Standard Operating Procedures for MECC Registries

Second Edition

December 2008

Edited by:

Hoda Anton-Culver Ph.D.
Professor and Chair, Department of Epidemiology
University of California Irvine
224 Irvine Hall
Irvine, California, 92697 USA

John L. Young, Jr. DrPH, CTR
Professor of Epidemiology
School of Medicine, Rollins School of Public Health, Emory University
1462 Clifton Road,
NE Atlanta, Georgia 30322 USA

The Second Edition has been discussed and amended by the directors of the MECC Registries at the MECC Steering Committee Meeting of October, 2008.

INTRODUCTION

Establishment of these standards is of major importance in enhancing the usefulness of cancer registry data. Collaborative studies and data comparisons are more feasible as data become more directly comparable. Electronic sharing of cancer data, edits, data recode routines, analysis software, documentation, and training aids are all increasingly possible once standards are established and adopted.

Cancer Registries are intended to provide a base for epidemiological research and case-control studies. Some registries emphasize cancer control and patient management considerations, and some focus on end results and survival.

Strictness: The standards presented vary in how strongly they are recommended. There are three levels of strictness to the standards:

- **MUST**s: There are certain registry characteristics that are necessary for the effective and efficient operation of a cancer registry. These are identified as **MUST**s in the standards. Although some existing registries may be able to function without these characteristics, it is the present consensus that any new registry should adopt these standards.
- **SHOULD**s: There are other characteristics that experience has shown are strongly recommended, but perhaps not as absolutely required as the **MUST**s above. These are designated as **SHOULD**s in the standards. Some of the problems addressed by the **SHOULD**s can be solved in alternate ways depending on local conditions, needs, and resources.
- **MAY**s: There are yet other characteristics that are highly desirable but not necessary. These are designated as **MAY**s.

I. ACCESS TO SOURCE DATA AND COMPLETENESS OF REPORTING

A. STRUCTURAL REQUIREMENTS

1. Legislation and Regulations

a) Definitions

Legislation: Legislation or statute refers to a form of law enacted by a government.

Regulation: Regulation or rule refers to a form of law created by administrative agencies of a government.

b) Introduction

The authority of a population-based cancer registry to collect data on cancer incidence is established through legislation for cancer reporting with or without regulations; and/or regulations/rules developed under general authorization for reporting non communicable disease, as specified health authorities.

Legislative authority should include specific components related to registry development and function, as well as specific directives for promulgation of regulations detailing these components. Often the authority is granted to the jurisdiction's health department, which in turn may delegate the authority to another agency. In other instances, the authority is granted directly to another agency, such as a university.

The guidelines of MECC include: reporting requirements, patient record access, data quality and data standards, confidentiality and disclosure of data.

c) Standards

(1) Standards for Reporting Requirements

General Considerations: Government legislation or a Ministerial decree **MUST** authorize a population-based registry. Legislation or decree **SHOULD** define reportable cancers, a reference date for registry operation, residency requirements of cases, who has the authority and responsibility for implementing and maintaining the database, who is responsible for reporting the data (i.e., physicians, hospitals, pathology laboratories, etc.), what geographic area is to be covered, timeliness of reporting, the type and format of data to be reported, to whom and under what circumstances the registry has authority to release the data, and it should address penalties for noncompliance.

Required Components of the Legislation and/or Ministerial Decree Regarding Reporting Requirements:

- Definitions: All terminology used in the text of the law or decree **MUST** clearly be defined.
- "Cancer" **SHOULD** include all neoplasms in the most recent edition of the International Classification of Disease for Oncology (ICD-O), with a behavior code of 2 or 3 (*in situ* or malignant). Possible exceptions to this **MAY** include many basal and squamous cell carcinomas of the skin and *in situ* carcinoma of the cervix uteri. Registries **MAY** collect benign tumors and this should be defined.
- "Reference date" refers to the effective date coverage starts in a specified population at risk. It is not the date the registry is organized or actually performs the work. Cases diagnosed on or after the reference date **MUST** be included. The reference date **SHOULD** be January 1 of a calendar year, but **MAY** be another date.
- All cancers occurring in the geographic region covered by the registry **SHOULD** be reportable. To allow for sharing of cases with

other population-based registries, to facilitate death clearance and other record linkages, and to allow for the preparation of reports to individual facilities that include all their cases, the registry **SHOULD** include all residents and nonresidents.

- The registry **MUST** be population-based. To assure maximum coverage of the designated population, cancers **SHOULD** be reported by, or case information obtained from, all hospitals or other facilities providing screening, diagnostic, or therapeutic services to patients with respect to cancer; and from physicians, surgeons, and all other health care providers diagnosing or providing treatment for cancer patients, except for cases directly referred to or previously admitted to a hospital or other facility providing screening, diagnostic, or therapeutic services to patients and reported by those facilities.
- The legislation or decree **SHOULD** state that cancer cases **SHOULD** be reported to the registry no more than 180 days from the date of admission or diagnosis in a format prescribed by the registry (the format itself **SHOULD** be incorporated by reference).

(2) Standards for Patient Record Access

General Considerations: The legislation and/or decree **SHOULD** provide for access to records of health care providers and facilities that would identify cases of cancer or would establish characteristics of the cancer, treatment of the cancer, or medical status of any identified cancer case by authorized representatives of the cancer registry. This access is necessary for meeting both initial reporting requirements and subsequent quality assurance activities.

(3) Standards for Data Quality

The legislation and/or decree **MUST** provide that data reported to the registry **SHOULD** meet standards of completeness timeliness and quality. Rules **SHOULD** follow standards set by the MECC and **SHOULD** provide for data quality audits conducted at reporting facilities by the cancer registry.

(4) Standards for Confidentiality and Disclosure of Data

General Considerations: The legislation and/or decree **MUST** specify the confidential nature of the database and provide for the confidentiality of all cancer patient data. The confidentiality directives of the legislation and/or decree **SHOULD** address how the data are to be released, to whom, and for what purpose. The legislation and/or regulations **SHOULD** state that aggregate data **SHOULD** be available to the public through published reports or through data access policies, but that access to confidential data or "raw data" is restricted. The guidelines **SHOULD** NOT be so strict that approved researchers are denied access to the raw data.

Components of the Legislation and/or Decree Regarding Confidentiality and Disclosure of Data:

- All data reported to the cancer registry **SHOULD** be available for use in aggregate form by cancer registry staff and authorized researchers for analyses and reports of the incidence, prevalence, management, survival, and risk factors associated with the registry's cancer experience.
- The cancer registry **MAY** exchange patient-specific data with the reporting facility, any other cancer-control agency, or clinical facility for the purpose of obtaining information necessary to complete a case record, provided these agencies and facilities comply with the cancer registry's confidentiality policies.
- For the purpose of complete case ascertainment, the cancer registry **MAY** exchange patient-specific data with other cancer registries if reciprocal data-sharing agreements that include confidentiality provisions are implemented.

- The cancer registry **MAY** grant researchers access to confidential information concerning individual cancer patients, provided those researchers comply with the cancer registry's

confidentiality policies and have the approval of the registry Institutional Review Board (IRB).

- Willful violation of confidentiality provisions **SHOULD** be punishable.

2. Reportability Definitions

a) Introduction

Precise definitions of cases that are reportable to the registry **MUST** be developed and promulgated.

b) Standards

(1) Standards for Reportable Diagnoses

The registry's reportable list **SHOULD** make reference to the International Classification of Diseases for Oncology. At a minimum, all neoplasms with a behavior code of 2 or 3 in ICD-O **SHOULD** be designated reportable. Any benign neoplasms or neoplasms of uncertain behavior that are reportable **SHOULD** clearly be identified with reference to their ICD-O codes (See Reportability in Appendix A).

(2) Standards for Multiple Primary Rules

To compare cancer rates for two registries, it is important that identical rules have been used for counting multiple tumors in the patient–whether in the same organ, opposite sides of paired organs, different sub sites, or different sites, and whether at the same or different times. Rules for determining multiple primaries in MECC registries are provided by IARC and IACR (see the MECC Coding Manual)

(3) Standards for Diagnostic Confirmation

To obtain complete incidence reporting and to have the registry's data accurately reflect the burden of cancer in the population at risk,

clinically diagnosed cases as well as microscopically confirmed cases **MUST** be designated as reportable. Microscopically confirmed cases include all cases with positive histopathology, including examinations of bone marrow and peripheral blood; and all cases with positive cytopathology, including peritoneal or pleural fluid, fine needle aspirations of cells, and bronchial washes. Clinically diagnosed cases include those without microscopic confirmation (i.e., those whose diagnoses are based only on diagnostic imaging, laboratory tests, or other clinical examinations).

(4) Standards for Ambiguous Terminology

Diagnoses and descriptions of patients' conditions often are described in the medical record with ambiguous terms such as "possible" and "rule out." For comparability, the registry **MUST** adopt rules for interpreting ambiguous terms. These rules are to be included in the MECC Coding Manual.

(5) Standards for Distribution

A copy of the reportable list and other rules **SHOULD** be provided to all reporting facilities or practitioners required to report; to all cancer registrars in the coverage area; to all medical records or cancer registrar training programs.

B. PROCESS STANDARDS

1. Participation of All Hospitals

a) Introduction

Participation of all hospitals in the reporting area that may diagnose or treat cancer is essential to ensure completeness of reporting.

b) General Standards

The registry **SHOULD** gain access to 100 percent of the hospitals that may diagnose or treat patients residing in the reporting area to ensure completeness. Letters of agreement **MAY** be useful for both the hospital and the registry. These letters **SHOULD** specify the responsibilities of the hospital, the responsibilities of the registry, and the timeframe for reporting.

2. Case Ascertainment in Non-hospital Sources

a) Introduction

Cancer patients can be seen for diagnosis or treatment in outpatient settings. Capturing these cases through an extended reporting system is important to ensure the completeness of cancer registration. Cancer registries **SHOULD** expand their coverage into non-hospital sources to facilitate complete reporting.

b) Standards

The registry **MUST** develop mechanisms to locate and obtain information on cases diagnosed or treated entirely outside of hospital settings. The usefulness of specific sources will vary across geographic areas and over time. However, experience has shown that at a minimum, the registry **SHOULD** obtain cases from the following types of facilities:

Histopathology and hematology laboratories.
Ambulatory surgery centers.
Radiation therapy or chemotherapy centers.
Screening programs.

Although cancer cases **MAY** be identified in pathology laboratories, the laboratory records often contain insufficient information for preparing a complete abstract. Information on the

patient's residence and/or health insurance number, for example, rarely is present. These cases **MUST** be followed back to the treating physician or facility for additional information (see also Section I.B.3.) if needed. The expansion of case ascertainment procedures into all types of nonhospital facilities would ensure complete reporting; however, the registry's ability to do so **MAY** be limited by its financial resources. Therefore, the registry **SHOULD** consider the following items when evaluating the expansion of case finding into non-hospital facilities such as chemotherapy treatment facilities, coroner's offices, private clinics, nursing homes, and hospices:

The cost of accessing each type of facility will depend on:

- The reporting law and which types of facilities and practitioners are required to report.
- The quality of the data and the number of new incidence cases that would be obtained from each type of facility.
- The impact on the future use of the data if a decision is made not to collect data from a specific type of facility.
- The impact of these requirements on each type of facility.

3. Reporting by Physicians

a) Introduction

Because not all persons diagnosed with cancer are hospitalized for diagnosis or treatment, a mechanism for registering cases from physicians' offices is necessary for complete case ascertainment. The registry **MAY** rely on reporting by physicians, or **MAY** have its own staff obtain the data from physicians' offices. The registry generally will require case information from an individual physician only when no report is obtained from a hospital or other reporting facility. However, the registry also might need to obtain demographic or treatment information on cases reported initially by other sources.

b) Standards

The registry **SHOULD** perform the following:

• Follow-back to physicians' offices to obtain reports on otherwise unreported cases identified in pathology laboratories, via consult-only reports from hospitals, or from death certificates.

• Develop an appropriate method to identify cases of and obtain information from oncologists, hematologists, dermatologists, dermatopathologists, and urologists, because these specialties are the ones most likely to diagnose malignancies that will not be identified through the active case finding methods used at hospitals and laboratories.

• Develop registration methods for physicians.

4. Out-of-area Coverage and Case Sharing

a) General Standards

The registry **SHOULD** include all residents and nonresidents diagnosed or treated in its coverage area to allow for sharing of cases with other population-based registries, facilitate death clearance and other record linkages, and allow for preparation of reports to individual facilities that include all their cases.

The registry **SHOULD** provide case information on a nonresident to the population-based registry covering the patient's place of residence when the required components listed below are in place.

b) Required Components

The following components generally will be required for the performance of case sharing between registries:

Case Sharing Agreements: These are written agreements between registries covering the usage and confidentiality of exchanged data. These **MAY** be informal agreements simply requesting data and

affirming the confidential nature of the data or they may be longer, more formal legal documents, depending upon the laws governing release of data.

Exchange Media: Data **MAY** be exchanged between registries across a variety of media. In order of preference, they are: electronic files of data on diskette, CD Rom, or tape; electronic files of data transferred via e-mail, or Web site; copies of paper abstracts; or printed reports generated from computer systems.

5. Informing Facilities and Practitioners

a) Introduction

To encourage compliance with cancer reporting requirements, the registry **SHOULD** notify facilities and practitioners that are required to report of their obligations.

b) Standards

Content of Notices

The notification **SHOULD** include:

- A brief description of the registry's history and purpose
- A description of and copy of the cancer reporting law
- The rationale for the registry's access to the source data
- The data items to be collected
- The procedures for reporting
- All relevant considerations for data handling and ensuring confidentiality.

Other Considerations:

- Support of the registry and its reporting methods from appropriate groups **MAY** be sought. Examples include medical societies, specialty colleges or boards, community groups. Citing such support or endorsements in the various

communications to medical professionals may encourage their compliance.

- Announcements **MAY** be made via professional organizations or societies regarding their members' cancer-reporting responsibilities. Mechanisms **MAY** include newsletters, direct mailings, journal articles, and presentations at scheduled meetings.
- In addition, the exact details of all expectations of and options available to the facilities and practitioners **SHOULD** be communicated via targeted contacts.

Means for accomplishing these steps include:

- Direct mailings to individuals
- Meetings with groups, such as staff of large clinics or specialty laboratories
- Presentations at scheduled meetings, such as hospital staff meetings or local medical society meetings
- Regional presentations and orientation workshops organized by the registry.
- All relevant physicians (e.g., pathologists, medical oncologists, dermatologists, general surgeons and surgical specialists, and radiation oncologists)
- All related facility personnel (e.g., hospital administrators, health information service administrators, and cancer registry managers).

6. Monitoring Use of and Changes in Facilities and Practitioners

a) Introduction

Registries **MUST** be able to document that they capture cases from the entire population at risk for their area. To do so, they **MUST** be able to document where residents of their population receive cancer

diagnoses and how those cancer cases are identified by the registry. Registries **SHOULD** monitor changes in the number and location of facilities and practitioners and where their area's patients are being diagnosed and treated. Facility openings, closings, and mergers and the establishment of new screening programs all can impact workload and procedures for the registry by influencing the number of cases diagnosed and the number and location of sources the registry needs to cover.

b) Standards

The registry **MUST** be aware of the flow of cancer patients outside the registry coverage area's borders for diagnosis and treatment; the closing of hospitals and clinics and the opening of new ones, including screening and treatment centers; mergers of facilities that impact the operation of hospital registries and the central registry; and shifts in utilization of screening, diagnostic, or treatment facilities that would impact where cases are diagnosed and treated. The registry **MAY** obtain information from governmental licensing agencies and also **SHOULD** conduct periodic surveys and review telephone directories, local newspapers, professional association publications, and the Internet.

7. Confidentiality Policies and Procedures: Issues in Data Collection and Management

a) Introduction

Confidentiality policies and procedures are required in all phases of the registry operations to: Protect the privacy of the individual patient Protect the privacy of the facilities reporting the cases Provide public assurance that the data will not be abused Abide by any confidentiality-protecting legislation or administrative rules that may apply. Although the cancer reporting regulations under which the registry operates may define only patient-specific data as

confidential, registries also **SHOULD** treat any information that specifically identifies a health care professional or an institution as confidential. Information that characterizes the caseload of a specific institution or health care professional also **SHOULD** be considered proprietary and confidential.

The Registry's Responsibilities

It is the responsibility of every registry to protect its data from unauthorized access and release. The cancer registry **MUST** maintain the same standards of confidentiality as customarily apply to the doctor-patient relationship; this obligation extends indefinitely, even after a patient's death. If data are maintained both on paper and in electronic formats, data security policies and procedures **MUST** address both types of data formats.

b) Standards

(1) Standards for Policies and Procedures for Data Security

The following components generally will be required to assure data security:

- The director of the registry **MUST** be responsible for data security.
- Suitable locks and alarm systems **MUST** be installed to control access to the registry, and a list of persons authorized to enter the registry **SHOULD** be maintained by the director.
- Registry staff **MUST** be responsible for the confidentiality of all data encountered during the collection of cancer data.
- Confidential data **MUST** NOT be transmitted by any means (mail, telephone, fax, electronic) without the explicit authority

from the director or a staff member to whom such authority has been delegated.

- Registries **SHOULD** consider the use of registered mail, overnight mail, or courier services for confidential data and **SHOULD** consider separating names from other data for transmission. When using mail services, registries **SHOULD** consider using double envelopes, with the confidential information in a separate envelope marked “confidential,” including a contact telephone number, and enclosed in the mailing envelope. Registries **SHOULD** consider using tear-free envelopes marked “confidential.”

- Precautions **MUST** be taken for both the physical and electronic security of confidential data sent via magnetic or electronic media.

- Computer use of confidential data **MUST** be controlled by electronic and, if possible, physical measures to enhance the security of the data, including the use of a separate room, use of passwords, automatic logging of all attempts to enter the system, and different levels of access to the data.

- Training and demonstrations of the computer system **SHOULD** be performed with separate fictitious or anonymous datasets.

- Consideration **MUST** be given to obtaining expert advice on security against unauthorized remote electronic access, if it is impossible to use isolated data processing systems.

- Measures **MUST** be taken to ensure the physical security of confidential data stored on paper, microfilm, microfiche, etc.

- A policy **MUST** be developed for the safe disposal of confidential waste. If a private document destruction company is used, the registry **MUST** have documented procedures for disposal and the security measures used for the company's employees. Data stored on laptops or other portable media **MUST** be encrypted.

(2) Standards for Personnel Policies and Procedures

Registry Staff Members/Employees:

- The registry staff **MUST** sign, as part of their employment agreement, a declaration that they will not release confidential information to unauthorized persons. This declaration **SHOULD** remain in effect after cessation of employment. The director **SHOULD** maintain a list of staff members indicating the nature and extent of their access to registry data.
- The training of all registry staff **MUST** include a comprehensive session concerning the confidentiality of data.
- Failure to observe the confidentiality policies **MUST** result in firm disciplinary action or dismissal. Some circumstances **MAY** warrant legal action against staff members who fail to comply with the registry's confidentiality policies. Depending on the jurisdiction, there also **MAY** be criminal penalties for failure to maintain the required confidentiality.
- Registry staff **SHOULD** be reminded annually about confidentiality policies.

Non registry Staff:

- Non registry staff, especially medical investigators, **MAY** request access to confidential registry data. Such requests **MUST** be in writing. All non registry staff who request access to these records **MUST**, at a minimum, agree to adhere to the same confidentiality safeguards practiced by registry staff.

- Requests **MAY** be adequately addressed without the release of confidential information. Whenever possible, it is preferable to respond to requests without the use of confidential information. The registry is not responsible for disclosure of individual clinical information to unauthorized requestors. Individuals requesting personal information from the Registry **SHOULD** be referred to the diagnosis and/or treating facility.

(3) Standards for Policies and Procedures for Release of Registry Data

Release of cancer registry data for clinical purposes, for research, and for health care planning is central to the utility of the registry, and the registry **MUST** develop procedures for data release that ensures the maintenance of confidentiality. See Section III.A.1. for a detailed discussion of confidentiality issues in research, reporting, and release of registry data. For the purpose of complete case ascertainment, the cancer registry **MAY** exchange confidential data with other registries if reciprocal case-sharing agreements that include confidentiality provisions are implemented. The cancer registry **MAY** permit the release of confidential data to treating hospitals in their own region for the purpose of patient follow-up. It is recommended that plans be made for the possible cessation of registry activity to maintain the subsequent utility of the database while safeguarding the confidentiality of its data.

8. Death Clearance

a) Introduction

Death clearance is an essential step in achieving complete population-based reporting. It serves as a check on the completeness of reporting from other sources and often identifies cases that should have been reported from those sources but were not. It also identifies patients known only to the physician. Furthermore, cases that remain as death certificate-only (DCO) cases after follow-back

MUST be included as incident cases by the registry. Death clearance for this purpose means identification of all deaths, with cancer given as a cause of death, that are not accounted for in the registry's files. Death clearance for the purposes of obtaining follow-up on cases already registered is not addressed by this section.

b) Standards

(1) Standards for Frequency and Timing

Death clearance for the purposes of case identification **SHOULD** be performed when the death files are complete for the calendar year being cleared, and with enough time for follow-back to be completed and the results incorporated in the registry's database before the registry publishes cancer incidence rates for the calendar year. Timing **must** carefully be planned. The goals are to link every cancer from the time period against every death from that period, avoiding unnecessary follow-back but distributing the follow-back workload across a reasonable time. In practice, death clearance usually is performed more than once for cases in a given time period. The death file for a given year may not be completed soon enough to meet the registry's needs, either because of coding delays at the vital statistics office or because not all deaths of a state's/province's residents occurring in other states have been incorporated (states/provinces exchange death records on residents of other states/provinces via the transcript exchange program). The registry's files also may be incomplete at the time of the initial linkage. Early linkages **MAY** be performed with incomplete death or registry files. Additional linkage or linkages then **MUST** be performed when the registry considers its case file to be complete and the death file is considered complete for the year by the vital statistics office.

(2) Required Components

The following components generally will be required for the performance of death clearance:

- The registry **SHOULD** establish a formal agreement with the appropriate vital statistics office covering access to computer records and paper files, subsequent use of death record information, and costs.
- The registry **MUST** have the ability to perform record linkage between the death files and cancer cases and identify matches, non matches, and potential matches with cancer as a cause of death.
- The registry **MUST** have staff adequate in number and trained in case-finding and abstracting to perform follow-back. A rough estimate for staffing is one full-time employee per 10,000 annual cases for a registry that has been performing death clearance for several years. It is important to keep in mind that this may represent a percentage of time for several different staff.
- The registry **SHOULD** have a system for tracking progress and results of follow-back. This system **SHOULD** preferably be automated, but **MAY** be manual.

(3) Other Standards

The registry also **SHOULD**:

- Include a tumor linkage comparison in its death clearance (i.e., verify that, for patients in both the registry file and the death file, the cancers are of the same primary sites), and if there are discrepancies, follow-back as necessary to determine if the patients had additional reportable cancers that should be registered.
- Analyze the results of death clearance, monitor them regularly, and use the information as feedback in the quality control

cycle to improve case finding and completeness of reporting from hospitals and other sources.

9. Training in Case Finding and Multiple Primary Determination

a) Introduction

To ensure that the personnel actually performing case ascertainment and abstracting are aware of the reporting rules and methods, it is important to make training available.

b) Standards

Before data collection for the registry begins, the registry **SHOULD** provide training in the following areas to all personnel who will be responsible for cancer case identification and abstracting:

- All criteria for case reportability
- Rules for multiple primary determination.

Training **SHOULD** be provided to registry staff and to staff in hospitals, laboratories, clinics, physicians' offices, and all other facilities where the staff may be identifying cases for the registry. Training **MAY** be offered via appropriate professional association meetings or at workshops scheduled by the registry. Professional publications and central registry newsletter articles also **MAY** be used to deal with reporting problems.

10. Monitoring Completeness of Reporting and Ensuring Compliance by All Facilities and Practitioners

a) Introduction

Monitoring the completeness of case finding is a required component of the central registry's quality control. Even where the reporting facilities are performing the case finding, it ultimately is the

central registry's responsibility to verify that the facilities are reporting all appropriate cases and to take corrective action when problems are discovered.

b) Standards

The registry **SHOULD** monitor the processing of the case finding sources on a regular basis. Frequent monitoring enables the registry to quickly identify problems and take corrective action. The registry **SHOULD** prepare and review various management reports such as the following to monitor the status of reporting:

- Completeness of reporting for each facility, each county, and the entire coverage area
- The status of screening of the case finding sources, such as each type of pathology report (i.e., surgical specimens, cytologies, autopsies, bone marrows, etc.), disease and operations indices, and radiation treatment logs for each facility
- Status of death clearance processing
- For each applicable facility and for the entire coverage area, counts of cases for primary sites often diagnosed and/or treated in an outpatient setting to identify potential underreporting from non hospital sources
- A report of the percent of histologically-confirmed cancers for each reporting facility to identify potential underreporting due to lack of screening non pathological sources. When the number of reported cases deviates widely from the number expected, the registry **SHOULD** undertake to determine the possible reasons. Cancer reporting may be late or incomplete, or the numbers may accurately reflect changes in occurrence or distribution of cancer. A hospital's census may be down, cases may have shifted to another hospital or clinic, or expected growth in population may not have occurred. If the

reporting law provides for a means of enforcing the reporting by facilities and practitioners, the central registry **MUST** undertake the necessary procedures to obtain complete reporting and compliance from all facilities.

11. Case Finding Audits

a) Introduction

Purpose and Definitions

Although DCO percentages, observed-to-expected ratios, and incidence-to- mortality ratios can provide some estimates of the level of completeness of registration, they only reflect how the registry performs as compared to an average. Cancer incidence and/or the diagnostic practices in a registry's area may or may not be the same as the "average." The only way to document the true level of completeness of ascertainment is through special studies, or audits, to identify and document deficiencies in the ascertainment system. The design of an audit will depend on the definition of "cancer," the reporting practices of the institutions in the area, the reporting requirements and policies, and the ascertainment methods used by the registry. Cancer registries **SHOULD** perform an independent review of case finding sources in reporting facilities to determine reporting completeness.

b) Standards

(1) Standards for Frequency of Audits

At least one audit to assess completeness of ascertainment **SHOULD** be performed per year.

(2) Standards for Types of Audits

More than one type of audit **SHOULD** be used to assess completeness. A rotating schedule **MAY** be set up for performing various types of audits.

12. Patient Follow-up

a) Introduction

Registries intending to provide end results (survival) data **MUST** follow all registered patients for life (often, carcinomas *in situ* of the cervix uteri and basal and squamous skin cancers, when they are registered, are not followed). Methods of obtaining follow-up will vary due to local considerations, such as the number of cases being followed by hospital cancer programs and the availability of databases against which the registry files can be linked. The goal is to achieve the highest possible success rates and avoid biases in the lost-to-follow-up group (see Section I.C.6. for a discussion of measuring success rates). Methods generally are classified as active or passive, with active meaning a direct contact with the patient, the patient's family, or the patient's physician; and passive referring to methods that do not require such contact. Central registries usually will need to employ a combination of complementary methods to achieve acceptable levels of success and avoid bias in the lost-to-follow-up group.

b) Standards

The choice of methods or sources for obtaining patient follow-up **SHOULD** be driven by:

- The availability of the method or source to the central registry.
- The effectiveness of the method or source.

The primary sources usually are:

Hospital Registries: The central registry **SHOULD** obtain reports on the results of their active follow-up activities from hospital registries. The central registry **SHOULD** establish a standardized electronic format for the submission of follow-up records. Follow-up is more of a collaborative activity if the central registry also provides hospital registries with the results of its own follow-up activities, especially the results of death clearance.

Death Clearance: The central registry **SHOULD** clear its case files against files of registered deaths to obtain vital status, dates of death, and causes of death.

C. OUTCOME MEASURES

1. Percent Death-Certificate-Only

a) Introduction

The percent of cases in a registry file for which the death certificate is the only reporting source traditionally has been used as a measure of registry completeness. In long-standing registries with very complete coverage, it is probably more a measure of assiduousness of follow-back. A more useful measure might be the proportion of cases initially identified through death certificates that would otherwise have been unreported, regardless of their eventual type of reporting source, but this is not a measure for which there is any consensus on codes or any history of collection. Registries continue to use percent death-certificate-only because it is simple and identifies registries that clearly are incomplete, although it does not discriminate well among relatively complete registries.

b) Standards

In the MECC registries DCO is an essential function. Based on the experience of the IARC Program, a rate of 5 percent or less DCO is acceptable. If 0 percent are DCO, death clearance has not been performed, and if the percent DCO is greater than 5, there probably is underreporting from other sources, or follow-back is incomplete, or both.

2. Observed Versus Expected Case Counts

a) Introduction

Incomplete ascertainment of cancer cases can result in artificially low incidence rates and can lead to incorrect conclusions about the cancer burden in the population. There are a number of ways cancer registry staff can determine the level of data completeness in the cancer registry: calculating the percentage of cases identified by death certificate only; analyzing collected data to be sure they follow known patterns (e.g., incidence > mortality); and, most importantly, conducting special studies or audits. Additionally, the comparison of the expected number of cancer cases for a given population with the observed number of unduplicated cases reported to the registry over a specified time period is very useful in determining whether standards of case ascertainment are being met and whether the data collected by the registry are complete enough for analysis.

Methodology for Calculating Observed Versus Expected Cases
Many methods **MAY** be used to calculate expected numbers of cases, from the simple to the very sophisticated. It is preferable that estimates be based on actual incidence data for the population at risk, or if those data are not available, on incidence data for a population similar in racial composition. Use of mortality rates is less useful. For the most accurate estimate of expected numbers, some

method of adjusting for time trends **MAY** be included, although this adds to the complexity of the calculations. One method that has been used is to multiply age- and sex-specific estimates of the population in the year of interest by the most recently published age- and sex-specific incidence rates for all cancers in the population covered. This method assumes that the registry has rates from previous years and that these rates are unbiased (based on complete and accurate information). If this is not a correct assumption, rates from another, high-quality registry covering a population whose demographics are similar to those in the registry's area, especially in its racial distribution, **MAY** be used. Age- and sex-specific numbers then are summed to determine the overall expected values for all cancers. All calculations and analyses addressed in this section and in Section I.C.3. assume that duplicate records for persons and tumors have been eliminated, that each case is counted only once, and that all patient and tumor information has been consolidated.

b) Standards

• The registry **SHOULD** compare observed and expected numbers at regular intervals during the year.

• If the size of the population is large enough to yield stable numbers, expected case counts **SHOULD** be compared to observed counts by county and/or region of the coverage area, by race/ethnicity if minorities make up an important part of the population, and by cancer site. Sites comprising the greater proportion of cancers reported to the registry **SHOULD** include breast, colon and rectum, lung, and prostate.

The expected number of cases **SHOULD** be evaluated and revised annually based on actual numbers of cases and other considerations, such as known trends toward increasing or decreasing rates of cancer of specific sites or changes in the population due to in- or outmigration.

Interpretation of observed versus expected counts requires a thorough knowledge of the underlying population. There **MAY** be specific reasons other than problems in data collection as to why observed numbers are higher or lower than expected.
Calculating and interpreting observed versus expected counts **SHOULD** not supplant other quality control activities, particularly case-finding audits.

3. Other Analyses

a) Introduction

Experience has shown that certain patterns occur in cancer data. Nonconformance with one or more of these patterns may indicate incorrect data. The registry **SHOULD** assign a qualified person to evaluate a year's data and use his or her judgment to determine whether or not data that deviate from these standards or norms are accurate.

b) Standards

Data **SHOULD** be analyzed for the following patterns:
Incidence rates and frequencies **SHOULD** be greater than mortality rates and frequencies. If mortality exceeds incidence for cancer of any site, the data for that site **MUST** be verified.
Lung, liver, and pancreas are typical sites for DCO cases. Investigation is required if there are no DCO cases for these sites.
Primary site of the cancer is unknown for about 5 percent of all cases.
Cancers of the male breast account for 0.5 to 1 percent of all breast cancers.

4. Timeliness of Registry Reporting

a) Introduction

Timely reporting of cancer information is an important goal for a registry. Epidemiology, cancer control, and clinical uses benefit from speedy access to the most current information. However, completeness and accuracy of data also are essential goals. Reports based on incomplete or inaccurate data can misinform scientists and the public about the true picture of cancer in the registry's area. The speed with which registry data can be collected, processed, analyzed, and reported depends on many factors, some of which are within the registry's control and others of which are not. Efficient data collection methods, use of computers and telecommunications, and adequate numbers of well-trained staff all can influence the timeliness of reporting of cases from hospitals, within limits. However, it also is true that the diagnostic work-up and treatment can occur over several months. Once cases have been received by the registry, a wide variety of activities take place, as outlined in Sections II, III, and IV of the manual. All of these processing steps take time, and some of them notably death clearance, sharing of cases with other registries, and establishment of population denominators, impose external delays on the registry.

b) Standards

Cases **MUST** be abstracted within 6 months of initial diagnosis. Within 12 months of the close of the diagnosis year, 90 percent of expected, unduplicated cases **SHOULD** be available to be counted as incident cases at the registry; and, within 24 months of the close of the diagnosis year, 95 percent of expected, unduplicated cases **SHOULD** be available to be counted as incident cases at the registry.

5. Case Finding Audit Results

a) Introduction

Case finding audits are studies involving independent re-ascertainment of cancer cases, usually in a sample of facilities and, within each facility, a sample of time periods. Cases identified during the audit are enumerated and matched against the registry's files. Unmatched cases are followed back to verify their reportability, and the percent of cases actually missed that should have been reported is calculated. Studies are designed for a variety of purposes and with varying degrees of statistical rigor. Most studies focus on hospital reporting, and thus provide an estimate of the completeness of reporting for hospitals only, and not a true registry completeness estimate. The following sources are problematic to review in a systematic way, and usually have not been incorporated into audit protocols:

Physician's offices
Clinics and outpatient facilities, including radiation therapy and surgery treatment centers
Free-standing pathology laboratories

Well-designed protocols with careful sampling plans and formal analysis plans are important when calculating an estimate of the registry's completeness that will be made public or used to assess registry completeness. If the goal is to identify possible ascertainment problems in facilities and to take corrective action, more informal methods **MAY** be appropriate; however, there are other advantages to a formal well documented protocol and written findings. It will allow repetition of the study at a later time or in another area or group of facilities, and findings can be compared over time and across samples if the same study design is used and results are well-documented.

b) Standards

Standards have not been established for the design of case finding studies or the statistical analysis of the results. However, it is important that such studies be designed by a statistician or epidemiologist familiar with cancer registries as well as sampling methods.

6. Follow-up Success Rates

a) Introduction

There are at least six different formulae used to calculate the percent successful follow-up. They vary by whether deceased individuals are included in the numerator and/or denominator and whether the month of follow-up is considered or only the year. Any standard established **MUST** specify the formula to be used. For the population-based registry's purpose of calculating patient survival based on accumulated follow-up data, the percent of cases successfully followed **SHOULD** be as high as possible and that the cases lost to follow-up be an unbiased group.

b) Standards

The requirement is for a success rate of at least 90 percent, preferably 95 percent or greater. The formula for calculating successful follow-up, applied separately to invasive and *in situ* cancers, is as follows: Assume that Y is the calendar year ending 19 months prior to the due date for an August data submission. The percent of patients diagnosed during the years prior to and who have current follow-up is defined as:

$$P = 100(D + A)/T$$

Where D is the number dead, A is the number alive with follow-up dates on or after January 1, Y + 1, and T is the total number of patients being followed. P can be calculated for individual years of

diagnosis up through Y-1 and for all years combined prior to Y. Systematic annual follow-up of patients is an important cancer registry function. A successful follow-up rate of 90 percent is required to use registry data for survival (outcome) analysis. The required rate for follow-up is calculated separately and is set at 80 percent. Cases are delinquent (lost) if the follow-up interval exceeds 15 months. The registry **SHOULD** apply the calculations to subgroups of patients to evaluate for bias. For example, calculation of follow-up rates by sex for three age groups, those under 15, 15 to 64, and those 65 and over, **MAY** show that, although the overall rate is very high, the registry is not successfully following its pediatric cancers, especially among females. An analysis by ethnic group or geographic area might identify other groups that have poor follow-up.

II. DATA QUALITY

A. STRUCTURAL REQUIREMENTS

1. Data Quality: General Requirements

a) Introduction

All aspects of the registry's operations impact data quality, including the laws and regulations under which the registry operates, relations with hospitals and physicians in the registry's coverage area, how the data collection system is designed, staff qualifications and training, review of data for analysis and reporting, and the capabilities of the computer system. The function usually termed "quality control" is limited to those personnel and activities that are directly focused on assessment of an improvement of quality of the data, but the quality control activity cannot function on its own.

b) Standards

The registry **MUST** have an overall program of quality assurance into which the specific activities fit. Quality control activities **MUST** be part of a planned whole, and not just a series of unrelated, sporadic activities. The quality assurance program **SHOULD** be formally defined, including the assignment of a specific responsible individual, the schedule for routine edits and reports, and steps to be taken when pre-specified conditions are not met. The registry **SHOULD** carefully document each of these activities, as well as procedural changes and any non routine dataset evaluations undertaken. The registry's budget **SHOULD** provide specified and adequate funding for quality control staff and activities. The registry **SHOULD** prepare written rules for identifying when action or further investigation is needed based on results of quality control activities, and **SHOULD** have predetermined procedures to follow under those conditions.

2. Staffing Guidelines for Data Quality

a) Introduction

Adequate registry staffing requires both the skills and the available personnel to conduct registry business in a timely, competent manner. Staff competencies necessary for quality control in a registry include content knowledge, analytic knowledge, training knowledge, and organizational skills. The numbers of persons required will depend on the size of the geographic area, the caseload of the registry, the number and detail of items collected, the method of data collection, and registry staff organization.

b) Standards

The following skills **MUST** be represented on a central registry staff for adequate quality control activities: Tumor Registrar (TR): One or

more TRs **SHOULD** be directly involved in monitoring abstract review, training the persons who abstract or edit data (both registry employees and staff at reporting facilities), and conducting quality control activities. In particular, TRs can contribute expertise with respect to diagnostic and treatment data, case finding, and follow-up. Statistical Analyst: Both systematic analysis and specific designed studies **MUST** be designed and evaluated with the active participation of an individual who is knowledgeable in statistical methodology and analysis, especially for determining appropriate samples, appropriate statistical measures, and criteria for taking remedial action. The person **MUST** be familiar with biostatistical, quality control, and sampling techniques. Abstractors and Coders: If the registry employs abstractors or coders in the office (for abstracting or routine editing) or in the field, they **MUST** be familiar with all item definitions and coding instructions used by the central registry, and well-trained in abstracting cancer data from patient records. Quality Control: One person on the staff **SHOULD** be identified as responsible for maintaining overall quality control for the registry. Often this will be the person in charge of training both registry staff and staff at contributing facilities. This is the person primarily responsible for interpreting the results of quality control auditing. Computer Expertise: The registry **MUST** have knowledgeable computer staff available to be involved in the design and implementation of edits and carrying out studies.

3. Procedure Manuals, Coding Manuals, and Other Documentation

a) Introduction

Permanent, current, widely distributed written documentation of all aspects of the registry's definitions and methods is essential to establish standardization, maintain continuity of meaning, document changes over time, develop training, and inform data users. The

documentation usually is in the form of procedure manuals, coding manuals, and other manuals.

b) Standards

Adequate staff and time **MUST** be provided to prepare and maintain high-quality, up-to-date documentation or manuals. The registry **MUST** formally document its dataset definitions, codes, coding rule interpretations, procedures, and decisions or recommendations of its medical advisors. The registry **MUST** have a mechanism for updating the documentation and keeping it current. The registry **MUST** incorporate standard manuals, such as MECC manuals whenever appropriate to ensure comparability. Documentation **MUST** be provided to all registry employees involved in data collection, data management, and data analysis, and also to employees of hospitals and facilities that are reporting data to the registry. Appropriate portions of the documentation **SHOULD** be provided to investigators and users of the data to explain definitions and methods. Traditionally, documentation has been in the form of printed manuals, including data dictionaries, coding manuals, and procedure manuals. Online electronic documentation is becoming increasingly important. The registry's documentation **MAY** be in printed form, online, or in a combination of media.

4. Edits and Data Processing Capabilities for Data Quality

a) Introduction and Definitions

Software engineering identifies repetitive manual processes that may be better performed by a computer program. Over the years, cancer registry software has been developed to address an increasing number of registry tasks, enabling staff to focus on tasks that require human judgment, analysis, or interaction, and in most cases bringing increased quality to the data. This pattern may be expected to continue for some time. Central registry computer software systems

MUST not only provide a repository for data and the tools to generate incidence reports, research data, or other registry end products, but also are a major focal point for quality control processes. A fundamental requirement of registry software is that the system maintains data integrity, through careful and effective data management and adequate system security. These functions are covered in this manual, titled Data Management. The present section covers aspects of the computer system that are directly related to quality control activities of the registry. Routine quality control functions that **SHOULD** be built into registry computer systems include:

Edits: Data edits are logical rules, typically embodied in a computer algorithm, that evaluate to "true," "false," or "maybe," for any value(s) of data item(s). In the cancer registry, edits are applied to all records to check for item validity, internal consistency, and inter record consistency. Data edits may involve a single field, multiple fields in a single record, multiple fields in different records within one database, or multiple fields in multiple databases.

Process Controls: When information from edit procedures is retained, it **SHOULD** be analyzed on a regular basis to identify trouble spots, for example, with data sources, coders, item code structure, or clarity of instructions in the manuals. The computer system **SHOULD** contain flags set to reflect the nature and disposition of edit failures and analytic routines for evaluating their contents. The data are summarized across time for individual data sources or item codes. Items **SHOULD** include dates each case was accessioned into the registry and later updated to evaluate delays between case reporting and accession.

Capabilities for Audits and Designed Studies: The system **SHOULD** support the conduct of audits and designed studies by facilitating the drawing of appropriate samples, efficient data entry for cases in the field, automated comparisons of original and re-abstracted or recoded data, and analysis of results.

Standardized Edits A subtle but important principle is that data that are edited differently will be statistically different, and non-comparable. To achieve comparable data, edits need to be standard across all registries for the following reasons:
The utility of any local data collection effort is substantially compromised when the categorization of data collected is not statistically comparable to other collections.
A standard edit serves as reminder and enforcer of standards; however, when an edit is nonstandard, it enforces non-comparable data.
Additional subsequent edits generally cannot remedy the effect of earlier, suboptimal edits.

b) Standards

(1) Required Components

The following components generally will be required for automated quality control procedures:

Computer Edits

The registry **MUST** have a system of computerized data edits with the following characteristics:

- Uses standard program code or algorithm wherever possible
- Performs single-field, multi-field, multi-record, and multi database edits as appropriate
- Is flexible enough to allow for changes
- Produces reports and error messages that are meaningful to those correcting errors and to everyone interpreting the data
- Is thoroughly documented as to logic and performance, with documentation and all tables used in the edits available and understandable to those correcting errors and everyone using the data

- Provides for edit output that **MAY** be passed back to individual facilities for resolution if necessary.

Process Controls: The registry **SHOULD** provide the capacity for process controls. The data items necessary to identify and store quality measures and the analytic routines for systematically evaluating them **SHOULD** be built into the computer system. These include:
A method of summarizing edit outcomes systematically.
Routine evaluation of edit outcomes, preferably presented in control charts or other easily interpretable forms such as graphs.

Designed Studies/Audits: The registry system **SHOULD** allow drawing of samples for quality control studies by any desired characteristic.
Staff: The registry **MUST** have sufficient staff trained in abstracting and coding to track and correct edit failures.

(2) Standards for Data Entry, Data Definition, Data Representation, Datasets, and Record Layout

Standardization of Data Entry: Standardization of output is facilitated by standardization of as many aspects as possible of the intervening steps in collecting and processing the data.
Standardization of the following aspects of registry software applications may improve data comparability:
• Prompts
• Coding choice lists
• Online help
• Edits: single field, multi-field, multi-record, or multi-database
• Error messages. Although convenient, auto-coding can be a dangerous feature, especially for variables such as histology where modifiers to a root word changes the code. Registries will vary in the extent to which they have control over these aspects, because some registries will obtain data collected by hospitals using a variety of

software applications. However, registries **SHOULD** encourage mechanisms for definition and promulgation of additional standards.

Standardization of Code Definitions: Historical continuity of the definition of data categories is required for trend analysis. Representation of meaning may be allowed to change over time, provided that translation tables are preserved. When certain categories must be discontinued, continuity of meaning may be preserved in some cases by an overlap in collection of the new and the old categories. When additional detail is desired, ensure that collapse into standard categories is feasible. In choosing a dataset, it is essential to identify the purpose to be served and then to choose appropriate subsets of items. Examples of specific purposes include:

- Patient care evaluation
- Descriptive epidemiology and surveillance
- Cancer control
- Research.
- Coding of data and data translations
- Item sequence and record layout
- Electronic media specifications.

Standardization of Data Exchange Format: Standardization of the electronic format for data exchange improves the quality of merged files.

(3) Standards for Frequency and Timing of Data Edits

Edits **SHOULD** be as physically close to the information source as possible to allow immediate verification/review upon edit failure. Edits **SHOULD** be as temporally close to the event as possible, to improve success of obtaining accurate clarification, and minimize permanent information loss. This also increases the value of the data. Item, internal consistency, and inter-report edits **SHOULD** be applied routinely as or before new records are added to the database, with

serious edit failures being withheld from incorporation into the analytic database until they are resolved. Analysis of edit failures **SHOULD** be performed continuously with special attention to results for new staff, new hospitals, new vendors, new procedures or other data-collection conditions that are not stabilized.

(4) Standards for Record Consolidation

Record consolidation is an important function of cancer registries. It ensures that all submitted cancer cases are counted only once. When records are not consolidated, over counts of cancer incidence occur.

B. PROCESS STANDARDS

1. Training for Improved Data Quality

a) Introduction

Training is an essential component for a population-based registry to assure that the data collected are accurate, consistent, and complete.

b) Standards

(1) Required Components

Training **MUST** be provided to employees of the central registry who are involved in data collection and quality control and to the employees of hospitals and other facilities that are reporting data to the registry. Training activities in the following areas are recommended: **Reporting Requirements:** Instruction on reporting requirements including frequency of reporting, mechanism of reporting, and required data items. Documentation **MUST** be provided that defines the reporting requirements. **Data Collection:** Instruction on reportable neoplasms, case finding procedures, abstracting requirements, ICD-O coding, staging, and, where

appropriate, treatment coding **MUST** be provided. The instruction **MUST** be based on the standardized reference manuals that the registry officially adopts. **Quality Control:** Instruction in visual and computer edits and feedback regarding edit results **SHOULD** be provided to the data collection staff and other staff from reporting facilities. **Data Processing:** Instruction regarding the use of computer software **SHOULD** be provided if computerized reporting is mandatory.

(2) Standards for Training Methods

A variety of methods **MAY** be utilized, including:

Formal Programs: These include introductory training classes, workshops, educational programs and symposia, plus regularly scheduled in-service training.

Audits: Identify areas that need additional training through the use of audits.

Feedback: Provide timely feedback to data collectors on the types and patterns of errors identified during quality control activities.

2. Quality Control Activities

a) Introduction

While it is appropriate and necessary to design a quality control program to fit the needs of a particular cancer registry and its users to the extent that registries adhere to the various standards addressed in this document, certain quality control activities will be universally applicable. These activities can be divided into three classes: inspection or acceptance sampling, process control, and designed studies. Inspection or acceptance sampling encompasses any form of regular, ongoing review to determine whether individual case abstracts or batches of case abstracts meet minimum standards

of acceptability. Process controls comprise all forms of monitoring the outcome of inspection to detect shifts from the “in control” to an “out of control” state. Designed studies are planned studies and generally are undertaken to address a particular problem, to examine the feasibility of change, or to independently quantify a complex registry characteristic. The three classes of activity are related hierarchically, with inspection being the simplest form of quality control, and the most appropriate for very young registries. Process controls can only be effectively implemented after the reporting process has stabilized and it is reasonably clear that the registry system is “in control.” Similarly, designed studies will not provide reliable results until a stable reporting system is in place.

Definitions

Quality Assurance Sampling: Inspection is the process of measuring, examining, or otherwise comparing the unit with the applicable requirements. Inspections become quality assurance sampling if failure to meet the requirements leads to rejection of the unit. There are several forms of quality assurance sampling of varying practical importance in cancer registries:

- Automated edit checks: high importance
- Visual review of text and codes: high importance
- Duplicate coding: minimal importance
- Duplicate abstracting: minimal importance. Process Control:

Statistical process control involves the prospective monitoring of rationally aggregated results of inspection. Process controls can involve both outcomes of acceptance sampling where errors in case abstracts (or batches) are detected (e.g., edit rejection rates) as well as other aspects of registry data and operation that do not necessarily represent errors, but that should exhibit stability over time or across regions (e.g., percent unknown primaries). Process

control design requires statistical expertise, including specification of an appropriate probability model, selection of a sampling plan and rational subgroups, selection of appropriate control charting procedures, and specification of control limits.

Designed Studies: Two types of designed studies are appropriate for cancer registries: optimization/feasibility studies and estimation studies. First, the feasibility or utility of substantive changes to registry design and operation (e.g., new forms, data items, sources of reporting, etc.) **SHOULD** formally be evaluated. Second, the key registry quality attributes of completeness (see also Section I) and accuracy **SHOULD** periodically be evaluated by a formally designed study. If these studies can be standardized and are executed on a routine basis, they become a form of acceptance sampling and the results **SHOULD** be monitored by appropriately designed process controls.

Re-abstracting Audits: Re-abstracting audits describes the process of independently re-abstracting cancer cases from the source patient records, coding the data, and comparing the abstracted and coded data to the data already in the registry. This type of study historically has been used in cancer registries, and the methods are well developed.

Recoding Audits: Recoding audits involve independently reassigning codes to abstracted text information but not reviewing the source documents. This type of study is done frequently, and is very useful in training new coders. It is easier and less expensive to perform than re-abstracting, but the method cannot detect problems with abstracting.

b) Standards

(1) Standards for Acceptance Sampling

General Considerations: Not all forms of acceptance sampling will be applicable to all registries. For example, duplicate data entry would not be appropriate for registries receiving data electronically. Regardless of which methods are used, the procedures, sampling plan, and intensity **SHOULD** be documented, and the results of inspection **SHOULD** be retrievable, either manually or preferably from an automated tracking system. There also **SHOULD** be a system to monitor progress in resolving errors. Unacceptable reports **SHOULD** be corrected and re-inspected. If inspection is on a sampling basis, documentation **SHOULD** include details of the sampling plan and specifications for batch rejection.

Automated Edit Checks

- **Application:** All registries.
- **Sampling: SHOULD** be 100 percent.
- **Comment:** Registries **SHOULD** attempt to implement all applicable standard edit checks and to develop additional edit checks to address data items and data structures unique to the registry. Errors **SHOULD** be documented and corrected.
- **Sampling: SHOULD** be applied broadly in young registries and selectively in well-established registries.
- **Comment:** Errors **SHOULD** be documented and **MUST** be corrected. If review is done on a sampling basis, then reports **SHOULD** be batched or stratified rationally, with entire batches being rejected. Some forms of errors **MAY** only be identified through visual review (or, review of the original medical records). For example, systematic misapplication of coding rules may only be detectable through comparison of codes and text.

• **Application:** Registries performing data entry, especially those using key-to-disk rather than interactive screen oriented data entry (the latter may not lend itself to cost effective duplicate data entry).
• **Sampling:** If used, either 100 percent or a sufficiently large sample/batch to allow detection of error rates in excess of 1 to 2 percent.
• **Comment:** Errors **SHOULD** be documented and **MUST** be corrected. If duplicate entry is done on a sampling basis, then records **SHOULD** be batched or stratified rationally, with entire batches being rejected.
• **Application:** Registries receiving backup documentation of submitted codes (e.g., computerized text, paper abstracts, or pathology reports).
• **Sampling: SHOULD** be applied selectively, if at all.
• **Comment:** Errors **SHOULD** be documented, categorized as to keying or coding errors, and **MUST** be corrected. If done on a sampling basis, records **SHOULD** be batched or stratified rationally, with entire batches being rejected (routine duplicate coding is different from special studies or recoding audits).
• **Application**: Registries receiving abstracts and having adequate access to source documents.
• **Sampling:** As an acceptance sampling technique **SHOULD** be applied selectively, if at all (but see Standard (3) below).
• **Comment:** Errors **SHOULD** be documented and **MUST** be corrected. If done on a sampling basis, then records **SHOULD** be batched and stratified rationally, with entire batches being rejected (routine duplicate abstracting is different from special studies or re-abstracting audits).
Visual Review of Text and Codes Application: Registries receiving backup documentation of submitted codes (e.g., computerized text, paper abstracts, or pathology reports). Duplicate Data Entry Duplicate Coding Duplicate Abstracting

(2) Standards for Process Controls

Process controls represent an additional level of sophistication, in which the aggregated results of inspection are tracked, usually over time, and used to determine objectively whether a process is “in control” or not. Design of statistical process controls require the specification of a sampling plan, selection of rational subgroups, computation of control limits, selection of a charting strategy (if control charts will be used), and specification of frequency of updates. These issues as well as actions to be taken **SHOULD** be fully documented. Measures of cancer registry quality that should benefit from formal development of process controls include, but are not limited to, the following:

• Visual review rejection rates
• Duplicate entry/coding/abstract rejection rates
• Edit check failure rates–overall and/or failure on the most important data items
• Missing data and use of unknown or ill-defined codes for data items considered critical to analysis by the registry
• Numbers of cases reported
• Lag time in reporting
• Percent death certificate only
• Re-abstracting agreement rates. Automated support for process controls is strongly recommended. For example, the computer can assist in the acquisition, management and charting of process control data and these functions can be built into registry software systems.

(3) Standards for Designed Studies

Cancer registries **SHOULD** periodically plan and execute case finding audits to assess overall completeness of reporting and re-abstracting audits or recoding audits to assess overall data reliability. Re-abstracting and recoding studies have a long history in cancer registries. The methodologies are well defined, and comparison data

may be available. Additional studies **MAY** be undertaken to address specific cancers, problem areas, or feasibility of proposed changes. All designed studies **SHOULD** be planned and executed according to a formal, written protocol. At a minimum, the protocol **SHOULD** address the following:

- Introduction and rationale
- Statement of purpose
- Sampling plan, including sample size considerations, stratifications, and randomization
- Eligibility criteria and study population
- Procedures to be followed for study execution
- Analysis plan, including data management, statistical analysis and summary statistics to be computed. Completed studies **SHOULD** be analyzed and the results communicated to management, data suppliers, and data users.

3. Dissemination of Quality-Control-Activity Results

a) Introduction

Identifying and correcting data errors is required to maintain quality data. In addition to correcting errors, it is essential that feedback be given to the data abstractor so that the quality of data will be maintained and recurring errors eliminated.

b) Standards

To reduce the number of data errors and avoid recurring problems, feedback **SHOULD** be provided in a timely manner. When abstracts are corrected or changed at the central registry, information about the changes **SHOULD** be returned to the abstractor for review. Discrepancy reports or error reports from edits also **MAY** be returned. The registry **SHOULD** provide results of recoding audits, case-finding audits, and re-abstracting audits to abstractors with analysis of discrepancies and recommendations for improvement.

Feedback on findings of audit studies and interpretation of the results **SHOULD** be given to all who participate in a study as well as the pool of individuals or organizations represented by the study participants. The feedback **SHOULD** indicate problems identified and recommended actions that will be undertaken to correct problems and improve data quality. Feedback **MAY** sometimes given via telephone calls or one-on-one meetings. Summary audit study results also **SHOULD** be made available to data users to assist in the interpretation of the data. The registry **SHOULD** incorporate the results of quality control activities as feedback to other aspects of registry functioning. For example, the registry **SHOULD**:

- Interpret the results of quality monitoring, and incorporate the conclusions in revising training, documentation, or item definition as needed
- Make public the more useful evaluative data, so that data users have an adequate context for interpreting their results.

C. OUTCOME MEASURES

1. Re-abstracting and Recoding Audits

a) Introduction

Re-abstracting audits and recoding audits are most often used to retrospectively assess accuracy (agreement with source medical records) and reproducibility (agreement among data collectors) of registry data. These audits are designed studies on a sample of cases and **SHOULD** be carried out following a study protocol that states the study objectives, describes the sampling scheme, and outlines plans for the analysis. These studies have a long history in cancer registries, and the methodologies are well developed. The objective of a re-abstracting study is to characterize the level of agreement between data already in the registry and data re-abstracted and recoded from

source records (the hospital medical records for most cases) by expert registrars. For each re-abstracted case and data item, codes are compared to determine if codes match exactly. If the codes do not match, the discrepancy is classified as to severity according to major and minor discrepancy definitions set up in advance for the specific study. Such studies require the establishment of an arbitration mechanism to determine which of the discrepant answers is correct for purposes of the study. Recoding audits help characterize the level of agreement within data records already in the registry. Samples of actual case abstracts in the registry are re-coded by expert registrars, based on the text contained in the abstract. As in a re-abstracting study, for each recoded case, the codes for each data item are compared for discrepancies with those assigned by the expert.

Study Results The registry can learn a variety of things from re-abstracting and recoding audits. Examples of information that can be gained include:

- Overall and item-specific agreement rates for the sample of cases studied, which **SHOULD** be expressed in terms of severity (see Appendix A used by SEER)
- Types of cases in which discrepancies occur more frequently
- Sources of variation (e.g., misinterpretation of source document information, information not available at initial abstracting, misinterpretation of coding rules, inadequate or erroneous computer consolidation of data between records)
- Effect of misclassifications on data analysis and use (e.g., are cases more frequently over-staged or under-staged?)
- Data quality with respect to other factors such as the newness of the registry, who collects the data (hospitals registrars versus non-registrars versus central registry), the training and skills of the registrars collecting the data, and difficulty of

abstracting and coding the specific data items. Where indicated, this information **SHOULD** be used to identify training needs and to modify registry processes and procedures to ensure future improvement in data quality.

b) Standards

(1) General Standards

Data quality standards for re-abstracting and recoding audits **SHOULD** be established as targets against which to assess and monitor how well the central registry and the individual reporting facilities are doing. Target agreement rates will vary from one data item to another, depending on the complexity and detail of the coding scheme and the quality of medical record information upon which coded information is based.

(2) Standards for Re-abstracting Studies

No known standards for agreement rates from re-abstracting studies have been published to date.

(3) Standards for Recoding Studies

Recoding studies usually are based on cancer abstract source documents and therefore remove abstracting differences as a possible source of code variation. Consequently, one would expect to achieve higher agreement rates from recoding studies than from re-abstracting studies. Recoding studies do not measure the accuracy of the coding with respect to the medical record; they measure the accuracy of coding as a function of the quality of the text justification submitted with the abstract. Poor performance on a recoding audit indicates a need for training on how to write informative text, in addition to training on how to code medical information.

2. Abstracting and Coding Reliability Studies

a) Introduction

In contrast to re-abstracting and recoding audits described above in which data already in the registry are compared with those collected by an expert registrar, reliability studies involve the abstracting and coding (or coding only) of a set of test cases by a group of reviewers such as abstractors or coders. These studies provide a measure of agreement among abstractors and coders of the "correct" or "right" answer. The method also can be used to compare computer procedures, such as data entry or abstracting software applications. The test-case method measures the quality of the abstracting/coding process in terms of reproducibility under special circumstances. Only to the extent that test cases are "like" those in the registry is anything learned about registry data quality. However, results from this study method help identify ambiguity or inadequacy of existing data definitions and rules, and areas for further registrar education and training. This method also is useful for testing whether new codes should be implemented as defined, and the degree to which there is likely to be consistency in coding. Two primary advantages of the test-case method are:

- Ease of comparing individual coders or groups of coders to some standard
- Relative simplicity and adaptability of the approach.

b) Standards

Kappa statistics measure agreement between reviewers. In quality control studies, the kappa statistic is a measurement that can be used to assess the proportion of agreement among two or more reviewers on specific data items. The kappa statistic treats all reviewers and data items symmetrically. A low inter-rater agreement rate (kappa statistic) for a specific data item, especially if one is

considering major coding differences, indicates questionable quality and usefulness of the data. To assess whether standard or best codes can be obtained reliably for particular data items; a case must be re-abstracted and coded by more than one reviewer. The maximum value of the kappa statistic is +1 if there is exact and complete agreement between the reviewers, and a minimum of -1 if there is not. For most targets, values greater than 0.75 represent excellent agreement beyond chance. Values below 0.40 represent poor agreement beyond chance. Values between 0.40 and 0.75 represent fair-to-good agreement beyond chance.

3. Unknown Values

a) Introduction

The proportion of cases coded unknown for various data items often is an indicator of the quality of the data. Unknown values can result from problems with the data collection system or access to necessary source documents, from problems with the way the item and the code values are defined, or from misapplication of coding rules. However, unknown values also can accurately reflect a limited workup or ambiguity in the medical record. A high proportion of unknown values for a data item can indicate that the item cannot be collected as defined, and that it may be appropriate to drop the item from the dataset. Modification of the definitions may decrease the proportion of unknown codes. The proportion of unknown values usually varies by primary site.

b) Standards

For a specific data item for a specific primary site, the percent coded unknown **SHOULD** be evaluated according to how analysis will be affected. Will incidence rates be affected, or survival rates? Will misleading conclusions from the data be possible because of the high

percent of unknown values? Depending on the analysis being performed, the percent unknown may be more or less problematic.

III. DATA ANALYSIS AND REPORTING

A. STRUCTURAL REQUIREMENTS

1. Confidentiality Policies and Procedures: Issues in Research, Reporting, and Release of Registry Data

a) Introduction

Confidentiality is of paramount concern to all cancer registries. There may be no greater threat to the operation and maintenance of a cancer registry than an actual or perceived breach of confidentiality. In fact, an actual or perceived breach of confidentiality in one registry threatens all registries.

Definition of Confidential Data

Although the cancer reporting laws and regulations under which the registry operates may define only patient-specific data as confidential, registries also **SHOULD** treat any information that specifically identifies a health care professional or an institute as confidential. Information that characterizes the caseload of a specific institution or health care professional also **SHOULD** be considered proprietary and confidential.

Other Resources Concerning Confidentiality

Although this document provides guidelines for developing a comprehensive confidentiality policy, registries are encouraged to consult the references cited for more information. In addition, examples of confidentiality policies may be obtained from established central cancer registries.

b) Standards

(1) Standards for Laws and Regulations Governing Confidentiality

Laws and regulations pertaining to confidentiality of cancer data vary by location. The registry **SHOULD** contact legal counsel to determine which rules govern the registry's area of coverage.

(2) Standards for Policies and Procedures for Release of Confidential Data

Confidential information about data subjects or data suppliers **MUST** NOT be released for purposes other than those specified by the registry, unless all parties concerned provide written consent for such release and agree in writing to adhere to all confidentiality polices. Confidential information **MAY** be released to health care providers and institutions directly involved in the care of the patient, for example:

- A hospital cancer registrar requests a list of all prostate cancer patients who have been treated at his or her facility.
- A physician requests a list of patients he or she has treated for breast cancer. Data **SHOULD** NOT be provided to individuals about themselves, except where required by law. Confidential information **MUST** NOT under any circumstances be published or made available to the general public. Inquiries from the press **MUST** be referred to the director or another member of the staff who has been delegated the authority to respond. Measures **MUST** be taken to eliminate the possibility that individuals might be identifiable from tables containing cells with very few entries. Registries **SHOULD** provide a document describing their procedures and criteria for release of registry data to researchers who request access to data. Inappropriate Uses of Confidential Information: Confidential cancer

registry data **MUST** NEVER be made available for uses such as the following:

- Businesses that are trying to market a product to cancer patients
- Health care institutions that are trying to recruit new patients
- Insurance companies that are trying to determine the medical status of a patient. If the registry is located within a governmental agency such as a health department, the registry **SHOULD** develop clear policies regarding access to data by other sections or programs of the department. Access by some other programs may jeopardize confidentiality and may be inappropriate.

(3) Standards for Suppression of Non-Confidential Data for Summary Statistics

Reports of summary statistics generally do not raise concerns of confidentiality. However, confidential information may be conveyed inadvertently through summary statistics. To avoid this situation, the central cancer registry **SHOULD** institute a policy to suppress the publication of summary statistics in some instances, especially when data are being presented for geographic areas with small populations. For example, some registries suppress the reporting of statistical data when there are fewer than 10 cases reported in a single cell of a table if the cell of the table represents a combination of variables, geographic area, race, age, and sex, which could inadvertently identify individuals. However, for straightforward breakdowns by age, sex, and large geographic areas such, cells with 0, 1, or a few cases normally need not be suppressed.

(4) Standards for Use of Registry Data for Research

Release of Confidential Data to Scientific Investigators: Often requests for registry data for research can be satisfied through

provision of a public use data file of non-confidential data. When non-confidential data are not sufficient, the registry is in the difficult position of determining who is and is not qualified to use registry data for research purposes. The cancer registry ultimately will suffer if it allows its data to be used for inappropriate or irresponsible purposes. Therefore, the registry **SHOULD** develop a set of guidelines to govern the accessibility of registry data to independent scientific investigators. The following criteria may be useful for developing such guidelines. **Registry data SHOULD be made available for scientific research only after the following criteria have been met:** Requests for data to be used for research **MUST** be in writing and include a suitable detailed outline of the proposed research and a justification of any need for confidential data. The registry is responsible for ensuring that researchers do not receive data that are more confidential than needed. The written research plan **SHOULD** be reviewed by appropriate registry staff. Requests for data **MUST** meet the registry's guidelines on confidentiality. The registry **SHOULD** determine that the research needs cannot adequately be addressed with non-confidential information. The proposed research **SHOULD** be approved by an appropriate committee, if necessary. The principal investigator **SHOULD** sign a written agreement to adhere to all confidentiality policies. Written agreements **SHOULD** include provisions for use of the information and for its return or destruction at the end of the study. The principal investigator **SHOULD** demonstrate adequate resources to conduct the research, including funding, staff, and technical expertise, and should demonstrate a history of having successfully conducted scientific research in the past. The scientific objectives of the study **SHOULD** be peer reviewed to ensure scientific validity. The registry **MUST** obtain evidence that researchers using registry data will adhere to the registry's guidelines on confidentiality.

Review of Research Results: Once the registry has granted an investigator access to confidential information for purposes of scientific research, the registry **MUST** ensure that confidential information is not, under any circumstances, published or displayed in reports that summarize the research results. The central registry **MUST** retain the right to review any reports prior to their dissemination to ensure that confidentiality has been respected.

Patient Contact for Participation in Epidemiologic Studies: Cancer registries sometimes serve to identify cancer patients as potential subjects for epidemiologic studies. In these instances, the investigators **MUST** meet all the criteria outlined above. Philosophies differ as to whether physician permission is needed prior to patient contact. Many patient advocacy groups maintain that only a patient has the right to decide study participation and his/her physician does not have the right to make that choice on the patient's behalf. Consequently, in many current epidemiologic studies, the physician is contacted to inform him/her that the patient will be contacted to participate in a study and to ask whether there are any contraindications to patient contact (patient too ill, patient unaware of diagnosis, etc.). Many investigators feel that this procedure protects the physician from any risk of adverse action on the part of the patient. Other investigators still insist on physician permission before contacting the patient.

2. Population Data

a) Introduction

Estimates of the number of persons in the population at risk covered by the registry, broken down by year, age, sex, race, and geographic subunits are a fundamental requirement for a population-based registry. The jurisdiction under which the registry operates may apply various constraints to population figures that are to be used.

b) Standards

(1) General Requirements

The amount of detail the registry will need to know about the population will vary, depending on the type of rates that are to be calculated. Crude rates can be calculated with an estimate of the size of the total population living within the registry's coverage area. However, knowledge of the age distribution of the population is required to calculate both age-specific and age-adjusted incidence rates. Often, incidence rates are calculated by other factors such as sex and race, which requires knowledge of the population's distribution by these factors.

(2) Standards for Sources of Population Estimates

The registry **MUST** identify the most appropriate sources of available population data for its area.

(3) Standards for Interpretation of Population Estimates

It is the responsibility of the registry staff to understand how the population estimates were derived, their limitations, and their potential impact on cancer rates. The registry staff **MUST** consult with local experts, especially demographers and members or representatives of special populations, to assure that the registry is collecting accurate racial and ethnic data. Further, the registry **MUST** work with these experts to assure that the data are reported in as accurate and sensitive a manner as is possible.

3. Staffing Guidelines for Data Analysis and Reporting

a) Introduction

The appropriate analysis, interpretation, use, and dissemination of cancer data is one of the primary functions of the cancer registry. The registry **MUST** identify staff members and consultants who are

qualified to conduct and interpret appropriate analyses of registry data.

b) Standards

(1) Standards for Number and Type of Staff

The registry **MUST** have the available expertise to conduct appropriate analyses and interpret results. This will include experts from the fields of oncology, pathology, epidemiology, biostatistics, and demography, and also may include programmers and analysts. The experts may be full-time or part-time, and they may be members of the registry staff or consultants. In any case, they **MUST** be readily available to answer questions that may arise. Data analysis staff and consultants **MUST** work closely with the registry's quality control and data management staff to ensure quality data are produced and disseminated. When appropriate, registry staff **SHOULD** conduct orientation sessions for expert consultants to ensure that they have adequate knowledge of registry operations and procedures. Sometimes it is not financially possible for a central cancer registry to retain a staff member for the sole purpose of data analysis and interpretation. In these instances, the registry may wish to develop the analysis skills of abstractors or other staff members so that they may assist consultants in the preparation of reports. Each registry **SHOULD** designate one or more staff members to serve as a liaison between the public and the registry. By centralizing the responsibility for these interactions, the registry cuts down on possible duplications of effort. This practice also minimizes the opportunity for misunderstandings that occur when information is obtained from multiple sources.

(2) Standards for Continuing Education

Staff involved in data analysis and reporting **SHOULD** be offered opportunities for continuing education so that they remain informed about analysis methods and trends in cancer data.

Continuing Education: Continuing education **SHOULD** be provided to data analysis staff to assure that they have up-to-date knowledge about trends in cancer diagnosis, management, incidence, and survival; statistical and epidemiological methods; demographic trends and methods; computer capabilities and other technologies; and cancer registries.

Access to Professional Literature, Online Services, and Other Activities: Data analysis staff **SHOULD** be supplied with appropriate references and literature to provide ongoing continuing education and to answer questions that arise. Current pertinent reference books and journals **SHOULD** be immediately available. The registry also **SHOULD** provide access to online services and bulletin board services so that staff have rapid access to the most current information.

Professional Associations and User Groups: Staff **SHOULD** be encouraged and funded to participate in local and national professional associations and user groups. The registry budget **SHOULD** include funds for participation by one or more persons at scheduled meetings. The registry **SHOULD** fund data analysis staff to attend scientific meetings, special symposia, conferences, courses, and appropriate trade shows that may occur from time to time.

B. PROCESS STANDARDS

1. Analysis Categories and Recoded Groups

a) Introduction

Many data items in a cancer registry are collected using code categories more numerous than are desirable or practical for analysis. Primary site, histological type, age, population subgroups, and extent of disease are all examples. To facilitate interpretation of data and comparisons across registries, the registry **SHOULD** use standardized grouping of these detailed codes into a smaller number of analysis categories. While conventional standards do exist, the choice of methods depends on many factors, including the number of cases available for study, the availability of comparison data, and the needs of the investigator. The selection of standard categories for analysis and presentation **MAY** depend on the choice and/or availability of comparison data.

b) Standards

(1) Standards for Grouping by Primary Site and Histologic Type

Cancer cases are commonly grouped by a combination of primary site and histologic type. A standard grouping is available in CanReg software which uses ICD-10 groupings. A listing is in Appendix A. A separate standard grouping is available through SEER. It would be optimal to have both groupings available. Pediatric cancers are so different in their site and histology distribution from adult cancers that they require a different set of analysis categories. The standard is that of the International Classification of Childhood Cancer based on ICD-O-3 (2005).

(2) Standards for Age Categories

The age distribution of cancer cases is most often summarized in 5- or 10-year age groups. The registry **SHOULD** use the recommended 5- year age groups beginning with the category 0 to 4 years, and continuing through ages 85 and older if available (i.e., 0-4, 5-9, 10-14, ... 75-79, 80-84, 85+). These are the standard groups used for population denominators. This grouping is fine enough to allow for reliability in the analysis. Pediatric cancers could be defined as those occurring under age 15. Some registries use “75+” years as the uppermost age category, but there is increasing interest in cancer in older age groups, and it is useful to provide data for the oldest groups. If a particular analysis does not use 5-year age groups (e.g., when the number of cases is small), the registry **SHOULD** choose age groups that allow for the identification of those aged 0 to 14 and those aged 65 and older.

(3) Standards for Time Period Categories

There are no standard time intervals that are used to present cancer data. However, one of the primary concerns in determining how best to summarize data by time period is the number of cases that are available for analysis. Thus the choice of time period intervals normally is based on the length of time the registry has been in existence and the size of the population covered. Because cancer is a chronic disease, analyses usually are based on calendar year of diagnosis. Cancer registries that have covered large populations may well have sufficient data to evaluate time trends in cancer statistics on a year-by-year basis. In contrast, registries with a small population base will have insufficient data to present time trends in such detail. Three- or 5-year averages are useful especially where the population covered is not large. Year-to-year comparisons can be misleading because of normal variations. Time trends over a decade or longer are more meaningful. Registry staff **SHOULD** consult with an

experienced epidemiologist, biostatistician, or demographer to determine how best to present temporal trends in cancer statistics

(4) Standards for Geographic Area Categories

Generally, geographic areas used are political entities, such as provinces, cities, and villages.

(5) Standards for Grouping by Stage of Disease

The MECC has recommended as a standard grouping the SEER summary stage.

2. Statistical Methods

a) Introduction

It is important to consider each of the methods outlined below in the context of three key elements of epidemiologic inquiry: person, place, and time. Analyses are usually based on cancer cases (i.e., independent primary cancers [a person can have multiple primary cancers]); however, some analyses focus on persons rather than cases.

Person: Reports of cancer data **SHOULD** document the demographic characteristics of the cases represented in the report. At a minimum, these characteristics **SHOULD** include sex, age, and race/ethnicity. A person may be represented by more than one case (i.e., more than one primary cancer) in the registry's files. **Place:** Reports of cancer data **SHOULD** specify the geographic area of coverage for the cases represented in the report. Typically, the area of coverage follows political boundaries such as provinces, cities, and villages.

Time: Reports **SHOULD** clearly state the relevant time period of study. Cancer statistics usually are reported annually, based on the diagnosis year, not the year reported. When studying trends of

cancer over time, it is useful to combine as many years as available to avoid misleading fluctuations arising from small numbers. Use of 3- or 5- year time periods centered on a census year is a useful strategy for dealing with population estimate problems.

b) Standards

(1) Standards for Counts

The most basic unit of measure for cancer registry data is the simple enumeration of cases. Knowledge of the number of cancer cases can be of great use for health planning purposes where it is important to measure the burden of cancer on existing health care resources and to assess the need for additional resources. However, simple counts of cases are of limited value as a measure of disease risk, for which incidence rates are preferable.

(2) Standards for Proportions

Simple Proportions: Simple proportions are useful for describing basic characteristics of registry data. Examples include:

- Percent distribution of cases by stage of disease at diagnosis
- The percentage of cases with histologically-confirmed diagnoses
- Percentage of cases which received a given treatment modality.

Percent Distribution by Site: A percentage distribution by site is useful for showing at a glance which cancers account for the majority of cases. Usually, cancer of the breast, lung, colorectum, and prostate will together account for well over half of all cancers, with each accounting for 12 to 15 percent of all cancers. This is a useful distribution to present so that non-population-based registries can assess whether their data represent a skewed distribution of cases.

Proportional Incidence: As outlined below, incidence rates are the measure of choice for expressing disease risk; however, appropriate

population estimates are not always available to serve as the denominators for rate calculations. In these instances, the proportional incidence ratio (PIR) may serve as a useful way to compare risk of disease in two populations. This measure compares the relative importance of a specific cancer in relation to all cancers in two groups. The PIR is not a rate, because the denominator is derived from the number of cancers and not from the population at risk. The PIR is calculated using the proportional distribution within a defined group to estimate the expected proportion in another group. The observed proportion then is compared to the expected proportion as an estimate of risk. Specifically, the proportion of all cases accounted for by a specific site is calculated for each age and sex group in the "standard" population. These proportions then are applied to the number of all cancers in each age and sex group in the comparison population to estimate the number of expected cases of that type by age and sex. Expected numbers are summed across age and sex groups to obtain an "age-adjusted" expected number of cases. The ratio of the observed cases compared to the expected cases gives the PIR. The PIR generally is multiplied by 100; a PIR of greater than 100 indicates that the observed proportion was greater than the expected proportion, and usually indicates an increased disease risk.

(3) Standards for Incidence Rates

General Considerations: The incidence rate is the most appropriate and useful measure of disease risk. Incidence rates express the number of new cases of disease diagnosed in a population with respect to the size of the population and the time period under study.

Case Selection Criteria: When selecting cases for incidence rate calculations:

- Include only resident cases first diagnosed during the selected time period.
- Count resident cases reported through death certificates only as incident at the date of death.
- Include resident cases discovered at autopsy.

Incidence rates are commonly expressed as follows:

• **Crude Incidence Rate**: The simplest incidence rate, obtained by dividing the number of new cases by the size of the population at risk of developing cancer during the study period. The crude rate does not take into account the age distribution of the population; therefore, they are not suitable for comparison across place and time.

• **Age-Specific Incidence Rate**: The age-specific incidence rate is the crude incidence rate for a defined age group.

• **Age-Adjusted Incidence Rate:** The age-adjusted incidence rate is a summary measure that is a weighted average of age-specific incidence rates with the weights being proportional to the number of people in each age group.

• **Cumulative Incidence Rate**: This term is a misnomer in that it is not a true “rate” of disease as outlined above. Rather, it is an estimate of the risk of developing cancer during a given time period, expressed as a percentage.

Standard Population: The choice of an appropriate standard population is an issue in the calculation of age-adjusted incidence rates. The choice of data for comparison may dictate the choice for standard population.

World Standard: Another common comparison population, and the one used in the WHO's Cancer Incidence in Five Continents, is the world standard used by the IARC. This is useful for international comparisons.

Other Standards: Use of a recent or midpoint standard population may be appropriate for special studies.

Denominator Data for Rate Calculation: One of the most important steps in calculating incidence or mortality rates is to obtain appropriate population estimates to serve as the denominator for the rate calculation. These estimates represent the population at risk. For a central cancer registry, these estimates would represent the population that resides within the registry's designated coverage area. For incidence rates, the population estimates should correspond to the population that resides within the registry's capture area for the time period during which the newly diagnosed cases of the disease were identified in the population.

Guidelines for Incidence Rate Calculations: When calculating incidence rates for the registry as a whole or for any geographic area within the registry's area, the registry **SHOULD**:

- Tabulate cases with unknown age, sex, or geographic area of residence separately. They **SHOULD** be excluded from rate calculations where appropriate, and the report **SHOULD** show the number of cases that were excluded because of the unknown data
- Explain in footnotes the variability in rates based on small numbers of cases. Units of Measure: Cancer incidence rates are most commonly expressed per 100,000 population per unit of time. Some rare cancers (childhood cancers, for example) are expressed per 1,000,000 population per unit of time.

(4) Standards for Mortality Rates

Mortality rates are most often reported by local health agencies or bureaus of vital statistics based on information reported through death registration. However, because of their expertise and focus on cancer, central cancer registries sometimes are called upon to calculate cancer mortality rates as well. If so, cancer mortality rates **SHOULD** be based on the underlying cause of death as reported through the death registration process. As with incidence rates, mortality rates can be expressed as crude, age-specific, and age-adjusted. The methods out-lined above for incidence rates also are applicable to mortality rates. The population estimates used **MUST** correspond to the same time period during which the deaths of interest occurred. The accuracy of mortality rates as a measure of cancer occurrence has been shown to vary by type of cancer. For this reason, caution **SHOULD** be exercised in the use and interpretation of cancer mortality rates.

(5) Standards for Survival Analysis

Survival analysis entails measuring the length of time between two events. Most frequently for cancer registries, the initial event is the date of cancer diagnosis, and the second event is a subsequent outcome, such as death. Survival rates often are used as an index of the quality of care following a diagnosis of cancer.

Data Requirements: The following data items are the minimal requirements for calculating survival rates:

- Date of Initial Event, usually Date of Cancer Diagnosis
- Date of Subsequent Outcome, such as Last Follow-up: The date of last follow-up represents the calendar time at which information was last obtained on the subject. If the patient is deceased, the date of last follow-up is the date of death. The accurate ascertainment of the date of last follow-up for all cancer patients is a key factor in the

validity of survival analyses. When survival to recurrence of cancer is being calculated, it is the date of recurrence that is used as the subsequent outcome.

• **Vital Status or Other Status:** Vital status describes the last known condition of the subject. At a minimum, this item **SHOULD** indicate whether the subject was alive or dead at the date of last follow-up. Some methods of survival analysis require knowledge of the cause of death. When survival to recurrence is being calculated, the patient's recurrence status is used instead of vital status.

Standard Methods: Four standard methods of survival analysis are described below.

• **Observed Survival Rate:** The observed survival rate is calculated by the life-table (actuarial) method. This method provides an estimate of the probability of an individual surviving to the end of a specified time interval, given that the person was alive at the beginning of this interval.

• **Relative Survival Rate:** The relative survival rate also is calculated by the life-table (actuarial) method. This method adjusts the observed survival rate to account for other causes of death that would be expected if the study subjects experienced the same mortality rates as the general population of similar age, race, sex, and calendar period of observation. By adjusting for other causes of death, this method attempts to estimate the effect of the cancer alone on survival. What this method does, in fact, is to measure the excess mortality that the cohort experiences in comparison to the general population. The accuracy of this method is then a function of how the study subjects differ from the general population. If the only difference is the fact of cancer, then this method works well. One notable exception is lung cancer, where the cancer cohort also is at

excess risk of death from heart disease compared to the general population due to a large number with a history of smoking.

• **Kaplan-Meier:** The Kaplan-Meier Method, also known as the product limit method, is a special case of the standard life table technique used for survival analysis. Kaplan-Meier is computationally similar to the standard life-table method, but the intervals of survival time are defined differently for the two methods. In the Kaplan-Meier Method, a calculation (of the observed survival rate) is done every time a patient dies rather than during a specific regular interval, such as a year or month. Thus, it gives a more exact description of the pattern of survival. The graphic display of survival rates derived from Kaplan-Meier is particularly useful for determining the median survival time and for comparing the survival experiences of two or more groups of patients. Because multiple calculations are required, the Kaplan-Meier Method generally is used when the number of patients is small, generally 25 to 30, as is usually the case in clinical trials. Statistics texts should be consulted for more details.

• **Cox Proportional Hazards Model:** The Cox Proportional Hazards Model allows for the comparison of survival rates between two or more groups, with simultaneous adjustment for potentially confounding variables.

Interpretation: The survival rate is a difficult measure to interpret. Survival from cancer is determined by many factors, including the patient's age, stage of disease at diagnosis, histologic type of cancer, treatment, and the presence of other illnesses. Comparison of survival rates among institutions or geographic areas may be difficult to interpret, especially if the respective patient populations differ with regard to prognostic factors. Survival data from cancer registries generally are considered inappropriate for assessing the efficacy of treatment modalities. Well-designed randomized clinical

trials are the most appropriate tools for evaluating the efficacy of cancer therapy. Calculation, interpretation, and reporting of survival rates **SHOULD** be undertaken only under the supervision of a qualified biostatistician or epidemiologist with expertise in survival analysis.

3. Reports

a) Introduction

The dissemination of cancer data is an important function of the central cancer registry, and is one of the primary means by which it is known in its community. Registry data may appear routinely in a standard format or may be prepared on an ad hoc basis in response to specific inquiries. The reputation and usefulness of a central cancer registry often is judged by the accuracy, timeliness, and clarity of its reports. In designing reports, it may be useful to compare one registry's experience with similar data from other cancer registries. Similarly, it may be helpful to design reports that are comparable within a registration system. Registries **SHOULD** obtain copies of reports and newsletters from established registries to use as models when developing their own publications. Most cancer registries are pleased to include other registries in the routine distribution of their reports and newsletters.

b) Standards

(1) Standards for Types and Frequency of Reports

Summary of Central Registry Data: Central cancer registries **SHOULD** assemble a comprehensive summary of the cancer experience within their area of coverage. At a minimum, the report **MUST** tabulate cases by primary site, by sex, race, and by age group, as well as possibly by stage, using standard recode groupings (analysis categories) for each. In addition, these reports **SHOULD** provide

population-based incidence and/or mortality rates, tabulated by site groupings, age, and sex. Survival rates also may be provided in these reports. Where possible, incidence, mortality, and survival rates **SHOULD** be displayed by ethnicity, race, and stage. If the registry has been in existence for a sufficiently long time period, and if the number of cases permits, the report **SHOULD** include temporal trends in cancer incidence, mortality, and survival rates. Some registries may elect to provide similar information by geographic area. Summaries of cancer registry data **SHOULD** be published annually.

Tables to be included in annual report:

(a) For all sites combined. Latest year of data by sex (including both sexes combined) and age. Show 18 (or 16 if populations not available) age groups. Show numbers and rates. For all ages combined – show crude rate and rate age-adjusted to the world standard. (For Israel, separate tables for Jews, Jews and Others, and Arabs, all other registries – residents only).

(b) For all sites combined. Same as above except for five-year average (or three-year) depending on available data.

(c) Table showing top 5 (or 10) sites in males and top 5 (or 10) in females.

(d) For last 5-year or 3-year average annual incidence rates By ICD-10 site group from CanReg and sex (including both sexes combined) rates age-adjusted to the world standard population (For Israel, separate tables for Jews and non-Jews, all other registries – residents only).

(e) For children age 0-19. For last 5-year or 3- year average annual rates By International Classifications of Childhood Cancer recodes Show 3 age-specific rates, crude rate and age-adjusted rate by sex (including both sexes combined).

(f) For each of top 5 sites in males and top 5 sites in females.

(2) Standards for Narrative Text

General Considerations: An important component of any report is the narrative text that accompanies the presentation of the data. As outlined below, the narrative guides the reader by documenting methods used to produce the report, highlighting important findings, and interpreting the results.

Documentation: One of the primary functions of the narrative is to document the methods by which the data were collected, compiled, and analyzed.

• The report **SHOULD** include an overview of the registry's data collection methods.

• The narrative **SHOULD** specify the classification systems which were used to collect, code, and tabulate the data (e.g., ICD-O for cancer diagnoses, and ICD-10 for mortality diagnoses).

• The report **SHOULD** clearly identify any recodes used and the statistical methodology that was used to conduct the analysis and prepare the report. References to more detailed descriptions of methods **SHOULD** be cited when the methodology cannot be fully described in the report.

• The report **SHOULD** identify the geographic area of coverage of the central cancer registry, as well as any geographic areas on which the report may specifically focus.

• The report **SHOULD** clearly state the time period for which cases are tabulated.

• When incidence and/or mortality rates are presented, the narrative **SHOULD** document the source of the population figures that were used to calculate the rates. If age-adjusted rates are included, the report **SHOULD** indicate the choice of standard population. A separate table of the relevant population figures, including the distribution of the standard population, **SHOULD** be provided.

Highlighting and Interpreting the Results: Few things are more daunting than a myriad of numbers in tabular format. Even the most interested observer may overlook some potentially important findings. Furthermore, central registry staff may be aware of facts and trends, unknown to the general readership, that may aid in presenting and interpreting the results presented in the report. For this reason, the narrative text **SHOULD** provide guidance to readers in the interpretation of the data. Artifacts that may lead to a misinterpretation of the data are especially important to document and **SHOULD** be noted in the report. These might include changes in data collection procedures and changes in disease classification. Similarly, changes in diagnostic methods or procedures may affect the numbers of cases being diagnosed or their classification. The reader **SHOULD** be cautioned against drawing definitive conclusions when the measures are based on small numbers.

Quality Indicators: The report **SHOULD** address what is known about the completeness and accuracy of the data in the report, and may specifically include information such as the following (see also Sections I.C. and II.C for other suggested measures that could be included):

- Percent DCO
- Percent microscopically confirmed
- Incidence/mortality ratios.

(3) Standards for Displaying

Data Tables: Numerical data often are displayed in tabular format. Tables **SHOULD** be able to stand alone; that is, they **SHOULD** be fully comprehensible apart from the narrative text. Descriptive titles, headings, and foot-notes are used to explain the contents of the table. Consideration **SHOULD** be given to rank ordering summary tables.

Graphs and Charts: The graphical presentation of data often is more intuitively appealing than a table full of numbers. However, 3-D charts or graphs **SHOULD** NOT be used when presenting data, because the depth of lines or bars can be misleading. Some of the most common types of graphs are listed below.

• **Line Graphs:** Line graphs are constructed by plotting the values for two variables on an x-y axis, and then connecting the points. Line graphs are most often used to display age-specific incidence rates and time trends in age-adjusted incidence rates. When choosing the scale of the y-axis for presenting time trends, a decision needs to be made whether the absolute change or the rate of change is of more interest. Rates of change can only be shown on a logarithmic scale.

• **Bar Graphs and Histograms:** Bar graphs and histograms use horizontal or vertical bars to represent frequencies or proportions. The differences between bar graphs and histograms are outlined below; however, these differences sometimes are obscured in practice. Bar graphs are used to present discrete data that are nominal or ordinal. Consideration **SHOULD** be given to rank ordering bar graphs, to increase the information being conveyed. Histograms are used to present grouped continuous data. The bars in a histogram are adjacent to one another, indicating the continuous nature of the data.

• **Pie Charts:** Pie charts are used to display percentages. To construct a pie chart, a circle is divided into segments, like slices of a pie, to represent various contributions to the whole.

• What the entries in the tables, charts, or maps are (e.g., number of cases, percents, rates, ratios, etc.)

(4) Standards for Review of Reports

The central cancer registry **MUST** designate staff members to review all routine reports and responses to requests for information before the information is released to assure that confidentiality is protected. All questions regarding the quality of the data **SHOULD** be brought to the attention of the quality control staff and **SHOULD** be resolved before the data are released. All questions regarding the appropriate interpretation of registry data **SHOULD** be brought to the attention of appropriate staff and **SHOULD** be resolved before the data are released. Because of the possible ramifications for the registry, participating facilities, and its parent organization, the registry director or designate **MUST** review and approve all information that is released to the news media. The registry director or designate also **SHOULD** inform the appropriate superiors and data providers before release so that they will be able to answer any subsequent questions from the press or the community.

IV. DATA MANAGEMENT

A. STRUCTURAL REQUIREMENTS

1. Data Management: General Requirements

a) Introduction

For cancer registries, the advancement in computer software and hardware has increased the efficiency of data collection, increased data quality and standardization, increased accessibility of data, and made large-scale national and international collaborative pooling of data possible. Computers have enhanced the users' ability to appropriately examine the rich source of data that registries represent. The full potential of these advances has been only partly employed by some registries and not used at all by others; yet the ability of cancer registries to remain cost-effective and affordable and to continue to supply relevant answers to important scientific, clinical, and policy questions depends on their continued exploitation of advances in computer technology and communications. Computers have allowed registries to perform more work with the same or fewer staff, and they have changed every aspect of their structure and operations. At the same time, cancer registries continue to face hiring freezes and cuts in their operating budgets while the number of cases they collect increases. In order to thrive, cancer registries must enthusiastically embrace creative uses of computer technology and exploit their potential in all its forms.

Section IV describes specific functional requirements, system design considerations, software and hardware requirements, and other features that are important to fulfilling the functions of a cancer registry and that any cancer registry **SHOULD** be able to perform. The

words “computer system” or “system” in this section generally refer to the complete system, including the hardware and software (i.e., the equipment and programs). This section will not recommend specific software or hardware. The technology will not remain static, and many future advances will be useful to cancer registries. Thus, it is the goal of this section to outline a set of general functional requirements that each registry **SHOULD** meet, and to encourage every cancer registry program not only to include these functions, but to go well beyond them. This section specifically addresses central registries at state and provincial levels, and those central registries at a regional level within a larger central registry system. Requirements for systems at a national level may vary somewhat from those stated here, and these differences are not addressed. This section does not address general-purpose computer tools such as word processing, accounting, spreadsheets, or desktop publishing; although it is assumed that the registry will require a wide variety of computer resources beyond those which are addressed in this section.

Overview of Major System Functions The utility of a cancer registry system **SHOULD** be measured by the ability of a given hardware and software combination to effectively accomplish those tasks assigned to a registry. A registry **SHOULD** be designed not only to collect accurate, error-free data, but to provide appropriate means for reporting and analysis and for communications with national collaborative efforts. A registry data processing system **SHOULD**:

- Provide multiple modes of data interfacing including data entry
- Support means for appropriately linking patient data with hospital and other data
- Help ensure data integrity, completeness, and accuracy
- Produce standard reports
- Provide tools for ad hoc analyses, lists, and reports

• Support appropriate security
• Be cost-effective and affordable
• Be dynamic (i.e., easily and inexpensively changed over time)
• Have adequate performance that supports timely data entry, analysis, and reporting.

Importance of Standards

For reasons of efficiency and comparability, it is important for registries to adopt existing standards where they exist, and to actually use existing resources in their systems. Few registries can afford the de novo development of their system, and even if they could, it would be less likely to be compatible with other systems. Idiosyncratic systems are more costly to maintain and enforce hidden costs in non-compatible data.

b) Standards

(1) Standards for Functional Requirements

The major functions of a central registry system are:

Support for All Registry Activities: The cancer registry's computer system **MUST** be able to support the efficient and effective execution of all of the tasks in Sections I, II, and III, including routine operations, analyses, reports, quality monitoring, communications with facilities and providers, etc.

Computerized Data Collection: Abstractors employed by the central registry and those in reporting facilities **SHOULD** use computerized data collection software for abstracting case data from source documents. The software **SHOULD** include standard features.

(2) Adherence to Standards

MECC Data Standards: The system **SHOULD** meet all of the standards specified in MECC Data Standards and Data Dictionary including the items collected and their codes and formats.

Standard Edits: The registry **SHOULD** use standard data edits.

Analysis Standards: The system **SHOULD** provide the capability to produce analyses using all of the standards described in Section III, including:

- Use of standard analysis categories
- Application of standard statistical methods
- Provision for use of multiple population standards
- Production of standard reports.

2. Staffing Guidelines for Data Management

a) Introduction

The computer and data management staff at the registry are in a crucial position to influence the overall success of the registry. The lead computer staff person **SHOULD** be considered a part of the registry's leadership and **SHOULD** be involved in planning and overall system design.

b) Standards

(1) Standards for Number and Type of Staff

The registry **MUST** provide data management staff sufficient in number and training to assure compliance with mandated reporting requirements, assure timely completion of all required tasks and reports, and meet all other standards. It is desirable that the data management staff have a background in health applications as well as the requisite technical knowledge. Registry personnel **MUST** be

sufficiently trained and cross trained in the operation of the system to protect against the possibility that the loss of a single person would cripple its operation.

(2) Standards for Continuing Education

Staff involved in data managing and data processing **SHOULD** be offered opportunities for continuing education so that the registry staff remains informed of technological changes.

Continuing Education: Continuing education **SHOULD** be provided to data management staff to assure that they have up-to-date knowledge about available technologies and cancer registries.

Access to Professional Literature, Online Services, and Other Activities: Data management staff **SHOULD** be supplied with appropriate references and literature to provide ongoing continuing education and to answer questions that arise. Current pertinent reference books, journals, and other periodicals **SHOULD** be immediately available. The registry also **MAY** provide access to online services and bulletin board services so that staff have rapid access to the most current information.

Professional Associations and User Groups: Staff **SHOULD** be encouraged and funded to participate in local and national professional associations and user groups pertinent to their technical area, and also in registry-oriented scientific meetings. The registry budget **SHOULD** include funds for participation by one or more persons at scheduled meetings. The registry **SHOULD** fund data management staff to attend trade shows, special symposia, conferences, and courses that may be offered from time to time.

B. PROCESS STANDARDS

1. Data Entry

a) Introduction

Data entry of cancer case abstracts is most often part of the process of abstracting directly onto a computer. Computerized data collection combines abstracting, coding, data entry, editing, and accessioning into one process. Some central registries provide software to reporting facilities to standardize this process. In addition, however, the registry probably will employ a variety of data entry methods for some new case abstracts; for corrections, deletions, or other transactions; or for physician and hospital data. These methods can include direct keying from source documents into the computer, key entry from data collection forms, and other methods. Regardless of the methods used, some form of verification of the keyed data **SHOULD** be in place.

b) Standards

To minimize keying errors, the registry **SHOULD** implement some form of verification of keyed data. The method will vary with the data entry method, and may include visual comparisons; duplicate keying when manual forms are used; extensive editing and analysis of input data; or other quality reviews.

A report detailing questions that arise during attempted correction of case data (e.g., edit failures that cannot be corrected at the central registry for lack of information) tumor status of a patient, that information **SHOULD** be sent to Dr John Young at Emery University.

Required Processing Functions: The registry system **SHOULD** have the capacity to perform the following functions regarding follow-up input files:

• **Linkage**: Provide the ability to link an incoming follow-up record with the appropriate database case.
• **Editing and Automatic Updating**: Provide the ability to automatically apply an incoming follow-up record to the database case, when appropriate, after editing for compatibility and consistency.
• **Error Reports**: Produce error reports for incoming follow-up records failing edits.
• **Management Information**: Provide the means to identify database cases where follow-up information has been changed and provide appropriate management reports.

(1) Standards for Correction Data Input Files

Definition: In addition to its own correction procedures for individual records, the registry **MAY** receive files of corrections from reporting facilities that have made changes to previously reported cases. These files contain the changes made to required data items after the case information has been transmitted to the registry.

Required Processing Functions: The registry system **SHOULD** have the capacity to perform the following functions regarding correction input files:
• **Linkage:** Provide the ability to link an incoming correction record with the appropriate database case.
• **Editing and Updating**: Provide the ability to either manually or automatically apply an incoming correction record to the database corresponding database case after editing for intra-field and inter-field consistency
• **Error Reports**: Produce error reports for incoming correction records which are failing edits.
• **Management Informatio**n: Provide the means to identify database cases where information has been changed and provide appropriate management reports.

(2) Standards for Deletion Data Input Files

Definition: Input files contains information on previously reported cases that have been deleted by the registry.

- **Linkage**: Provide the ability to link an incoming deletion record with the appropriate database case.
- **Reports**: Produce reports from incoming deletion records containing case identifiers and reason for deletion.
- **Manual Processing**: Provide the ability to manually delete a database case. Do not reuse deleted numbers.
- **Management Information**: Provide the means to identify deleted database cases and provide appropriate management reports.
- **Restore**: Provide the ability to restore a case mistakenly deleted.

(3) Standards for Death Clearance Input Files

Definition: Death clearance processing involves use of data about residents for whom death certificates were filed. The purpose is to provide new information about previously reported cases (follow-up) and to obtain new case information for previously unreported patients or cancers (follow-back).

Required Processing Functions: The registry system **SHOULD** have the capacity to perform the following functions regarding death clearance whenever possible:

- **Linkage**: Provide the ability to link an incoming death certificate record to the appropriate database case.
- **Editing and Updating**: For death certificate records that link to database cases, provide the ability to automatically apply the incoming death information to the database cases, when appropriate, after editing for compatibility and consistency; and to update other items coded in the death record, such as race and birthplace when the database case contains unknown or nonspecific values and the death record is more specific

• **Error Reports**: For linked death records failing edits, produce error reports.
• **Suspense**: For death certificate records that do not link to database cases but are cases that should have been reported, provide the ability to suspend the death records in the database for further follow-back investigation
• **Management Information**: Provide the means to identify cases where death information has been applied to the case or entered in a suspense file and provide appropriate management reports.

(4) Standards for Limited-Case-Information Input Files

Definition: These files contain limited information about cancer cases. The case may not have been reported yet because it is not yet complete (e.g., a case identified through rapid case ascertainment); the case may have been ascertained from a source with limited information necessitating follow-back to other sources (e.g., a case identified through a pathology laboratory); or the case may have been overlooked by the facility responsible for reporting it. Required File Processing Functions: The registry system **SHOULD** have the capacity to perform the following functions regarding limited-case-information input files:
• **Editing:** Edit the incoming data for very basic content.
• **Suspense:** Provide the ability to suspend the case records in the database for further investigation.
• **Reports**: Provide reports of the suspected cases according to the source to which they need to be followed back and prepare inquiries to the appropriate sources.
• **Linkage**: Provide the ability to periodically link the limited information records with the database cases so that the limited information records can be deleted if the cases have been added to the database from another source.
• **Deletion:** Provide the ability to delete a limited information record if the case is found to be non-reportable.

• **Management Information**: Provide the means to identify disposition of limited information cases and provide appropriate management reports.

2. Outputs

a) Introduction

In addition to analytical reporting covered in Section III and input processing covered in Section IV.B.1., the registry's computer system **SHOULD** be able to provide several different types of outputs: Management reports that allow for monitoring of the database and registry operations. Standard reports to give feedback to or request information from reporting sources. Output that responds to ad hoc queries from quality control operations, management staff, and others.

b) Standards

(1) Standards for Management Reports

The registry **SHOULD** produce management reports with a frequency that will enable monitoring the operations of the registry. Examples of possible reports include: A table presenting the number of cases reported for each reporting facility and for other sources of cases (such as DCO cases, or physician-only cases whenever available) by month and year of admission (or, for DCO cases, month and year of death). A table presenting the difference between the number of cases expected from each reporting facility and the number received. By ordering the table in descending order with the facility with the largest deficit on top, this report helps to allocate registry resources to the area with the greatest impact. A table presenting the cases from all sources by month and year of diagnosis. A table presenting the distribution of cases by year of diagnosis by site for comparison with other registries. To monitor workflow, a table presenting the

number of cases by process completed (e.g., number inspected or visually reviewed, number in suspense, etc.), by date received in the registry. To show timeliness of abstracting, tables showing the interval between diagnosis date and date abstracted, and between diagnosis date and the date the case was entered in the central registry system, by facility. For registries collecting patient follow-up, tables showing the status of follow-up by facility and by diagnosis year, and for subpopulations of interest (e.g., specific age groups).

(2) Standards for Reports to Facilities

The registry's data processing system **SHOULD** provide for a variety of reports to be routinely prepared and distributed to the facilities submitting cases to the registry. These reports can be transmitted to the facilities electronically or in hardcopy.

Reports for Monitoring Workflow and Completeness: To provide information to the reporting facilities about their caseload, or about the completeness of their reporting, reports such as the following are useful:

- Immediate or very rapid acknowledgment of the central registry's receipt of a case submission, so that the facility can verify that its cases were received and were readable.
- A table presenting the number of cases from that facility by month and year of admission.

(3) Standards for Ad Hoc Queries

The system **MUST** allow for easy routine querying of the database by management and quality control staff at the registry, without programmer intervention. The results from ad hoc queries may take the form of interactively displayed reports on the screen or printed output. The types of output available **SHOULD** include:

Listings: The system **SHOULD** be able to provide listings of rows in the database that meet specified criteria and are sorted as needed by the user. On a screen display, the user **SHOULD** have the ability to scroll through the rows. As an example, in revolving linkage problems manually, it often is necessary to query the database using alternate spellings, phonetic compression, or incomplete values for given fields, and to review the records retrieved.

Patient-Tumor-Admission Displays: The system **MUST** allow display on the screen all the data values that are stored for a specific patient, tumor, or admission.

Frequencies: The system **SHOULD** allow easy output of frequencies or counts by any variable or combination of variables. To prevent users who do not fully understand the organization of the data from obtaining misleading results, it is useful to require that the user provide answers to a series of questions before the count is generated, specifically:

- Should the results be limited to a certain time period?
- Should the results count patients, tumors, or hospital reports?
- Should the results include *in situ* diagnoses, invasive diagnoses, or both?
- Should the results be limited to residents of the registry's coverage area?
- Should DCO cases be included?

3. Record Linkage

a) Introduction

When data are added to the registry's database, whether in cases of adding data to an existing record, or in cases of adding new records, a suitable record linkage mechanism is needed to assure that the additional data are correctly associated with the existing data. If a

record is added to the database without adequate checking for redundancy, incidence rates may be overestimated due to the fact that a single tumor may be reported by multiple institutions. However, if efforts to prevent duplicate records are overzealous, then truly distinct records can be linked together mistakenly, resulting in under estimation. In any of the above situations, the probabilities of falsely matching records increases, diminishing the quality of the database and resulting in incorrect incidence rates. Statistically speaking, an erroneous record linkage increases the type I error and type II errors that are associated with it (the probability of accepting a match given it is the wrong match and the probability of rejecting a match given it is a true match, respectively).

Types of record linkage A record linkage can be performed deterministically or probabilistically. A deterministic record linkage involves the comparison of two records on several key fields (e.g., last name, first name, etc.). A match is achieved if and only if all of the key fields coincide on both records. Any typographical errors or missing information in any of the fields results to a non match. Therefore, a deterministic record linkage is suitable for records with no errors or missing data. A probabilistic record linkage involves the comparison of two records on several key fields as well. Additionally, a probability is associated with a correct and a false match. This usually is achieved by building a scoring algorithm based on the number of fields that coincide in both records and the degree of trust in these fields. In essence, this type of linkage assimilates an individual's thought process if the linkage were to be performed manually. At the same time, it allows assessing a degree of trust in the linkage.

Linking patients versus linking tumors.

The key fields used for the record linkage should be analyzed before use to ensure that they are reliable. Items like names, sex, social

security numbers, phonetic comparison indices, date of birth, or county of residence can be used for record linkage at the patient level. Additional information like address, marital status, etc., can be used for questionable linkages that need to be reviewed manually. Multiple submitted cancer records for the same patient need to be linked as well. Records that describe the same tumor must be identified so that they can be consolidated; records describing separate tumors for the same patient need to be stored as separate cases. The task of tumor consolidation is harder to fully automate; it involves comparisons of the primary sites and the dates of diagnoses. Complications that have to do with assigning a morphology to a tumor and the ambiguous rules in determining the date of initial diagnosis can make this procedure cumbersome and may require more manual intervention. Several registries are working on site/morphology tables to overcome this obstacle.

Software Commercial record linkage software is available. Some commercial packages provide a score that reflects the degree of certainty for a possible linkage, while allowing for the manual review of the questionable linkages. The selection of the key fields and the compilation of the algorithm are determined by the user.

b) Standards

The registry must have an effective record linkage system for linking patients and cancers. Record linkages can be done manually, by computer, or by a combination of both. Small and well-funded registries can afford the employees necessary for manually linking their cancer records. However, for large or underfunded registries, this is an impossible task.

4. Edits

a) Introduction

Computer edits are a key aspect of the registry's overall computer system. Edits are a part of quality control.

b) Standards

The registry system **SHOULD** employ standard edits whenever possible. Edits **SHOULD** be applied as physically close to the information source as possible, and as temporally close to the collection of the data as possible. Edits **MAY** be performed interactively, as a batch process, or both, and **MUST** be applied at several points in the data flow: To newly submitted case records before they are linked against the database To database cases after linkage To database cases after any changes have been made. The registry's edits **SHOULD** allow for some override flags for situations in which the edit identifies a rare condition that needs review but may be correct. The override flag prevents the condition from continuing to be identified as an error. In error reports and discussions with abstractors and coders, it **MAY** be helpful to label data failing edits as "inconsistencies" rather than "errors," because the data are not necessarily incorrect.

5. Record Consolidation

a) Introduction

Consolidation refers to the process of reconciling or compiling data obtained from more than one source on the same person or tumor. The sources can include multiple abstracts from hospitals, clinics, or other providers, or they can include information from the death records or from other registries. Values for the same data items for the same patient and tumor may be identical from each source, but they also may be contradictory or complementary. A large task of the

central registry system is to prepare a composite set of values for each patient and tumor, incorporating information from a variety of sources. This composite set of values then can be stored and managed in a variety of ways, either as a separate consolidated record, or with the individual values in different records flagged as those to be used for the consolidated record. In any case, the original records always **SHOULD** be kept intact. It is important to recognize the difference between record consolidation and the identification of multiple tumors for the same patient.

b) Standards

Standards for item-specific consolidation rules, either for computer application or manual application, have not been developed, but many existing systems can be used as models. Some general principles can be stated; Where it can be ascertained (in a cost effective manner), the best, or true, value for each item is the one that **SHOULD** be retained. The system **SHOULD** perform automatic consolidation whenever possible, and produce a report of the computer's actions for manual review, but also **SHOULD** be able to identify instances where the computer cannot determine the correct value. Known values are preferred over unknown values, and more specific values are preferred over less specific values.

6. Guidelines for Processing Follow-up, Correction, and Deletion Transactions

a) Introduction

The cancer registry database is a dynamic one. Unlike some applications, the data are never final. Cases continuously are added, changed, and deleted as long as the registry continues, even after patients have expired and the data have been included in reports. The registry's system will need to process follow-up, correction, and deletion transactions. Good data collection software for abstractors

automatically will generate these records for the central registry when the changes are made to the local database. Ideally, the central system should handle these automatically; however, some problems arise when conducting automatic updating, especially when combining data from multiple hospitals and multiple software systems.

7. Linkages with External Files

a) Introduction

Linkage of the registry database with non-registry files serves several purposes for the registry. For example, there may be external files that can provide follow-up for the registry's cases, or there may be special research studies requiring the linking of a cohort against the registry database.

b) Standards

The registry **MUST** develop the technical, procedural, and administrative capacity to perform linkages with external files. Examples include: Linkage With Death (Mortality) Files and other files such as population registries.

8. Documentation

a) Introduction

Good documentation is an essential aspect of a well-designed system. It is necessary for system maintenance, training, quality control, and security; yet it often is incomplete and out-of-date. Documentation deserves to be high among the registry's priorities.

b) Standards

Adequate staff and time **MUST** be provided to prepare and maintain high quality, up-to-date system documentation. The system

documentation **SHOULD** include a management-level, functional description of the system, including a comprehensive narrative and flow diagrams. In addition, manuals or subsets of the documentation **SHOULD** be produced for the system as follows:

User Manual: The user manual **SHOULD** describe the user interface with the input, processing, and output of the system.

Technical Manual: The technical manual **SHOULD** provide information to computer-trained personnel about the design and software of the system. It **SHOULD** contain system flowcharts defining major components of the system, definitions of individual programs, numerical analyses defining special calculations, definition of inputs and outputs, and definitions of reports.

Operator Manual: The operator manual **SHOULD** describe the database and security and recovery procedures for the system. It **SHOULD** contain error codes/messages and handling procedures, computer run instructions, definition of file retention and backup procedures, and definition of data security. Documentation **MAY** be available online as well as in hardcopy form.

V. APPENDICES

A. REPORTABILITY

REPORTABLE DIAGNOSES

1. In Situ and Malignant/Invasive Histologies

a. All histologies with a behavior code of /2 or /3 in the *International Classification of Diseases for Oncology,* Third Edition (ICD-O-3).

b. *Exceptions:* Malignant and invasive histologies not required by MECC

i. Skin primary (C440-C449) with any of the following histologies:

1. Malignant neoplasm (8000-8005)
2. Epithelial carcinoma (8010-8046)
3. Papillary and squamous cell carcinoma (8050-8084)
4. Basal cell carcinoma (8090-8110)

ii. Carcinoma **in situ** of **cervix** (/2) or cervical intraepithelial neoplasia (**CIN III**) of the cervix (C530-C539)

iii. Prostatic intraepithelial neoplasia (**PIN III**) of the prostate (C619)

2. Benign/Non-Malignant Histologies

a. **Pilocytic/Juvenile astrocytomas** are reportable; code the histology and behavior code 9421/3.

b. **Benign** and **borderline** primary **intracranial** and **CNS** tumors with a behavior code of /0 or /1 in ICD-O-3 are collected for the following sites;

Cerebral meninges C700
Spinal meninges C701
Meninges, NOS C709

Cerebrum C710
Frontal lobe C711
Temporal lobe C712
Parietal lobe C713
Occipital lobe C714
Ventricle, NOS C715
Cerebellum, NOS C716
Brain stem C717

Overlapping lesion of brain C718
Brain, NOS C719

Spinal cord C720
Cauda equine C721
Olfactory nerve C722
Optic nerve C723
Acoustic nerve C724
Cranial nerve, NOS C725
Overlapping lesion of brain and
central nervous system
C728
Nervous system, NOS C729

Pituitary gland C751
Craniopharyngeal duct C752
Pineal gland C753

Note: Benign and borderline tumors of the cranial bones (C410) are **not reportable.**

B. MAJOR-MINOR DISCREPANCY DEFINITIONS

Data Item Major Difference Minor Difference

Date of Diagnosis

C Different calendar year

C Difference > 1 month

C Unknown (99) versus known month or year

Same calendar year, but difference of 1 month

Sequence Number

Any Difference

Primary Site (ICD-O-3)

C Difference in first three digits

C C34.0 versus C34.1- C34.9

C Exception: C76._ versus C80.9

C C34.1-C34.9

C C76._ versus C80.9 C

For all other sites:

Difference only in third digit

Laterality at Diagnosis

Any difference

Histologic Type

C (9650-9667) versus (9590-9596) or (9670-9729) or (9730-9758)

C For all other histologies: Difference in first three digits

C (9650-9667)

C (9590-9596, 9670- 9729, 9730-9758)

C For all other histologies: Difference in the fourth digit

Behavior

Any difference

Grade

Any difference

For the following sites, a coding difference in grade that results in a change in AJCC staging is a major difference: bone, soft tissue, prostate, brain, thyroid. Review the sites and TNM classifications to determine if a discrepancy is major or minor.

Morphology

Summary

Major coding difference(s) occurred in histologic type and/or behavior

Only minor coding difference(s) occurred in histologic type, behavior, and/or grade

Source: SEER Program.

C. SAMPLE CASE SHARING AGREEMENT

Agreement for the Exchange of Cancer Data Between

the _____ (name of submitting registry) _____

and

_____ (name of receiving registry)

(1) Services: By signing this agreement, the parties state their intention to exchange information concerning cancer patients who are residents of the other's state, province, or county. This exchange is based on the mutual assurance that the identifying information on the patient(s) exchanged are protected and shall be kept strictly confidential. This exchange does not pertain to any data collected as part of special morbidity or mortality studies or other research projects. In addition, the parties agree to:

a) Provide the information electronically.

b) Provide the full exchange record.

c) Provide the information within 20 months of the date of diagnosis. g) Terminate this agreement immediately upon the written notification of either party to terminate the agreement.

(2) Confidentiality:

a) The parties understand and agree that any and all data which may lead to the identification of any patient, research subject, physician, other person, or reporting facility is strictly privileged and confidential, and agree to keep all such data strictly confidential.

b) The parties further agree to require all officers, agents, and employees to keep all such data strictly confidential; to communicate the requirements of this section to all officers, agents, and employees; to discipline all persons who may violate the requirements of this section; and to notify the originating party in writing within 2 working days (48 hours) of any violation of this section, including full details of the violation and corrective actions to be taken.

c) The parties further agree that all data provided under the provisions of this agreement may only be used for the purposes named in this agreement.

(3) Amendments: This agreement may not be amended without prior written approval of both parties to the agreement.

(4) Assignment: The parties understand and agree that this agreement may not be sold, assigned, or transferred in any manner and that any actual or attempted sale, assignment, or transfer shall render this agreement null, void, and of no further effect.

(5) Term: This agreement shall be in effect from the date of execution until terminated by either of the parties. Termination shall be in writing sent pursuant to Section (6).

(6) Notices: All notices required or desired to be made by either party to this agreement shall be sent by certified mail to the following respective addresses: (Provide address and contact for each party to this agreement)

(7) Signatures:

(Provide name, title, agency, date and appropriate signatures for each registry)

D. ACCESS TO NON-CONFIDENTIAL MECC DATA

MECC may grant access to its data to persons who have a valid scientific interest in the data, are engaged in demographic, epidemiological, or other similar studies related to health, and agree in writing to meet the guidelines established by MECC Executive and Steering Committees. Individuals seeking access must provide information sufficient to justify the request as described in the Agreement for Access to MECC Data. The signed agreement must be submitted to the Executive Director of MECC. The MECC Executive Committee must review the request, determine that the applicant is qualified, the scientific question is appropriate, and approve the agreement in writing prior to granting access to the individual.

Disclosure of MECC Data

Guidelines

Disclosure of MECC data means the granting of the right to examine the data for the scientific question proposed by the requester. Further disclosure of the data by the requester to any other entity or the use of the data for a purpose other than that for which the data was requested is prohibited. Only data necessary for the stated purpose of the request will be disclosed. The data may be used only for the approved purpose. Disclosure must be requested by an individual (usually a principal investigator) and the institution with which the principal investigator is affiliated. As part of the request, both the principal investigator and an authorized representative of the institution must sign the Agreement for Disclosure of MECC Data. If the request is approved, the disclosure will be transmitted to the principal investigator.

Format and transmittal of data

Data will be formatted in a mutually agreed upon file format. Files will be encrypted using a strong encryption (such as the Advanced Encryption Standard) and put on CD ROM's. The CD ROM's will be shipped overnight to the Principal Investigator via a company that provides tracking information on the shipment. In addition to the general requirements, the researcher is required to demonstrate that their research has scientific merit, and that their institution has acceptable data security policies and procedures.

Application procedure

The following materials must be submitted:

1. Description of the project and a list of requested data items. Materials submitted as part of a request may include a letter of support from a mentor.
2. Documentation that the project was reviewed by an internal committee of the submitting institution.
3. Documentation of adequate financial and institutional resources to complete the proposed project. If the research study is externally funded, a copy of the Notice of Grant Award or similar document should be provided.
4. A statement describing how long and in what form MECC data will be kept and if it will be returned or destroyed at the end of the project.

5. Agreement for Disclosure of MECC Data signed by the principal investigator and an authorized representative of the recipient institution.

E. AGREEMENT FOR ACCESS TO MECC DATA

Name of applicant: _____

Title: _____

Organizational affiliation: _____

Address: _____

_____

1. Applicant requires access to MECC data to engage in the following demographic, epidemiological or other similar studies related to health:

_____
_____
_____

2. The specific purpose for which Applicant will use MECC data and the data files to be accessed (e.g. type(s) of cancer, patient characteristics, diagnosis years, geographical areas) and other relevant information are:

_____
_____
_____

3. Applicant's qualifications to engage in these activities are as follows:

_____
_____
_____

4. In consideration for approval of this application, Applicant represents, warrants, and agrees as follows:

a. For purposes of this agreement, "MECC data" means all information relating to cases of cancer collected by the MECC member Countries.

b. Applicant agrees to access and use the requested MECC data in strict conformity with the specific purposes set forth in his or her application. Applicant agrees not to use the data for any other purpose. Applicant agrees not to copy or reproduce the MECC data in whole or in part, in any manner or format, or permit others to do so.

c. Applicant may describe the results of Applicant's use of MECC data in professional journals, public reports, presentations, press releases and other publications, provided that a copy is provided to the institution from which Applicant receives access and that all publications contain acknowledgment and disclaimer (see below)

d. If Applicant becomes aware that any person or institution not authorized to access MECC data has attempted to gain access or gained access to the MECC data, Applicant agrees to immediately notify MECC Executive Director.

e. MECC Executive Director reserves the right to withdraw Applicant's right to access and use MECC data at any time.

By my signature I declare as follows: I have read the foregoing agreement. By signing below I make the agreements, representations and warranties contained therein. I understand that these agreements, representations and warranties are material representations of fact upon which reliance was placed when this transaction was entered into.

_____ _____ Signature

_____ Printed Name and Title

_____ Dated

The _____ ("Recipient Institution") (Name of institutional recipient)

APPROVAL BY

MECC Executive Director

_____ _____ Signature

_____ Printed Name and Title

_____ Dated

F. ACKNOWLEDGEMENT AND DISCLAIMER

All publications shall contain the following acknowledgment and disclaimer: "The collection of cancer incidence data used in this study was supported by the U.S. National Cancer Institute as part of the MECC Registries program. The ideas and opinions expressed herein are those of the author and endorsement by the MECC members is not intended nor should be inferred."

INDEX