

# Public 12 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: 10/5/2023

## Quick Summary

	summary String	all_fields_counts Double	all_fields_pct Double	quasi_fields_counts Double	quasi_fields_pct Double		
1	total_rows	101,224,479	NaN%	101,224,479	NaN%		
2	total_columns	12	NaN%	3	NaN%		
3	total_cells	1,214,693,748	100.0%	303,673,437	100.0%		
4	suppressed_fields	82	0.0%	82	0.0%		
5	missing_fields	276,722,511	22.8%	4,138,542	1.4%		
6	unknown_fields	82,963,400	6.8%	31,639,768	10.4%		
7	non_blank_fields	855,007,755	70.4%	267,895,045	88.2%		

## Field Level Utility Summary

	variable String	suppressed Long	suppressed_pct String	missing Long	missing_pct String	unknown Long	unknown_pct String
1	sex	14	0.0%	492,894	0.5%	1,024,361	1.0%
2	age_group	54	0.0%	1,020,000	1.0%	0	0.0%
3	race_ethnicity_combined	14	0.0%	2,625,648	2.6%	30,615,407	30.2%
4	records_with_any_quasi_identifier	54	0.0%	3,830,519	3.8%	30,932,036	30.6%

