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.

Versions:

Dataset version: 7/1/2022

Quick Summary

	summary	all_fields_counts	all_fields_pct	quasi_fields_counts	quasi_fields_pct
1	total_rows	77,544,202	NaN%	77,544,202	NaN%
2	total_columns	19	NaN%	8	NaN%
3	total_cells	1,473,339,838	100.0%	620,353,616	100.0%
4	suppressed_fields	39,340,844	2.7%	34,463,413	5.6%
5	missing_fields	337,552,440	22.9%	52,178,462	8.4%
6	unknown_fields	70,601,079	4.8%	36,419,306	5.9%
7	non_blank_fields	1,025,845,475	69.6%	497,292,435	80.2%

Field Level Utility Summary

	variable	suppressed	suppressed_percent	missing	missing_percent	unknown	unknown_percent
1	case_month	13	0.0%	0	0.0%	0	0.0%
2	res_state	1,121	0.0%	0	0.0%	0	0.0%
3	res_county	4,876,310	6.3%	0	0.0%	0	0.0%
4	sex	1,877,582	2.4%	263,117	0.3%	948,098	1.2%
5	age_group	631,380	0.8%	730,191	0.9%	0	0.0%
6	ethnicity	12,927,437	16.7%	4,985,327	6.4%	15,249,457	19.7%
7	race	11,561,752	14.9%	6,315,186	8.1%	10,687,519	13.8%
8	death_yn	2,587,818	3.3%	39,884,641	51.4%	9,534,232	12.3%
9	records_with_any_quasi_identifier	18,672,362	24.1%	42,420,788	54.7%	25,773,856	33.2%