

# COVID-19 Case Surveillance Restricted Access Detailed 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: 2021-05-10

Total records in dataset: 25,009,120

## Quick Summary

	all_fields	quasi_fields
total_fields	32	6
total_records	25,009,120	25,009,120
total_cells	800,291,840	150,054,720
missing_fields	403,315,032	18,371,208
missing_pct	50%	12%
complete_fields	396,976,808	131,683,512
complete_pct	50%	88%
unknown_fields	66,174,390	12,155,799
unknown_pct	8%	8%
suppressed_fields	812,204	809,721
suppressed_pct	0%	1%
available_fields	329,990,214	118,717,992
available_pct	41%	79%

## Field-level Utility Summary

	suppressed	suppressed_percent	missing	missing_percent
race_ethnicity_combined	305,008	1.2%	603,909	2.4%
sex	99,773	0.4%	45,338	0.2%
age_group	44,918	0.2%	170,302	0.7%
hc_work_yn	355,754	1.4%	16,869,681	67.5%
res_county	2,134	0.0%	681,958	2.7%
res_state	2,134	0.0%	20	0.0%
records_with_any_field	355,754	1.4%	16,909,877	67.6%