoing studies are underway in the Southwest Oncology Group. With comprehensive assessment of epidemiological factors during and following therapy, and banked blood specimens as well as tissues that can be accessed, such studies can provide excellent data on predictors of cancer outcomes.

A second major limitation is that the clinical trial setting does not reflect cancer care in “real life.” Only 3% of adult U.S. cancer patients currently participate in clinical trials. These are usually conducted at academic institutions with a greater expertise in cancer care and usually much better treatment facilities. Thus, studies in the context of clinical trials play an important role, yet they need to be complemented by research in a more community-based setting.

Intervention Studies or Randomized Controlled Trials for Secondary Prevention

These studies are uniquely suited to test hypotheses about a prognostic factor. In particular, randomized controlled trials are considered the gold standard and the only study design that can without doubt establish causality. Note that these trials are distinguished from those in the previous section in that they are not studies testing the efficacy or toxicity of cancer drugs themselves. Rather, they randomize cancer survivors to specific lifestyle activities or factors, such as physical activity and its ability to directly influence a prognostic outcome. For example, physical activity may reduce cachexia...
Molecular Epidemiological Designs for Prognosis

...in the context of more homogeneous subsets of disease of the nature of the drugs and dosages. Originally monitored for toxicities experienced, a standardized scale. All of and minimize some studies of the effects from archived normal treatment outcomes. Groups to collect and provide an in vitro serum may be samples from around the world. Variability in time could introduce some variability in time. With comprehensive patient data, and banked blood samples from around the world. Variability in time could introduce some variability in time.

Statistical Methodologies Used in Studies of Cancer Prognosis

Standard statistical tools may be applied to molecular epidemiological studies of cancer prognosis. Predictors of treatment toxicities may be assessed using standard methods for binary data (chi-square tests, logistic regression models), with a focus on specific toxicities (blood counts, cardiac, diarrhea, fatigue, febrile neutropenia, liver function, mucositis, nausea and vomiting, sensory neuropathy, and pain) or on all combined toxicities, usually grades 3 and 4. To adjust for other known prognostic factors, logistic regression models can be applied. To determine effects of predictors on recurrence and survival, standard time-to-event methods are usually used for the analysis of disease-free survival, such as log-rank tests, Cox models, and Kaplan-Meier estimates. Cox regression models are generally used to adjust for other known prognostic factors.

If the study is in the context of a therapeutic clinical trial, toxicities and/or disease-free survival may differ by treatment arm, and these differences may impact relationships between epidemiological factors, genetic polymorphisms, and treatment outcomes. Thus, careful analyses should be conducted to first determine if relationships differ by treatment arm, and, if so, treatment arm should be considered in the analysis.

In many studies, the effects of tumor characteristics, of genetic variability, and of epidemiological factors on treatment outcomes are examined singly, without consideration of the potential interactions among these factors. For a more comprehensive assessment of the molecular epidemiology of cancer prognosis, more sophisticated analytic techniques need to be implemented. Kattan first developed "nomograms," using primarily clinical data to model cancer outcomes (39,40), in which a patient's predicted probability of disease-specific survival is assumed to be a function of both the baseline hazard function shared by all patients and a linear combination of the individual patient's predictor variable values. It would be of interest to build nomograms incorporating genetic and epidemiological data, as well as clinical characteristics.

The use of Classification and Regression Trees (CART) analysis may be particularly useful for studying the combined effects of genetic polymorphisms, tumor
characteristics, and clinical and epidemiological factors on treatment outcomes. The model is fit using binary recursive partitioning whereby the data are successively split along coordinate axes of the predictor variables so that at any node, the split that maximally distinguishes the response variable in the left and the right branches is selected. Splitting continues until nodes are pure or data are too sparse; terminal nodes are called leaves, while the initial node is called the root. In practice, to avoid overfitting, typical decision tree systems then "prune" the tree to get a smaller tree that is nearly consistent with the data, though not necessarily completely consistent. Each leaf then makes the majority class prediction among data points that end at that leaf. Using approaches such as these may lead to a better understanding of the multiple factors that impact treatment outcomes among cancer patients.

PROGNOSTIC STUDIES—OPPORTUNITIES AND OBSTACLES

There is a growing interest in the epidemiological community to focus on the molecular epidemiology of cancer prognosis. To date, this has been a highly understudied area, and there are few good data on which lifestyle recommendations to cancer patients can be made. With efforts toward personalized medicine based on tumor characteristics and genetic profiles and application of molecular epidemiology to cancer prognosis, it is likely that this area will grow, and through multidisciplinary research, a better understanding of predictors of treatment outcomes will be had. However, there are numerous obstacles that the research community will have to confront, many of which are discussed above. Approaches to ascertaining and consenting patients, and reviewing medical records, will need to be compliant with the Health Insurance Portability and Accountability Act (HIPAA) and more stringent requirements from institutional review boards (IRBs), while still enabling research. As in studies of cancer risk, methods for assessment of exposures and behaviors need to be refined, and rigorous study design applied. Most importantly, researchers from multiple fields, as well as pathologists and clinicians, will need to communicate well with each other, so that novel approaches to studying the multiple factors impacting treatment outcomes can be developed, leading to an elucidation of the molecular epidemiology of cancer prognosis.

REFERENCES

Molecular Epidemiological Designs for Prognosis


