REGISTRAR PIP

Visit SEER*Educate: A comprehensive training platform for registry professionals

Unknowns: An Opportunity for Improvement Just Waiting to Happen

Introduction

Given that the value of a cancer registry lies in the ability of its data to be used to carry out various research and public health planning and evaluation activities, discussions regarding quality control procedures are commonplace in our business. Quality control is often referred to as a process through which we attempt to ensure that data accuracy, completeness, and timeliness is maintained or improved by creating quality benchmarks and checking for significant performance variation.

The development of a quality control program implies a comprehensive approach to the maintenance of quality that should involve every aspect of registry operations. Quality data can be used for its intended purposes in operations, research, decision making and planning. Without quality data and processes, we would likely be unable to generate any interest in what we work so hard to collect.

Data Quality Profile (DQP)

While there are many registry-related quality control efforts that could be discussed, for this edition we are going to focus on one that also happens to be an area of interest to the Surveillance, Epidemiology, and End Results (SEER) Program. The DQP is a report generated annually by SEER for each of its registries, so we (and they) know how we are doing in terms of meeting critical contractual obligations and our standard setter's expectations related to data quality.

As part of our database reporting package, we can also generate this report locally any time we want to. An important part of any quality control program is an awareness of current performance. The reason we choose to run this report daily is to stay informed as to our progress in reaching the defined goals. Nobody wants to learn about a data quality issue at the time of a required submission when it is too late to take corrective action.

In addition to evaluating the completeness of incidence reporting and successful follow-up rates by age group, the DQP also shows us the completeness levels and the performance benchmarks established for the five data items listed in Table 1. The registry submits preliminary data to SEER every February. By the time we make our final data submission in November, we need to reach the completeness quality goals for each data item. As you can see, we still have a bit of work to do to identify the race for patients who currently are in the database with unknown race.

Table 1
February 2022 Status for Cases Diagnosed in 2020
CSS Percent Unknown versus SEER Goal

Data Item	Percent Unknown	Goal
Unknown/Ill-Defined Site	0.75%	< 2.5%
Unknown Laterality	2.73%	< 6.0%
Non-Specific Histology	0.24%	< 1.5%
Unknown Derived Summary Stage 2018	4.72%	< 10.0%
Unknown Race	7.13%	<= 3.0%

CSS is funded by the National Cancer Institute's SEER Program, Contract Number HHSN261201800004I

Some data items are more difficult to obtain when there is sole reliance on medical records to obtain information. The lack of high-quality documentation within the medical record can produce incompletely abstracted data elements which results in higher than desired levels of unknown data values being recorded. For example, we notice a much higher level of unknown race codes for non-hospital cases as compared to hospital cases. Non-hospital records are much more limited in scope than that found in hospital records. There is also little registrar contact in most non-hospital settings which unfortunately negatively impacts the abstracting process because it is the hospital registrar who plays a pivotal role in explaining the need for certain data items and encourages their documentation in the medical record by medical personnel, which improves data item completeness at hospitals.

Table 2 indicates there are some significant issues with data completeness for cases diagnosed 2018-2020 at some hospitals in our region given the range of unknown values for four of these fields. Only laterality meets the SEER DQP completeness goal at the time of original abstraction at all hospitals. For the other four data items, one of two things is going on:

- Information is missing from the medical record.
- Information is in the medical record but the abstractor missed capturing it.

Table 2
Hospitals Percent Unknown versus SEER Goal
Diagnosis Date 2018 2020

Data Item	Goal	2018		2019		2020	
		Range	Average	Range	Average	Range	Average
Unknown/Ill-Defined Site	< 2.5%	0 - **4.76%**	1.06%	0 - **5.62%**	1.08%	0 - **4.14%**	1.06%
Unknown Laterality	< 6.0%	0 - 3.17%	1.05%	0 - 4.49%	1.26%	0 - 3.91%	1.34%
Non-Specific Histology	< 1.5%	0 - **1.55%**	0.27%	0 - 1.26%	0.25%	0 - **1.76%**	0.25%
Unknown Derived Summary Stage 2018	< 10.0%	0 - **23.97%**	3.39%	0 - **44.45%**	4.70%	0 - **29.82%**	3.16%
Unknown Race	<= 3.0%	0 - **7.69%**	2.46%	0 - **6.14%**	2.39%	0 - **6.56%**	2.84%

While achieving 100% completeness is difficult, as Table 2 indicates it is not impossible to achieve for a data item. For each of the data items reviewed there was one or more hospitals that achieved 100% completeness. That's impressive.

Why these data items?

Sometimes knowing the "why" associated with an activity (or a choice) helps us understand its purpose and allows us to focus our attention on the standard setter's priorities. So, why are these five data items of particular interest to SEER and why did assessing their completeness on the DQP become a measure of the quality of the registry's data? Here's a brief description of the "why" for these five data items.

Race - There are distinct differences in cancer morbidity and mortality seen in different racial groups. Without complete and accurate race data, underserved populations are not accurately represented. Complete and accurate race information helps to better define the cancer burden in racial/ethnic minorities and medically underserved populations and supports research, applications and surveillance involving the diversity of our population.

Stage - Cancer staging is a fundamental element necessary in patient care (determining prognosis and treatment options including clinical trial eligibility). The accurate collection of this data item also allows researchers to evaluate the impact of cancer in a population and facilitate the development of cancer programs for screening and treatment.

CSS is funded by the National Cancer Institute's SEER Program, Contract Number HHSN261201800004I

Laterality/Ill-Defined Site/Non-Specific Histology – Can you imagine a world in which the auto-determination of the applicable Heme and Solid Tumor Rule might become our reality? I don't know about you, but some of those rules really bend my mind when trying to figure out how to apply them. If by more accurately and completely coding site/histology/laterality moves us in that direction, I'm all for it. This is the reason these three data items are being evaluated annually by SEER.

Conclusion

If stumped about how to initiate a quality check on the registry data items, certainly assessing the percentage of unknown values recorded for each of the five data items discussed in this article is a great starting point. Those with extremely high levels of unknown values recorded need to take action to improve their data quality. Until a check of the hospital data is done, it is impossible to know with certainty whether a data collection issue exists at a hospital.

If this quality control project identifies abstracting or coding issues, then outline a strategy to improve the data for the hospital. Do you need to go to administration or IT to request approval for a data item be consistency recorded in a specific location in the medical record to facilitate abstractors finding it? Do you need to work with physicians to encourage documentation of the needed information? Do you need to review staff members' abstracts to ensure available information is correctly documented?

If you discover you met the performance benchmarks for these initial data items, give yourself a pat on the back for this quality control effort and share the information internally. Everyone likes to know they've done a good job in an area of performance, especially after it has been confirmed during a quality control review.

We need to remind ourselves we shouldn't use the ability to code an unknown value in a field as a way to avoid the extra effort to obtain information that is not immediately available. The usefulness of a registry depends on not only its completeness in terms of incidence reporting but, equally important, is complete information on key data items. All cancer registries should be able to routinely provide some objective indication of the quality of the data they have collected. Remember, the worst thing that can happen to any of us is learning about an issue with the data only after someone starts using it!

CSS is funded by the National Cancer Institute's SEER Program, Contract Number HHSN261201800004I