

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: 61,267,087

Quick Summary

	all_fields	quasi_fields
total_fields	12	3
total_records	61,267,087	61,267,087
total_cells	735,205,044	183,801,261
missing_fields	161,648,957	2,355,736
missing_pct	22%	1%
complete_fields	573,556,087	181,445,525
complete_pct	78%	99%
unknown_fields	52,751,214	20,405,847
unknown_pct	7%	11%
suppressed_fields	96	96
suppressed_pct	0%	0%
available_fields	520,804,777	161,039,582
available_pct	71%	88%

Field-level Utility Summary

	suppressed	suppressed_percent	missing	missing_percent
race_ethnicity_combined	9	0.0%	1,486,239	2.4%
sex	9	0.0%	145,443	0.2%
age_group	78	0.0%	724,054	1.2%
records_with_any_field	78	0.0%	2,196,010	3.6%