

COVID-19 Case Surveillance Public Use Data Utility Summary

Users should consider the level of completeness, including suppression levels when planning their analyses and use of public datasets. Privacy protections will suppress field values to reduce reidentification risks. Completeness varies by jurisdiction (i.e., state, local, and territorial) and time period. Variables are consistently coded to the value “Unknown” when jurisdictions specify in the case data submitted to CDC that the value is unknown, the value “Missing” when jurisdictions do not provide a value, and the value “NA” when the value is suppressed as part of privacy protections.

Dataset version: 2021-10-12

Total records in dataset: 51,447,566

Quick Summary

	all_fields	quasi_fields
total_fields	12	3
total_records	51,447,566	51,447,566
total_cells	617,370,792	154,342,698
missing_fields	132,677,696	1,834,539
missing_pct	21%	1%
complete_fields	484,693,096	152,508,159
complete_pct	79%	99%
unknown_fields	43,507,256	16,609,064
unknown_pct	7%	11%
suppressed_fields	122	122
suppressed_pct	0%	0%
available_fields	441,185,718	135,898,973
available_pct	71%	88%

Field-level Utility Summary

	suppressed	suppressed_percent	missing	missing_percent
race_ethnicity_combined	11	0.0%	1,223,370	2.4%
sex	18	0.0%	103,526	0.2%
age_group	93	0.0%	507,643	1.0%
records_with_any_field	93	0.0%	1,718,348	3.3%