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: 6/6/2022

Quick Summary

	summary	all_fields_counts	all_fields_pct	quasi_fields_counts	quasi_fields_pct
1	total_rows	74,849,225	NaN%	74,849,225	NaN%
2	total_columns	19	NaN%	8	NaN%
3	total_cells	1,422,135,275	100.0%	598,793,800	100.0%
4	suppressed_fields	37,983,300	2.7%	33,296,709	5.6%
5	missing_fields	324,695,600	22.8%	49,767,622	8.3%
6	unknown_fields	69,023,724	4.9%	36,060,378	6.0%
7	non_blank_fields	990,432,651	69.6%	479,669,091	80.1%

Field Level Utility Summary

	variable	suppressed	suppressed_percent	missing	missing_percent	unknown	unknown_percent
1	case_month	10	0.0%	0	0.0%	0	0.0%
2	res_state	1,062	0.0%	0	0.0%	0	0.0%
3	res_county	4,685,529	6.3%	0	0.0%	0	0.0%
4	sex	1,823,234	2.4%	146,487	0.2%	1,267,620	1.7%
5	age_group	609,962	0.8%	725,814	1.0%	0	0.0%
6	ethnicity	12,497,148	16.7%	4,587,692	6.1%	14,868,269	19.9%
7	race	11,155,477	14.9%	5,924,521	7.9%	10,458,411	14.0%
8	death_yn	2,524,287	3.4%	38,383,108	51.3%	9,466,078	12.6%
9	records_with_any_quasi_identifier	18,028,057	24.1%	40,570,046	54.2%	25,250,605	33.7%