

Public 19 Reporting

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: 8/4/2023

Quick Summary

	summary String	all_fields_counts Double	all_fields_pct Double	quasi_fields_co... Double	quasi_fields_pct Double		
1	total_rows	99,566,261	NaN%	99,566,261	NaN%		
2	total_columns	19	NaN%	8	NaN%		
3	total_cells	1,891,758,959	100.0%	796,530,088	100.0%		
4	suppressed_fields ...	53,352,621	2.8%	46,534,157	5.8%		
5	missing_fields	434,485,609	23.0%	67,791,712	8.5%		
6	unknown_fields	90,107,651	4.8%	45,240,891	5.7%		
7	non_blank_fields	1,313,813,078	69.4%	636,963,328	80.0%		

Field Level Utility Summary

	variable String	suppressed Long	suppressed_pct String	missing Long	missing_pct String	unknown Long	unknown_pct String		
1	case_month	18	0.0%	0	0.0%	0	0.0%		
2	res_state	1,574	0.0%	0	0.0%	0	0.0%		
3	res_county	6,816,890	6.8%	0	0.0%	0	0.0%		
4	sex	2,745,378	2.8%	420,159	0.4%	791,926	0.8%		
5	age_group	947,991	1.0%	935,017	0.9%	0	0.0%		
6	ethnicity	17,330,609	17.4%	6,189,651	6.2%	18,276,624	18.4%		
7	race	15,498,737	15.6%	7,700,365	7.7%	13,013,536	13.1%		
8	death_yn	3,192,960	3.2%	52,546,520	52.8%	13,158,805	13.2%		
9	records_with_any_quasi_identifier ...	25,143,272	25.3%	55,458,352	55.7%	32,459,042	32.6%		