

# 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: 12/7/2023

## Quick Summary

	summary String	all_fields_counts Double	all_fields_pct Double	quasi_fields_co... Double	quasi_fields_pct Double		
1	total_rows	102,731,244	NaN%	102,731,244	NaN%		
2	total_columns	19	NaN%	8	NaN%		
3	total_cells	1,951,893,636	100.0%	821,849,952	100.0%		
4	suppressed_fields   ...	56,537,076	2.9%	49,282,910	6.0%		
5	missing_fields	449,776,765	23.0%	70,176,773	8.5%		
6	unknown_fields	92,253,319	4.7%	46,093,868	5.6%		
7	non_blank_fields	1,353,326,476	69.3%	656,296,401	79.9%		

## Field Level Utility Summary

	variable String	suppressed Long	suppressed_pct String	missing Long	missing_pct String	unknown Long	unknown_pct String		
1	case_month	12	0.0%	0	0.0%	0	0.0%		
2	res_state	1,750	0.0%	0	0.0%	0	0.0%		
3	res_county	7,252,416	7.1%	0	0.0%	0	0.0%		
4	sex	3,052,553	3.0%	421,726	0.4%	791,428	0.8%		
5	age_group	1,044,027	1.0%	1,010,514	1.0%	0	0.0%		
6	ethnicity	18,316,617	17.8%	6,266,432	6.1%	18,593,695	18.1%		
7	race	16,296,293	15.9%	7,861,380	7.7%	13,169,631	12.8%		
8	death_yn	3,319,242	3.2%	54,616,721	53.2%	13,539,114	13.2%		
9	records_with_any_quasi_identifier   ...	26,566,527	25.9%	57,552,473	56.0%	33,181,453	32.3%		

