

COVID-19 Case Surveillance Restricted Access Detailed 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: 2021-05-11

Total records in dataset: 25,120,922

Quick Summary

	all_fields	quasi_fields
total_fields	32	6
total_records	25,120,922	25,120,922
total_cells	803,869,504	150,725,532
missing_fields	405,263,979	18,464,035
missing_pct	50%	12%
complete_fields	398,605,525	132,261,497
complete_pct	50%	88%
unknown_fields	66,410,196	12,206,728
unknown_pct	8%	8%
suppressed_fields	812,562	810,100
suppressed_pct	0%	1%
available_fields	331,382,767	119,244,669
available_pct	41%	79%

Field-level Utility Summary

	suppressed	suppressed_percent	missing	missing_percent
race_ethnicity_combined	305,190	1.2%	604,536	2.4%
sex	99,791	0.4%	45,382	0.2%
age_group	44,901	0.2%	170,477	0.7%
hc_work_yn	355,980	1.4%	16,955,471	67.5%
res_county	2,119	0.0%	688,149	2.7%
res_state	2,119	0.0%	20	0.0%
records_with_any_field	355,980	1.4%	16,995,732	67.7%