

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: 2021-06-21

Total records in dataset: 26,887,803

Quick Summary

	all_fields	quasi_fields
total_fields	19	8
total_records	26,887,803	26,887,803
total_cells	510,868,257	215,102,424
missing_fields	108,390,058	13,710,006
missing_pct	21%	6%
complete_fields	402,478,199	201,392,418
complete_pct	79%	94%
unknown_fields	25,372,880	13,257,304
unknown_pct	5%	6%
suppressed_fields	16,145,283	14,312,406
suppressed_pct	3%	7%
available_fields	360,960,036	173,822,708
available_pct	71%	81%

Field-level Utility Summary

	suppressed	suppressed_percent	missing	missing_percent
case_month	0	0.0%	0	0.0%
res_state	650	0.0%	0	0.0%
res_county	1,832,227	6.8%	0	0.0%
age_group	305,828	1.1%	172,222	0.6%
sex	801,502	3.0%	32,974	0.1%
race	4,573,908	17.0%	2,136,626	7.9%
ethnicity	5,483,645	20.4%	1,702,110	6.3%
death_yn	1,314,646	4.9%	9,666,074	35.9%
records_with_any_field	7,750,081	28.8%	10,807,337	40.2%