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: 5/2/2024

Quick Summary

	summary String	all_fields_counts Double	all_fields_pct Double	quasi_fields_co... Double	quasi_fields_pct Double	
1	total_rows	105,869,141	NaN%	105,869,141	NaN%	
2	total_columns	19	NaN%	8	NaN%	
3	total_cells	2,011,513,679	100.0%	846,953,128	100.0%	
4	suppressed_fields •••	59,696,313	3.0%	52,138,114	6.2%	
5	missing_fields	464,579,635	23.1%	72,287,351	8.5%	
6	unknown_fields	94,305,800	4.7%	46,901,453	5.5%	
7	non_blank_fields	1,392,931,931	69.2%	675,626,210	79.8%	

Field Level Utility Summary

	variable String	suppressed Long	suppressed_pct String	missing Long	missing_pct String	unknown Long	unknown_pct String	
1	res_county	7,556,227	7.1%	0	0.0%	0	0.0%	
2	case_month	3	0.0%	0	0.0%	0	0.0%	
3	res_state	1,972	0.0%	0	0.0%	0	0.0%	
4	sex	3,552,932	3.4%	423,155	0.4%	793,914	0.7%	
5	age_group	1,137,051	1.1%	1,085,405	1.0%	0	0.0%	
6	ethnicity	19,275,208	18.2%	6,363,046	6.0%	18,828,708	17.8%	
7	race	17,177,726	16.2%	7,994,096	7.6%	13,264,624	12.5%	
8	death_yn	3,436,995	3.2%	56,421,649	53.3%	14,014,207	13.2%	
9	records_with_any_quasi_identifier •••	27,824,257	26.3%	59,380,812	56.1%	33,904,670	32.0%	