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: 9/7/2023

Quick Summary

	summary String	all_fields_counts Double	all_fields_pct Double	quasi_fields_co... Double	quasi_fields_pct Double	
1	total_rows	100,325,980	NaN%	100,325,980	NaN%	
2	total_columns	19	NaN%	8	NaN%	
3	total_cells	1,906,193,620	100.0%	802,607,840	100.0%	
4	suppressed_fields •••	54,179,553	2.8%	47,272,360	5.9%	
5	missing_fields	438,061,969	23.0%	68,335,879	8.5%	
6	unknown_fields	90,717,442	4.8%	45,483,834	5.7%	
7	non_blank_fields	1,323,234,656	69.4%	641,515,767	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	3	0.0%	0	0.0%	0	0.0%	
2	res_state	1,600	0.0%	0	0.0%	0	0.0%	
3	res_county	6,905,593	6.9%	0	0.0%	0	0.0%	
4	sex	2,828,737	2.8%	420,631	0.4%	792,918	0.8%	
5	age_group	971,849	1.0%	957,147	1.0%	0	0.0%	
6	ethnicity	17,640,364	17.6%	6,224,226	6.2%	18,343,354	18.3%	
7	race	15,701,806	15.7%	7,723,702	7.7%	13,052,778	13.0%	
8	death_yn	3,222,408	3.2%	53,010,173	52.8%	13,294,784	13.3%	
9	records_with_any_quasi_identifier •••	25,543,127	25.5%	55,929,000	55.7%	32,662,128	32.6%	