Research Data Requests

Communicating with requestors

Providing data, both in aggregate data summaries and at the individual level, is a very important role of the central cancer registry. In addition to the annual calls for data, NAACCR, NPCR, SEER, and StatCan, registries field requests from a large variety of sources, including academic researchers, health care professionals, community outreach organizations, as well as the public. The response to a data request will depend on the nature of the request and the role of the requestor.

When communicating with the requestor, it is important to first identify what their need is—not all requestors know what data are available and sometimes requestors do not know what data are appropriate for their research question. Once the analyst understands the requestor's need, then the analyst can determine what type of data to release.

If the request needs an aggregate data summary, a simple response via memo with summary tables and figures may be the most appropriate response. In most registries, further approval from review committees is not required when providing aggregate data. ***It is considered a best practice to have another colleague review the data prior to release.*** Release of data should be in line with registry standards, and consideration should be given for masking cells with small counts and avoiding providing data for small geographic areas to protect the confidentiality of the cancer patients. Additionally, small counts and/or small populations, can lead to unstable estimates. Every registry will have different policies, so it is important to know what regulations apply for your own organization. For example, in NAACCR's Cancer in North America (CiNA) data products, estimates based on less than 6 cases are suppressed. ***Additional resources on data confidentiality and data release best practices are located on the Data Security & Confidentiality Issues landing page:*** https://www.naaccr.org/data-security-confidentiality-issues/.

If a researcher needs individual-level data, Institutional Review Board (IRB) and/or committee approval is usually required. It is recommended to meet with the researcher prior to IRB submission to ensure the data sought after are available and to facilitate a smoother approval process. Each registry's IRB will have their own approach for assessing and approving research using registry data. The registry's IRB will consider the type of variables being requested when approving a data release. Each IRB will have slightly different standards about what they will approve, especially for studies using a waiver of consent, so communication with the IRB and researcher about potentially identifiable data is essential to getting requests approved. For example, a request for full dates, or even month and year, may be harder to get approved than a request for year only. However, your IRB may have workarounds to maintain confidentiality, such as the registry calculating the number of days between two events rather than releasing the full event dates. Meeting with a researcher before they submit the IRB application can help them identify any potential pitfalls in their request.

At the onset of a data request, it is important to define the cohort of cancer cases. The cohort definition should include the dates of diagnosis, geographic area, as well as cancer sites of interest. Other commonly used variables to define a cohort include race and ethnicity, age at diagnosis, sex, histology, stage, behavior, and site-specific data items. Ask the requestor for specifics for the cohort definition, as they may not consider important variables when making a request. The registry should consider using standardized language for specific types of requests and developing a data request application to eliminate confusion with the request. ***It is considered a best practice to have a data request application form (either online or PDF), which is a helpful way of capturing information about data requests including cohort definition.***

As registry data experts, our role is to make sure we are communicating the necessary information to ensure the data are interpreted properly. Directing researchers to online resources such as the NAACCR Data Dictionary or the Site-Specific Data Items (SSDI) lists can help the researcher understand the coded values and evaluate the variables appropriately. Researchers may not understand the nuances of data changes over time, so it is important to let them know of any relevant changes to data collection for their request. For instance, if they are requesting data from 1990 to 2020 and are interested in SEER Summary Stage, there are several different summary stage variables that would be applicable to those diagnosis years. Providing context for the time frame of the data changes can help the researcher use the data effectively for their research. Requestors should also be made aware of any known data quality issues for requested variables or if there are any issues with case ascertainment for the years requested.

It is considered a best practice that a written summary of the data that is extracted and prepared should be included with every data request. This summary should give enough detail to replicate the request. A full definition of the cohort used should be included, and enough detail to identify which variables were included. For instance, "stage" can mean a large number of variables in the NAACCR Data Dictionary, so any written documentation should identify exactly which stage variable was used, generally by identifying the NAACCR Item #. This written summary should be tailored to the level of expertise of the requestor. An academic researcher may want a lot more detail and foot notes to the tables and figures provided, while a member of the public may be interested in a broad overview to address a concern. On occasion, requests may come in from the media, local journalists, or members of the state or national legislature. When these requests are received, make sure any responses are consistent with your own registry and agency policies and procedures. These special requests may require additional review or coordination with colleagues outside the cancer registry before the request can be released.

With any request, your registry may have standard exclusion criteria with the cohort, such as removing any Veteran's Administration only cases or Interstate Data Exchange only cases (although most registries have updated policies to allow such data release). If you do remove these cases, make sure you communicate to the requestor that the final information sent does not include the restricted release cases.