

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: 69,664,983

Quick Summary

	all_fields	quasi_fields
total_fields	12	3
total_records	69,664,983	69,664,983
total_cells	835,979,796	208,994,949
missing_fields	186,848,443	2,700,975
missing_pct	22%	1%
complete_fields	649,131,353	206,293,974
complete_pct	78%	99%
unknown_fields	58,868,524	23,216,443
unknown_pct	7%	11%
suppressed_fields	97	97
suppressed_pct	0%	0%
available_fields	590,262,732	183,077,434
available_pct	71%	88%

Field-level Utility Summary

	suppressed	suppressed_percent	missing	missing_percent
race_ethnicity_combined	9	0.0%	1,729,448	2.5%
sex	9	0.0%	204,723	0.3%
age_group	79	0.0%	766,804	1.1%
records_with_any_field	79	0.0%	2,519,565	3.6%