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: 40,911,354

Quick Summary

	all_fields	quasi_fields
total_fields	12	3
total_records	40,911,354	40,911,354
total_cells	490,936,248	122,734,062
missing_fields	105,355,604	1,413,279
missing_pct	21%	1%
complete_fields	385,580,644	121,320,783
complete_pct	79%	99%
unknown_fields	31,274,039	12,943,732
unknown_pct	6%	11%
suppressed_fields	107	107
suppressed_pct	0%	0%
available_fields	354,306,498	108,376,944
available_pct	72%	88%

Field-level Utility Summary

	suppressed	suppressed_percent	missing	missing_percent
race_ethnicity_combined	6	0.0%	952,467	2.3%
sex	13	0.0%	80,428	0.2%
age_group	88	0.0%	380,384	0.9%
records_with_any_field	88	0.0%	1,319,759	3.2%