

C3I Data Considerations

Missing Data

Note that some data are missing by design, as some assessments were conducted only at baseline or particular waves (as described in the next section). Other data are missing because the center did not provide relevant data in a given survey wave. Some data are missing due to removal of data that were logically inconsistent with previous data (e.g., if the count of patients reached by the TTP was greater than the total number of adult patients who use tobacco, we attempted to obtain corrected data from centers; if this was not possible, we set the inconsistent variables to missing).

Flagged Errors in Patient Data

The Coordinating Center provided templates for grantee TTP patient and effectiveness data reporting, which encouraged data consistency checks. However, during final harmonization in 2024, the Coordinating Center identified discrepancies in the data reported in addition to the errors corrected or treated as missing during the initial cleaning process (see the Missing Data section above). To aid interpretation of the data, discrepancies in key variables related to TTP reach and effectiveness have been categorized and flagged with discrete fields, which are described in Table 1 below. These flags do not indicate the magnitude or direction of the discrepancies, which vary markedly across centers and waves of data collection but rather serve as signals that careful consideration is needed when using these data. We encourage particular care when comparing values collected across variables at different population levels (e.g., when comparing the demographics of all patients who smoke with the demographics of all patients who receive TTP services). In the table below, the flag descriptions tell you how variables were checked against their “parent group” to identify discrepancies. For example: the “smok_female,” “smok_male,” and “smok_othsex” variables are demographic sub-groups of the “smoke” parent group, which refers to all current adult patients who smoke. Although the aggregate of the number of patients in the female, male, and other sex category should not be greater than the total number of patients who smoke, it is sometimes higher.

C3I Data Considerations

Flag	Description
flag_anyTTP_gt_smoke	The number of patients who received TTP (anyTTP) is greater than (gt) the number of patients who smoke (smoke).
flag_screened_gt_all	The number of patients screened for tobacco use (screened) is greater than the total patient count (all)
flag_smoke_gt_all	The number of patients reported as smoking (smoke) is greater than the total patient count (all)
flag_smoke_gt_screened	The number of patients reported as smoking (smoke) is greater than the number of patients reported as screened for tobacco use (screened)
flag_internal_demo_mismatch_ anyTTP enroll_all effect_all smoke	The non-zero sums of the demographic sub-groups within the parent group do not match each other. For example, the sum of male and female patients does not equal the sum of all patients of any ethnicity (Hispanic or Latino, non-Hispanic or Latino, and ethnicity-unknown patients). Separate flags are provided for 4 sub-populations: anyTTP (all patients who received any TTP service) enroll_all (all patients who completed at least 1 session of treatment and provided smoking status data at TTP enrollment), effect_all (all patients with 6-month follow-up smoking status data), and smoke (all patients who smoked).
flag_parent_mismatch_ anyTTP enroll_all effect_all smoke	The largest sum of the non-zero demographic sub-groups does not match the non-zero value found in the parent group. For example, this was flagged if the sum of patients in different age groups who smoke was larger than the total number of patients who smoke in a given wave. Separate flags are provided for discrepant sums with anyTTP, enroll_all, effect_all, and smoke.
flag_subgroup_missing_sum_ ttp/enroll /effect/smoke_	Demographic sub-group aggregate is 0 despite other demographic sub-groups in a parent group aggregating to > 0. For example, the sum of male and female

C3I Data Considerations

Flag	Description
age_group ethnicity insurance race	patients (used as the reference group in these comparisons) is greater than 0, but the total number of patients across age groups is 0. Separate flags are provided (where applicable) for discrepant sums with anyTTP, enroll_all, effect_all, and smoke and for aggregates across age group, race, ethnicity, and insurance type.
flag_all_subgroups_missing_ anyTTP enroll_all effect_all smoke	All demographic groups sum to 0 while the group aggregate has a value > 0. For example, anyTTP is non-zero but the counts for TTP-treated patients in each demographic category are all 0. Separate flags are provided for anyTTP, enroll_all, effect_all, and smoke.
flag_parent_missing_ anyTTP enroll_all effect_all smoke	A parent variable has a value of 0 but at least one demographic subgroup is greater than 0. For example, the count of female patients who received any TTP (anyTTP_female) is >0, but anyTTP=0. Separate flags are provided for anyTTP, enroll_all, effect_all, and smoke.
flag_demo_item_gt_parent_ ttp/enroll/effect/smoke_ female/male black/white/raceunk hisp/nonhisp/ethunk age25/age45/age65 priv/caid/care/insunk	The count for a specific demographic item/group has a value greater than the associated non-zero parent aggregate. For example, a true value on flag_demo_item_gt_parent_enroll_age45 indicates that the count of patients age 45-64 years fully enrolled in a TTP and eligible for effectiveness tracking is greater than the total enrolled in the TTP (enroll_all). Only flags with at least one true value (i.e., a case in which a discrepancy was found) were retained in the dataset; this is why some demographics groups do not have a flag in this category.

C3I Data Considerations

Flag	Description
flag_demo_mismatch_parent_ ttp/enroll/effect/smok_ gender age_group ethnicity insurance race	The sum of the demographic sub-group has a value that does not match the associated parent, and both values are greater than 0. For example, the non-zero sum of patients who smoke across ethnicity subgroups (Hispanic or Latino, Non-Hispanic or Latino, and ethnicity unknown) does not equal the non-zero count of all patients who smoke (smoke).

Table 1. Flagged Inconsistencies in Patient Data

Other Data Considerations

Additional examinations of the cumulative data identified outliers and unexpected values in some fields. For example, some centers showed marked variation in the number and characteristics (e.g., racial backgrounds) of patients who smoke from wave to wave. Because we do not have access to the source EHR and TTP program data that C3I programs used to prepare the aggregate data they submitted, we did not correct these data. We also did not flag or set an arbitrary cutpoint to identify outliers, as users may want to take different approaches to defining and dealing with such outliers.