

# 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: 5/4/2023

## Quick Summary

	summary String	all_fields_counts Double	all_fields_pct Double	quasi_fields_co... Double	quasi_fields_pct Double	
1	total_rows	98,273,331	NaN%	98,273,331	NaN%	
2	total_columns	19	NaN%	8	NaN%	
3	total_cells	1,867,193,289	100.0%	786,186,648	100.0%	
4	suppressed_fields   •••	51,946,036	2.8%	45,221,278	5.8%	
5	missing_fields	429,557,665	23.0%	67,153,769	8.5%	
6	unknown_fields	90,441,245	4.8%	45,050,401	5.7%	
7	non_blank_fields	1,295,248,343	69.4%	628,761,200	80.0%	

## Field Level Utility Summary

	variable String	suppressed Long	suppressed_pct String	missing Long	missing_pct String	unknown Long	unknown_pct String
1	case_month	24	0.0%	0	0.0%	0	0.0%
2	res_state	1,333	0.0%	0	0.0%	0	0.0%
3	res_county	6,723,425	6.8%	0	0.0%	0	0.0%
4	sex	2,595,143	2.6%	420,989	0.4%	792,823	0.8%
5	age_group	869,649	0.9%	930,387	0.9%	0	0.0%
6	ethnicity	16,896,103	17.2%	6,202,202	6.3%	18,239,204	18.6%
7	race	15,097,545	15.4%	7,715,799	7.9%	13,003,508	13.2%
8	death_yn	3,038,056	3.1%	51,884,392	52.8%	13,014,866	13.2%
9	records_with_any_quasi_identifier	24,575,822	25.0%	54,759,341	55.7%	32,308,787	32.9%