

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: 2021-11-09

Total records in dataset: 37,532,072

Quick Summary

	all_fields	quasi_fields
total_fields	12	3
total_records	37,532,072	37,532,072
total_cells	450,384,864	112,596,216
missing_fields	96,399,423	1,323,719
missing_pct	21%	1%
complete_fields	353,985,441	111,272,497
complete_pct	79%	99%
unknown_fields	29,121,517	12,249,731
unknown_pct	6%	11%
suppressed_fields	108	108
suppressed_pct	0%	0%
available_fields	324,863,816	99,022,658
available_pct	72%	88%

Field-level Utility Summary

	suppressed	suppressed_percent	missing	missing_percent
race_ethnicity_combined	13	0.0%	875,896	2.3%
sex	19	0.0%	76,966	0.2%
age_group	76	0.0%	370,857	1.0%
records_with_any_field	76	0.0%	1,232,005	3.3%