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: 2022-02-28

Total records in dataset: 63,272,363

Quick Summary

	all_fields	quasi_fields
total_fields	12	3
total_records	63,272,363	63,272,363
total_cells	759,268,356	189,817,089
missing_fields	167,521,153	2,433,660
missing_pct	22%	1%
complete_fields	591,747,203	187,383,429
complete_pct	78%	99%
unknown_fields	54,090,839	21,022,013
unknown_pct	7%	11%
suppressed_fields	95	95
suppressed_pct	0%	0%
available_fields	537,656,269	166,361,321
available_pct	71%	88%

Field-level Utility Summary

	suppressed	suppressed_percent	missing	missing_percent
race_ethnicity_combined	9	0.0%	1,534,510	2.4%
sex	9	0.0%	153,366	0.2%
age_group	77	0.0%	745,784	1.2%
records_with_any_field	77	0.0%	2,267,768	3.6%