

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: 11/3/2022

Quick Summary

	summary String	all_fields_counts Double	all_fields_pct Double	quasi_fields_co... Double	quasi_fields_pct Double	
1	total_rows	90,183,148	NaN%	90,183,148	NaN%	
2	total_columns	19	NaN%	8	NaN%	
3	total_cells	1,713,479,812	100.0%	721,465,184	100.0%	
4	suppressed_fields ***	46,366,180	2.7%	40,171,078	5.6%	
5	missing_fields	397,618,698	23.2%	62,431,904	8.7%	
6	unknown_fields	82,228,078	4.8%	42,296,097	5.9%	
7	non_blank_fields	1,187,266,856	69.3%	576,566,105	79.9%	

Field Level Utility Summary

	variable String	suppressed Long	suppressed_pct String	missing Long	missing_pct String	unknown Long	unknown_pct String	
1	res_county	6,193,814	6.9%	0	0.0%	0	0.0%	
2	case_month	9	0.0%	0	0.0%	0	0.0%	
3	res_state	1,288	0.0%	0	0.0%	0	0.0%	
4	sex	2,145,400	2.4%	416,404	0.5%	1,115,106	1.2%	
5	age_group	723,020	0.8%	753,470	0.8%	0	0.0%	
6	ethnicity	14,954,285	16.6%	5,818,326	6.5%	17,642,870	19.6%	
7	race	13,326,965	14.8%	7,614,949	8.4%	12,360,181	13.7%	
8	death_yn	2,826,297	3.1%	47,828,755	53.0%	11,177,940	12.4%	
9	records_with_any_quasi_identifier ***	22,094,945	24.5%	50,772,623	56.3%	29,920,155	33.2%	