Topic Summaries for Registry of the Future Workshop

Topic	Author(s)
New and Emerging Data Sources	Lynne Penberthy, National Cancer Institute Chuck Wiggins, New Mexico Tumor Registry

Definition/Abstract

Central cancer registries historically relied on medical and vital records as their primary sources of information. Additional sources of useful information emerged in recent decades, including large datasets such as social security beneficiary records, the National Death Index, and patient billing records from private oncology practices. Additionally the complexity of cancer care must include more detailed treatment information from sources such as pharmacies to understand outcomes in the context of ongoing treatment. Further with the growing evidence for precision medicine, genomics data is becoming essential to characterizing the cancer – just as histology and stage have been traditional means for characterizing each cancer case. Cancer registries must actively pursue novel sources of information to augment and complete their data collection efforts in the current era and the foreseeable future as it will not be possible nor sustainable to require the registrars to manually collect these critical data.

Background

Medical records and vital records, the latter primarily death certificates, have served as a primary source of information for central cancer registries, both historically and in the present era. In recent decades, additional sources have been utilized to gain access to information on vital status (e.g., National Death Index, social security beneficiary files, and records of driver licenses), cancer-directed therapy (e.g., patient billing records and in-patient/out-patient discharge records from hospitals), and diagnostic/prognostic testing (e.g., pathology reports and records from providers of genomic testing).

There are two key components to this issue: 1.) Identifying new sources of data; and 2.) Identifying new methods for obtaining such information, both from new and existing sources.

We recognize that there is some overlap with other "topics" in this conference, especially Item #3, "Expanding the Scope of Cancer Surveillance."

Rationale/Vision

Central cancer registries are presently charged with systematic collection of a large number of data items relevant to cancer-directed therapy. However, information regarding many of these data items is incomplete in the medical records and, therefore, is poorly documented in central registries. In addition, registries have characterized cancers according to histopathologic characteristics traditionally, but now, with the advent of genomics information that enables more precise targeting of treatment, it is essential that registries include these data. Cancer registries need to identify new sources of information, as well as new methods for acquiring information from both existing and novel data sources.

Types of Information:

Treatment

- Information in the central registry regarding some treatment modalities is relatively complete (i.e., surgery and radiation therapy) but only for the initial course of therapy. In contrast, information for other key treatment modalities (e.g., chemotherapy and hormonal therapy) is often incomplete. In particular for treatments that are not provided in the hospital setting such as orally administered agents or treatments received in the outpatient oncology practice or other freestanding specialty such as urology or dermatology practices.
- Need to fully document information on cancer-directed therapy in a timely manner.
- Need to identify and accommodate data collection for new treatment modalities as they emerge.
- Need to capture both initial and subsequent courses of therapy – survival is no longer the only relevant outcome and it is critical to begin to capture data on multiple courses of treatment (both systemic and radiation)

Potential sources:

- Claims
- Pharmacy data
- EMR data
- Other?

Genomics

- Relatively few data items regarding cancer genetics are presently ascertained in central cancer registries (specific biomarkers such as PSA , CEA, Oncotype dx 21, etc)
- The growing relevance of single biomarkers and genomic panels to understand the cancer and to assist in therapeutic decision making make these data essential in the registry mandate to characterize each cancer case. This is essentially an extension of prior systems used to characterize a cancer such as histology, stage etc.
- Cancer registries need to define policies and procedures for prioritizing which data items should be included in the registry
- Is it possible to be nonspecific due to the rapid evolution and addition of new genomic tests that are increasingly used in practice?

Potential Sources

- Genomic labs (Foundation, GHI, Guardent Genome Dx, Academic Molecularpath labs)
- Genomic lab intermediaries (Syapse?)
- EMRs

Outcomes, including recurrence

Sources of Information:

Patient-Generated Data

- Central cancer registries should take advantage of new methods for ascertaining cancer-related information directly from cancer patients.
- There are certain data that can only be captured directly from the patient
- Also provides a potentially opportunity to give data back to the patient

Recurrence

- Recurrence is a critical piece of information that is now an essential outcome necessary to support research but perhaps more importantly to provide data to patients and clinicians to help them understand their risk (of recurrence and response etc.)
- This goes hand in hand with the capture of longitudinal treatment data
- Likely initial focus on distant recurrence?
- Methods of diagnosis and physicians diagnosing will vary by cancer site, thus capture of recurrence is complex and will require multi-layered compilation of data from multiple sources
- Likely will require new methods (e.g. NLP and machine learning and new linkage sources)
- Opportunity to test/leverage coding/taxonomy system for recurrence (PRISM)?

Sources

- Patient reported?
- Pathology reports could identify subsequent biopsies for selected cancer sites
- Longitudinal claims could identify recurrence based on sequence and timing as well as specific agent (i.e. used for recurrence)
- Radiology reports often initial diagnostic method for distant recurrence
- EMR data (clinical notes)
- Hospital abstracts capture recurrence- sensitivity low but PPV high

Challenges/Resolutions

The greater cancer surveillance community must strictly follow and enforce standards for adding and deleting data items to the cancer registry workload. Thus the focus should be on automation and linkages where feasible when adding these new sources and new data. It is not sustainable to rely on additional manual data abstraction work to collect, edit, and maintain these new but essential data.

References

Objectives of discussion:

1. Review/Discuss potential sources
2. Identify barriers to implementing and problem solve for solutions.
3. Identify additional sources that could represent important components to enhance the surveillance data
4. Discussion of potential issues with population based versus limited enhancement with these various data sources. (e.g. Medicare is all patients over 65, but other claims sources might cover oncology providers by practice or a specific commercial insurer)
    a. How does that impact the surveillance mission?