North Carolina Department of Health and Human Resources, Division of Mental Health - Drug Control Unit – Data Guidelines

SECTION I	INTRODUCTION

CSRS data: no PII can ever be released except to authorized parties (i.e. NCMB, NCBoN, SBI). Researchers can request de-identified data if their need is substantiated. There is a full definition of de-identified health information and a table of restricted/available CSRS variables below.

Note on age: researchers cannot be provided with Date of Birth (DOB). Age group will be provided instead. For de-identified datasets, the variable is calculated based on the patients' ages on January 1 of that year. This is necessary because DOB could be calculated with age and the Filled_at variable. Any patients 80 years of age or older will be grouped as "80+."

The CSRS data can be requested in the form of summary statistics or as a de-identified dataset.

SECTION II	SUMMARY STATISTICS

1. SUMMARY STATISTICS: the most common type of request. Includes the count/sum of requested metrics (pills/prescriptions/MMEs) broken down by requested dimensions (year/county/drug/gender). This sort of table would be delivered in the format shown below.

Dimensions				Metrics			
Year	County	Class	Gender	Number of prescriptions	Unique patients	MME average	NC Population
2012	Alamance	Opioid	M	10,000	2,000	52.5	73,236
2012	Alamance	Opioid	F	12,000	2,500	50.2	80,684
2012	Alamance	Benzo	M	8,000	1,500	N/A	73,236
2012	Alamance	Benzo	F	10,000	2,000	N/A	80,684
2013	Alamance	Opioid	M	11,000	2,000	53.3	73,726
2013	Alamance	Opioid	F	13,000	2,500	52.8	80,652

Only metrics with counts between 1-4 need to be suppressed (as "<5") and the counts must be suppressed only if there are fewer than 500 residents within the selected dimensions. For example, there were 73,236 residents in Alamance County in 2012 who were male; we don't need to suppress any of those metrics, as it is unlikely that someone could identify 1-4 people from such a large group. Conversely, there were only 375 residents in Hyde County in 2012 in the 70-79 age group; we would have to suppress any counts between 1-4, as it is possible that someone could determine who those 1-4 patients are. No row or column totals are provided with the data release, as doing so would require suppression of additional cells to prevent discovery of the suppressed cells by subtraction or addition.

A. POPULATION:	Population suppression will be based on the CDC's annual population dataset, which is stratified by age, gender, county, and year.
B. GROUPING:	Grouping to prevent cell suppression, groups with populations below 500 can be combined. For instance, small counties would be "Other," gender would be "Both," and age groups would be "All." Which groups to combine is flexible depending on the intended purpose of the data.

SECTION III   DE-IDENTIFIED DATASET

The De-Identified Dataset includes prescription/transaction level data without any PHI. This type of data will be requested by researchers with advanced statistical analysis software who want to perform their own analyses.

A. ELEMENTS

The following elements are available in the de-identified dataset:

- AGE_GROUP
- AUTHORIZED_REFILL_COUNT
- COMPOUND_DRUG
- DATE_DISPENSED
- DAYS_SUPPLY
- DISPENSATION_ID
- DRUG_NAME
- PARTIAL_FILL
- PRODUCT_ID
- QUANTITY
- REFILL_NUMBER
- UNITS
- VETERINARIAN_PRESCRIPTION
- PATIENT_GENDER
- PATIENT_STATE
- DISPENSARY_STATE
- PATIENT_ENTITY_ID
- PATIENT_COUNTY
- PATIENT_COUNTY_FIPS
- PRES_PROV_ENTITY_ID
- PRESCRIBER_COUNTY
- PRESCRIBER_COUNTY_FIPS
- PRESCRIBER_STATE
- DISPENSARY_ENTITY_ID
- DISPENSARY_COUNTY
- DISPENSARY_COUNTY_FIPS

B. POPULATIONS BELOW 500

We want to suppress the same information in the de-identified dataset as we do in summary statistics. To achieve this, we need to identify any groups with populations below 500. For such groups, we must suppress the following variables for all prescriptions:

- AGE_GROUP
- AUTHORIZED_REFILL_COUNT
- COMPOUND_DRUG
- DAYS_SUPPLY
- DRUG_NAME
- PARTIAL_FILL
- PRODUCT_ID
- QUANTITY
- REFILL_NUMBER
- UNITS
- PATIENT_GENDER
- SPECIES
- PATIENT_ENTITY_ID
- PRES_PROV_ENTITY_ID
- DISPENSARY_ENTITY_ID

Age groups will be used in the de-identified dataset in order to provide more information to researchers. In 2018, if we provided specific age as well as sex, 69% (11,845/17,200) of groups containing 24% of the population would have their data suppressed. By combining ages into 10-year groups (and an 80+ group), we are only forced to suppress 5% (98/1,800) of groups containing 0.3% of the population.
Other groupings are not available as differently grouped datasets would allow cells to be calculated.