

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: 2021-05-24

Total records in dataset: 25,470,408

Quick Summary

	all_fields	quasi_fields
total_fields	19	8
total_records	25,470,408	25,470,408
total_cells	483,937,752	203,763,264
missing_fields	102,822,847	13,270,903
missing_pct	21%	7%
complete_fields	381,114,905	190,492,361
complete_pct	79%	93%
unknown_fields	24,031,914	12,344,844
unknown_pct	5%	6%
suppressed_fields	15,512,350	13,639,117
suppressed_pct	3%	7%
available_fields	341,570,641	164,508,400
available_pct	71%	81%

Field-level Utility Summary

	suppressed	suppressed_percent	missing	missing_percent
case_month	0	0.0%	0	0.0%
res_state	724	0.0%	0	0.0%
res_county	1,872,509	7.4%	0	0.0%
age_group	283,936	1.1%	162,065	0.6%
sex	744,028	2.9%	33,201	0.1%
race	4,343,541	17.1%	2,057,595	8.1%
ethnicity	5,202,456	20.4%	1,672,904	6.6%
death_yn	1,191,923	4.7%	9,345,138	36.7%
records_with_any_field	7,457,970	29.3%	10,495,418	41.2%