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: 44,083,363

Quick Summary

	all_fields	quasi_fields
total_fields	12	3
total_records	44,083,363	44,083,363
total_cells	529,000,356	132,250,089
missing_fields	112,929,939	1,547,740
missing_pct	21%	1%
complete_fields	416,070,417	130,702,349
complete_pct	79%	99%
unknown_fields	35,223,487	13,942,941
unknown_pct	7%	11%
suppressed_fields	107	107
suppressed_pct	0%	0%
available_fields	380,846,823	116,759,301
available_pct	72%	88%

Field-level Utility Summary

	suppressed	suppressed_percent	missing	missing_percent
race_ethnicity_combined	6	0.0%	1,063,336	2.4%
sex	19	0.0%	84,709	0.2%
age_group	82	0.0%	399,695	0.9%
records_with_any_field	82	0.0%	1,449,706	3.3%