

SUPPORT Datasets | [Datasets Archived at ICPSR](#)

The `support` dataset is a random sample of 1000 patients from Phases I & II of SUPPORT (Study to Understand Prognoses Preferences Outcomes and Risks of Treatment). This dataset is very good for learning how to fit highly nonlinear predictor effects, imputing missing data, and for modeling 4 types of response variables:

- Binary (hospital death)
- Ordinal (severe functional disability - see below)
- Continuous (total hospital costs or possibly study length of stay)
- Time until event (death) with waning effects of baseline physiologic variables over time. Patients are followed up to 5.56 years.

See the following reference for more information about the SUPPORT study (funded by the Robert Wood Johnson Foundation): Knaus WA, Harrell FE, Lynn J *et al.* (1995): The SUPPORT prognostic model: Objective estimates of survival for seriously ill hospitalized adults. *Annals of Internal Medicine* 122:191-203.

You may want to use the following normal values that have been found to be satisfactory in imputing missing baseline physiologic data:

<i>Baseline Variable</i>	<i>Normal Fill-in Value</i>
Serum albumin	3.5
PaO ₂ /FiO ₂ ratio (<code>pafi</code>)	333.3
Bilirubin	1.01
Creatinine	1.01
BUN	6.51
White blood count	9 (thousands)
Urine output	2502

The ordinal functional disability variable `sfdm2` has the following levels (in order):

<i>Level</i>	<i>Meaning</i>
1	Patient lived 2 months to be able to get 2 month interview, and from this interview there were no signs of moderate to severe functional disability
2	Patient was unable to do 4 or more activities of daily living at month 2 after study entry. If the patient was not interviewed but the patient’s surrogate was, the cutoff for disability was ADL ³ 5
3	Sickness Impact Profile total score at 2 months ³ 30
4	Patient intubated or in coma at 2 months
5	Patient died before 2 months after study entry

There are 159 patients surviving 2 months for whom there were nopatient or surrogate interviews. These patients have missing `sfdm2`.

support2 Dataset

Data for all 9105 SUPPORT patients are available. These files contains all variables in the above `support` plus the following (see variable labels for more documentation). As before, most of the patient assessments are made on the patient’s third study day.

<i>Variable</i>	<i>Meaning</i>
<code>sps</code>	SUPPORT day 3 physiology score
<code>aps</code>	APACHE III day 3 physiology score
<code>surv2m</code>	SUPPORT model 2-month survival estimate
<code>surv6m</code>	SUPPORT model 6-month survival estimate
<code>hday</code>	Day in hospital at which patient entered study
<code>diabetes</code>	Diabetes as a comorbidity
<code>dementia</code>	Dementia as a comorbidity
<code>ca</code>	Patient has cancer
<code>prg2m</code>	Physician’s 2-month survival estimate for pt.
<code>prg6m</code>	Physician’s 6-month survival estimate for pt.
<code>dnr</code>	Patient had DNR order
<code>dnrday</code>	Day of DNR order (<0 if before study)

To develop models without using findings from previous models, be sure not to use `aps`, `sps`, `surv2m`, `surv6m` as predictors. You also will probably not want to use `prg2m`, `prg6m`, `dnr`, `dnrday`.