

COVID-19 Case Surveillance Public Use Data with Geography 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: 2021-07-06

Total records in dataset: 27,145,726

Quick Summary

	all_fields	quasi_fields
total_fields	19	8
total_records	27,145,726	27,145,726
total_cells	515,768,794	217,165,808
missing_fields	109,361,899	13,862,453
missing_pct	21%	6%
complete_fields	406,406,895	203,303,355
complete_pct	79%	94%
unknown_fields	25,370,460	13,148,961
unknown_pct	5%	6%
suppressed_fields	16,437,080	14,599,904
suppressed_pct	3%	7%
available_fields	364,599,355	175,554,490
available_pct	71%	81%

Field-level Utility Summary

	suppressed	suppressed_percent	missing	missing_percent
case_month	0	0.0%	0	0.0%
res_state	652	0.0%	0	0.0%
res_county	1,836,524	6.8%	0	0.0%
age_group	313,607	1.2%	173,941	0.6%
sex	820,988	3.0%	32,920	0.1%
race	4,681,773	17.2%	2,211,675	8.1%
ethnicity	5,605,423	20.6%	1,731,172	6.4%
death_yn	1,340,937	4.9%	9,712,745	35.8%
records_with_any_field	7,881,313	29.0%	10,855,906	40.0%