

# 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: 6/1/2023

## Quick Summary

	summary String	all_fields_counts Double	all_fields_pct Double	quasi_fields_co... Double	quasi_fields_pct Double	
1	total_rows	98,629,359	NaN%	98,629,359	NaN%	
2	total_columns	19	NaN%	8	NaN%	
3	total_cells	1,873,957,821	100.0%	789,034,872	100.0%	
4	suppressed_fields   •••	52,274,555	2.8%	45,561,065	5.8%	
5	missing_fields	430,488,030	23.0%	67,316,657	8.5%	
6	unknown_fields	90,691,423	4.8%	45,081,577	5.7%	
7	non_blank_fields	1,300,503,813	69.4%	631,075,573	80.0%	

## Field Level Utility Summary

	variable String	suppressed Long	suppressed_pct String	missing Long	missing_pct String	unknown Long	unknown_pct String
1	res_county	6,712,042	6.8%	0	0.0%	0	0.0%
2	case_month	9	0.0%	0	0.0%	0	0.0%
3	res_state	1,448	0.0%	0	0.0%	0	0.0%
4	sex	2,637,316	2.7%	421,069	0.4%	790,127	0.8%
5	age_group	891,435	0.9%	930,412	0.9%	0	0.0%
6	ethnicity	17,024,759	17.3%	6,192,341	6.3%	18,241,758	18.5%
7	race	15,232,831	15.4%	7,685,221	7.8%	13,001,033	13.2%
8	death_yn	3,061,225	3.1%	52,087,614	52.8%	13,048,659	13.2%
9	records_with_any_quasi_identifier	24,696,077	25.0%	54,961,624	55.7%	32,345,567	32.8%