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: 2021-12-14

Total records in dataset: 40,314,956

Quick Summary

	all_fields	quasi_fields
total_fields	12	3
total_records	40,314,956	40,314,956
total_cells	483,779,472	120,944,868
missing_fields	103,802,764	1,396,279
missing_pct	21%	1%
complete_fields	379,976,708	119,548,589
complete_pct	79%	99%
unknown_fields	30,734,448	12,761,844
unknown_pct	6%	11%
suppressed_fields	113	113
suppressed_pct	0%	0%
available_fields	349,242,147	106,786,632
available_pct	72%	88%

Field-level Utility Summary

	suppressed	suppressed_percent	missing	missing_percent
race_ethnicity_combined	10	0.0%	938,496	2.3%
sex	17	0.0%	79,693	0.2%
age_group	86	0.0%	378,090	0.9%
records_with_any_field	86	0.0%	1,303,269	3.2%