Data Use Agreement for Delay Databases with Individual Registry Identifiers
November 2018 Data Submission

This data use agreement is for access to databases that use delay adjustment factors and contain individual registry identifiers. When using a database that contains individual registry identifiers, caution is necessary because some estimates for small populations may be unstable or biased. In requesting access to these databases, you agree to read and adhere to the cautions stated in this data use agreement.

The idea behind modeling reporting delay is to adjust the current case count to account for anticipated future corrections (both additions and deletions) to the data. These adjusted counts and the associated delay model are valuable in more precisely determining current cancer trends. The submissions for the most recent diagnosis year are, in general, about four percent below the number of cancers that will be submitted for that year eventually (i.e. a "delay factor" of 1.04) , although this varies by cancer site and other factors. For information about the background and methodology, see Cancer Incidence Rates Adjusted for Reporting Delay at https://surveillance.cancer.gov/delay/.

Cautions

Individual delay factors are estimated to strike a balance between two goals: (1) capture the unique pattern of delay for each registry; (2) provide stable estimates that are not too noisy. There is a tradeoff between stability and local relevance. This is achieved by estimating delay factors for sparse cancer site/registry combinations (defined as fewer than 50 cases per year) from groups of registries that have similar delay factors. Additionally, for each registry/cancer site combination, age, race, and ethnicity strata need to have an average of 25 cases per year or they are pooled across these groups before modeling. Thus, for some smaller cancer site/registry combinations, delay factors may have been estimated from groups of "similar" registries, rather than from single registries. Moreover, age/race/ethnicity-specific estimates for some of the smaller groups in any particular registry may have been estimated from pooled age, race, and/or ethnic group-specific strata. Delay factors for American Indians/Alaska Natives are estimated by the Contract Health Service Delivery Areas (CHSDA) for all cancer sites combined, lung, female breast, and colorectal cancers, and at the US level only (from counties designated as CHSDA), for all other cancer sites (see https://seer.cancer.gov/seerstat/variables/countyattribs/#chsda)

The statistical modeling estimates individual delay factors by cancer site, registry, diagnosis year, age, race, and delay time. However, composite delay-adjusted rates for any desired grouping are computed in SEER*Stat by weighting each group appropriately. Some individual delay factors stored in SEER*Stat may be unstable or biased - perhaps because of some of the pooling that is performed before analysis to obtain sufficient sample sizes—may mostly reflect the dominant groups in the pooled estimate. Composite delay-adjusted rates estimated across large groups of registries (e.g. SEER9, SEER13, SEER18, SEER21) are generally very stable. Delay factors were developed with the

idea that they could be used to estimate delay-adjusted trends for individual or small groups of registries. However, caution should be applied in computing registry-specific delay-adjusted rates and trends, especially for the small population racial/ethnic groups the rarer cancer sites. While the delay modeling team felt that reasonable tradeoffs were made between stability and the local/subpopulation relevance of estimates, these estimates are by their nature tradeoffs that are not perfect. Please report any estimates that appear spurious to delay_adjustment@imsweb.com.

I have read and understand the limitations of the Delay Databases with Individual Registry Identifiers described above. I will include a description of relevant limitations in any analyses published using the SEER data. I acknowledge that the SEER Program has advised me that there are substantive concerns about using these data to address certain research questions as described above. I understand that any findings from such analyses may be inaccurate or misleading.

Print Name _____

SEER*Stat Username/SEER ID _____

Signature _____ Date _____

Please print, sign, and date the agreement. Send the form to The SEER Program:

- **By fax to 301-680-9571**
- **Or, e-mail a scanned form to seercustomdata@imsweb.com**