

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: 1/4/2024

Quick Summary

	summary String	all_fields_counts Double	all_fields_pct Double	quasi_fields_co... Double	quasi_fields_pct Double		
1	total_rows	103,535,146	NaN%	103,535,146	NaN%		
2	total_columns	19	NaN%	8	NaN%		
3	total_cells	1,967,167,774	100.0%	828,281,168	100.0%		
4	suppressed_fields ...	57,184,886	2.9%	49,864,847	6.0%		
5	missing_fields	453,709,983	23.1%	70,725,255	8.5%		
6	unknown_fields	92,769,119	4.7%	46,345,895	5.6%		
7	non_blank_fields	1,363,503,786	69.3%	661,345,171	79.8%		

Field Level Utility Summary

	variable String	suppressed Long	suppressed_pct String	missing Long	missing_pct String	unknown Long	unknown_pct String		
1	case_month	12	0.0%	0	0.0%	0	0.0%		
2	res_state	1,799	0.0%	0	0.0%	0	0.0%		
3	res_county	7,318,240	7.1%	0	0.0%	0	0.0%		
4	sex	3,105,306	3.0%	421,833	0.4%	793,678	0.8%		
5	age_group	1,063,611	1.0%	1,025,806	1.0%	0	0.0%		
6	ethnicity	18,534,191	17.9%	6,304,614	6.1%	18,672,316	18.0%		
7	race	16,493,228	15.9%	7,908,655	7.6%	13,217,435	12.8%		
8	death_yn	3,348,460	3.2%	55,064,347	53.2%	13,662,466	13.2%		
9	records_with_any_quasi_identifier ...	26,847,365	25.9%	58,015,433	56.0%	33,389,954	32.2%		

