

# 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: 10/3/2022

## Quick Summary

	summary String	all_fields_counts Double	all_fields_pct Double	quasi_fields_co... Double	quasi_fields_pct Double	
1	total_rows	87,306,976	NaN%	87,306,976	NaN%	
2	total_columns	19	NaN%	8	NaN%	
3	total_cells	1,658,832,544	100.0%	698,455,808	100.0%	
4	suppressed_fields ***	43,845,792	2.6%	38,362,437	5.5%	
5	missing_fields	383,917,073	23.1%	60,152,853	8.6%	
6	unknown_fields	80,116,760	4.8%	41,032,248	5.9%	
7	non_blank_fields	1,150,952,919	69.4%	558,908,270	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	7	0.0%	0	0.0%	0	0.0%	
2	res_state	1,208	0.0%	0	0.0%	0	0.0%	
3	res_county	5,482,147	6.3%	0	0.0%	0	0.0%	
4	sex	2,074,188	2.4%	410,957	0.5%	1,085,807	1.2%	
5	age_group	702,467	0.8%	751,987	0.9%	0	0.0%	
6	ethnicity	14,430,522	16.5%	5,690,628	6.5%	17,086,733	19.6%	
7	race	12,911,390	14.8%	7,410,589	8.5%	11,964,725	13.7%	
8	death_yn	2,760,508	3.2%	45,888,692	52.6%	10,894,983	12.5%	
9	records_with_any_quasi_identifier ***	20,832,600	23.9%	48,783,774	55.9%	29,078,603	33.3%	