

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: 2021-05-10

Total records in dataset: 25,009,120

Quick Summary

	all_fields	quasi_fields
total_fields	12	3
total_records	25,009,120	25,009,120
total_cells	300,109,440	75,027,360
missing_fields	61,477,791	855,939
missing_pct	20%	1%
complete_fields	238,631,649	74,171,421
complete_pct	80%	99%
unknown_fields	20,862,880	8,967,561
unknown_pct	7%	12%
suppressed_fields	110	110
suppressed_pct	0%	0%
available_fields	217,768,659	65,203,750
available_pct	73%	87%

Field-level Utility Summary

	suppressed	suppressed_percent	missing	missing_percent
race_ethnicity_combined	19	0.0%	628,767	2.5%
sex	19	0.0%	54,525	0.2%
age_group	72	0.0%	172,647	0.7%
records_with_any_field	72	0.0%	795,987	3.2%