Statistical Approach to Evaluating Completeness

Establishment and Composition of Statistical Expert Panel

The Statistical Expert Panel includes representatives from CDC, NACDD, NAACCR, SEER, Information Management Services, Inc. (IMS), individual registries, and outside consultants. Participants were invited based on their expertise in quantitative aspects of cancer surveillance, particularly past involvement with defining and refining completeness measures for North American registries.

- Dr. Kevin Ward of the Georgia Comprehensive Cancer Registry has served as the primary curator of the existing NAACCR completeness method and has chaired a work group designed to address some of its shortcomings.
- Dr. Hannah Weir of the CDC has served on this work group and was the co-chair of the group that initially developed the NAACCR method. She brings a historical background to the group.
- Drs. Eric Feuer, Li Zhu, and Huann-Sheng Chen of the National Cancer Institute's SEER program have led the development of the modeling approach and internal methods for measuring completeness, as well as the implementation of delay adjustment for a variety of cancer surveillance purposes.
- Dr. Barnali Das of the National Center for Health Statistics (NCHS) previously worked on the modeling approach when she was at SEER.
- Dr. Robert Anderson from NCHS is a national expert on mortality data and frequently collaborates with the cancer surveillance community.
- Dr. Charles Wiggins is the director of the New Mexico Tumor Registry and past president of NAACCR. He has been outspoken about the implications of the constant incidence-to-mortality ratio assumption. New Mexico has consistently appeared to be underestimated using the NAACCR method.
- Dr. Lihua Liu, representing the Los Angeles Cancer Surveillance Program, is a national expert on demographics and cancer, particularly among minority populations. The Los Angeles Cancer Surveillance Program also has frequently appeared to be underestimated for completeness using the NAACCR method.
- Dr. Francis Boscoe of Pumphandle is a national cancer surveillance expert with nearly 20 years' experience at the New York State Cancer Registry.
- NAACCR was represented by Betsy Kohler, the Executive Director, and Lori Havener provided logistical support at the in-person meetings.
- The CDC additionally was represented by Drs. Paul Sutton, Trevor Thompson, and Manxia Wu. Dr. Paulette Valliere represented NACDD.

- Andy Lake, Rick Firth, Don Green, and Joe Zou variously represented IMS in the role of providing programming and IT support and access to NAACCR databases.
- Kathy Huamani of SCG served as notetaker during both the conference calls and in- person meetings.

Description of Completeness Models Considered

Incidence to Mortality Rate Ratio Method

Measuring Completeness

Cancer surveillance relies on complete, unduplicated case capture within a defined catchment area and during a defined time period to accurately enumerate incident cancer cases and calculate age-adjusted cancer incidence rates. Any disease surveillance system requires complete case capture to accurately measure disease burden within a defined population. However, it is often difficult to estimate how many cases go undetected and, thus, not enumerated. Cancer is distinguished by having relatively clear diagnostic criteria, generally microscopic confirmation of malignant cells, and sometimes diagnoses are made based on other clinical findings. Yet still, cases may remain unidentified by the healthcare system. In order to estimate the accuracy with which cancer registries are able to identify all cases within their catchment areas, several completeness estimates have been developed and evaluated. These estimates form the foundation by which compliance the NPCR 12-month data standard is measured. Yet, it is understood that the measures currently in use are not ideal; thus, this project evaluated the existing methods to determine if the measures could be improved or replaced with a more accurate measure.

Description of Completeness Measures Considered

The Expert Panel considered several different approaches for measuring completeness. The methods considered were those that have been in use by cancer surveillance organizations in the United States and also in other parts of the world. They represent all of the most commonly used methods to measure completeness of cancer registry reporting. Each of the methods is discussed in detail below. The Expert Panel recommends that two of these methods be considered for continued development and assessment—one not to be further considered and one to be considered to the extent that it informs the delay-adjustment methodology.

More information on the deliberations of the Statistical Expert Panel may be found in Appendix C.

Incidence-to-Mortality Ratio Method

This is the method currently used by NAACCR and NPCR to measure registry completeness. There are minor differences between the methods used by the respective organizations; a recent comparison found that the estimate differed by more than 1% for only five registries and by more than 2% for just one registry. This summary considers only the NAACCR version of the method.

The method defines completeness as the ratio of observed-to-expected incidence rates for a registry. As is characteristic of such ratios, the average value across all registries is 1, or 100%; values are distributed around this average so that roughly half of registries have values above 1, and roughly half have values below 1. The observed incidence rate is calculated by summing age-adjusted rates stratified by sex, race, and cancer site. Many race and site classifications have been assessed. Currently, the method considers 18 sites for men and 15 for women, including nearly all of the most common sites but excluding the two most common sites—the breast and prostate. Currently, race comprises whites and blacks, but Hispanics and “other” categories also have been proposed for inclusion. For minority race groups, a rule of thumb has been that such groups must comprise at least 10% of the population in a registry to be included in a calculation.

The expected number of cases is obtained by multiplying the registry’s mortality rate by the national incidence-to-mortality rate ratio (IMRR), again stratified by sex, race, and cancer site. Thus, if the national IMRR is 1.5 for white male bladder cancer, for example, then the incidence rate of white male bladder cancer in that state would be expected to be 1.5 times the mortality rate. In order to achieve a more stable measure, the IMRR makes use of 5 years of data, and the registry’s mortality rate uses 2 years of data (3 years for small registries with fewer than 500,000 people). National IMRRs use national mortality but incidence from 11 SEER registries (See Glossary for explanation). (The NPCR method uses incidence data from all NPCR registries instead of restricting to the SEER 11 registries.)

The core assumption of this method is that incidence tracks mortality in a constant and universal manner by sex, race, and cancer site. This is not the case, of course, and indeed much of cancer surveillance is concerned with showing how this relationship is not constant insofar as it is driven by factors such as screening, healthcare access, and quality of care. One way of correcting for this issue has been to introduce an adjustment term to the method that increases the weight that each registry gives to its own data, effectively smoothing the calculated IMRRs toward 100%, raising them for registries with low-measured completeness and lowering them for registries with high-measured completeness. At the extreme, raising the adjustment term to its maximum value would mean that every registry was using only itself as a reference, and all registries would be 100% complete. The appropriateness of such an adjustment term is unknown, and it lacks any empirical basis.

The Statistical Expert Panel discussed this measure at length and identified advantages and disadvantages to the method. The primary advantages are its long tenure and familiarity within the registry community and its transparency; it is a simple matter to independently verify the calculation using routine surveillance data. In addition, a spreadsheet is available that allows any registry or researcher to assess the implications of varying parameters such as sex, race, and site stratification and the adjustment term.

The primary disadvantages include the constant IMRR assumption, instability of the measure for small registries, exclusion of the most common cancer sites, and a seeming systematic underestimation of completeness in areas with heavily Hispanic populations.

This measure has been used as a criterion in NAACCR certification for more than 30 years. Registries that achieve 95% completeness (and meet other criteria) are certified at the gold level. In recent years, nearly all NAACCR registries have been certified gold. For cases diagnosed in 2015, for example, 49 U.S. registries were certified gold, six were silver, and two were uncertified. The registries certified silver or uncertified most often in recent years have included Nevada, Arizona, New Mexico, Los Angeles, and Minnesota. (Minnesota's status was never attributed to completeness.) Four of these five are in the Southwest, and each of these four has a large Hispanic population, representing the registries with the first, second, sixth, and seventh largest percentages of Hispanics in the country. However, adding a Hispanic stratum into the method does not improve the completeness of these registries substantially, and this issue has not seemed to affect Texas with the third highest Hispanic population.

Another limitation is that a registry's completeness estimate can be sensitive to whether and how sex, race, and site are stratified and which adjustment term is chosen. At the second in-person meeting in Vancouver, Canada, there was a demonstration of how one registry (Arizona) could have fallen into any of these three categories depending on what assumptions were made. Other registries straddled two categories. Prior work by one member of the Expert Panel showed that this characteristic is not only a property of parameter selections but of sampling variability (Das et al., 2008). The Vancouver presentation was limited to four registries, all of which have had difficulty meeting the NAACCR completeness standard consistently (New Mexico, Arizona, Nevada, Los Angeles). Members of the Expert Panel agreed that it would be useful to expand this analysis to all registries to see if the conclusions are broadly applicable or confined to these negative outliers.

Recall that under the IMRR method, the average completeness value across all registries is 100%; values are distributed around this average so that roughly half of them have values above 100% and roughly half have values below 100%. In fact, the distribution of values appears to closely approximate a normal distribution or bell-shaped curve, with most of the values clustered around 100% and fewer values farther from 100%. In recent years, the range of values has been from about 85% to about

115%. Let us focus momentarily on states at the high end of this range. There are only three possible ways that a state can have a score this high. One is that they truly have found more cases than actually exist. This could happen if there are a large number of duplicates or if nonreportable cases are being counted. We are confident that between the duplicate protocol, edits, and well-defined rules for counting invasive tumors, this is not a likely explanation, certainly not at the grand scale that a score of 115% implies. A second is random error. Even if every state were to capture its cases to the exact same degree, there would still be some variation in the measurement stemming from the collective imprecision of all of the various inputs into the method (population age, sex and race composition, mortality rate, and so on). It is hard to know the exact size of this error, but members of the Statistical Expert Panel do not feel this is a major source of variation—at most a few percentage points.

That leaves an unexplained variation as the dominant explanation. There must be factors other than the incidence-to-mortality ratio that account for variation in completeness. If a state is coming in at 115%, it is because its expected count is artificially low. Why this is important is because of what happens when states on the other side of the distribution are considered. If three states can have a score of 115% for reasons have nothing to do with registry quality, then on these same grounds we would expect three states to have a score of 87% (the reciprocal) for reasons having nothing to do with registry quality. We currently assume that such state has too many unreported cases when in fact the chances are quite good that it is due to methodological imprecision.

Modeling Method

The modeling method is adapted from the method used to predict current cancer counts for the nation that was jointly developed by the National Cancer Institute and the American Cancer Society (Pickle et al., 2007; Das et al., 2008; Zhu et al., 2012). This method uses a hierarchical Poisson regression model, which includes spatial and temporal random effects across counties, and years of diagnosis. Using county-level cancer incidence counts from the Cancer in North America (CiNA) Deluxe file—stratified by age, sex, race, and diagnosis year as an input, it models incidence as a function of cancer mortality, sociodemographic variables for each county (urban/rural status, household characteristics, income, education, medical resources), and behavioral risk factors (smoking, obesity, health care coverage, cancer screening).

Completeness is then taken to be the ratio of the observed counts submitted by registries to the expected counts from the model. Like the IMRR method, this method is a relative method that implicitly assumes that completeness is 100% for the reference population, which is this case in the entire nation. Half of the population will belong to registries with completeness below 100%, and half will belong to registries with completeness above 100%.

Preliminary results for cases diagnosed in 2015—using 2015 as the reference year—were presented at the in-person meeting in Vancouver and revealed a moderate correlation with the IMRR method and a moderately narrower range of estimates, as indicated in **Table 1**. Since the method uses data from CiNA Deluxe, registries that did not meet the standards for inclusion in this volume or that opted not to have their data included are not reflected in the table. Additionally, states with multiple registries (California, Washington, Michigan) were grouped.

Table 1. Completeness score comparison, IMRR and Modeling Methods, 2015 diagnosis year.

Completeness Score	Incidence to Mortality Rate Ratio Method (NAACCR Version)	Modeling Method
90.0–94.9%	3	4
95.0–104.9%	25	30
105.0%–109.9%	15	11
110.0%–114.9%	—	1
Total	**46**	**46**

A major advantage of this method is that all cases are counted equally, regardless of site or race. This would presumably remove the temptation for registries to delay the processing of some cases intentionally because they do not count toward completeness. It also does not depend on the problematic assumption that incidence and mortality are perfectly correlated but instead incorporates factors known to influence cancer rates for which data are available, including demographic, behavioral and institutional data.

The major disadvantage is that the model is complex and is not transparent, and at present, its expected case counts are not independently reproducible by the registries. A spreadsheet similar to that developed for the IMRR method could be developed to potentially ameliorate this problem. Another issue is that deriving both observed and expected counts from the same year of data means that changes in absolute case counts cannot be captured. This will likely be an issue when estimating completeness of 2018 data, where reporting reductions of more than 5% are anticipated. The modeling method will assign an average completeness of 100% to the nation, with individual registries mainly distributed between 95% and 105%, as if nothing had changed. One potential way of addressing this problem would be to use an internal method (see Section C.5 below) as an additional validity check.

The modeling approach has the same limitation that was seen with the IMRR method; the values are normally distributed around 100%. If values well above 100% are seen,

they must be largely due to unmodeled variation, which means that values well below 100% can also be attributed to unmodeled variation, with equal probability. Based on preliminary analyses, modeled estimates are falling within the range of 90–110%, a narrower range than seen with the IMRR method. However, the values at the low end of the range are still close to current quality and certification standards. A realistic goal might be to reduce model uncertainty to the point were no states exceed 105% estimated completeness, which would mean that scores below 95% would more likely be related to incompleteness than to modeling imprecision. This could be achieved by incorporating additional covariates such as interstate migration, foreign-born composition, survival, and environmental variables. There will always be some variation that will not lend itself to being modeled, because the necessary data do not exist and there are no adequate proxies.

Flow Method

The flow method is a method for measuring completeness that was developed in Great Britain about 20 years ago and was subsequently adopted by several European registries (Bullard et al., 2000). The term “flow” comes from the way that the computation draws upon the flow of cases through a registry as part of its routine operation. It categorizes all cancer cases into one of seven different categories. Five of them are easily counted:

- Patients alive at the time of interest and registered
- Patients deceased at the time of interest and registered, with cancer recorded on the death certificate
- Patients deceased at the time of interest and registered, with cancer not recorded on the death certificate
- Patients deceased at the time of interest but not registered, with cancer recorded on the death certificate and with cancer information obtained through follow-back (“death certificate-initiated” cases)
- Patients deceased at the time of interest but not registered, with cancer recorded on the death certificate, but without cancer information obtained through follow-back (“death certificate only” cases)

The remaining two cannot be counted and must be estimated:

- Patients alive at the time of interest and not registered (“missing” cases)
- Patients with cancer deceased at the time of diagnosis, with cancer not recorded on the death certificate, and not registered (“lost” cases)

Estimating the missing and lost patients is accomplished by estimating the probability that a patient is registered while alive, the probability that cancer is accurately mentioned on a death certificate, and the expected patient survival.

The Expert Panel discussed several drawbacks with this method. It assumes that the survival of missing and lost cases matches those of recorded cases, when they would be expected to be quite different (Tervonen et al., 2017). It also requires that death certificates are timely and of high quality. Because, in general, U.S. death records require more than a year for acquisition, linkage, and processing, it would be impossible to use the flow method to estimate completeness for periods of 1 year or less, a crucial consideration for this project. The method also requires registries to identify "death certificate initiated" cases, which is not a property that U.S. registries routinely record. The Expert Panel felt that while some registries could likely deduce this information, others would find it difficult or impossible. On the positive side, completeness obtained from the flow method is intuitive—what you have is an estimate of the ratio of recorded cases to total cases, with an upper limit of 100%.

Capture-Recapture Method

The Expert Panel also discussed the capture-recapture[1] method, whereby completeness is ascertained by comparing reporting to different entities (Brenner et al., 1995). In its simplest form, it involves comparing cases reported to a central registry and cases reported on a death certificate. Assuming the two are independent, the number of cases not reported to either location (D) can be derived algebraically as

$$D = (ABC + B^2C + BC^2) / (A^2 + AB + AC),$$

where variables A through D correspond as illustrated in **Table 2**:

Table 2. Variables in the Equation to Calculate the Number of Cases Not Reported to Death Certificates or Cancer Registry

		Reported on Death Certificate	
		Yes	No
Reported to Cancer Registry	Yes	A	B
	No	C	D

Completeness is then simply 1 minus D.

A test of this approach on a past NAACCR data submission revealed immediate problems. One state reported zero death certificate-only (DCO) cases (cell C), implying a completeness of 100%. Another registry reported very few cancer deaths (cell B), implying poor completeness. In both instances, the limiting factor was not the cancer registry data but the timeliness and accuracy of the mortality data. Whether that was

[1] We note that the term "capture-recapture" was not actually used during the summit—no one in attendance made the connection to the 1990s work of Hermann Brenner and others—and the notes from the meeting refer to this as the "naïve method," because the assumption that central cancer registries and vital records were independent and that the vital records data was error-free seemed naïve.

because the mortality data were incomplete or in error in its original form or because the registry did not process them correctly is not known. In any case, “completeness” as measured this way ends up being a hybrid measure of both incidence and mortality completeness that is not interpretable.

Any method that relies heavily on death certificate-only rate is also subject to the problem that as the rate diminishes toward zero—as has been the general trend in North American cancer registries in recent years—the proportion of cases that are not true DCOs increases. Cause of death coding, while extremely good, is not perfect, and an unpublished study conducted by one registry found that a significant share of the DCOs actually died of other causes, the miscoding occurring sometimes through what appeared to be simple typographical errors.

When capture-recapture is expanded to include more sources (for example, hospitals, outpatient cancer centers, labs, physician offices, and so on), the interactions between these source types result in a loss of precision and accuracy. The capture-recapture method requires sources to be independent and equally likely to be reported, and it is easy to see how these requirements are grossly violated in the cancer data collected in the United States. The mix of source types depends on the patient’s age, the survivability of the cancer, whether the cancer is screen-detectable, the availability of laboratory tests to identify molecular subtypes, and so on. In a test case, the capture-recapture method was applied to data from a state that has had consistently highly complete data and it suggested completeness was actually low but with wide confidence intervals.

Internal Method

The internal method refers to the practice of using registries’ past case counts as the sole input to predict future case counts and assess the completeness of current counts. The advantage of this method is that it is straightforward to calculate and does not depend on external demographic or mortality data. SEER uses it to estimate completeness in its internal February (14-month) data submission. Many other registries do this implicitly when they provide mid-year progress reports back to facilities. For example, if a facility is told that they had 93 cases reported at this time last year but 76 cases this year, there is an implication that this number may be too low, that they may be behind in their submissions, because this year’s number is expected to be equal or greater.

The disadvantage is that it can be thought of as more of a measure of consistency than quality. As a simple example, imagine a national registry in a developing country where there are 10 hospitals, only 2 of which report to the registry. As long as this year’s case counts are similar to or greater than last year’s case counts, which will be true as long as the same two hospitals continue to report, completeness will appear high. The method implicitly assumes that a registry was virtually complete at least one time in the past. This may be a reasonable assumption for U.S. registries, but it is difficult to prove.

A registry that has been consistently 90% complete over its entire existence has the same issue as in the previous example.

Other Topics Considered—Using Delay Adjustment to Develop National Rates from 12-Month Data

During the course of their deliberations, the Expert Panel considered options to utilize existing data to produce national cancer incidence rates at 12 months following diagnosis. Knowing that the data available at that time will be incomplete to some degree, the Panel considered statistical solutions that could account for this "missingness" and still produce robust and reliable rates for public health purposes.

The Expert Panel finds that it is feasible to use data from existing 12-month data submissions to project national cancer rates. This can be accomplished by utilizing the cases reported in the 12-month submissions and projecting the data to the anticipated final counts using the delay-adjustment methodology already in widespread use in U.S. cancer surveillance (Lewis et al., 2018). Delay-adjustment uses the ratios of current to past case counts to anticipate cases still to be reported. The method developed by SEER statisticians considers 11 years of data, although in practice virtually all the cases are received within 3 years. **Table 3** below shows a typical matrix of case counts, where the 24-month data are inflated by 3.2% and previous years are inflated by smaller amounts.

Table 3. Matrix of Case Counts: 24-Month Data Inflated by 3.2% and Previous Years by Smaller Amounts

	Submission Year											
Diagnosis Year	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	Delay Factors	Expected Count After 11 Years
2006	17,083	17,422	17,603	17,638	17,725	17,754	17,763	17,765	17,780	17,787	1.000	17,787
2007	—	17,262	17,621	17,701	17,885	17,901	17,940	17,996	18,015	18,027	1.000	18,027
2008	—	—	18,076	18,255	18,482	18,522	18,561	18,568	18,583	18,596	1.000	18,596
2009	—	—	—	18,137	18,782	18,856	18,908	18,940	18,971	18,990	1.000	18,990
2010	—	—	—	—	19,440	19,802	19,886	19,923	20,383	20,409	1.001	20,429
2011	—	—	—	—	—	19,882	20,113	20,161	20,600	20,643	1.002	20,684
2012	—	—	—	—	—	—	20,673	20,828	21,216	21,278	1.003	21,342
2013	—	—	—	—	—	—	—	20,294	20,939	21,059	1.005	21,164
2014	—	—	—	—	—	—	—	—	21,149	21,356	1.014	21,655
2015	—	—	—	—	—	—	—	—	—	21,251	1.032	21,931

With 24-month data, delay factors are typically under 5%, although this varies by cancer site and registry. If we were to apply this methodology to 12-month data, delay factors would likely be closer to 20%. Nationwide, in the most recent NAACCR data submission, about 80% of the cases reported in the 24-month data submission were also present in the 12-month submission (see **Figure A**). We chose to stratify the counts by type of reporting source to illustrate that the distribution of source types between the 2 years does not change dramatically—the hypothesis that the second-year submissions are fundamentally different from the first-year submissions and thus a source of bias is not supported by the evidence in the NAACCR file. While there are more death certificate-only cases after the second year, and proportionally fewer cases from free-standing surgery and other hospital outpatient centers, these are small shares of the total. It may be true that individual registries may exhibit bias, but the goal here is to estimate rates for the nation, not individual registries, so the influence of any individual registry is diluted. Furthermore, registries with highly incomplete data will be excluded from this calculation, as will be explained shortly.

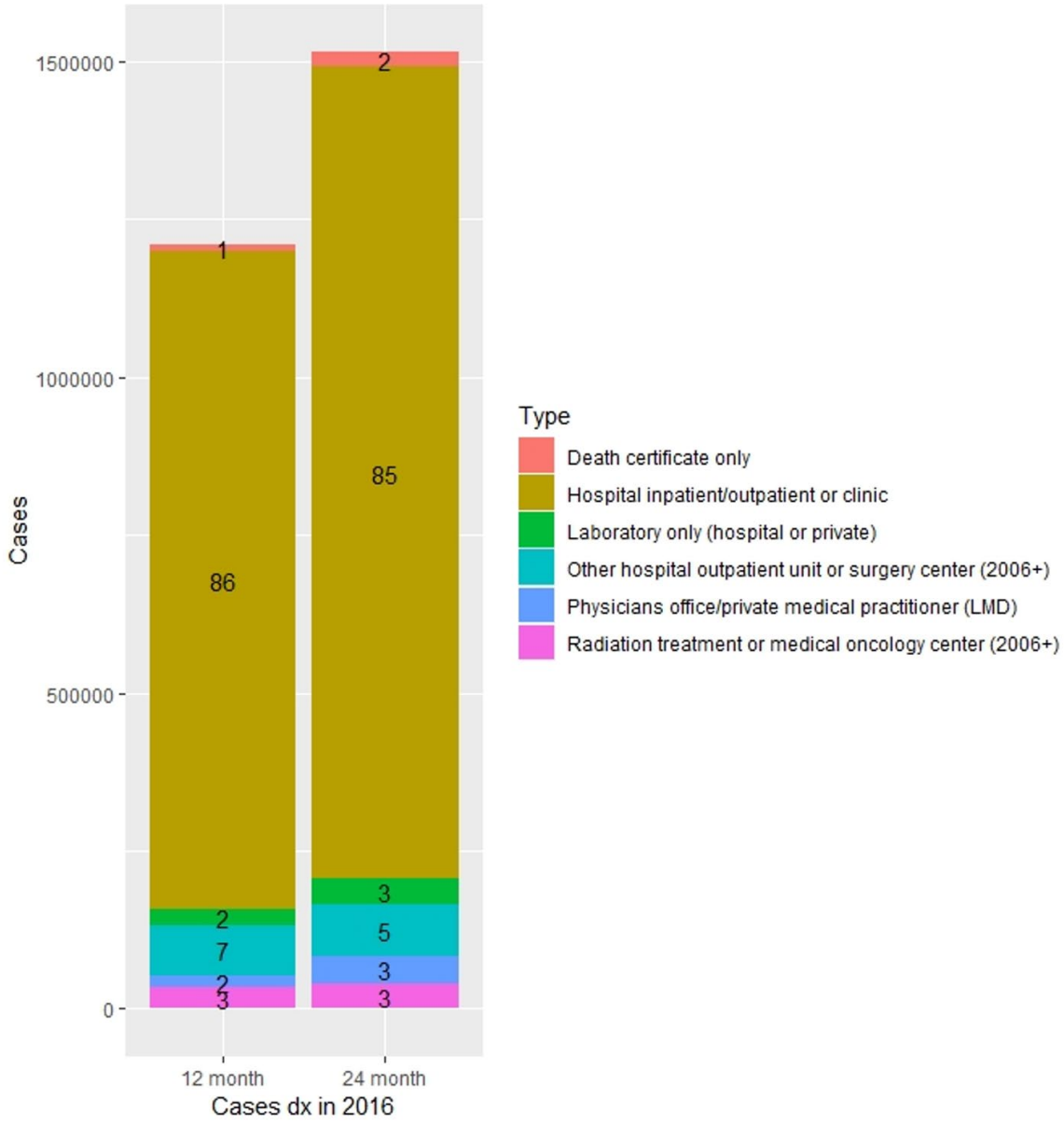

Figure A

The ratio of 12-month case counts to 24-month case counts varies by cancer site (see **Figure B**). Ignoring a few rare sites, the range is seen to extend from about 67% for chronic lymphocytic leukemia to 88% for uterine cancer (these figures were obtained by subtracting the value on the y-axis from 1). For certain sites such as CLL and possibly melanoma and prostate, the projection of national rates may be less feasible given higher uncertainty. Knowing which cancer sites may be projected and under which circumstances will require further analysis.

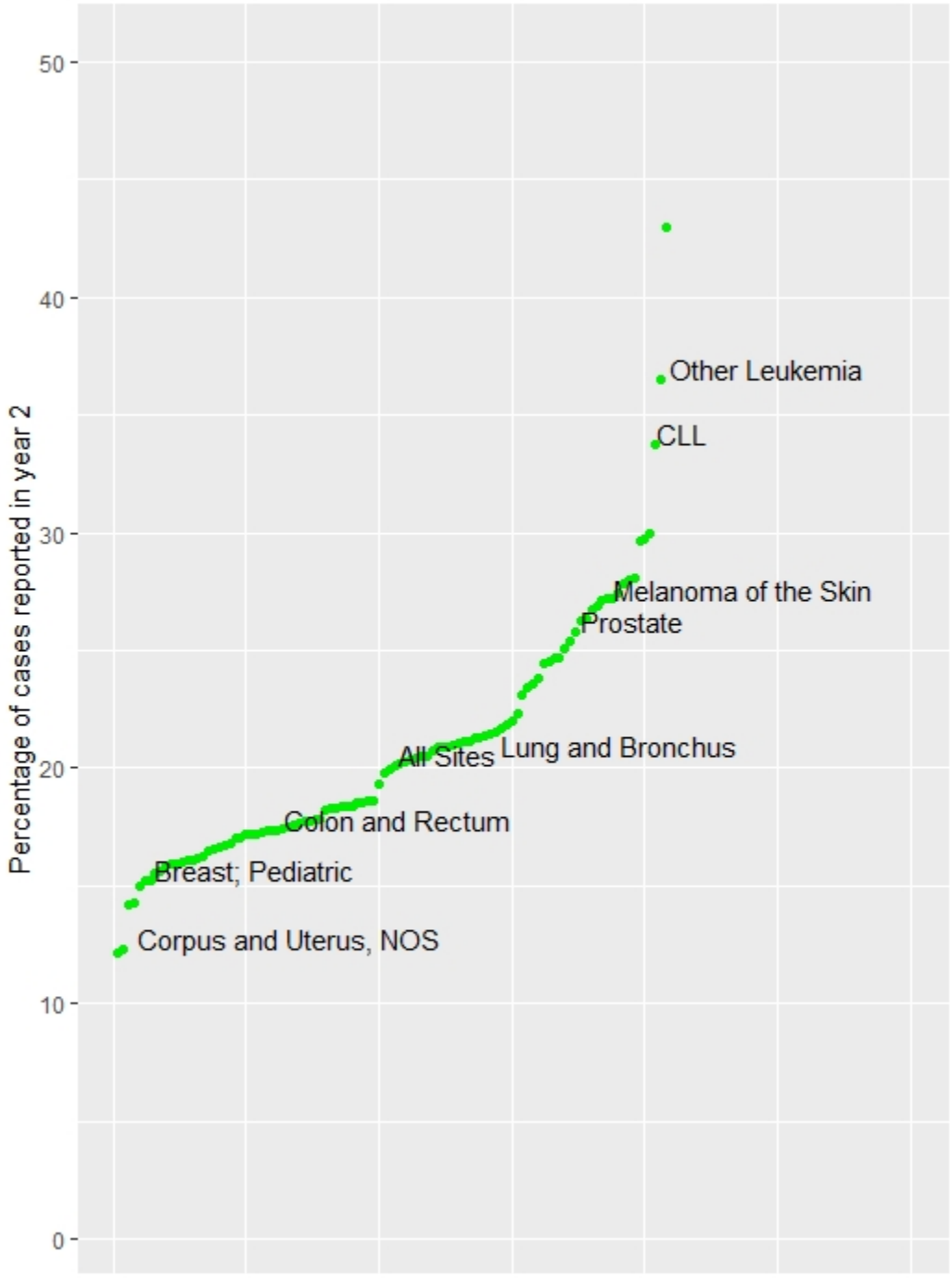

Figure B

The above figures made use of all data available to the Expert Panel—specifically, data for cases diagnosed in 2016 submitted to NAACCR between November 2017 and January 2018 (what we will henceforth refer to as 12-month or one-year data, even though technically it ranged between 11 and 13 months), and data for cases diagnosed in 2016 submitted to NAACCR in November 2018 (what we will call 24-month or 2-year data, even though technically it was 23-month data). However, the picture is better than

this if registries with an unusually poor 12-month reporting performance are excluded. Drawing upon prior analysis among SEER registries, the Expert Panel is proposing to limit the basis for national projections to only those registries that had a 12-month case count to 24-month case count ratio of at least 0.8 in at least three of the four most recent diagnosis years and to those registries that were certified gold or silver in all four of these years.

Applying these criteria to the NAACCR data submissions for cases diagnosed between 2013 and 2016, 36 registries would meet these criteria (see **Figure C**). These are the registries with at least three points above the horizontal yellow line, representing years in which at least 80% of the 24-month data was reported within 12 months. We anticipate that the picture would improve still further if NPCR rather than NAACCR and SEER submissions were used for the analysis. For many registries the submissions are the same, but some registries separately submit 12-month data to NAACCR in November and then again to NPCR in January. In one state, the difference between these submissions is substantial —typically below 80% complete in the first submission and above 90% complete in the second. In addition, there are several states that only submit their 12-month data to NPCR and not to NAACCR. The total number of states included in the generation of national rates could potentially be raised to at least 40.

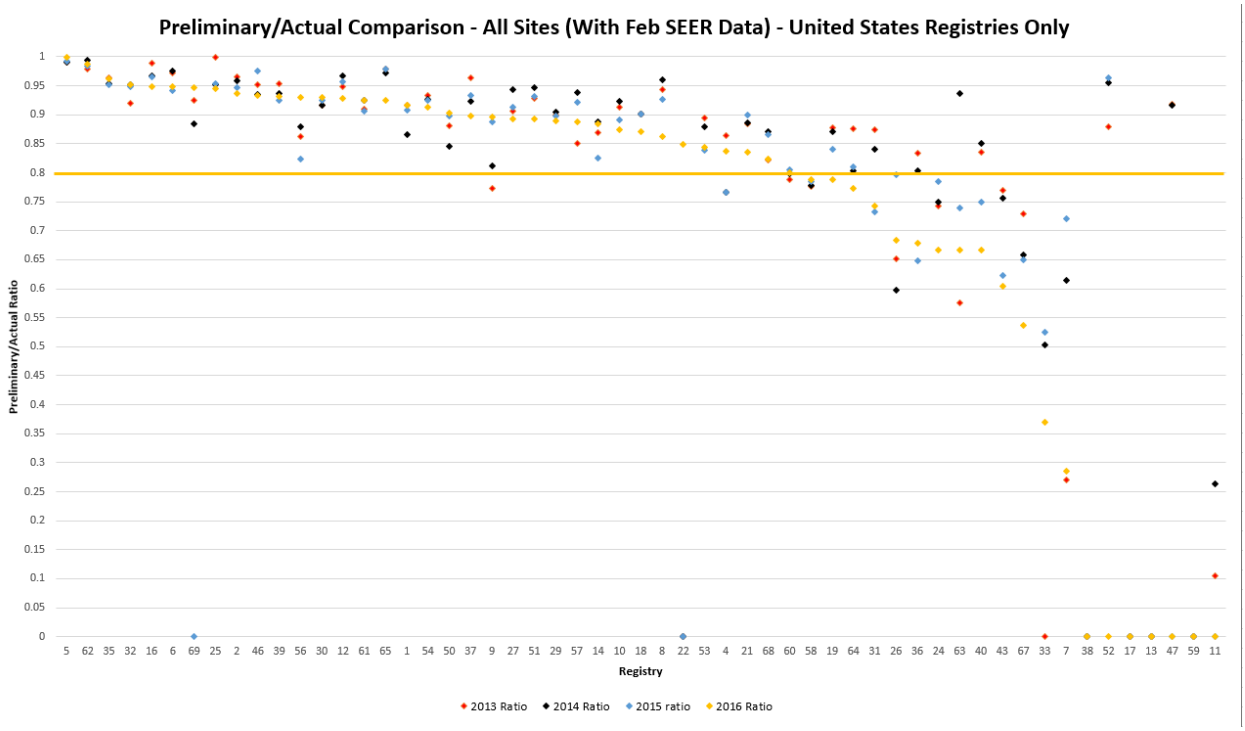

Figure C

We note that if 12-month data are to be used going forward to generate national incidence rates, there are obvious advantages to placing all registries on the same calendar. Based on the 22-registry sample used in our in-depth assessment, 11 registries responded that they typically report 12-month data to NAACCR in January, 10

registries report in November, and one registry does either. Whether the registries that report in November make a separate submission in January to NPCR was not assessed. While it is possible to continue to analyze a mixture of 11-, 13-, and 14-month data, it complicates the analysis and delays the eventual release of the data by 3 months.

Summary of Statistical Expert Panel Findings

The Statistical Expert Panel carefully evaluated several models for estimating completeness of cancer reporting by central cancer registries. All of the methods considered had disadvantages. In addition to the disadvantages described above, there appear to be some aspects of the measures that systematically disadvantage some registries, making it difficult to achieve acceptable completeness scores because of factors beyond their control (See Factors Related to the Completeness Measure that Influence Compliance with the 12-Month Data Standard). In short, no single method has emerged as a satisfactory measure. However, the group wishes to continue working on this problem and conduct further analysis and modeling, with the intent of improving the methods currently in use.

In addition, the Panel believes that it is possible to produce reliable cancer incidence rates for the nation and states by applying statistical modeling to data collected by registries at 12–13 months after diagnosis.

Recommendations and Next Steps

The Statistical Expert Panel recommends further analysis of the modeling approach to measuring the completeness of cancer reporting. The Panel thinks that the modeling approach will provide the most accurate estimate of completeness, because it takes into account many local factors that influence cancer rates such as behavioral risk factors, socioeconomic status measures, etc. The model behaves and performs like the existing IMRR method, but with reduced variance, which will reduce the likelihood of a “false negative”—that a registry will be deemed not to have met a quality standard when, in fact, it has. The reduction in variance is a consequence of including information on demographics, health care systems, and behavioral risk factors absent from the IMRR approach. The modeling approach needs further refinement as described below.

Regarding the existing incidence to mortality rate ratio method, the Panel intends to compute completeness measures using this method for all registries over multiple years with a wide range of parameters. This exercise was completed for several states and presented at the Vancouver in-person meeting, and it showed that at least for some registries, completeness scores varied widely when minor parameter adjustments were made. The results of this sensitivity analysis will be useful in illustrating why the modeling approach is the preferred method. In addition, comparing these results with those from the modeling approach is of interest to see how completeness scores might have changed in the past or will change in the future.

The flow and capture-recapture methods studied are not recommended for further analysis at this time due to significant problems uncovered with these methods. The internal method is being kept in the conversation only insofar as it may prove useful in evaluating 2018 data should there be evidence of reduced completeness on the nationwide basis. The modeling approach scales the national completeness level to 100%, but the internal method makes use of year-to-year comparisons.

The Panel also recommends developing and refining methods to use the data submitted at 12 months to accurately project incidence rates for the nation. Using data from NPCR states that are at least 80% complete and representing approximately 70% of the population, the group believes that reasonably accurate incidence rates could be derived for the nation and states for public health purposes. Using appropriate statistical techniques, the data from the sample would be adjusted to account for the cases not yet reported. These techniques would be based on what has been learned through delay adjustment modeling conducted by NCI, CDC, and NAACCR over the past several years.

1. Short-term plans for improving the completeness estimate are as follows:

    - Generate historic completeness estimates for diagnosis years 2013 through 2016 using the modeling approach and compare these with existing completeness estimates.
    - Seek ways to reduce residuals in the model through the inclusion of other covariate terms. This may include variables that capture interstate migration, international migration, survival, and environmental variables.
    - Consider the implications of the false-negative rate for certification that is implied by the distribution of model residuals. That is, making reasonable assumptions about the shape of the distribution of completeness estimates around 100% implies that some states should be in the lower range of completeness estimates (under the IMRR method and, to a lesser degree, in the modeling method as it currently exists) for reasons entirely beyond the control of the registries.
    - Ascertain which states, if any, tend to be systematically overpredicted or underpredicted.
    - Discuss the development of communication materials around the modeling approach so that NAACCR members will have a clear understanding of how it works.
    - Expand the presentation showing the sensitivity of the IMRR method to varying parameters to include all registries.

2. Long-term plans for developing incidence projections from 12-month data submissions are as follows:

- Further develop and refine the delay adjustment-based method for projecting 12-month data into national incidence rates. Thus far, we have demonstrated that by using a lower completeness threshold of 80%, data from a large majority of registries would be included. The resulting over-80% sample that the 12-month data represents has many properties in common with the 24-month data, including similar site and reporting source distribution.

Appendix B: Expert Panels and Participating Registries

Registry Operations Summit Expert Panel

Wendy Aldinger	Pennsylvania Cancer Registry
Lynn Giljahn	Ohio Cancer Incidence Surveillance System
Lori Havener	NAACCR
Mona Highsmith	Minnesota Cancer Surveillance System
Ann Marie Hill	Consultant
Stephanie Hill	New Jersey State Cancer Registry
Leslie Hoglund	Virginia Cancer Registry
Mei-Chin Hsieh	Louisiana Tumor Registry
Deborah Hurley	South Carolina Central Cancer Registry
Mary Jane King	Ontario Cancer Registry
Lori Koch	Illinois State Cancer Registry
Betsy Kohler	NAACCR
Sue Lai	Kansas Cancer Registry
Gary Levin	Florida Cancer Data System
David O'Brien	Alaska Cancer Registry
Winny Roshala	Consultant/Cancer Registry of Greater California
Frances Ross	Kentucky Cancer Registry
Colleen Sherman	New York State Cancer Registry
Valerie Somma	Colorado Central Cancer Registry
Melanie Williams	Texas Cancer Registry

Statistical Summit Expert Panel

Bob Anderson	NCHS
Francis Boscoe	Pumphandle
Huann-Sheng Chen	SEER
Barnali Das	NCHS
Rocky Feuer	SEER
Rick Firth	IMS
Don Green	IMS
Ann Marie Hill	Consultant
Betsy Kohler	NAACCR
Andy Lake	IMS
Lihua Liu	Los Angeles Cancer Surveillance Program - USC
Paul Sutton	CDC
Trevor Thompson	CDC
Paulette Valliere	NACDD
Kevin Ward	Metropolitan Atlanta SEER Registry
Hannah Weir	CDC
Chuck Wiggins	New Mexico Tumor Registry
Manxia Wu	CDC
Li Zhu	SEER
Joe Zou	IMS

List of Participating Registries

Alaska Cancer Registry

California Cancer Registry

Cancer Data Registry of Idaho

Colorado Central Cancer Registry

Florida Cancer Data System

Georgia Comprehensive Cancer Registry

Illinois State Cancer Registry

Kansas Cancer Registry

Kentucky Cancer Registry

Louisiana Tumor Registry

Maine Cancer Registry

Massachusetts Cancer Registry

Minnesota Cancer Surveillance System

New Jersey State Cancer Registry

New York State Cancer Registry

North Dakota Statewide Cancer Registry

Ohio Cancer Incidence Surveillance System

Oregon State Cancer Registry

Pennsylvania Cancer Registry

South Carolina Central Cancer Registry

Texas Cancer Registry

Virginia Cancer Registry

Appendix C: Statistical Summary

Introduction

Following a series of conference calls, an in-person statistical summit was convened in Gaithersburg, Maryland, on April 8 and 9, 2019. There were 21 attendees in all, including representatives from CDC, NACDD, NAACCR, SEER, individual registries, and outside consultants. Most of the meeting was spent carefully evaluating the pros and cons of various methods for measuring registry completeness, then selecting the most promising of these for more rigorous analysis, focused on cases diagnosed between 2013 and 2016. IMS made the NAACCR data submissions from these years available to selected group members following the summit. Then, following additional conference calls in April and May, a second in-person meeting was held at the NAACCR annual conference in Vancouver, on June 12 to report on progress.

The workgroup considered a number of methodological approaches to solving both the problem of measuring completeness and the additional problem of using the 12-month data submission to develop national incidence rates. We describe the latter topic first, since it informs the discussion of the completeness measures.

Using delay-adjustment to develop national rates from 12-month data

The workgroup finds that it is feasible to use data from existing 12-month data submissions to project national cancer rates. This can be accomplished by taking the cases reported in the 12-month submissions and projecting them to the anticipated final counts using the delay-adjustment methodology already in widespread use in U.S. cancer surveillance (Lewis et al. 2018). Delay-adjustment uses the ratios of current to past case counts to anticipate cases still to be reported. The method developed by SEER statisticians considers 11 years of data, though in practice virtually all the cases

have been received within 3 years. Figure 1. shows a typical matrix of case counts, where the 24-month data are inflated by 3.2% and previous years by smaller amounts.

Figure 1. Typical matrix of case counts

	Submission Year											
Diagnosis Year	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	Delay Factors	Expected Count After 11 Years
2006	17083	17422	17603	17638	17725	17754	17763	17765	17780	17787	1.000	17787
2007		17262	17621	17701	17885	17901	17940	17996	18015	18027	1.000	18027
2008			18076	18255	18482	18522	18561	18568	18583	18596	1.000	18596
2009				18137	18782	18856	18908	18940	18971	18990	1.000	18990
2010					19440	19802	19886	19923	20383	20409	1.001	20429
2011						19882	20113	20161	20600	20643	1.002	20684
2012							20673	20828	21216	21278	1.003	21342
2013								20294	20939	21059	1.005	21164
2014									21149	21356	1.014	21655
2015										21251	1.032	21931

With 24-month data, delay factors are typically under 5%, though this varies by cancer site and registry. If we were to apply this methodology to 12-month data, delay factors are likely to be closer to 20%. Nationwide, in the most recent NAACCR data submission, about 80% of the cases reported in the 24-month data submission were also present in the 12-month submission (Figure 2). We chose to stratify the counts by type of reporting source to illustrate that the distribution of source types between the two years does not change dramatically – the hypothesis that the second year submissions are fundamentally different from the first year submissions and thus a source of bias is not supported by the evidence. While there are more death certificate only cases after the second year, and proportionally fewer cases from freestanding surgery and other hospital outpatient centers, these are small shares of the total. It may be true that individual registries may exhibit bias, but the goal here is to estimate rates for the nation, not individual registries, so the influence of any individual registry is diluted. Furthermore, registries with highly incomplete data will be excluded from this calculation, as will be explained shortly.

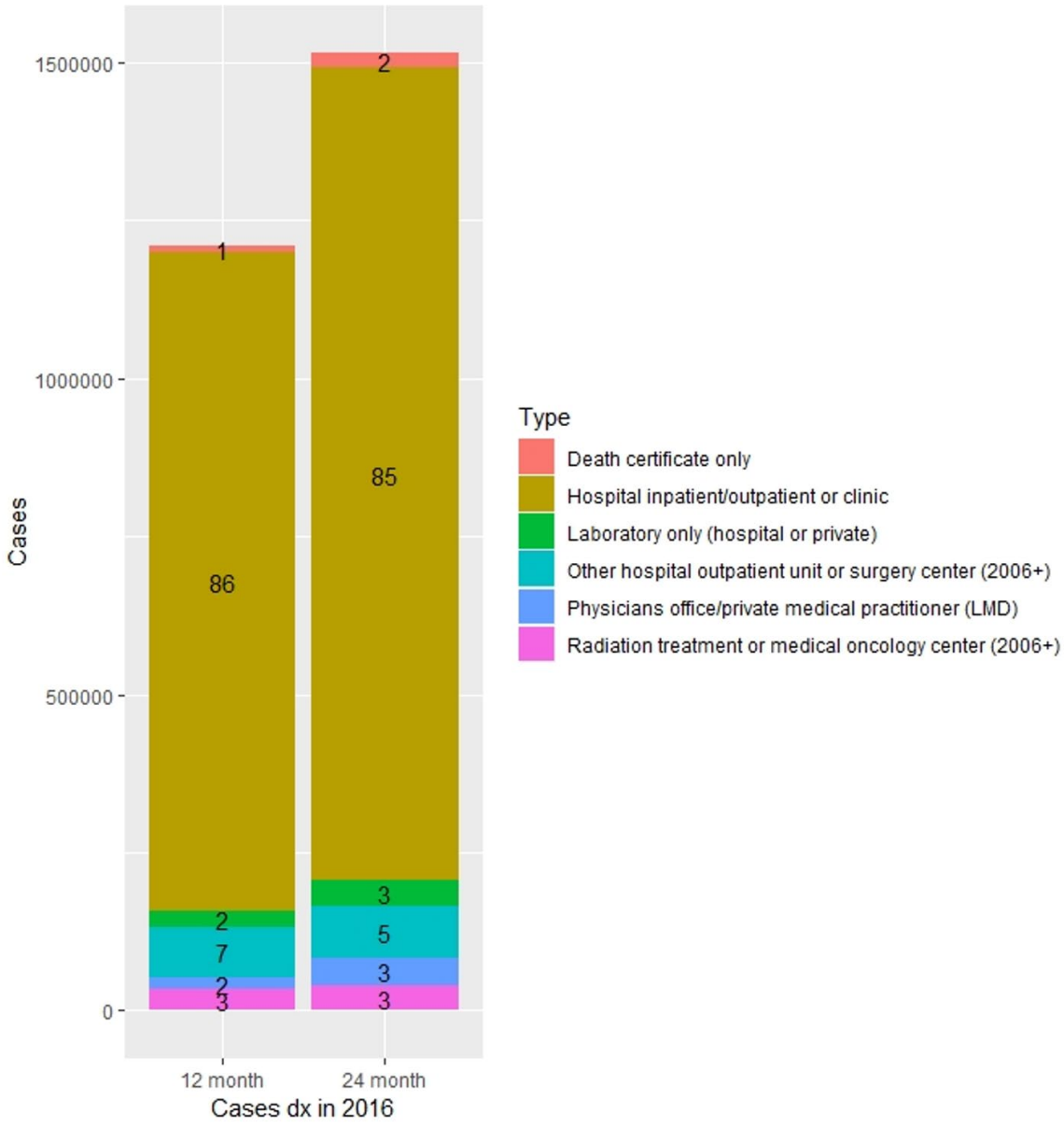

Figure 2. Cases reported in the 24-month and 12-month data submission

The ratio of 12-month cases counts to 24-month case counts varies by cancer site (Figure 3). Ignoring a few rare sites, the range is seen to extend from about 67% for chronic lymphocytic leukemia to 88% for uterine cancer (these figures were obtained by subtracting the value on the y-axis from 1). For certain sites such as CLL and possibly melanoma and prostate, the

Figure 3. Ratio of 12-month cases counts to 24-month case counts

projection of national rates may be less feasible given higher uncertainty. Knowing which cancer sites may be projected and under which circumstances will require further analysis. The above figures made use of all data available to the work group – specifically, data for cases diagnosed in 2016 submitted to NAACCR between

November, 2017 and January, 2018 (what we will henceforth refer to as 12-month or one-year data, even though technically it ranged between 11 and 13 months), and data for cases diagnosed in 2016 submitted to NAACCR in November, 2018 (what we will call 24-month or 2-year data, even though technically it was 23- month data). However, the picture is better than this if registries with an unusually poor 12-month reporting performance are excluded. Drawing upon prior analysis among SEER registries, the work group is proposing to limit the basis for national projections to only those registries that had a 12-month case count to 24-month case count ratio of at least 0.8 in at least 3 of the 4 most recent diagnosis years, and to those that were certified gold or silver in all four of these years.

Applying these criteria to the NAACCR data submissions for cases diagnosed between 2013 and 2016, 36 registries would meet these criteria (Figure 4). We anticipate that the picture would improve still further if NPCR rather than NAACCR and SEER submissions were used for the analysis. For many registries these are the same, but some registries separately submit 12-

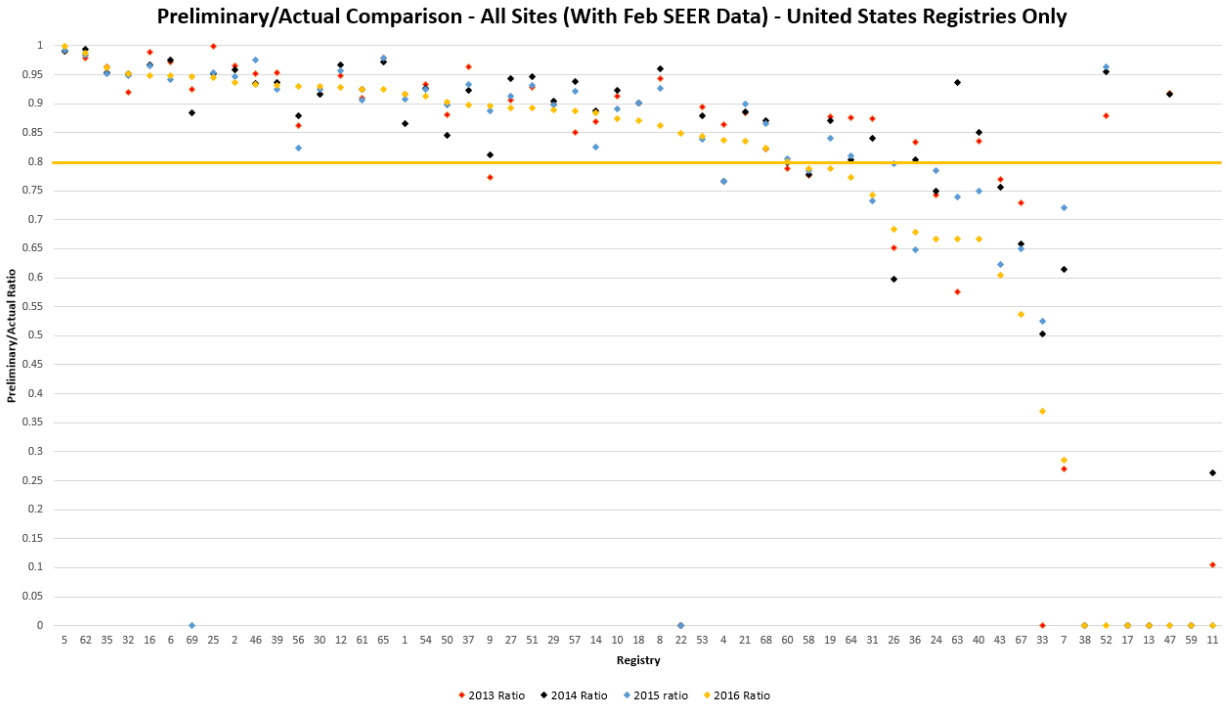

Figure 4. 2013-2016 Preliminary/Actual Comparison All Sites

month data to NAACCR in November and then again to NPCR in January. In one state the difference between these submissions is substantial – typically below 80% complete in the first submission, and above 90% complete in the second. In addition, there are several states that only submit their 12-month data to NPCR and not to NAACCR. The total number of states included could potentially be raised to at least 40.

Based on the 22-registry used in our in-depth assessment, 11 registries responded that they typically report 12-month data to NAACCR in January, 10 report in November, and one does either. Whether the registries who report in November make a separate submission in January to NPCR was not assessed. Regardless, if 12-month data are to be used going forward, there are obvious advantages to placing all registries on the same calendar. While it is possible to continue to analyze a mixture of 11-, 13-, and 14-month data, it complicates the analysis and delays the eventual release of the data by three months.

Measuring Completeness

The workgroup considered four different approaches for measuring completeness. The workgroup recommends that two of these be considered for continued development and assessment, one not to be further considered, and one to be considered to the extent that it informs the delay-adjustment methodology just described.

Incidence-to-mortality ratio method

This is the method currently used by NAACCR and NPCR to measure registry completeness. There are minor differences between the methods used by the respective organizations; a recent comparison found that the estimate differed by more than 1% for only five registries, and by more than 2% for just one registry. This summary considers only the NAACCR version of the method.

The method defines completeness as the ratio of observed to expected incidence rates for a registry. As is characteristic of such ratios, the average value across all registries is 1 or 100%; values are distributed around this average so that roughly half have values above 1 and roughly half have values below 1. The observed incidence rate is calculated by summing age-adjusted rates stratified by sex, race, and cancer site. Many race and site classifications have been assessed. Currently, the method considers 18 sites for men and 15 for women, including nearly all of the most common sites but excluding the two most common, breast and prostate. Currently, race comprises whites and blacks, but Hispanics and “other” categories also have been proposed for inclusion. For minority race groups, a rule of thumb has been that such groups must comprise at least 10% of the population in a registry to be included in a calculation.

The expected number of cases is obtained by multiplying the registry’s mortality rate by the national incidence-to-mortality rate ratio (IMRR), again stratified by sex, race, and cancer site. Thus if the national IMRR is 1.5 for, say, white male bladder cancer, then the incidence rate of white male bladder cancer in that state would be expected to be 1.5 times the mortality rate. In order to achieve a more stable measure, the IMRR makes use of 5 years of data, and the registry’s mortality rate uses 2 years of data (3 years for small registries with less than 500,000 people). National IMRRs use national mortality but incidence from 11 SEER registries.

The core assumption of this method is that incidence tracks mortality in a constant and universal manner by sex, race, and cancer site. This is not the case, of course, and indeed much of cancer surveillance is concerned with showing how this relationship is not constant, insofar as it is driven by factors such as screening, health care access, and quality of care. One way of addressing this issue has been to introduce an adjustment term to the method which effectively smooths the calculated IMRRs toward 100%, raising them for registries with low measured completeness and lowering them for registries with high measured completeness. However, the existence or magnitude of this adjustment term is only an estimate and lacks any empirical basis.

The primary advantages of this method are its long tenure and familiarity within the registry community and its transparency - it is a simple matter to independently verify the calculation using routine surveillance data. In addition, a spreadsheet is available that allows any registry or researcher to assess the implications of varying parameters such as sex, race, and site stratification and the adjustment term.

The primary disadvantages include the indefensibility of the constant IMRR assumption; instability of the measure for small registries, particularly the most sparsely-populated Canadian provinces and territories; exclusion of the most common cancer sites; and a seeming systematic underestimation of completeness in areas with heavily Hispanic populations. In recent years, nearly all NAACCR registries have been certified gold. For cases diagnosed in 2015, for example, 49 U.S. registries were certified gold, 6 silver, and 2 were uncertified. The registries certified silver or uncertified most often in recent years have included Nevada, Arizona, New Mexico, Los Angeles, and Minnesota. (Minnesota’s status was not due to completeness.) Four of these five are in the Southwest and each of these four has a large Hispanic population, representing the registries with the first, second, sixth and seventh largest percentages of Hispanics in the country. However, adding a Hispanic stratum into the method does not improve the completeness of these registries substantially, and this issue has not seemed to affect Texas, with the third highest Hispanic population.

A final limitation is that a registry’s classification as gold, silver, or uncertified can be sensitive to whether and how sex, race, and site are stratified, and which adjustment term is chosen. At the second in-person meeting in Vancouver, there was a demonstration of how one registry (Arizona) could have fallen into any of these three categories depending on what assumptions were made. Other registries straddled two categories. Prior work by one workgroup member showed that this characteristic is not only a property of parameter selections but of sampling variability (Das et al, 2008). The Vancouver presentation was limited to four registries, all of which have had difficulty meeting the NAACCR completeness standard consistently (New Mexico, Arizona, Nevada, Los Angeles). Work group members agreed it would be useful to expand this analysis to all registries to see if the conclusions are broadly applicable or confined to these negative outliers.

Modeling method

The modeling method is adapted from the method used to predict current cancer counts for the nation that was jointly developed by the National Cancer Institute and the American Cancer Society (Pickle et al, 2007; Das et al., 2008; Zhu et al, 2012). That method uses a hierarchical Poisson regression model which includes spatial and temporal random effects across counties and years of diagnosis. Using county-level cancer incidence counts from the CiNA Deluxe file stratified by age, sex, race, and diagnosis year as an input, it models incidence as a function of cancer mortality, sociodemographic variables for each county (urban/rural status, household characteristics, income, education, medical resources) and behavioral risk factors (smoking, obesity, health care coverage, cancer screening). To adapt this model to measure completeness, the work group agreed to remove the spatial and temporal random effects to minimize the problem of overfitting the data for large registries.

Completeness is then taken to be the ratio of the observed counts submitted by registries to the expected counts from the model. Like the IMRR method, this method is a relative method that implicitly assumes that completeness is 100% for the reference population, which is this case is the entire nation. Half of the population will belong to registries with completeness below 100%, and half will belong to registries with completeness above 100%.

Preliminary results for cases diagnosed in 2015 using 2015 as the reference year were presented at the in-person meeting in Vancouver and revealed a moderate correlation with the IMRR method and a narrower range of estimates (Figure 5). Since the method uses data from CiNA Deluxe, five registries that did not meet the standards for inclusion in this volume or that opted not to have their data included are not reflected in the figure. Additionally, states with multiple registries (California, Washington, Michigan) were grouped.

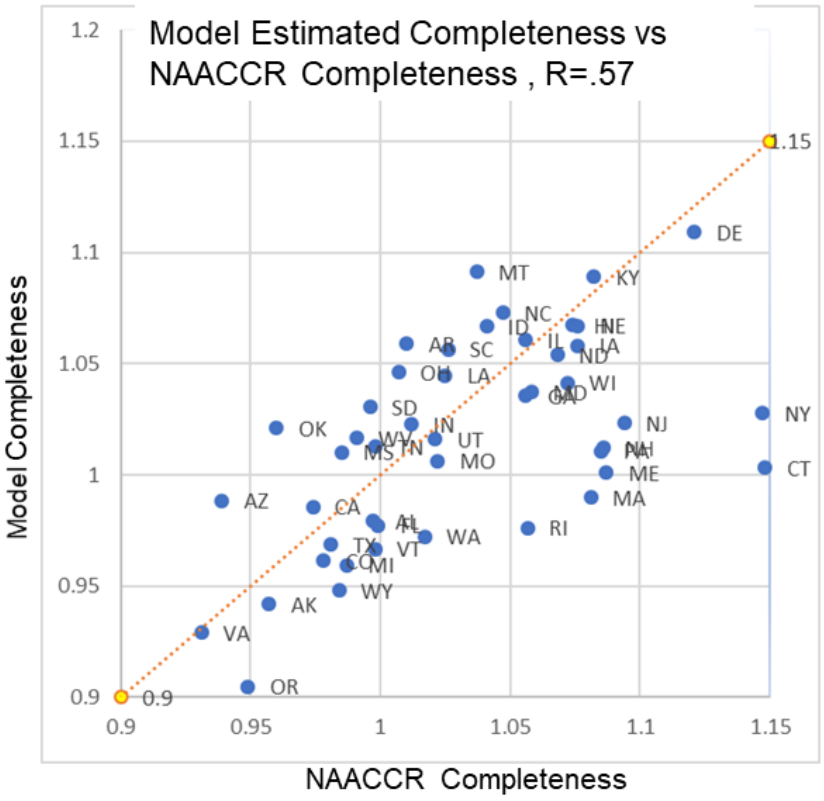

Figure 5. NAACCR Estimated vs. Actual Completeness

	NAACCR	Model
<88.9%	2	2
89.0-93.9%	2	1
94.0-99.9%	23	25
100.0-104.9%	10	17
105.0-109.9%	9	1
110.0%+	5	2
Total	51	48

Table 1.

Advantages of this method include the fact that all cases are counted equally, regardless of site or race. This would presumably remove the temptation for registries to delay the processing of some cases intentionally because they do not count toward completeness. It also does not depend on a problematic assumption (i.e. that incidence and mortality are perfectly correlated), but instead incorporates factors known to

influence cancer rates for which data are available, including demographic, behavioral and institutional data.

The major disadvantage is that the model is something of a black box and its expected case counts are not independently reproducible. A spreadsheet similar to that developed for the IMRR method could be developed to potentially ameliorate this problem. Another issue is that deriving both observed and expected counts from the same year of data means that changes in absolute case counts cannot be captured. This will likely be an issue when estimating completeness of 2018 data, where reporting reductions of 5% or even 10% are anticipated. The modeling method will assign tan average completeness of 100% to the nation, with individual registries mainly distributed between 95% and 105%, as if nothing had changed. Reintroducing temporal random effects to the model could solve this problem.

Flow and capture-recapture methods

The flow method is a method for measuring completeness that was developed in Great Britain about 20 years ago and was subsequently adopted by several European registries (Bullard et al. 2000). The term “flow” comes from the way that the computation draws upon the flow of cases through a registry as part of its routine operation. It categorizes all cancer cases into one of seven different categories. Five of these are easily counted:

- Patients alive at the time of interest and registered
- Patients deceased at the time of interest and registered, with cancer recorded on the death certificate
- Patients deceased at the time of interest and registered, with cancer not recorded on the death certificate
- Patients deceased at the time of interest but not registered, with cancer recorded on the death certificate, and with cancer information obtained through follow-back (“death certificate initiated” cases)
- Patients deceased at the time of interest but not registered, with cancer recorded on the death certificate, without cancer information obtained through follow-back (“death certificate only” cases)

The remaining two cannot be counted and must be estimated:

- Patients alive at the time of interest and not registered (“missing” cases)
- Patients deceased at the time of diagnosis, with cancer not recorded on the death certificate, and not registered (“lost” cases)

Estimating the missing and lost patients is accomplished by estimating the probability that a patient is registered while alive, the probability that cancer is accurately mentioned on a death certificate, and expected patient survival.

This method has a number of obvious drawbacks. It assumes that the survival of missing and lost cases matches those of recorded cases, when they would be expected to be quite different (Tervonen et al. 2017). It also requires that death certificates are timely and of high quality. Since, in general, death records require more than a year for

acquisition, linkage, and processing, it would be impossible to use the flow method to estimate completeness for periods of one year or less, a crucial consideration for this project. The method also requires registries to identify “death certificate initiated” cases, which is not a property U.S. registries record. Workgroup members felt that while some registries could likely deduce this information, others would find it difficult or impossible. On the positive side, completeness obtained from the flow method matches people’s intuitive sense of the concept – what you have is an estimate of the ratio of recorded cases to total cases, with an upper limit of 100%.

Silcocks and Robinson (2007) attempted to validate the flow method by creating the most realistic simulated data set they could, then removing up to 3% of the data for three different cancer sites and seeing if the method would correctly identify completeness values of 97% and above.

The group also discussed the capture-recapture method, whereby completeness is ascertained by comparing reporting to different entities (Brenner et al. 1995). In its simplest form, it involves comparing cases reported to a central registry and cases reported on a death certificate. Assuming the two are independent, then the number of cases not reported to either location (D) can be derived algebraically as:

$$D = (ABC + B2C + BC2) / (A2 + AB + AC)$$

Where variables A through D correspond to the following:

		Reported on death certificate	
		yes	no
Reported to cancer registry	yes	A	B
	no	C	D

Completeness is then simply 1 minus D.

A test of this approach on a past NAACCR data submission revealed immediate problem, however. One state reported zero death-certificate-only cases (cell C), implying a completeness of 100%. Another registry reported very few cancer deaths (cell B), implying poor completeness. In both instances the limiting factor was not the cancer registry data but the timeliness and accuracy of the mortality data. Whether that was because the mortality data was incomplete or in error in its original form or because the registry did not process it correctly is not known. In any case, “completeness” as measured this way ends up being a hybrid measure of both incidence and mortality completeness that is not interpretable.

Any method that relies heavily on death-certificate-only rate is also subject to the problem that as the rate diminishes toward zero, as has been the general trend in North American cancer registries in recent years, the proportion of cases that are not true DCOs increases. Cause of death coding, while extremely good, is not perfect, and an unpublished study conducted by one registry found that a significant share of the DCOs actually died of other causes; sometimes through what appeared to be simple typographical errors.

Given these issues and problems, the flow method and the two-source capture-recapture method were not given further consideration by the work group. We note that the term “capture-recapture” was not actually used during the summit – no one present made the connection to the earlier work of Hermann Brenner and others – and the notes refer to this as the “naïve method” because the assumptions that central cancer registries and vital records were independent and that the vital records data was error-free seemed naïve.

Internal method

In this approach, registries’ past case counts are taken as the sole input to predicting future case counts and assess the completeness of current counts. The logic is the same used to calculate delay-adjusted rates (see section 1), wherein rates are inflated by small amounts, depending on cancer site and other variables, based on historic patterns of delayed case reporting beyond the 24-month data submission. The advantage of this method is that it is straightforward to calculate and does not depend on external demographic or mortality data. SEER uses it to estimate completeness in its internal February (14-month) data submission. Many other registries do this implicitly when they provide mid-year progress reports back to facilities. For example, if a facility is told they had 93 cases reported at this time last year, but 76 cases this year, there is an implication that this number may be too low, that they may be behind in their submissions, because this year’s number is expected to be equal or greater.

The disadvantage is that it can be thought of as more of a measure of consistency than quality. To take a naïve example, imagine a national registry in a developing country where there are 10 hospitals, only 2 of which report to the registry. As long as this year’s case counts are similar to or greater than last year’s case counts – which will be true as long as the same two hospitals continue to report – completeness will appear high. The method thus assumes that a registry was virtually complete at least one time in the past. This is probably a reasonable assumption for United States registries, but it is difficult to know just which ones. A registry that was consistently 90% complete has the same issue as the registry in the developing country. For this reason, this approach was not considered for further development as a completeness measure.

Recommendations and next steps

No method of measuring completeness is perfect, but the modeling approach holds the most promise at present. Unlike any of the other approaches, it accounts for variability in health care systems and behavioral risk factors that explain much of the state-level variation in incidence, mortality, and their ratio. However, there is still much to do before it is considered ready. It must be run using data from all NAACCR member registries, including those excluded from the CiNA file. The manner in which prior years of data can inform the completeness measurement of the present data year must be worked out. Ways of making the method more transparent, through some combination of a spreadsheet tool, journal article or white paper, and education through webinars or conference workshops, must be determined.

For the existing IMRR method, we need to extend the analysis presented in Vancouver to all registries, in order to see how sensitive the method is to different choices of reference files, stratifications, categorizations, and adjustment terms.

With respect to estimating 12-month cancer incidence rates, we need to continue our work pinning down the inclusion criteria and applying the delay adjustment factors.

We believe that each of these activities can reasonably be completed during 2019. Pending feedback from the recipients of this report, the work group will reconvene conference calls as needed, if funding permits.

References

Brenner H, Stegmaier C, Ziegler H. Estimating completeness of cancer registration: an empirical evaluation of the two source capture-recapture approach in Germany. *Journal of Epidemiology and Community Health* 1995; 40(4): 426-430.

Bullard J, Coleman MP, Robinson D, Lutz J-M, Bell J, Peto J. Completeness of cancer registration: a new method for routine use. *British Journal of Cancer* 2000; 82: 1111-1116.

Das B, Clegg LX, Feuer EJ, Pickle LW. A new method to evaluate the completeness of case ascertainment by a cancer registry. *Cancer Causes and Control* 2008; 19(5): 515-525.

Lewis DR, Chen HS, Cockburn ME, Wu XC, Stroup AM, Midthune DN, et al. Early estimates of cancer incidence for 2015: expanding to include estimates for white and black races. *Cancer* 2018; 124(10): 2192-2204.

Pickle LW, Hao Y, Jemal A, Zou Z, Tiwari RC, Ward E, et al. A new method of estimating United States and state-level cancer incidence counts for the current calendar year. *CA: A Cancer Journal for Clinicians* 2007; 57: 30-42.

Silcocks PBS, Robinson D. Simulation modelling to validate the flow method for estimating completeness of case ascertainment by cancer registries. *Journal of Public Health* 2007; 29: 455-462.

Tervonen HE, Roder D, Morrell S, You H, Currow DC. Does exclusion of cancers registered only from death-certificate information diminish sociodemographic disparities in recorded survival? *Cancer Epidemiology* 2017; 48: 70-77.

Zhu L, Pickle LW, Ghosh K, Naishadham D, Kim H-J, Portier K, Chen H-S, et al. Predicting US- and state-level cancer counts for the current calendar year. Part I: Evaluation of temporal projection methods for mortality. *Cancer* 2012; 118(4): 1091-1099.

We would like to express our deep gratitude to the members of the Statistical Work Group who gave so generously of their time and expertise to this project. The group worked with purpose and thoughtfulness on finding the best way to measure registry completeness. It also assisted in the development of a potential time delay projection model for 12-month reporting. Your generous support is very much appreciated.

Frank Boscoe, Pumphandle

Huann-Sheng Chen, SEER

Barnali Das, NCHS

Rocky Feuer, SEER

Rick Firth, IMS

Donald Green, IMS

Kathy Huamani, SCG

Betsy Kohler, NAACCR

Andrew Lake, IMS

Lihua Liu, Los Angeles Cancer Surveillance Program - USC

Paul Sutton, CDC

Trevor Thompson, CDC

Paulette Valliere, NACDD

Kevin Ward, Metropolitan Atlanta SEER Registry

Hannah Weir, CDC

Chuck Wiggins, New Mexico Tumor Registry

Manxia Wu, CDC

Li Zhu, SEER

Joe Zou, IMS

Appendix D: Statistical Summit Notes

Marriott Washingtonian Center
Gaithersburg, MD
April 8–9, 2019

The North American Association of Central Cancer Registries (NAACCR) and the National Association of Chronic Disease Directors (NACDD) convened a Statistical Summit on April 8–9, 2019, to discuss the merits of various methods for estimating case completeness at NAACCR cancer registries and approaches to developing an improved method. Dr. Frank Boscoe, Founder of Pumphandle LLC, and previous NAACCR President-Elect, presided over the meeting. Ann Marie Hill of Rutgers University facilitated the meeting. NAACCR, NACDD, the Centers for Disease Control and Prevention's (CDC) National Program of Cancer Registries (NPCR), National Cancer Institute's (NCI) Surveillance, Epidemiology, and End Results Program (SEER), Information Management Services, Inc. (IMS), and cancer registry representatives participating in the Summit included:

Robert Anderson, National Center for Health Statistics (NCHS), CDC

Frank Boscoe, Pumphandle LLC

Kathy Brown-Huamani, the Scientific Consulting Group, Inc.

Huann-Sheng Chen, NCI

Barnali Das, NCHS, CDC

Eric (Rocky) Feuer, NCI

Don Green, IMS

Lori Havener, NAACCR

Ann Marie Hill, Rutgers University

Betsy Kohler, NAACCR

Andy Lake, IMS

Lihua Liu, Los Angeles Cancer Surveillance Program - USC

Danny Miller, IMS

Trevor Thompson, CDC

Paulette Valliere, NACDD

Kevin Ward, Metropolitan Atlanta SEER Registry

Hannah Weir, CDC

Charles (Chuck) Wiggins, New Mexico Tumor Registry

Manxia Wu, CDC

Joe Zhou, IMS

Li Zhu, NCI

Day 1: April 8, 2019

Action Items

- Dr. Ward agreed to update the NAACCR workbook by loading new data.
- Dr. Zhu agreed to re-run the data for the American Cancer Society (ACS) model with 1996–2015 data with the five missing uncertified registries included. She might run the model without random effects but will discuss the analysis approach with Drs. Das and Feuer before proceeding.

- Dr. Feuer agreed to redo the validation analyses of the 1-year NAACCR data for diagnosis years 2015–2018 with some modifications (e.g., only for major cancer sites).
- Dr. Thompson agreed to check on the best source of tumor record number.
- Dr. Boscoe agreed to test his Naïve Method using data for all NAACCR registries.
- Dr. Boscoe agreed to send Outlook notifications for planned calls on May 2, 2019, (1:30-3:00 pm) and May 31 (1:00-2:30 p.m.).

Background

Betsy Kohler

Ms. Kohler discussed the history behind the development of a method for measuring the completeness of cancer registry data. NAACCR first developed a data completeness measure approximately 25 years ago to aggregate data from registries across the United States and Canada in preparation for publishing Cancer in North America ,1996–2000 (CiNA). Completeness standards helped to clarify when registries had data of sufficient quality to publish incidence rates and eventually served as criteria for registry certification. NCI's SEER program and the CDC also have developed completeness measures. NAACCR, SEER, and CDC measures all have weaknesses but provide an indicator of whether a registry is capturing all cases in its jurisdiction.

Measures of completeness periodically have been reviewed, and attempts have been made to improve them. Dr. Ward recently worked on improving the NAACCR measure by including more racial/ethnic groups. Currently, NAACCR is evaluating the completeness of data submitted approximately 12-months after the end of the diagnosis year (early submission) and determining whether the measure now used to evaluate the completeness of data submitted approximately 24 months after the end of the diagnosis year (regular submission) also can be used to measure completeness at 12 months. Ms. Kohler would like the completeness measure to allow registries to determine how many cases in their coverage area have been captured at a specific point in time and whether completeness is adequate to use the aggregated data for public health and surveillance research efforts and publication.

Defining Completeness

Frank Boscoe and Hannah Weir

Efforts to certify cancer registries began in the late 1990s with the implementation of NAACCR Data Evaluation and Certification Committees. The Evaluation Committee first evaluated registry data using the NAACCR Data Evaluation and Publication Committee method of case completeness. The incidence/mortality rate ratio (IMRR) used in this method is problematic because case completeness often exceeds 100 percent, a result that is difficult to interpret. Efforts to improve the method have involved adding race/ethnicity and decreasing mortality. In modifying the completeness measure, an

important goal is to create a tool that registries can use to determine where to focus their case completeness improvement efforts.

Presentation

Frank Boscoe

Dr. Boscoe presented a working definition of completeness for the purposes of discussion at this Summit. He highlighted the difference between an idealized definition of completeness, with the goal of capturing every incident cancer case in the registry's jurisdiction, versus the practical definition of completeness with the goal of capturing incident cases that realistically can be captured because the cases have been captured by the medical system. The idealized definition likely would always lead to lower-than-expected completeness because of the "dark number" of unreported cases. The dark number concept comes from the field of criminology; analysts in this field attempt to estimate the dark number of cases from unofficial sources, such as self-report surveys, and then retroactively calculate estimated crime rates. Lyme disease surveillance offers an example of a practical definition of completeness. They are not trying to count all cases, just those that have been diagnosed. Reviews of clinical laboratory and private insurance claim data have been used for this purpose; estimates are only really to the nearest 100,000 cases.

The Flow Method (discussed on a prior call) and Dr. Boscoe's Naïve Method are potential measures of ideal completeness. Measures of practical completeness would have registerable cases as the denominator.

Measures of completeness also can be divided into external and internal approaches. An external measure can use either registries considered "high quality" or an average across all registries as the reference. External approaches, therefore, measure relative rather than absolute completeness. Measures of relative completeness include the current NAACCR IMRR and the SEER modeling methods. Internal approaches would measure completeness based on a comparison with a registry's own past performance. This approach assumes the registry achieved very high case completeness at some point, otherwise a registry that was consistently poor could appear to be doing ok. The current SEER completeness method is an example of an internal approach.

The current external NAACCR method measures completeness based on a comparison with other registries considered to have high quality data. The completeness of the external reference data is assumed to be 100 percent, which causes completeness rates for individual registries to be distributed around 100 percent. Some registries, therefore, would be expected to have above 100 percent completeness. NAACCR might need to better communicate the reason for 100 percent being the standard.

Delay adjustment is not a completeness measure, but can be a component of a completeness measure that considers data timeliness. Delay adjustment is not based on assumptions regarding case counts or trends, it simply estimates the increase in cases counts between points in time based on historical data.

Discussion

Participants agreed to focus on the practical definition of completeness, although consideration of an idealized definition completeness was not ruled out. Special studies might be used to retroactively improve estimates for case completeness, although ideal completeness likely is unattainable. Special studies of certain cancers that might be underreported in the medical system, such as thyroid cancer or chronic lymphocytic leukemia (CLL), might be informative. Some studies have compared self-reported cancer history to registry data. For example, Dr. Lynn Penberthy examined self-reported CLL and found that a large proportion of CLL cases were missing from cancer registries. Participants agreed that examining unreported cancers would be useful but should not be the focus of this meeting.

Initially, completeness measures were validated based on manual case finding. An independent party was responsible for looking for new cases that might have been missed by the central registry. No current statistical method has been validated based on manual case finding. This approach likely would not have found all cases but might have been more accurate, especially in an era when most cases were seen at hospitals. Another advantage of manual case finding was the ability to quantify the number of cases identified through case finding compared to regular facility reporting. NAACCR continues to rely primarily on hospitals to report cases, but the cases most often missed are those diagnosed outside of hospitals. Registries have information on the proportion of their cases received from hospitals and the proportion received from other sources. Participants suggested using these proportions to estimate the proportion of cases that might be missed.

Participants noted the need to distinguish between registries that might not meet standards because of reporting delays and registries with low case counts that never capture those missing cases. Both situations occur.

The IMRR is not constant across all populations. Attempts have been made to adjust for race/ethnicity, but the assumption that the IMRR is constant continues to be false.

Participants discussed whether the goal should be to develop a method for estimating overall completeness or completeness across different cancer sites. Participants agreed that completeness should be measured for different cancer sites, but they also might want to measure overall completeness. The barrier to estimating overall completeness in the past was that the NAACCR method excluded major cancer sites. When breast and prostate cases were added back into the measure, completeness estimates moved closer to 100 percent, which reduced completeness estimates for some registries. Participants agreed it is important to include counts for the major cancer sites when estimating completeness.

Participants emphasized the need to consider variance when examining completeness methods. Larger registries have smaller variances, which might put smaller registries at a disadvantage. One option would be to adjust standards based on registry size. Other

factors that generate variability in completeness between registries is health care access for different populations within the coverage area and care seeking patterns of those populations. The NAACCR algorithm was designed to be the measure that works best for most of the registries. Participants discussed whether registries that cover large populations with challenges to obtaining cancer care might improve their completeness counts over a longer period as death certificates (DCs) for missing cases are received. Registries handle DCs differently.

The current NAACCR method is a model that includes gender, race, and cancer site. Participants discussed other ecologic covariates that might be included in the model at the county or state level to adjust for variation between registries. Source of registry data might be one important covariate. Many registries are using a wider range of reporting sources than in the past, but reporting sources vary widely across registries. Registries with a wider range of reporting sources can be expected to have better completeness. Different registries also might have different proportions of cases for different cancers from different reporting sources. If this is the case, it might be worthwhile to examine whether these differences result from different local healthcare systems or different practices in case-finding and use of reporting sources. Participants suggested examining reporting source by cancer site. Most registries identify death certificate only (DCO) as a reporting source, but other reporting sources might not be documented. More detailed information on reporting sources would allow NAACCR to provide guidance to registries about stronger and weaker data sources overall and for certain cancer sites.

Reporting practices are changing with changes in hospital structure and a greater number of auxiliary care facilities taking over care for cancer patients. This trend merits consideration because it likely affects case completeness. NAACCR would need to examine local changes in health care systems to determine what system variables might affect case completeness for specific registries.

NAACCR will need to consider actions to take when a rapid change occurs at the national or registry level that has the potential to affect completeness.

Recapitulation of Conference Call Presentations—Strengths and Weaknesses of Each Method

Current NAACCR Completeness Method

Kevin Ward and Chuck Wiggins

Dr. Ward reviewed the findings of a task force organized to compare SEER, NPCR, and NAACCR completeness methods and examine modeling approaches for estimating completeness. A key strength of the NAACCR method is its transparency, which allows registries to evaluate their own data. The method also is robust and has worked well with minor modifications over time. Weaknesses of the method include the tendency to inflate estimates, leading to estimates over 100 percent; underestimation of

completeness in most states where Hispanics are a large proportion of the population; and problems measuring completeness for small populations, such as those in Canadian territories.

In the United States, 90 percent of NAACCR registries were gold certified in 2014 and 2015. Only two registries did not achieve certification in 2015 and only one in 2014. In both years, the four registries where Hispanics made up the largest proportion of the population received silver or no certification.

The current NAACCR Method initially calculates separate completeness estimates within each race strata (weighted proportion to race distribution). The race strata are White and Black... Completeness across cancer sites (with breast and prostate excluded) is calculated as the ratio of observed cancer incidence to expected cancer incidence within each race/gender strata. The ratio is weighted by gender within each race strata and age-adjusted to the 2000 U.S. population. The expected incidence rate is the 5- year SEER11 incidence rate divided by the 5-year U.S. mortality rate (i.e., IMRR) multiplied by the 2-year registry mortality rate. An adjustment for Hispanic ethnicity has been explored using alternate race strata of non-Hispanic Black, non-Hispanic White, non-Hispanic Other, and Hispanic, but only strata representing at least 10 percent of the population in the registry area are included in the calculation

The Method assumes a constant IMRR within each subgroup. The NAACCR Method attempts to adjust for case fatality, but registries with higher-than-expected mortality relative to incidence still have lower completeness estimates. Survival disparities, therefore, create a problem for registries serving large populations affected by those disparities. Incorporating survival or stage into the model might resolve this problem.

Dr. Wiggins demonstrated how increases in mortality could substantially reduce completeness using the NAACCR algorithm with a simulated data set. He argued that the NAACCR Method IMRR is strongly affected by cancer disparities when the populations affected by those disparities represent a relatively large proportion of the total population covered by a registry. For example, in New Mexico, non-Hispanic Whites are the minority, and Hispanics and Native Americans combined comprise nearly 60 percent of the population. Research has shown that these populations tend to access care differently than most of the U.S. population. Hispanic populations also tend to have a lower median age when compared to non-Hispanic White populations. Participants noted that some attempt is made to control for disparities by calculating a separate IMRR for each racial/ethnic group examined, but this approach might not be sufficient to eliminate the effect of disparities on completeness estimates. The four racial/ethnic groups used in the NAACCR Method might not be sufficient to capture other disparities, particularly disparities within the Hispanic group. For example, cancer outcomes in Cuban Americans in Florida might differ markedly from outcomes in Mexican Americans in New Mexico. In addition, Dr. Wiggins and colleagues performed analyses demonstrating profound differences in cancer incidence between different Native American populations across the U. S. Dr. Wiggins concluded that the

assumption of a constant IMRR was invalid because that ratio can vary by race/ethnicity, urban versus rural residence, and other factors. A measure that does not rely on this assumption might be more accurate.

The fact that SEER11 incidence is used to calculate expected incidence creates an independence problem because some NAACCR registries that are certified are part of SEER11. Normally, the group examined should not comprise a large portion of the reference group. Some NAACCR registries that were among the SEER11 registries contribute a large proportion of the reference population, which violates statistical assumptions. Participants discussed whether a registry that comprises a large proportion of the SEER11 sample receives an advantage when completeness is calculated using the NAACCR algorithm. NAACCR could conduct analyses to examine this question or address concerns by replacing the SEER11 incidence rate in the model with another incidence rate, such as that of all U.S. NAACCR registries combined. Participants appeared to favor the suggestion to replace SEER11 incidence with incidence for all U.S. registries to improve fairness, particularly in view of the fact that most NAACCR registries now have a level of completeness that is acceptable for research purposes.

Another major limitation of the current NAACCR Method is the exclusion of the most common cancers—breast and prostate—from calculations. Participants generally agreed that breast and prostate cancers should be included in future completeness estimates.

Modeling Methods

Li Zhu

Dr. Zhu discussed a modified version of Dr. Das's spatial-temporal model developed in collaboration with the ACS to produce Cancer Facts and Figures 2019. This model was run using data from 1996 to 2015 for all but five registries, county by county, for the ACS report. The explicit statistical model includes mortality, demographic, lifestyle, and other variables that have been shown to affect incidence and these variables have been updated periodically. Participants noted that identifying the best covariates would be important. Dr. Zhu and colleagues used the expected case counts from this model and compared them with observed case counts for NAACCR-certified U.S. state registries. The completeness rates generated by this model are correlated with completeness rates generated by the current NAACCR Method, but the spatial-temporal model generates fewer completeness estimates above 100 percent. The advantages of the model are that it borrows strength across years and space; is based on county-level data; and includes all racial/ethnic groups, both genders, and all cancer sites. The model mimics observed counts very well, which might create problems when used to evaluate data quality. In smaller areas, observed and expected incidence are almost exactly aligned across several diagnosis years. Dr. Feuer explained that this method generates the proportion of covariates contributing to completeness for each year in addition to the completeness estimate. The problem with the model is that large

registries with small variance would have high completeness estimates regardless of performance. The goal would be to use only the portion of the model that is a function of the covariates. This alternative model would be run each year; retain covariates on mortality, demographic factors, and lifestyle factors (e.g., smoking prevalence, screening behaviors); and exclude spatial and temporal effects. Dr. Zhu clarified that the covariates currently in the model are based on county-level data and are divided by race or gender. Dr. Feuer recommended adding stage distribution to the model (at either the state or county level). County-level covariates might work better than state-level covariates for capturing variation. If state-level data are used for the covariates, dividing covariates into race/gender categories might improve accuracy.

The model has been used to predict counts and was validated. The model now is being updated, and cross-validation will be performed. The fact that the model considers random effects improves accuracy but might make the model overly complex (and therefore less transparent) for the purpose of measuring registry case completeness.

Delay Adjustment Methods

Rocky Feuer

Dr. Feuer discussed the use of the reporting delay adjustment method to measure completeness for 1-year data. Questions to consider regarding the use of this method include: (1) Should NAACCR certify 1-year data? and, (2) Can delay adjustment produce 1-year data of adequate quality? Delay adjustment has allowed SEER to produce trends using 1-year data from registries. The method uses delay factors based on historical data to predict how much case counts will increase over time. More recent data are more heavily weighted. The method assumes that data are complete 11 years after the diagnosis year; therefore, the case count 11 years post diagnosis represents the expected count. Delay factors by themselves are not a viable measure of completeness but can serve as an indicator of registry quality. 1/delay is a metric measured on a scale similar to that for completeness that represents the underreporting of current cases compared to what might be found eventually. 1/delay is the observed number of cases after a set number of years of delay (usually 2 years) divided by the expected number of cases after an 11-year delay. The delay method measures additional cases that are expected to be received in the future (usually 3 to 11 years later), whereas the NAACCR Method measures expected cases after 2 years. Reporting delay can be used as component of completeness. One component would be the expected number of cases for a certain number of years after submission. Expected cases would be based on an average of all registries or SEER registries for the submission, adjusted for gender, race, mortality, fatality, and other relevant characteristics. A second component could be the additional expected number of cases after 11 years, if NAACCR wanted to examine this number. This component would use a composite delay factor for the group of registries used to compute the expected number in the first component. If a component of the delay adjustment method was

included in the NAACCR algorithm, completeness standards might need to be lowered because the adjustment is based on what would be expected after 11 years rather than 2 years. Alternatively, the standard could remain the same, but fewer registries would be certified. The delay adjustment method also could be used to estimate the number of cases that never would be found, but Dr. Feuer did not recommend this approach. IMRRs are based on 5-year estimates, and 2-year mortality is used in the NAACCR Method, so delay adjustment might not be necessary.

All registries experience delayed reporting of some cases, and should estimate the proportion of cases still missing at the time of a submission. Participants asked what is known about these cases that are received late. Others responded that most late reporting of cases is linked to patient care patterns, which may vary by cancer site but is likely to be consistent across registries. Other factors associated with reporting delays, such as a new reporting facility, affect individual registries and are difficult to predict.

Dr. Feuer discussed the use of delay adjustment for SEER registries, which have an early data submission date in February of each year, approximately 1-year after the end of a diagnosis year. Delay adjustment is used to predict the shortfall in the February submission compared to the November submission, which is approximately 2 years after the diagnosis year. Four February submissions were used to predict the delay for the SEER registries. The delay factors for the February submission are more than twice as large as those for the November submission, but produce highly accurate adjustments. This level of accuracy allows delay-adjusted data from the February submission to be used to calculate trends. These trends are posted on a web site and are presented as preliminary. It is not clear how delay adjustment would work for subsets of data from the SEER submissions.

NAACCR has been receiving early data submissions for enough years to perform delay adjustment of the next January submission. Dr. Feuer analyzed NAACCR January submissions and noted that their completeness was improving over time. NAACCR registries are allowed to submit the 1-year data before January, which might explain improvements in completeness. A quality program might help improve the ratio of the January to December case counts. Data quality criteria could be examined after one more 1-year data submission is received by NAACCR, but significant, sudden improvements in data quality would make it difficult to produce accurate delay factors.

Flow and Naïve Methods

Frank Boscoe

Dr. Boscoe discussed the Flow Method developed at European registries to estimate missing living cases and lost deceased cases for whom the cancer is not noted on the DC. Missing and lost cases are derived from the time distribution of three probabilities: (1) survival, (2) registration during the patient's life, and (3) the cancer being mentioned on the DC. An attempt was made to validate the Flow Method with a simulated data set and the results of this analysis were published. With this method, over time the estimate

approaches the true completeness value. The advantage of the Flow Method is its simplicity and ability to be easily explained. A major weakness of the Flow Method is the time needed to achieve a high level of confidence regarding the estimate, which would be a minimum of 5 years after the diagnosis year. Another weakness is that the method appears to depend on assumptions that cannot always be supported.

Dr. Boscoe also presented a Naïve Method for measuring completeness. This simple method would calculate missing cancer cases not on the DC or reported to the registry based on the assumption that these represent independent events. The DCOs used in this calculation would be for the most recent diagnosis year, and historical submissions could be examined to predict case counts and DCOs. Participants discussed the fact that DCs can include as many as 20 cause of death (CoD) fields and cancer might be mentioned on any of them. It is unclear whether state registries examine all fields to identify a reportable cancer that might have been missed.

Data Collected by NAACCR

Kevin Ward

Dr. Ward presented a list of data elements included in the NAACCR submission. These elements include patient care variables; stage at diagnosis; county variables; state and other census tract information; race/ethnicity including NAACCR Hispanic Identification Algorithm, Asian/Pacific Islander, and Indian Health Service variables; sex; age at diagnosis; cancer-specific clinical codes; diagnostic confirmation (indicating whether cancer was pathologically confirmed); type of reporting source (sources are prioritized by best expected source, so hospital takes precedence over other reporting sources); histology and behavior; primary payer at diagnosis; socioeconomic status (SES) measures if available (including one census tract-level poverty measure); tumor, node, and metastasis (TNM) staging variables and summary stage; treatment variables (dates of initial treatment course, surgery variables, radiation and adjuvant therapy variables); date of last contact; vital status and source of this information; collaborative stage variables; and some site-specific factors.

Participants expressed particular interest in the reporting source variable. Relatively low diversity of reporting sources could be an indicator of poor data quality and might be incorporated into the completeness estimation method for some cancer sites.

Participants also expressed interest in survival data. The NAACCR data set includes one census tract–level survival field. Most states have performed linkages to obtain survival data. Stage distribution data would be expected to be related to survival, but a large proportion of stage fields are coded as “unknown.” “Unknown” stage could indicate missing stage information or information that is not sufficient for distinguishing between regional and distant metastases. NAACCR collects data on cause-specific death, which would permit cause-specific survival analysis.

Group Discussion

Participants joined one of the following four separate discussion groups as indicated below:

Topic	Leaders	Members
Existing NAACCR Method	Kevin Ward Chuck Wiggins	Lihua Liu, Andy Lake, Hannah Weir, Manxia Wu
Modeling Method	Li Zhu	Barnali Das, Trevor Thompson, Joe Zhou
Other Methods (SEER Internal using historical data, Flow, Naïve)	Frank Boscoe Rocky Feuer	Robert Anderson, Huann-Sheng Chen, Betsy Kohler

Participants agreed to discuss delay adjustment in all groups. Delay adjustment could be incorporated into any method.

Participants were encouraged to interact across groups to cross-pollinate ideas. Ideally, the final method selected by the group would be informed by multiple methods. Groups also were asked to consider approaches for testing the methods they discussed, as well as modifications to those methods. The goal is to generate four to 10 measures of completeness that IMS can test and compare using NAACCR data. Each group was asked to recommend at least one method. Groups were asked to consider the following questions: (1) Should we continue to consider each method for measuring NAACCR registry case completeness? (2) Are there improved versions of each method that should be tested? (3) Is each method and modified method testable at this point? (4) If certain methods are not viable for measuring completeness, could a component of those methods be integrated usefully into another acceptable method? (5) How should NAACCR move forward with testing each selected method?

Specifically, the NAACCR Method group was asked to develop a recommendation regarding whether to continue using the current algorithm or make recommendations for a modified version of the algorithm. The Modeling Method group was asked to determine which method is most viable for measuring case completeness and make recommendations for testing models. The Other Methods group was asked to consider how the other methods could be integrated into the NAACCR or Modeling Methods and how these could be tested. Groups met separately for approximately 1 hour.

NAACCR Method Report Back

Kevin Ward

The NAACCR Method group proposed some enhancements to the existing method alone and in combination with other methods. For example, delay adjustment might be used to help validate models. Dr. Wiggins raised the question of possible variation in the

quality of data provided by the Bureau of Vital Statistics in different states. This type of variation could affect the completeness of mortality data and, in turn, affect the IMRR and registry case completeness. Other participants explained that mortality reporting does not vary much between states, overall, but can vary at different points in time because of delays in death reporting. Timeliness of death reporting, however, varies substantially by CoD. For example, reporting of deaths caused by drug overdose, suicide, and homicide are delayed about 6–9 months relative to other CoDs. States also vary in lag time for reporting deaths, but cases are reported within about 1 year, so the lag should not affect cancer registry completeness. State Bureaus of Vital Statistics now auto code CoD in approximately 75 percent of cases and expect to auto code CoD for 95 percent of cases soon. Dr. Wiggins suggested matching the cancer registry data to death data to determine what percentage of registry cases die of cancer.

Participants noted that certain states are likely to have higher mortality rates. For example, states where substantial numbers of people retire and subsequently die might experience higher mortality rates. Dr. Feuer's analyses of incidence-based mortality for lung cancer found discrepancies between registry cases and DC-reported cases of lung cancer. When researchers examined all cancer cases, however, they found that many DC-reported lung cancers were originally other cancers that had metastasized to the lungs. This finding suggested that DC-based lung cancer mortality was not as accurate as incidence-based mortality for this cancer. Incidence-based mortality, which represents only cases diagnosed in the state, might be used to examine migration of cancer cases between registries.

The NAACCR Method group also discussed additional adjustments related to case fatality. One concern was variation in case fatality within racial/ethnic these groups. The group discussed cause-specific survival data in the SEER registries. The two registries with the largest Hispanic populations had the lowest cause-specific survival among Hispanics. One of these registries also had the lowest cause-specific survival among Asians. Review of the SEER data also revealed that cause-specific survival varies substantially within racial/ethnic groups. These findings raise the question of how to adjust for these variations in survival in the NAACCR algorithm. Dr. Ward also examined case fatality at the four registries with the largest Hispanic populations that were not gold certified in the past 2 years. He found that raising the adjustment factors for these registries increased completeness estimates by four or five percentage points for two of the registries but did not change case completeness rates for the other two registries. Completeness estimates might not have changed for the latter two registries because the states are poorly resourced and, as a result, could have lower actual completeness rates.

The group considered whether the model would be improved with the use of age-specific rather than age-adjusted rates. NAACCR formerly used age-specific rates. Dr. Zhu's approach that adjusts to the registry rather than the U.S. standard for a specific race/gender group might achieve the same effect. Participants suggested adjusting a (i)

rates for all registries with the i value determined by the state-specific survival. Dr. Ward suggested raising the value to improve completeness for registries with high mortality rates relative to the United States as a whole. He also suggested testing methods that adjust for case fatality by comparing to NAACCR Method estimates for highly fatal cancers.

A combination of methods might be used for registry certification purposes. Components of certain methods could improve other methods. Dr. Ward supported the IMRR because it does not rely on historical incidence data (it uses mortality data to estimate expected incidence). He suggested combining this external method with an internal method that relies on historical data, such as the SEER method, and use weighting to ensure appropriate contributions from each method.

The group could not identify a way to use reporting source in the NAACCR Method. Other participants suggested examining reporting source to determine ways to improve actual completeness. Registries will be able to better assess completeness with real-time pathology reporting and the ability to distinguish incident and prevalent pathology. Optimal pathology reporting should identify approximately 94 percent of cancers in the United States. Assessment of completeness also could improve with knowledge about the proportion of clinically diagnosed cases and variability in this proportion across registries.

The group considered the possibility of using stage data but confirmed that too many cases had stage coded as “unknown.” Stage is the strongest predictor of survival, but some other survival adjustment factor might be needed until the quality of stage data improves. Survival data are fairly complete, but completeness varies across registries.

Participants proposed using a national reference rather than using SEER11 as the reference. If this change causes completeness to decline for some registries, however, it could be problematic. First, NAACCR might want to investigate what factors drive differences in completeness.

Participants agreed to upload new data and possibly modify the NAACCR workbook. One variable that could be added is “region” to allow for regional comparisons.

Dr. Ward suggested running all proposed completeness models using NAACCR data for different populations to see differences in completeness results and trends. It might be informative to rank registries using the NAACCR and modeling methods. Differences in ranking would highlight potential discrepancies in methods. Participants generally agreed that ranks would be useful.

Participants also suggested running different permutations of the NAACCR Method using a single registry’s data. The New York registry might be ideal because of the diversity of populations and the ability to compare New York City with the demographically different portions of the state outside New York City. Dr. Ward noted

that every registry's completeness increases when race is removed from the I/M model, so investigating the effects of race on completeness would be worthwhile.

Participants noted that cases from hospitals run by the Veteran's Health Administration of the U.S. Department of Veterans Affairs (VA) are missing for many registries because these hospitals do not report in many locations. VA hospital cases are excluded from submissions in California because the proportion of cases contributed by VA hospitals varies substantially across registries. Participants discussed approaches for handling missing VA hospital cases at other NAACCR registries. VA hospital cases cannot be removed retrospectively. If most registries are missing VA hospital cases, no registries would be penalized because their completeness estimates would be compared to the U.S. average. CDC collects data on the number of VA hospital cases by state, so it might be possible to estimate the proportion of missing cases that are accounted for by VA facilities. Some participants believed that VA hospital cases do not substantially affect differences in completeness across registries.

Adjusting for Case Fatality in the NAACCR Method

Barnali Das

Case fatality is not constant across registries, so failure to adjust for case fatality can produce inaccurate completeness estimates. If a registry has low case fatality compared to the average, it will have a lower- than-expected incidence rate and higher completeness. Conversely, higher-than-average case fatality leads to a higher-than-expected incidence rate and lower completeness. Some ad hoc survival adjustments already are included in the NAACCR algorithm to address differences in case fatality. For example, breast and prostate cancer are excluded because these cancers have different case fatality at different locations because of differences in screening rates. The IMRR also is constant for a fixed geographic unit. The use of SEER incidence divided by U.S. incidence rather than SEER incidence divided by SEER mortality is another adjustment for case fatality differences, although this adjustment might not work for all registries.

Drs. Das and Tom Tucker performed an in-depth examination of different approaches for calculating case fatality using data from SEER registries. This analysis was done before survival data was available for many other NAACCR registries. Dr. Tucker decided to use differences in site- and sex-specific 12-month survival rates to adjust for differences in case fatality in place of the ad hoc adjustments employed by the NAACCR Method. Dr. Tucker decided not to use race-specific survival, but this could be added to the model. Drs. Das and Tucker calculated mortality adjustment factor alpha with SEER as the target. If the alpha for a registry differed from the SEER standard, the mortality rate was adjusted. Dr. Das ran this new adjusted model excluding breast and prostate cancers, using SEER incidence divided by SEER mortality rather than SEER incidence divided by U.S. mortality for the IMRR. This analysis generated completeness rates closer to 100 percent for the registries. Notably, the analysis generated higher completeness for the New Mexico and Utah registries when compared to the NAACCR

Method. Dr. Das then ran the same model, still excluding breast and prostate cancers, using SEER incidence divided by U.S. mortality for the IMRR. This approach over adjusted resulting in completeness estimates over 100 percent for all registries. Dr. Das tried this model including breast cancer but excluding prostate cancer and including both. These analyses revealed that including breast and prostate cancers was not problematic when the model was adjusted for case fatality. Dr. Das concluded that it is important to adjust for case fatality, which brings completeness estimates closer to 100 percent. She also recommended not using SEER incidence divided by U.S. mortality rates for the IMRR in this adjusted model because of over adjustment. In addition, the same geographic units should be used in calculating the IMRR.

Day 2: April 9, 2019

Modeling Report Back

Li Zhu

Advantages of modeling include the ability to add covariates, including county-level small area variables related to behavior and sociodemographic, and the flexibility to remove random effects and add data sources and delay factors. The current ACS model takes the delay adjusted observed counts divided by delay adjusted expected counts. Participants discussed whether standards should be changed if this model is used, because the model lowers completeness rates, with rates for many registries falling below 95 percent, and rates for some currently certified registries falling below the certification standard. The model still needs fine tuning, so the completeness rates might change. If NAACCR has confidence in the delay factors for this model and those factors are consistent over time, the model might help to measure delays in reporting for registries. Model results should be compared to the completeness rates generated by the NAACCR Method when delay adjustment is performed. Participants proposed examining how much control registries have over delays in reporting, reasons for those delays, how much delay should be expected 2 years after the diagnosis year, and, on average, how many more cases can be expected to be added between 2 and 11 years after diagnosis. Ms. Kohler expressed interest in this proposal. Participants discussed whether to use registry-specific delay or average delay across all registries. Participants suggested examining average case count after 11 years, after delay factors and registry-specific delay adjusted counts have been put in the model, and after adjusting for covariates. Participants also suggested using delay factors to rank registries for cancers that historically have more delayed reporting. If registries demonstrate no delay, they either are not finding new cases after a submission or are capturing cases in a timely manner. The model might be measuring expected case counts rather than completeness and, therefore, might need to be modified before it can be used as a measure of completeness.

Dr. Das recently developed a model to measure the variability of completeness rates at each registry and demonstrated how the difference between predicted and observed

case counts is greater in states with small populations. The model included mortality as a fixed covariate; other covariates were selected through a stepwise process. The model presented by Dr. Zhu also can incorporate a measure of registry variability. Registry variability of completeness estimates would be presented as a confidence interval that might cross different certification thresholds. Ms. Kohler suggested that registries might be certified at the highest level reached by their confidence interval. Dr. Das' paper discusses approaches for calculating the probability of a registry's completeness rate falling into each certification category, so an alternative approach would be to certify the registry in the category with the highest probability of being correct. Participants discussed covariates that might be useful to include in this type of model, such as reporting source, county-level sociodemographic, stage distribution, or survival. Coefficients in the model would provide some information about the variables that are most likely to influence completeness, and the degree to which different covariates influenced the final completeness rate could be reported. A disadvantage of this model is its complexity, but Dr. Das believes that the model is transparent in terms of what goes into it and how it arrives at the completeness estimate. The model also is flexible and allows for appropriate weighting and easy adjustment of the weights once the model is built. There also are strong rationales for the model's components. In addition, the model generates expected case counts by cancer site. If the model is used, a decision would need to be made regarding whether to use state or county level data. A participant noted that models built on national data would give less weight to smaller registries. This is a problem for rarer cancers.

The greatest challenge for modeling is determining how to validate a model. The modeling group proposed validating models using data from a test registry. Past projected counts for that registry, perhaps obtained from Cancer Facts and Figures, would be compared to observed counts for a given year. Another method would be needed to validate how well a model measures completeness, perhaps through reabstraction of cases at the test registry. Participants indicated that reabstraction probably was not feasible. An important consideration when selecting models would be whether they could be run with 1-year data.

Dr. Feuer proposed comparing the different models with different covariates with different years of data. Dr. Das expressed concerns about overfitting but agreed this approach was feasible. This testing also would generate a measure of average observed completeness across registries, which would inform NAACCR standard setting. The average would be based on the assumption that 2015 was a normal diagnosis year. The American Community Survey, ongoing annual Census, and other new data sources provide useful covariates for testing. Covariates likely will not change much from year to year. Modeling methods would be useful for determining causes of poor completeness or timeliness.

Canadian registries present a challenge for modeling. Race data are not collected in Canada and the population and health system are different compared to the United

States. Participants discussed whether to use a separate model to calculate completeness for the Canadian registries.

Delayed Reporting Issues

Participants emphasized the need to recognize that not all registries operate under the same conditions, so reporting criteria might need to differ depending on those conditions. Differences between registries have meaning beyond completeness and timeliness that is worth investigating. Reporting delay might be more within the control of the registry than other factors that affect completeness. Delay also might be a better indicator of quality than completeness. Delay factors could be used to validate models. Participants suggested identifying differences between registries with larger and smaller delay factors and comparing how well models perform for these two groups of registries. Delay factors for SEER and NPCR registries differ. Participants expressed interest in examining the types of cancer cases that are reported late and possibly those that might never be reported to registries.

Dr. Boscoe mentioned another NAACCR project that involves interviews with registry staff to determine reasons for problems in timeliness of reporting. Some information collected suggests that cases diagnosed outside hospitals eventually go to a hospital, often several years after diagnosis. NAACCR could have a sample of registries examine cases that they receive after submission for reporting source, pathologic confirmation, and other variables that might explain the delay. SEER*DMS can show when a case came in and how many times it was modified before it became a Cancer/Tumor/Case. This information might point to registry processes that explain delayed reporting.

Dr. Zhu is able to examine which cases are delayed, and IMS already has examined this issue and has data on reporting source. It would be useful to construct a data set to identify very delayed cases and their reporting sources for each state. Currently, no information is available on whether a case came from another state. The number of cases that are never reported also is unknown.

Other Methods Report Back

Frank Boscoe and Rocky Feuer

The NAACCR registry reporting source categories include hospital, radiation facility or medical oncology center, physician's office, nursing home, DC, autopsy only, and other hospital outpatient unit. Dr. Boscoe confirmed the high variability in reporting sources across registries but did not see a strong correlation between reporting sources and registry completeness measures. High levels of cases reported by hospitals suggest some cases might have been missed, although this could simply reflect the state's healthcare system. Because of Dr. Boscoe's findings, participants questioned the utility of reporting source as a covariate. Reporting source might have more effect on the model if categories were collapsed. Participants recommended including reporting source in models, but it might be dropped in stepwise analysis because of low variation. Reporting source might be useful in predicting the distribution of delayed cases reported

at different points in time after the submission. The New York registry might serve as useful case study of the relationship between reporting source and completeness.

Dr. Boscoe described the distribution of completeness rates if the U.S. average rather than SEER11 was the incidence reference. He explained that outliers with completeness rates well above the average would make registries with below-average rates less likely to be certified because they would move farther from the average.

SEER Internal Method

Huann-Sheng Chen

Dr. Chen discussed the SEER Internal Method, which can predict case counts up to 11 years. The model assumes that delay is constant, but the expectation is that delayed reporting will decrease. SEER examines changes in delay by comparing delay for earlier and later years. If delay improves, separate analyses are run for years with less delay. Improvements in delay for smaller registries might not be identified because of insufficient power to measure the change. The new SEER method uses Joinpoint to analyze trends unlike the older linear SEER method. The year that SEER switched to the new method, analysts compared registry rates using both methods. SEER has modified the method based on feedback from the registries. A weakness of the SEER Method is that registries with consistently poor completeness would not be identified because each year's completeness rates are compared to the registry's own history of completeness. Over time poor completeness would become clear based on information from DCs, but this would take too long. For most registries, delay is improving or remaining stable. On average, about three to four percent of cases are delayed at SEER registries.

Participants discussed whether to test the SEER method for all NAACCR registries. This method likely would inflate completeness for some registries. It would be informative to see the outliers on a scatter plot of SEER Method results for all registries. The SEER Method might be useful for examining delay patterns. For example, it is expected that small numbers of delayed cases are going to be received over time. A problem arises when a registry submits a large amount of delayed cases at one time. In these cases, Ms. Kohler would like registries to examine the delayed cases for reporting source and other factors that might explain the delay. Participants suggested presenting a measure of delay/timeliness separate from the completeness measure.

Naïve and Flow Methods

Frank Boscoe

Dr. Boscoe discussed the Naïve Method, which uses only a few variables. This simple method might not work for every cancer site and every registry because of small cell sizes. Dr. Boscoe recommended including the Naïve Method in method comparison analyses, because correlation with other completeness measures might be informative.

A problem with the Naïve Method is that some cases in a specific diagnosis year will not fall neatly into any of the following four categories: (1) cancer captured by the registry and reported on the DC, (2) cancer not captured by the registry but reported on the DC, (3) cancer captured by the registry but not reported on the DC or no DC is available, and (4) cancer not captured by the registry and not reported on a DC. For example, some cancer deaths might have been diagnosed in another state and should not be in the registry. Participants discussed the possible effect of cancer patients who are diagnosed in one state and die in another state. Some "snowbird" states have agreements with registries in northern states. Participants also noted that the four categories might not be independent. The number in the fourth category likely is unverifiable.

The Other Methods group did not see the Flow Method as a feasible measure of completeness. This method, as published, depends on data items that are not collected or not collected well by NAACCR registries. Dr. Boscoe did not recommend including this method in the comparison analyses.

Testing the Methods

Participants recommended testing different models using 2015 data with different sets of covariates, with and without random effects, at the state and county level. Recommended covariates included survival and stage distribution. They recommended testing and comparing different variations of the NAACCR Method, the new SEER Method, and modeling methods. Participants recommended using data from all NAACCR registries, not just the certified registries. An alternative for testing purposes would be to use SEER data only. IMS will perform the testing of most methods.

Participants discussed the timetable for the analyses. They recommended a monthly check in to provide updates and share and resolve problems together. The deadline for completing testing of the methods is the end of July 2019, but there is an opportunity for a no-cost extension. The Canadian model is low priority and likely will take more than six months to develop and test. Ms. Kohler would like some analyses to be run before the NAACCR conference so that the results can be discussed at that meeting. The NAACCR conference is scheduled for June 9–13, 2019.

One task is comparing completeness estimates produced by the different methods, another is examining the ability of different methods to accurately estimate 1-year trends. Dr. Ward recommended a simple approach to estimating these trends: using 2-year estimates when they are received, and NAACCR has confidence in them. The question that needs to be answered is: "How well can they estimate rates and then trends based on 1-year NAACCR data?" Once NAACCR has an improved measure of completeness, the Association can examine the possibility of certifying 1-year data. In addition, if NAACCR plans to release 1-year incidence rates at some point, those rates will need to be produced earlier than 15 months after the diagnosis year. Participants asked about potential users of 1-year data. These data would be useful for identifying

sudden changes in cancer trends early, conducting certain kinds of research (especially on childhood cancer), and informing policy. A more realistic goal might be to have NAACCR 1-year data available on January 1 of the second year after diagnosis instead of April 15. SEER already has been able to estimate rates and produce accurate trends using 1-year data from the February submission. SEER still does not release a data file of the 1-year data because of the larger delay factor for these data. Instead, SEER releases preliminary delay-adjusted rates and trends based on the early submission on a website with many caveats. SEER is working to release these preliminary statistics earlier.

Dr. Feuer presented the 1-year delay modeling approach used for SEER18. He noted that delay factors need to be aggregated across registries for stability. Delay factors are generated for all cancer sites in the February submission, which are validated against the November submission. NAACCR has three years of validated 1-year data to date (diagnosis years 2013-2015). Dr. Feuer now is comparing NAACCR 1-year data to data from the December submissions for the respective years by cancer site. To be included in this analysis, registries had to be NAACCR certified for at least three of the four years. Dr. Feuer also set a completeness threshold for inclusion in the analysis. Participants recommended redoing these analyses using the four major cancer sites only. Once Dr. Feuer validates the 1-year data against the December submission, he might try the analysis using Joinpoint. Participants asked whether the 1-year data identified the change in prostate cancer incidence. If not, a more useful approach might be to use actual counts from the previous two years with Joinpoint to project incidence.

Priority Actions

Ann Marie Hill

Ms. Hill led a discussion with the goal of prioritizing actions toward developing a new NAACCR completeness measure. She asked participants to categorize actions as either important or less important and as either short-term or long-term. Participants agreed that identifying an item as long-term means that the action can be completed after the NAACCR meeting. They also agreed that all analyses should be run on data from all NAACCR registries (certified and uncertified) for diagnosis years 2013 and 2015.

Important/Short Term

NAACCR Method

- Update the workbook with the new NAACCR incidence data and mortality data.
    - Include non-certified registry data.
    - Include historical delay adjustment factors.
    - Possibly include cases received since the last submission date.
    - Add SEER mortality data.
- Run the NAACCR Method using the revised workbook and assess results.
- Determine permutations of the NAACCR model.

    - Consider Dr. Das' approach that adjusts for case fatality. SEER mortality should be examined as a denominator.
    - Include race/ethnicity and gender in the model.
    - Calculate registry-specific case fatality ratios.
    - Test using 5-year U.S. incidence rates, SEER11 incidence rates, and SEER18 incidence rates with the associated mortality rates.
    - Determine geographic areas to test.
- Compare age-adjusted to age-specific survival modeling.
    - Consult with Dr. Zhu on this effort.

Modeling Methods

- Obtain the ACS model and data for test years.
    - Test the model with specific cancers sites using the NAACCR workbook structure.

Other Methods

- Run the old linear SEER Data Completeness Estimation method in SEER*Edits using 10 years of data in the NAACCR workbook.
- Create scatter plots to compare Naïve, SEER, and current ACS methods.

Important/Long Term

NAACCR Method

- Run 1-year delay adjustment rates and trends. Test the effect of delay on the IMRR using the last year of data.
    - Decide what diagnosis years to include in the IMRR.
- Estimate completeness for the southwest region only, using regional incidence and mortality as the reference.

Modeling Methods

- Run the ACS model for each cancer site in the NAACCR workbook.
    - Decide whether to run state- or county-level data for each registry.
    - Use the Black/White/Other race categories initially. Then add other races/ethnicities (Hispanics are high priority).
    - Run with and without random effects.
    - Add covariates such as data source, percent foreign born, stage distribution, case fatality ratio, and survival. The last three variables may be run by race/ethnicity.
    - Determine how to add a delay factor.
    - Validate model.

Other Methods

- Redo comparisons of NAACCR 1-year data for diagnosis years 2015-2018 to data from the December submissions for the respective years by the four major cancer sites.
- Run the current and new SEER Data Completeness Estimation methods in SEER*Edits using the NAACCR workbook data (10 years of data). Create scatter plots.
    - Assemble data for each method.
    - Determine the criteria.
    - Obtain delay factors.

- Apply the new SEER Internal Joinpoint Method to every NAACCR registry. Use the 11-year observed count to calculate completeness.
- Examine the cause of delayed cases qualitatively (Where do they come from? What cancer sites? How could they be obtained sooner?).
    - Identify the data set with the tumor record numbers (NPCR or NAACCR?).

Less important/short term

Other Methods

- Run the Naïve Method. Use all NAACCR registry data.
    - Test the relationship between reporting source and completeness.
    - Use underlying CoD.

Less important/long term

- Develop a Canadian model in collaboration with Canadian registrars.
- Create a toolbox for the new model.

Next Steps

Participants agreed to monitor progress toward completing each of the prioritized tasks. They agreed to conduct two calls before the NAACCR meeting to provide updates on progress and troubleshoot. The first call was scheduled for May 2, 2019 (1:30–3:00 p.m. EDT) and the second for May 31 (1:00–2:30 p.m. EDT). NAACCR also will notify stakeholders regarding model changes in advance of implementing the final model, particularly if some registries might experience a decline in completeness as a result.

Registry Certification Standards

The primary goal of this meeting was to examine methods for measuring completeness, but standards and the interpretation of completeness estimates were frequent discussion topics. Dr. Kohler emphasized the need to differentiate between the estimate of completeness and the interpretation of that estimate. For example, if 100 percent completeness is expected 11 years after the diagnosis year, 95 percent might be an appropriate standard at two years. In addition to completeness, registry data quality can be evaluated based on reporting delays or the proportion of DCOs. SEER notes these criteria in its data quality profile. Participants agreed that another important task, after developing the completeness measure, is to identify methods for diagnosing the reasons for delay. They agreed that guidelines and best practices for improving completeness would be useful.

The purpose of standard setting and measuring data quality and completeness is to help registries improve. Almost all registries now have annual data submissions that are sufficiently complete to be used in research, so certification standards might no longer be necessary. Completeness rates could serve simply as information to help registries improve. Data quality metrics also are important for motivating registries to improve and to assure data users that the data meet their needs. If standards are eliminated, it would

be important to monitor trends in completeness and other data quality measures at each registry. Metrics also can help identify important changes in the registry area (demographic, environmental, etc.).

Many states with low completeness rates have large Hispanic populations, and the number of reported cancer cases for Hispanics have been lower than expected in these states. Hispanic populations, however, historically have had low cancer rates. Cancer mortality in this population, however, tends to be higher than among non-Hispanic Whites. These trends might be explained by the younger average age of most Hispanic populations, low screening rates, and lower access and utilization of the healthcare system. Both Native American and Hispanic populations have lower screening rates. Lower screening and healthcare utilization rates are particularly low among low-SES and foreign-born populations. Another possible problem is racial misclassification on DCs, particularly among Native Americans.

In California, large proportions of cases now are covered by ePath, and a new law will increase reporting from smaller laboratories, so completeness might improve in California registries. The law took effect in January 2019, but the implementation of the new law will take time. Small laboratories already are contacting the Los Angeles registry. Many of these laboratories only send PDFs currently, and registries do not have the resources to manually abstract from the PDFs. Participants added that cases missed as a result of non-reporting likely are captured through DCs.

Data runs will help reveal possible causes of lower data completeness rates at some registries. Participants discussed other approaches for evaluating the reasons for low data completeness at registries that have taken all possible steps to ensure completeness. Participants suggested hiring an outside evaluator to examine registry practices and other factors, such as different cancer patterns at locations struggling with low completeness rates.

Participants recommended examining cases initiated by DCs as an indicator of what registries might be missing. For example, many registries initiate follow back if cancer is mentioned as any CoD on the DC. This practice identifies several thousand missing cases. Participants also suggested examining in- and out-migration in states and its effect on mortality rates and completeness.

The number of duplicate cases across state registries might be significant. States with large numbers of cases that were diagnosed in other states but who died in their state could negatively affect completeness, particularly in states with smaller populations. Registries currently send cases to the National Death Index for linkage, and this linkage allows them to identify cases who died in another state. Registries generally do not contact the other state registry, however, to inform them that they have a DCO in their file that links back to a case in a different state registry. The Virtual Pooled Registry (VPR) will allow the identification of duplicate cancer cases across registries. DCO cases should decline once registries start using the VPR. Prior to submitting their data,

registries could use the VPR to determine what proportion of their cancer deaths were incident cases in other states. The VPR might be ready before the December 2019 submission.