

Public 19 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: 7/6/2023

Quick Summary

	summary String	all_fields_counts Double	all_fields_pct Double	quasi_fields_co... Double	quasi_fields_pct Double	
1	total_rows	98,952,628	NaN%	98,952,628	NaN%	
2	total_columns	19	NaN%	8	NaN%	
3	total_cells	1,880,099,932	100.0%	791,621,024	100.0%	
4	suppressed_fields ...	52,637,477	2.8%	45,897,185	5.8%	
5	missing_fields	431,848,367	23.0%	67,498,357	8.5%	
6	unknown_fields	89,830,620	4.8%	45,078,622	5.7%	
7	non_blank_fields	1,305,783,468	69.5%	633,146,860	80.0%	

Field Level Utility Summary

	variable String	suppressed Long	suppressed_pct String	missing Long	missing_pct String	unknown Long	unknown_pct String	
1	case_month	9	0.0%	0	0.0%	0	0.0%	
2	res_state	1,522	0.0%	0	0.0%	0	0.0%	
3	res_county	6,738,770	6.8%	0	0.0%	0	0.0%	
4	sex	2,674,087	2.7%	420,143	0.4%	791,766	0.8%	
5	age_group	911,669	0.9%	930,376	0.9%	0	0.0%	
6	ethnicity	17,137,548	17.3%	6,210,621	6.3%	18,215,283	18.4%	
7	race	15,334,083	15.5%	7,711,978	7.8%	12,960,976	13.1%	
8	death_yn	3,099,497	3.1%	52,225,239	52.8%	13,110,597	13.2%	
9	records_with_any_quasi_identifier ...	24,843,115	25.1%	55,125,920	55.7%	32,357,963	32.7%	

