

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: 4/6/2023

Quick Summary

	summary String	all_fields_counts Double	all_fields_pct Double	quasi_fields_co... Double	quasi_fields_pct Double	
1	total_rows	97,799,772	NaN%	97,799,772	NaN%	
2	total_columns	19	NaN%	8	NaN%	
3	total_cells	1,858,195,668	100.0%	782,398,176	100.0%	
4	suppressed_fields ...	51,438,391	2.8%	44,712,891	5.7%	
5	missing_fields	428,385,394	23.1%	67,042,282	8.6%	
6	unknown_fields	88,862,867	4.8%	45,047,422	5.8%	
7	non_blank_fields	1,289,509,016	69.4%	625,595,581	80.0%	

Field Level Utility Summary

	variable String	suppressed Long	suppressed_pct String	missing Long	missing_pct String	unknown Long	unknown_pct String
1	res_county	6,724,273	6.9%	0	0.0%	0	0.0%
2	case_month	25	0.0%	0	0.0%	0	0.0%
3	res_state	1,227	0.0%	0	0.0%	0	0.0%
4	sex	2,509,917	2.6%	420,897	0.4%	1,202,402	1.2%
5	age_group	845,228	0.9%	928,800	0.9%	0	0.0%
6	ethnicity	16,685,882	17.1%	6,212,001	6.4%	18,219,243	18.6%
7	race	14,902,535	15.2%	7,837,128	8.0%	12,892,323	13.2%
8	death_yn	3,043,804	3.1%	51,643,456	52.8%	12,733,454	13.0%
9	records_with_any_quasi_identifier	24,382,788	24.9%	54,634,509	55.9%	32,022,532	32.7%

