

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: 9/6/2022

Quick Summary

	summary String	all_fields_counts Double	all_fields_pct Double	quasi_fields_co... Double	quasi_fields_pct Double	
1	total_rows	85,937,106	NaN%	85,937,106	NaN%	
2	total_columns	19	NaN%	8	NaN%	
3	total_cells	1,632,805,014	100.0%	687,496,848	100.0%	
4	suppressed_fields ...	42,949,007	2.6%	37,610,942	5.5%	
5	missing_fields	378,139,047	23.2%	59,195,689	8.6%	
6	unknown_fields	78,397,376	4.8%	40,379,152	5.9%	
7	non_blank_fields	1,133,319,584	69.4%	550,311,065	80.0%	

Field Level Utility Summary

	variable String	suppressed Long	suppressed_percent String	missing Long	missing_percent String	unknown Long	unknown_percent String
1	case_month	3	0.0%	0	0.0%	0	0.0%
2	res_state	1,185	0.0%	0	0.0%	0	0.0%
3	res_county	5,336,880	6.2%	0	0.0%	0	0.0%
4	sex	2,017,741	2.3%	414,438	0.5%	1,045,688	1.2%
5	age_group	679,228	0.8%	751,878	0.9%	0	0.0%
6	ethnicity	14,167,350	16.5%	5,621,615	6.5%	16,912,241	19.7%
7	race	12,665,345	14.7%	7,303,243	8.5%	11,852,522	13.8%
8	death_yn	2,743,210	3.2%	45,104,515	52.5%	10,568,701	12.3%
9	records_with_any_quasi_identifier ...	20,434,614	23.8%	47,940,333	55.8%	28,591,141	33.3%

