

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: 5/3/2022

Quick Summary

	summary	all_fields_counts	all_fields_pct	quasi_fields_counts	quasi_fields_pct
1	total_rows	71,387,132	NaN%	71,387,132	NaN%
2	total_columns	19	NaN%	8	NaN%
3	total_cells	1,356,355,508	100.0%	571,097,056	100.0%
4	suppressed_fields	36,471,861	2.7%	31,899,418	5.6%
5	missing_fields	308,445,375	22.7%	47,426,657	8.3%
6	unknown_fields	66,772,559	4.9%	34,604,739	6.1%
7	non_blank_fields	944,665,713	69.6%	457,166,242	80.1%

Field Level Utility Summary

	variable	suppressed	suppressed_percent	missing	missing_percent
1	case_month	11	0.0%	0	0.0%
2	res_state	1032	0.0%	0	0.0%
3	res_county	4571411	6.4%	0	0.0%
4	sex	1754500	2.5%	145856	0.2%
5	age_group	583790	0.8%	729439	1.0%
6	ethnicity	11916411	16.7%	4246597	5.9%
7	race	10606835	14.9%	5647715	7.9%
8	death_yn	2465428	3.5%	36657050	51.3%
9	records_with_any_quasi_identifier	17316306	24.3%	38585576	54.1%