

Public 12 Reporting

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: 10/3/2022

Quick Summary

	summary String	all_fields_counts Double	all_fields_pct Double	quasi_fields_counts Double	quasi_fields_pct Double
1	total_rows	87,306,976	NaN%	87,306,976	NaN%
2	total_columns	12	NaN%	3	NaN%
3	total_cells	1,047,683,712	100.0%	261,920,928	100.0%
4	suppressed_fields	90	0.0%	90	0.0%
5	missing_fields	239,032,194	22.8%	3,614,030	1.4%
6	unknown_fields	72,250,157	6.9%	28,777,999	11.0%
7	non_blank_fields	736,401,271	70.3%	229,528,809	87.6%

Field Level Utility Summary

	variable String	suppressed Long	suppressed_pct String	missing Long	missing_pct String	unknown Long	unknown_pct String
1	sex	11	0.0%	472,210	0.5%	1,275,313	1.5%
2	age_group	75	0.0%	780,038	0.9%	0	0.0%
3	race_ethnicity_combined	4	0.0%	2,361,782	2.7%	27,502,686	31.5%
4	records_with_any_quasi_identifier	75	0.0%	3,328,699	3.8%	27,999,909	32.1%

