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: 2021-10-12

Total records in dataset: 34,203,283

Quick Summary

	all_fields	quasi_fields
total_fields	12	3
total_records	34,203,283	34,203,283
total_cells	410,439,396	102,609,849
missing_fields	84,413,564	1,226,339
missing_pct	21%	1%
complete_fields	326,025,832	101,383,510
complete_pct	79%	99%
unknown_fields	27,033,127	11,356,933
unknown_pct	7%	11%
suppressed_fields	101	101
suppressed_pct	0%	0%
available_fields	298,992,604	90,026,476
available_pct	73%	88%

Field-level Utility Summary

	suppressed	suppressed_percent	missing	missing_percent
race_ethnicity_combined	11	0.0%	800,657	2.3%
sex	17	0.0%	72,100	0.2%
age_group	73	0.0%	353,582	1.0%
records_with_any_field	73	0.0%	1,139,091	3.3%