Emergency Medical Services Authority

CEMSIS Data Requests

Information Technology
11120 International Drive, Suite 200
Rancho Cordova, CA 95670
Phone: (916) 323-9875 | Fax: (916) 324-2875
www.emsa.ca.gov

The Emergency Medical Services Authority (EMSA) Data Use Agreement (DUA) policy is compliant with the Information Practices Act (California Civil Code § 1798, et. seq.). This form is for use by organizations or individuals interested in obtaining non-public data held by the Health Information and Data Technologies Branch under the Data Technologies Unit in the California EMS Information System (CEMSIS) repository.

Pursuant to Civil Code Section 1798.24, qualifying entities may request CEMSIS data elements containing personal information if they have obtained research protocol approval from the Committee for the Protection of Human Subjects (CPHS).

LIMITATIONS

The CEMSIS Database is a large convenience sample— it consists solely of data submitted by participating EMS agencies and providers, and it is not a population-based data set. In addition, the CEMSIS database inherits any individual deficiencies originating from its contributing entities. The EMS Authority is continually cleaning and standardizing the data to improve data quality. Data files received from contributing EMS agencies and providers are checked for completeness, logical consistency, and proper formatting. Any data files not passing the CEMSIS validation and data cleaning processes are rejected or flagged; based upon the seriousness of the discovered errors. A data profile report is generated for each submitted file from a submitting entity allowing the opportunity to review the quality of submitted data, correct errors, and resubmit their data if needed. Any remaining data limitations, errors, or inconsistencies are allowed into the database. Errors remain for two reasons:

1. It would be extremely difficult to flag individual errors and request corrections from hundreds of agencies submitting millions of records, and

2. Allowing local EMS Agencies (and associated EMS providers) to view data as submitted, will hopefully facilitate efforts to further refine data collection techniques to increase the accuracy of data collected.

The CEMSIS database is not a population-based dataset. The CEMSIS database is subject to the limitations of any "convenience sample." In addition, users of the National EMS Information System (NEMSIS) database should understand that CEMSIS data are "event-based" and not "patient-based." That is, a single patient may be represented in more than one record for a variety of reasons. A patient may request EMS assistance frequently and, therefore, be represented in the dataset more than once. In addition, several agencies may respond to the same event (i.e., one patient) with each agency submitting a patient care record to the CEMSIS database. Thus, as NEMSIS refers to the national EMS database, the CEMSIS database is a registry of "EMS activations."

This data cannot be loaded or shared with a system that utilizes GenAI. If GenAI is going to

be utilized, it must be assessed and approved by both EMSA and CDT. ([Executive order | GenAI](#))

The California EMS Database is not a Population-Based Dataset
The CEMSIS database is subject to the limitations of any "convenience sample." Since inception, not all local EMS agencies (LEMSAs) have submitted all their EMS data. In addition, users of CEMSIS data should understand that the data is "event-based" and not "patient-based." That is, a single patient may be represented in more than one record for a variety of reasons, as discussed above.

Selection Bias and Information Bias
As a "convenience sample," the CEMSIS database is subject to various forms of bias. The most obvious forms of this are selection bias and information bias, such as inconsistency with which clinical variables can be measured or documented and inter-agency differences in treatment and transport practices.

Selection bias refers to an apparent difference between two groups that is caused by different inclusion criteria. For example, if one LEMSA includes all patients treated and transported by EMTs in its registry and another LEMSA does not, the LEMSA not including this patient population may appear to have a greater amount of "higher acuity" events compared to the LEMSA including patients treated and transported by EMTs. Any difference in inclusion/exclusion criteria could produce a selection bias.

Information bias refers to an apparent difference between two groups that is caused by a difference in the data available to compare them. Many data fields shared between NEMSIS/CEMSIS allow for null values. Differences in the proportion of cases with missing (or null) data may be responsible for apparent differences among EMS agencies between or within LEMSAs. If one LEMSA demonstrates incomplete data on patients when reporting treatment complications, for example, it may falsely appear to be delivering better quality care than a LEMSA that diligently requires every treatment complication to be recorded.

Missing Data in CEMSIS
The proportion of missing data varies across data elements in the CEMSIS database, and it is important to decide how to deal with missing data when doing analyses. In most cases, NEMSIS/CEMSIS data are not missing at random and analyses, therefore, are subject to bias if missing data are ignored. That is, the results may be misleading when excluding all observations with missing data or null values. Excluding observations with missing values is the default for most software programs when running statistical analyses. Another available option is to provide plausible values for the missing data, either by single value or multiple value imputation. A single imputation of a value may be an educated guess at the value, substitution of the mean value, or substitution based on a regression equation using other (observed) values. For example, one might assume that a patient has suffered an injury, (eSituation.02 = Yes), if the complaint reported to dispatch was "Traumatic Injury". Most statistical software packages can do imputations without much difficulty. However, it is important to explore the impact of missing data with sensitivity analyses. That is, repeat an analysis with and without imputation and see whether there are any important differences.