Specimen Banks for Cancer Prognostic Factor Research

Harry B. Burke, MD, PhD, Donald E. Henson, MD

• **Prognostic factors are necessary for determining whether a patient will require therapy, for selecting the optimal therapy, and for evaluating the effectiveness of the therapy chosen. Research in prognostic factors has been hampered by long waiting times and a paucity of outcomes. Specimen banks can solve these problems, but their implementation and use give rise to many important and complex issues. This paper presents an overview of some of the issues related to the use of specimen banks in prognostic factor research.**

(*Arch Pathol Lab Med.* **1998;122:871–874**)

Prognostic factors are important for assessing the natural history of cancer, for selecting the optimal therapy, and for evaluating the effectiveness of treatment.[1] Two issues are central to the evaluation of prognostic factors. The first concerns the time from diagnosis to the analysis of outcomes (eg, mortality). The longer this interval is, the longer the prediction time interval becomes. To provide, for example, 10-year survival predictions, a patient population must be followed for 10 years. The 10-year information is used to assess the predictive accuracy of a prognostic factor and to provide 10-year outcome predictions to future patients. The second issue is the accrual of a sufficient number of outcomes so that the assessment of the factor is statistically reliable. Reliable means that a similar result would be observed if the analysis were repeated.

A human specimen bank that contains abnormal and normal tissue, white cells, serum, and plasma facilitates prognostic factor research because it eliminates the waiting time problem and the outcome accrual problem by collecting specimens from a defined patient population and following the patient population for a sufficient number of years. When a new putative prognostic factor is discovered, the stored material can be used to assess its predictive power.

Because specimen banks have all the difficulties of traditional data analysis as well as new difficulties, the development of a well-designed and useful specimen bank presents formidable challenges, especially for prognostic factor research. Furthermore, because specimen banks wait in silence for future use, an initial error may not become apparent for many years.

Although specimen banks will prove extremely useful, they do not solve the problem of how to validate a putative prognostic factor for a new therapy, that is, for a therapy that is not represented in the specimen bank population, nor do they solve the problem of the absence of agreed upon methods for validating prognostic factors and a consequent inability to replicate results.

Specimen banks are most commonly created for cancers whose initial therapy is surgical resection of the tumor. Specimen banks often collect more than the primary tumor. At surgery, blood, adjacent "normal" tissue, and metastatic tissue may also be collected. Although difficult and sometimes not feasible, blood should be collected after surgical therapy at regular intervals over many years and, if possible, tissue should be collected at recurrence, including from metastases, to assess a prognostic factor's predictive ability over time.

Clinical follow-up information and status, which are critical data for prognostic factor research, should be collected regularly, and investigators who have used the specimen bank should regularly update their data sets. A mechanism should exist for tracking patients who change physicians, move, or who for any other reason are lost to follow-up. Specimen banks for relatively slow-growing tumors (eg, breast and prostate cancer) and for rare tumors should be maintained for at least 20 years and preferably longer. Computer-based databases should be created and maintained for the life of the specimen bank.

A specimen bank must provide investigators with the relevant information regarding its data and tissue so that investigators can decide whether the tissue is appropriate for their task. Because tissue is to be distributed to many investigators for different purposes, the specimen bank's collection and reporting methods are critical. The better the specimen collection and reporting methods are, the more effectively the specimen bank can be used.

What follows is an overview of some of the important issues related to the use of specimen banks for prognostic factor research. Many of these issues are difficult to deal with and expensive to solve, and some may be insoluble, but they must be explicitly recognized by those creating and maintaining specimen banks.

MULTI-INSTITUTION COLLECTION

It has been the common practice of cancer investigators to collect and store human tissue for their own research.

Accepted for publication June 4, 1998.

From Bioinformatics and Health Services Research, Department of Medicine, New York Medical College, Valhalla (Dr Burke), and the Early Detection Branch, Division of Cancer Prevention, National Cancer Institute, Rockville (Dr Henson).

Reprints: Harry B. Burke, MD, PhD, Bioinformatics and Health Services Research, Department of Medicine, New York Medical College, Valhalla, NY 10595.

The collecting investigator usually knows the characteristics of the patient population (including demographic and clinical data, treatment, and outcome) and of the tissue acquisition, storage, and retrieval process, and therefore the strengths and limitations of the specimen bank. To hasten the evaluation of molecular genetic putative prognostic factors, granting agencies have recently been providing funds for the coordinated collection and storage of human tissue from large numbers of patients.[2,3] For breast cancer alone there are currently at least 57 specimen banks.[4] Because of the number of patients required, these specimen acquisition efforts usually involve multiple investigators and multiple institutions.

An essential difference between multi-investigator, multi-institutional specimen banks and individual investigator specimen banks is that the investigators using multi-institutional specimen banks must rely on patient information and tissue supplied by the specimen bank. These investigators need detailed information regarding the patient selection criteria; demographic and clinical variable definitions; data acquisition methods; data coding system; tissue acquisition technique; and tissue preparation, storage, and retrieval methodology. Effective use of the tissue depends critically on the ability of the specimen bank to acquire, organize, and disseminate this information accurately and in a timely manner and, where possible, to standardize the patient acquisition, tissue collection, and clinical reporting process across collecting institutions to reduce interinstitutional variance.

A central issue for multi-institutional specimen banks is uniformity of methods. Surgical technique, specimen handling, tissue preparation, and quality assurance should be as uniform as possible across physicians and institutions. Some degree of standardization can be achieved by standardized procedures and training programs. Most cancer prognostic factors, including new molecular-genetic factors, are relatively weak predictors.[5] Therefore, the less uniformity that exists among institutions, the greater the loss of prognostic power and the greater the likelihood is that a strong factor will become weak and that a factor that is important for a small number of patients will be lost. It is through the combination of relatively weak factors in prognostic models that accurate predictions become possible.[1]

For multi-institutional data, the agreement among institutions in terms of clinical and laboratory variables and tissue characteristics should be assessed and reported to investigators. For continuous variables (eg, tumor size), a measure of central tendency, such as the mean (if the variable is normally distributed), and its variance should be calculated and the statistical differences assessed. For categorical variables (eg, race), the χ^2 statistic can be calculated, and for ratings, Cohen's κ statistic[6] can be calculated. If there is high interinstitutional variance, an investigator may, depending on the research issue and study design, restrict the source of tissue to one institution or to certain variables.

PATIENT SELECTION

The representativeness of a population of cancer patients is of vital importance because it determines the generalizability of the research results. For this reason, patient representativeness information should be provided to investigators prior to the use of the tissue. An unbiased patient selection process is necessary for population representativeness, since a bias in patient selection may yield results that are not generalizable. Sometimes a bias is minor and can be ignored, sometimes it is major but can be dealt with in terms of a reasonable assumption, and sometimes it is major and cannot be dealt with. In the last case, the investigator may wish to explore other more representative specimen banks. For example, tissue could be collected from a special group of women, and the frequency of *BRCA1* mutations could be determined and related to the incidence of breast cancer in that special population. It cannot be assumed, however, that the same quantitative relationship will hold true for all populations of women.

Population representativeness depends on many factors. One factor is the recruitment of patients. For example, clinical trials rarely provide a representative population of cancer patients because their entry criteria usually exclude certain patient groups. Thus, clinical trial populations are usually a biased sample because their entry criteria operate as a selection bias mechanism. This mechanism can limit the generalizability of the prognostic factor results. If an investigator expects the prognostic factor results to apply only to the patients that met the clinical trial's entry criteria, that is, to be in a position to perform only conditional prediction tasks, then the analysis may proceed. If the investigator expects the study results to apply to all the patients, however, additional assessment should be performed to determine if this generalization is reasonable given the study being contemplated. For example, for a study performed in a single institution, the investigator can determine whether the institution's patient population is representative of a more general patient population by comparing the characteristics of the institution's patient population with those of the larger patient population.

Patient populations may be biased by the method used for identifying incident cases. For example, some collection methods miss, in a nonrandom manner, approximately 18% of incident cases.[7] Patient populations may be biased by the clinical setting in which the cases are detected. For example, it is known that there are differences in the TNM stage frequencies reported by different types of hospitals.[8] Patient populations may be biased by where the patients are treated. Oncology clinic populations may differ from hospital populations. For example, in situ cancers may be more common in oncology clinics than in hospitals.[7] Patient populations may be biased by not distinguishing between incident and prevalent cases because prevalence depends on survival.[7] Patient populations may be biased by the incomplete collection of representative patient data. For example, incompleteness is an issue in data sets that have not existed for at least 40 to 50 years because the sample is not representative of the full spectrum of prevalent cases.[9] Finally, definitions may change over time. For example, changes in the TNM variable definitions[10–14] make it difficult to compare outcomes in terms of extent of disease.[15]

VARIABLE DEFINITIONS AND CODING

It is a nontrivial task to standardize the definition and coding of the clinical variables essential for prognostic factor research. Tumor registrars, for example, continue to refine the definitions and coding of the commonly collected variables. Prior to any data or tissue collection, representatives from each collecting institution should agree on a list of variables to be collected and create explicit definitions and a coding system for each variable. As the

specimens are collected and stored, an active quality assurance process should be in place at each institution. The quality assurance process will determine whether the definitions and coding system are being followed, and if not, whether to increase institutional vigilance or change the definitions and coding so that they can be adhered to.

DATA ACQUISITION METHODS

There are 3 data acquisition problems. The least difficult is the veracity of the original data source and the data entry. These errors include either incorrect entry of patient data in the patient's medical record or an error by the data collector. Data entry errors can be minimized through proper quality control. Two more difficult problems are missing data, that is, unknown variable values, and patients lost to follow-up, a type of censoring.

If a data set contains nonrandomly missing variable data,[16] then the missing data mechanism should be explicitly considered.[17] For example, in the National Cancer Data Base data set for the year 1983 with follow-up until 1990, of the 19147 cases listed, only 10357 are TNM staged (H.B.B., unpublished data, 1997). An investigator using tissue from these cases would have to ascertain whether the unstaged cases are the "same" as the staged cases. The missing values for some variables should be reacquired, while those of other variables should be filled in, depending on the importance of the variable. Basic information that should be included in the data set, such as therapy, can be easily lost,[7] and that type of data should be recaptured. If missing data are to be filled in, Little and Rubin's[17] book should be consulted.

Censoring (eg, patients lost to follow-up) is of critical importance to investigators because the primary purpose of the specimen bank is to provide outcome information that is useful for prognostic factor research. The 8-year censoring rate in the 1983 National Cancer Data Base data set is more than 25% (unpublished data). Censored cases usually have a less favorable outcome than noncensored cases.[7,18] Consequently, investigators should be informed of the specimen bank's censoring rate. If it is high (eg, greater than the 5% observed in the Surveillance, Epidemiology, and End Results [SEER] Program[19]), measures should be taken to systematically recapture these patients or the censoring mechanism should be explicitly considered. In addition, all patient variables should be compared with existing national databases. For example, the SEER Program reports an 84% 5-year survival rate,[19] and the National Cancer Data Base reports a similar 80% 5-year survival rate for patients with breast cancer (unpublished data). A survival rate in this range should be observed in breast cancer specimen banks. Using the existence of outcome data as a criterion for tissue selection by an investigator would not be appropriate if the patients who were lost to follow-up were lost because of their disease. In other words, if patients were lost because they were too sick, then selecting only those patients who were followed could bias a prognostic factor analysis.

It should be noted that there can be problems with survival estimation if the ascertainment of vital status is not random.[20] Finally, using cancer-specific survival as an outcome may be problematic because it assumes (1) that physicians accurately code death certificates and (2) that every patient dies of a single identifiable disease that can be correctly ascertained without autopsy. These assumptions are doubtful in the real world of competing risks.[21]

TISSUE ACQUISITION, PREPARATION, STORAGE, AND RETRIEVAL

In addition to clinical data collection, there are important issues related to tissue acquisition, preparation, storage, and retrieval. The manner in which the tissue has been collected, classified, stored, and retrieved may limit the types of studies that can be conducted or the "yield" of the tissue, or may affect the results of the study.[22-24] For each case, a sufficient amount of tumor should be available to investigators. If appropriate and possible, matching nonneoplastic tissue should accompany the tumor specimen. Because cellular degradation begins soon after removal, specimens should be kept as cool as possible and processed rapidly.

The 2 most common types of tissue processing are freezing and formalin/paraffin. Each has advantages and disadvantages for molecular genetic analysis. Ideally, tissue should be snap-frozen in liquid nitrogen as soon as possible after excision and should be stored at −70°C.[25] Delays in freezing or inadequate freezing can result in artifactual genetic changes. Some analyses can only be performed on frozen tissue. For example, if the concordance between abnormalities detected by immunohistochemistry and cDNA is important, then, depending on the antigen, paraffin-embedded, formalin-fixed tissue may not be useful.[26] Frozen tissue allows the polymerase chain reaction to be carried out on long stretches of DNA (1000 base pairs), since the DNA remains intact. Unfortunately, frozen tissue is difficult to handle, expensive to store, hard to provide in small amounts, and difficult to distribute to multiple investigators.[27] Procedures for processing tissue for molecular pathology have been published.[28]

For routine pathologic examination, tissues are usually fixed in formalin, dehydrated, and embedded in paraffin. Formalin forms cross-links between the reactive amine groups on adjacent proteins and between DNA and proteins, which makes the DNA rigid. Because of its rigid structure, subsequent tissue processing may cause fragmentation of the DNA. With conventional 10% formalin fixation, numerous variables are involved, including the duration of fixation, days in fixation that may result in DNA fragmentation by nucleases, size of the tissue, fixation gradients, and pH. These variables are difficult to standardize across institutions.[29]

Polymerase chain reaction amplification studies may be performed on archival formalin-fixed tissue even after years of storage.[30-32] In contrast to the situation with frozen tissue,[33] the restriction fragments will be relatively small, usually 100 to 300 base pairs in length. Cross-linking of the DNA can interfere with hybridization.[22] However, in spite of these effects of fixation, experience indicates that archived tissue is an invaluable resource for research in molecular genetics. Investigators have reported successful amplification of DNA extracted from 40-year-old specimens.[34] Fixatives that contain heavy metals, such as zinc or mercury (Zenker's fixative), will destroy DNA and are not suitable for tissue that is destined for DNA analysis. There is evidence that fixation in nonbuffered formalin may also degrade DNA.[35] The labile nature of RNA makes it difficult to recover in formalin-fixed tissue. Therefore, investigators interested in mRNA should use specimen banks that have the relevant cDNA library or that contain frozen tissue.

Highly suitable for tissue banks, immunohistochemical

methods can be used to detect the presence of specific gene products in tissue sections. For some antigens that deteriorate, the proper storage of tissues for subsequent immunohistochemistry is critical. Inappropriate fixation and tissue processing can affect the results as much as the variation in the antibody.[29] Fixed tissue may be stored as unstained cut sections mounted on glass slides or in paraffin blocks. Fixed tissues cut into sections, mounted on slides, and stored unstained may not always be suitable for immunohistochemistry. It has been shown, for example, that immunostaining intensity for p53 and other antigens will decay over time if the sections are cut from paraffin-embedded tissue and stored unstained on glass slides.[36,37] On the other hand, antigens usually do not decay if tissues are maintained in paraffin blocks. There is no decline in p53 staining intensity in tissues stored in paraffin blocks for more than 13 years.[38]

COMMENT

Prognostic factors are necessary for determining whether a patient will require therapy, for selecting the optimal therapy, and for evaluating the effectiveness of the therapy. Research in prognostic factors has been hampered by long waiting times and a paucity of outcomes. Although the implementation and use of specimen banks give rise to many important and complex issues, including the possibility of population biases and problems related to specimen handling, storage, and retrieval, specimen banks, because they solve the problems of prediction time and reliability, are a major advance in the field of prognostic factor research.

This work was supported in part by a research grant from the US Army Medical Research and Development Command Breast Cancer Research Program (DAMD 17-94-J-4383). We thank William E. Grizzle, MD, PhD, and Sheila Taube, PhD, for their helpful comments.

References

1. Burke HB. Increasing the power of surrogate endpoint biomarkers: the aggregation of predictive factors. *J Cell Biochem.* 1994;19S:278–282.
2. Clausen KP, Grizzle WE, LiVolsi VA, Newton WA, Aamodt R. The cooperative human tissue network. *Cancer.* 1989;63:1452–1455.
3. LiVolsi VA, Clausen KP, Grizzle W, Newton W, Pretlow TG, Aamodt R. The cooperative human tissue network: an update. *Cancer.* 1993;71:1391–1394.
4. Anton-Culver H, Bringman DA, Taylor TH, Kim S. *National Action Plan on Breast Cancer: National Biological-Resource Banks Working Group Final Report, May, 1997.* Irvine, Calif: Epidemiology Division, Department of Medicine, University of California, Irvine; 1997.
5. Burke HB, Hoang A, Iglehart JD, Marks JR. Predicting response to adjuvant and radiation therapy in early stage breast cancer. *Cancer.* 1998;82:874–877.
6. Cohen J. A coefficient of agreement for nominal scales. *Educ Psyol Measurement.* 1960;20:37–46.
7. McClish DK, Penberthy L, Whittemore M, et al. Ability of Medicare claims data and cancer registries to identify cancer cases and treatment. *Am J Epidemiol.* 1997;145:227–233.
8. Steele GD, Winchester DP, Menck HR, Murphy GP. *National Cancer Data Base Annual Review of Patient Care 1993.* Chicago, Ill: Commission on Cancer; 1993:12.
9. Capocaccia R, DeAngelis R. Estimating the completeness of prevalence based on cancer registry data. *Stat Med.* 1997;16:425–440.
10. *Manual for Staging of Cancer, 1977.* Chicago, Ill: American Joint Committee for Cancer Staging and End-Results Reporting; 1977.
11. Beahrs OH, Myers MH, eds. *Manual for Staging of Cancer.* 2nd ed. Philadelphia, Pa: JB Lippincott; 1983.
12. Beahrs OH, Henson DE, Hutter RVP, Myers MH, eds. *Manual for Staging of Cancer.* 3rd ed. Philadelphia, Pa: JB Lippincott; 1988.
13. Beahrs OH, Henson DE, Hutter RVP, Kennedy BJ, eds. *Manual for Staging of Cancer.* 4th ed. Philadelphia, Pa: JB Lippincott; 1992.
14. Fleming ID, Cooper JS, Henson DE, eds. *AJCC Cancer Staging Manual.* 5th ed. Philadelphia, Pa: JB Lippincott-Raven; 1997.
15. Henson DE, Ries L, Shambaugh EM. Survival results depend on the staging system. *Semin Surg Oncol.* 1992;8:57–61.
16. Choi SC, Lu IL. Effect of non-random missing data mechanisms in clinical trials. *Stat Med.* 1995;14:2675–2684.
17. Little RJA, Rubin DB. *Statistical Analysis With Missing Data.* New York, NY: John Wiley & Sons; 1987.
18. Elfstron J, Stubberod A, Troeng T. Patients not included in medical audit have a worse outcome than those included. *Int J Quality Health Care.* 1996;8:153–157.
19. Ries LAG, Kosary CL, Hankey BF, Miller BA, Harras A, Edwards BK, eds. *SEER Cancer Statistics Review, 1973–1994.* Bethesda, Md: National Cancer Institute; 1997.
20. Hu P, Tsiatis AA. Estimating the survival distribution when ascertainment of vital status is subject to delay. *Biometrika.* 1996;83:371–380.
21. Ebrahimi N. The effects of misclassification of the actual cause of death in competing risks analysis. *Stat Med.* 1996;15:1557–1566.
22. Moerkerk PT, Kessles HJ, ten Kate J, de Goeji AF, Bosman FT. Southern and dot blot analysis of DNA from formalin-fixed, paraffin-embedded tissue samples from colonic carcinomas. *Virchows Arch B Pathol Mol Pathol.* 1990;58:351–355.
23. Floss RD, Guha-Thakurta N, Conran RM, Gutman P. Effects of fixative and fixation time on the extraction and polymerase chain reaction amplification of RNA from paraffin-embedded tissue: comparison of two housekeeping gene mRNA controls. *Diagn Mol Pathol.* 1994;3:148–155.
24. Arnold MM, Srivastava S, Fredenburgh J, Stockard CR, Myers RB, Grizzle WE. Effects of fixation and tissue processing on immunohistochemical demonstration of specific antigens. *Biotech Histochem.* 1996;71:224–230.
25. Farkas DH, Kaul KL, Wiedbrauk DL, Kiechle FL. Specimen collection and storage for diagnostic molecular pathology investigation. *Arch Pathol Lab Med.* 1996;120:591–596.
26. Sjogren S, Inganas M, Norberg T, et al: The p53 gene in breast cancer: prognostic value of complementary DNA sequencing versus immunohistochemistry. *J Natl Cancer Inst.* 1996;88:173–182.
27. Naber SP. Continuing role of a frozen-tissue bank in molecular pathology. *Diagn Mol Pathol.* 1996;5:253–259.
28. Kiechle FL, Chambers LM, Cox RS Jr, et al: *Reference Guide for Diagnostic Molecular Pathology/Flow Cytometry.* Fascicle VII. Northfield, Ill: College of American Pathologists; 1996.
29. Grizzle WE, Myers RB, Manne U, Stockard CR, Harkins LE, Srivastava S. Factors affecting immunohistochemical evaluation of biomarker expression in neoplasia. In: Hanausek M, Walaszek Z, eds. *Methods in Molecular Medicine: Tumor Marker Protocols.* Totowa, NJ: Humana Press; 1998.
30. Pavelic J, Gall-Troselj K, Bosnar MH, Kardum MM, Pavelic K. PCR amplification of DNA from archival specimens: a methodological approach. *Neoplasma.* 1996;43:75–81.
31. Shibata D. Extraction of DNA from paraffin-embedded tissue for analysis by polymerase chain reaction: new tricks from an old friend. *Hum Pathol.* 1994;25:561–563.
32. Mies C. Molecular biological analysis of paraffin-embedded tissues. *Hum Pathol.* 1994;25:555–560.
33. Farkas DH, Drevon AM, Kiechle FL, et al. Specimen stability for DNA-based diagnostic testing. *Diagn Mol Pathol.* 1996;5:227–235.
34. Shibata D, Martin WJ, Arhneim N. Analysis of DNA sequences in forty-year-old paraffin-embedded thin-tissue sections: a bridge between molecular biology and classical histology. *Cancer Res.* 1988;48:4564–4566.
35. Goelz SE, Hamilton SR, Vogelstein B. Purification of DNA from formaldehyde fixed and paraffin embedded human tissue. *Biochem Biophys Res Commun.* 1985;130:118–126.
36. Jacobs TW, Prioleau JE, Stillman IE, Schnitt SJ. Loss tumor marker immunostaining intensity on stored paraffin slides of breast cancer. *J Natl Cancer Inst.* 1996;88:1054–1059.
37. Shin HJC, Kalapurakal SK, Lee JJ, Ro JY, Hong WK, Lee JS. Comparison of p53 immunoreactivity in fresh-cut versus stored slide with and without microwave heating. *Mod Pathol.* 1997;10:224–230.
38. Manne U, Myers RB, Srivastava S, Grizzle WE. Stability of p53 and Bcl-2 antigens in paraffin blocks. *J Natl Cancer Inst.* 1997;89:585–586.