

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: 8/1/2022

Quick Summary

	summary String	all_fields_counts Double	all_fields_pct Double	quasi_fields_... Double	quasi_fields_... Double	
1	total_rows	82,422,585	NaN%	82,422,585	NaN%	
2	total_columns	12	NaN%	3	NaN%	
3	total_cells	989,071,020	100.0%	247,267,755	100.0%	
4	suppressed_fields ...	90	0.0%	90	0.0%	
5	missing_fields	225,283,281	22.8%	3,430,917	1.4%	
6	unknown_fields	67,303,579	6.8%	27,301,475	11.0%	
7	non_blank_fields	696,484,070	70.4%	216,535,273	87.6%	

Field Level Utility Summary

	variable String	suppressed Long	suppressed_percent String	missing Long	missing_percent String	unknown Long	unknown_percent String
1	sex	10	0.0%	431,725	0.5%	1,172,988	1.4%
2	age_group	76	0.0%	773,463	0.9%	0	0.0%
3	race_ethnicity_combined	4	0.0%	2,225,729	2.7%	26,128,487	31.7%
4	records_with_any_quasi_iden... ...	76	0.0%	3,143,755	3.8%	26,579,226	32.2%