# 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: 2021-05-24

Total records in dataset: 25,470,408

## Quick Summary

|  | all_fields | quasi_fields |
|---|---|---|
| total_fields | 12 | 3 |
| total_records | 25,470,408 | 25,470,408 |
| total_cells | 305,644,896 | 76,411,224 |
| missing_fields | 62,698,708 | 858,265 |
| missing_pct | 21% | 1% |
| complete_fields | 242,946,188 | 75,552,959 |
| complete_pct | 79% | 99% |
| unknown_fields | 21,200,197 | 9,104,874 |
| unknown_pct | 7% | 12% |
| suppressed_fields | 101 | 101 |
| suppressed_pct | 0% | 0% |
| available_fields | 221,745,890 | 66,447,984 |
| available_pct | 73% | 87% |

## Field-level Utility Summary

|  | suppressed | suppressed_percent | missing | missing_percent |
|---|---|---|---|---|
| race_ethnicity_combined | 11 | 0.0% | 630,926 | 2.5% |
| sex | 16 | 0.0% | 55,007 | 0.2% |
| age_group | 74 | 0.0% | 172,332 | 0.7% |
| records_with_any_field | 74 | 0.0% | 798,093 | 3.1% |