

Example Data Structure for the Core Digital Measures of Sleep

When data are collected and annotated, they can be stored in a variety of ways. For epoch level data, such as that described in the [Core Measures: Sleep Measurement System](#), one convenient way of storing data for later analysis is described here.

Although it is not a requirement for a technology developer to supply data in this format, or for a researcher to use this format for their analysis, thoughtful collaboration involves the ability for one party to be able to use a tool employed by a previous party, with minimum friction.

By using the structure presented here, or collaborating with the other stakeholders in the project, users can ensure that the data transfer and later analysis is a smooth and fruitful process without unnecessary barriers to research.

Epoch level labels

The example datasets presented here are in long format. This format designates a single record (row) for each epoch, meaning there are multiple records for each patient, labelled with the date, start time, and end time of the epoch, along with other useful information, such as the date the study began (when conducting research).

The epoch level labels in the Core Measures Sleep System can be stored as 0 and 1 values for each epoch, representing the presence or absence of a given state or event. An extract of an example dataset is shown and annotated below:

[illegible]

COPD: chronic obstructive pulmonary disease, SAMD: Software as a Medical Device, CGM: Continuous Glucose Monitoring, ECG: Electrocardiogram, MD: Physician, NP: Nurse Practitioner, PA: Physician Assistant, PT: Physical Therapist, OT: Occupational Therapist, SLP: Speech-language Pathologist, CSW: Certified Social Worker.

Variables

The epoch level labels can be used to create additional datasets. In line with the Core Measures: Sleep Measurement System, a useful dataset for creation contains the variables listed in this system. These variables include start times, end times, and durations of the periods outlined in the system (e.g., primary sleep period, primary in bed period, time attempting to sleep, etc.).

The data can continue to be presented in long format. One reason for this formatting choice is some variables could have multiple records in a given study day (such as rest periods, wake events, and nap events). A long format dataset can handle multiple records of this nature.

The start and end times may be displayed in multiple ways, and could include the date and time in hours, minutes, and seconds.

USUBJID	DAY	Date	Variable	Start Time	End Time	Duration
1	1	20240124	Primary_In_Bed	20:15	7:05	650
1	1	20240124	Time_Attempting_to_Sleep	20:20	7:00	640
1	1	20240124	Primary_Sleep_Period	20:35	7:00	625
1	1	20240124	Wake_Event	21:15	21:35	20
1	2	20240125	Rest_Period	7:50	8:05	15
1	2	20240125	Nap_Event	7:55	8:05	10

Here, duration is shown in minutes. It is likely that duration will be held in the dataset in seconds, but displayed in minutes or other interpretable values in the analysis output.

Core Measures

A final dataset can build on the epoch level and the variable level dataset to define the Core Digital Measures of Sleep.

A unit column can be added to describe the units used, or units can be defined in a specification document.

USUBJID	AVISITN	Date	Core Measure	Value	Unit
1	1	20240124	Total Napping Time	0	minutes
1	1	20240124	Total Sleep Time	605	minutes
1	1	20240124	Initial Sleep Onset Latency	15	minutes
1	1	20240124	Wake After Sleep Onset	20	minutes
1	1	20240124	Number of Wake Events in the Prim	1	count
1	1	20240124	Sleep Efficiency	94.50%	percent
1	2	20240125	Total Napping Time	10	minutes

As above, minutes are presented here for illustrative purposes, but a finer resolution (seconds) will likely be held in the dataset, using minutes for the presentation of results.

