

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: 1/12/2023

Quick Summary

	summary String	all_fields_counts Double	all_fields_pct Double	quasi_fields_co... Double	quasi_fields_pct Double	
1	total_rows	94,138,905	NaN%	94,138,905	NaN%	
2	total_columns	19	NaN%	8	NaN%	
3	total_cells	1,788,639,195	100.0%	753,111,240	100.0%	
4	suppressed_fields ...	48,766,766	2.7%	42,240,913	5.6%	
5	missing_fields	413,085,187	23.1%	64,623,974	8.6%	
6	unknown_fields	86,254,554	4.8%	44,295,590	5.9%	
7	non_blank_fields	1,240,532,688	69.4%	601,950,763	79.9%	

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	1,233	0.0%	0	0.0%	0	0.0%
3	res_county	6,524,620	6.9%	0	0.0%	0	0.0%
4	sex	2,304,697	2.4%	417,767	0.4%	1,172,230	1.2%
5	age_group	773,888	0.8%	762,807	0.8%	0	0.0%
6	ethnicity	15,704,085	16.7%	5,946,867	6.3%	18,138,527	19.3%
7	race	14,018,478	14.9%	7,773,546	8.3%	12,734,299	13.5%
8	death_yn	2,913,903	3.1%	49,722,987	52.8%	12,250,534	13.0%
9	records_with_any_quasi_identifier	23,188,233	24.6%	52,707,171	56.0%	31,381,820	33.3%

