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: 2/1/2024

Quick Summary

	summary String	all_fields_counts Double	all_fields_pct Double	quasi_fields_co… Double	quasi_fields_pct Double	
1	total_rows	104,544,006	NaN%	104,544,006	NaN%	
2	total_columns	19	NaN%	8	NaN%	
3	total_cells	1,986,336,114	100.0%	836,352,048	100.0%	
4	suppressed_fields •••	58,165,102	2.9%	50,746,848	6.1%	
5	missing_fields	458,608,926	23.1%	71,398,762	8.5%	
6	unknown_fields	93,495,108	4.7%	46,677,264	5.6%	
7	non_blank_fields	1,376,066,978	69.3%	667,529,174	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	1,822	0.0%	0	0.0%	0	0.0%	
3	res_county	7,416,432	7.1%	0	0.0%	0	0.0%	
4	sex	3,304,992	3.2%	422,263	0.4%	794,268	0.8%	
5	age_group	1,083,508	1.0%	1,046,219	1.0%	0	0.0%	
6	ethnicity	18,808,019	18.0%	6,345,292	6.1%	18,764,985	17.9%	
7	race	16,753,429	16.0%	7,963,804	7.6%	13,256,718	12.7%	
8	death_yn	3,378,634	3.2%	55,621,184	53.2%	13,861,293	13.3%	
9	records_with_any_quasi_identifier •••	27,220,583	26.0%	58,592,760	56.0%	33,687,215	32.2%	