

PREDICTING MORTALITY OF ICU PATIENTS: THE PHYSIONET/COMPUTING IN CARDIOLOGY CHALLENGE 2012

Quick Links:

- Read the [papers](#) [papers/] presented at CinC 2012 by participants in this challenge.
- Training set A [set-a] [tarball] [set-a.tar.gz] [zip archive] [set-a.zip] [Outcome-related descriptors] [Outcomes-a.txt]
- Test set B [set-b] [tarball] [set-b.tar.gz] [zip archive] [set-b.zip] [Outcomes not provided]
- Software [for MATLAB] [matlab-software] [for C] [c-software]
- Challenge rules and deadlines [rules-and-dates]

Updates:

15 March: Sets A and B have been corrected (see [details](#) [data-correction]). The sample entry is now available. The [scoring criteria](#) [scoring] for events 1 and 2 have been defined.

23 March: [Registration](#) [registration] is now open. The software provided for challenge participants has been updated.

27 March: Register and submit a preliminary entry by 25 April if you wish to participate in the Challenge.

26 April: Phase 1 has ended and no further entries may be submitted until Phase 2 begins on 1 June.

8 May: The data sets have been revised for Phase 2 (see [details](#) [data-correction]).

30 August: [Final scores for both events](#) [top-scores.shtml] have been posted.

The development of methods for prediction of mortality rates in Intensive Care Unit (ICU) populations has been motivated primarily by the need to compare the efficacy of medications, care guidelines, surgery, and other interventions when, as is common, it is necessary to control for differences in severity of illness or trauma, age, and other factors. For example, comparing overall mortality rates between trauma units in a community hospital, a teaching hospital, and a military field hospital is likely to reflect the differences in the patient populations more than any differences in standards of care. Acuity scores such as APACHE and SAPS-II are widely used to account for these differences in the context of such studies.

By contrast, the focus of the PhysioNet/CinC Challenge 2012 is to develop methods for patient-specific prediction of in-hospital mortality. Participants will use information collected during the first two days of an ICU stay to predict which patients survive their hospitalizations, and which patients do not.

DATA FOR THE CHALLENGE

See the Quick Links at the top of this page to download the Challenge data!

The data used for the challenge consist of records from 12,000 ICU stays. All patients were adults who were admitted for a wide variety of reasons to cardiac, medical, surgical, and trauma ICUs. ICU stays of less than 48 hours have been excluded. Patients with [DNR](#) [http://en.wikipedia.org/wiki/Do_not_resuscitate] (do not resuscitate) or CMO (comfort measures only) directives were not excluded.

On this page:

- Data for the Challenge [data-for-the-challenge]
- General descriptors [general-descriptors]
- Time Series [time-series]
- Outcome-related Descriptors [outcome-related-descriptors]
- Entering the Challenge [rules-and-dates]
- Sample entries, test frameworks, and sample results [sample-entries-test-frameworks-and-sample-results]
- Challenge scoring [challenge-scoring]
- Frequently asked questions about the Challenge [frequently-asked-questions-about-the-challenge]
- Can I submit an entry written in ...? [can-i-submit-an-entry-written-in-]
- Would it still be possible to participate if we send you code that is compiled and tested to run on the destination hardware and OS? [would-it-still-be-possible-to-participate-if-we-se]
- If I don't submit all 5 preliminary (Phase 1) entries, can I add the unused ones to my quota of 5 Phase 2 entries? [if-i-dont-submit-all-5-preliminary-phase-1-entries]
- Will it be possible to outperform the sample entries based on SAPS-II? [will-it-be-possible-to-outperform-the-sample-entri]
- Why are the scores for the C and m-code sample entries different? [why-are-the-scores-for-

Four thousand records comprise *training set A*, and the remaining records form *test sets B and C*. Outcomes are provided for the training set records, and are withheld for the test set records.

Up to 42 variables were recorded at least once during the first 48 hours after admission to the ICU. Not all variables are available in all cases, however. Six of these variables are *general descriptors* (collected on admission), and the remainder are *time series*, for which multiple observations may be available.

Each observation has an associated *time-stamp* indicating the elapsed time of the observation since ICU admission in each case, in hours and minutes. Thus, for example, a time stamp of 35:19 means that the associated observation was made 35 hours and 19 minutes after the patient was admitted to the ICU.

Each record is stored as a comma-separated value (CSV) text file. To simplify downloading, participants may download a zip file or tarball containing all of training set A or test set B. Test set C will be used for validation only and will not be made available to participants.

Update (8 May 2012): The extraneous ages that were present in the previous versions of some data files have been removed, and a new general descriptor (*ICUType*, see below) has been added in each data file.

Five additional *outcome-related descriptors*, described below, are known for each record. These are stored in separate CSV text files for each of sets A, B, and C, but only those for set A are available to challenge participants.

All valid values for *general descriptors*, *time series variables*, and *outcome-related descriptors* are non-negative (≥ 0). A value of -1 indicates missing or unknown data (for example, if a patient's height was not recorded).

GENERAL DESCRIPTORS

As noted, these six descriptors are collected at the time the patient is admitted to the ICU. Their associated time-stamps are set to 00:00 (thus they appear at the beginning of each patient's record).

- *RecordID* (a unique integer for each ICU stay)
- *Age* (years)
- *Gender* (0: female, or 1: male)
- *Height* (cm)
- *ICUType* (1: Coronary Care Unit, 2: Cardiac Surgery Recovery Unit, 3: Medical ICU, or 4: Surgical ICU)
- *Weight* (kg) . [#weight]

[#weight]

The *ICUType* was added for use in Phase 2; it specifies the type of ICU to which the patient has been admitted.

TIME SERIES

These 37 variables may be observed once, more than once, or not at all in some cases:

- | | | |
|--|--|--|
| • <i>Albumin</i> [http://en.wikipedia.org/wiki/Human_serum_albumin] (g/dL) | • <i>HCT</i> [http://en.wikipedia.org/wiki/Hematocrit] [Hematocrit (%)] | • <i>PaCO2</i> [http://en.wikipedia.org/] pressure of arterial CO ₂ |
| • <i>ALP</i> [http://en.wikipedia.org/wiki/Alkaline_phosphatase] [Alkaline phosphatase (IU/L)] | • <i>HR</i> [http://en.wikipedia.org/wiki/Heart_rate] [Heart rate (bpm)] | • <i>PaO2</i> [http://en.wikipedia.org/wiki/partial_pressure_of_oxygen] pressure of arterial O ₂ (r |
| • <i>ALT</i> [http://en.wikipedia.org/wiki/Alanine_transaminase] [Alanine transaminase (IU/L)] | • <i>K</i> [http://en.wikipedia.org/wiki/Hypokalemia] [Serum potassium (mEq/L)] | • <i>pH</i> [http://en.wikipedia.org/wiki/A14]] |
| • <i>AST</i> [http://en.wikipedia.org/wiki/Aspartate_transaminase] [Aspartate transaminase (IU/L)] | • <i>Lactate</i> [http://en.wikipedia.org/wiki/Lactic_acid] (mmol/L) | • <i>Platelets</i> [http://en.wikipedia.org/wiki/Platelet] |
| • <i>Bilirubin</i> [http://en.wikipedia.org/wiki/Bilirubin] (mg/dL) | • <i>Mg</i> [http://en.wikipedia.org/wiki/Magnesium#Biological_role] [Serum magnesium (mmol/L)] | • <i>RespRate</i> [http://en.wikipedia.org/wiki/Respiration_rate] [Respiration rate (bpm)] |
| • <i>BUN</i> [http://en.wikipedia.org/wiki/BUN] [Blood urea nitrogen (mg/dL)] | • <i>MAP</i> [http://en.wikipedia.org/wiki/Mean_arterial_pressure] [Invasive mean arterial blood pressure] | • <i>SaO2</i> [http://en.wikipedia.org/wiki/Saturation_of_hemoglobin] saturation in hemoglobin |
| • <i>Cholesterol</i> [http://en.wikipedia.org/wiki/Cholesterol] (mg/dL) | | • <i>SysABP</i> [http://en.wikipedia.org/wiki/Systolic_blood_pressure] [Invasive systolic arterial |

the-c-and-m-code-sample-ent]

- Why didn't you choose [my favorite statistic] as a scoring metric? [#why-didnt-you-choose-my-favorite-statistic-as-a-sc]
- What is the precise deadline for Phase 1? What happens if I can't figure out how to submit a valid entry by then? [#what-is-the-precise-deadline-for-phase-1--what-hap]
- Suggested readings on acuity scores and mortality prediction [#suggested-readings-on-acuity-scores-and-mortality-]

- *Creatinine* [http://en.wikipedia.org/wiki/Serum_creatinine#Plasma_creatinine] [Serum creatinine (mg/dL)]
- *DiasABP* [http://en.wikipedia.org/wiki/Diastolic_blood_pressure] [Invasive diastolic arterial blood pressure (mmHg)]
- *FiO2* [http://en.wikipedia.org/wiki/FiO2] [Fractional inspired O₂ (0-1)]
- *GCS* [http://en.wikipedia.org/wiki/Glasgow_coma_score] [Glasgow Coma Score (3-15)]
- *Glucose* [http://en.wikipedia.org/wiki/Serum_glucose] [Serum glucose (mg/dL)]
- *HCO3* [http://en.wikipedia.org/wiki/Bicarbonate#Diagnostics] [Serum bicarbonate (mmol/L)]
- *(mmHg)*
- *MechVent* [http://en.wikipedia.org/wiki/Mechanical_ventilation] [Mechanical ventilation respiration (0:false, or 1:true)]
- *Na* [http://en.wikipedia.org/wiki/Serum_sodium] [Serum sodium (mEq/L)]
- *NIDiasABP* [http://en.wikipedia.org/wiki/Diastolic_blood_pressure] [Non-invasive diastolic arterial blood pressure (mmHg)]
- *NIMAP* [http://en.wikipedia.org/wiki/Mean_arterial_pressure] [Non-invasive mean arterial blood pressure (mmHg)]
- *NISysABP* [http://en.wikipedia.org/wiki/Systolic_blood_pressure] [Non-invasive systolic arterial blood pressure (mmHg)]
- *Temp* [http://en.wikipedia.org/wiki/Temperature_(°C)] [Temperature (°C)]
- *TropI* [http://en.wikipedia.org/wiki/Troponin_I] [Troponin I (ng/mL)]
- *TropT* [http://en.wikipedia.org/wiki/Troponin_T] [Troponin T (ng/mL)]
- *Urine* [http://en.wikipedia.org/wiki/Urine] [Urine (mL)]
- *WBC* [http://en.wikipedia.org/wiki/White_blood_cell_count] [White blood cell count (10⁹/L)]
- *Weight* (kg)^{*} [#weight]

The time series measurements are recorded in chronological order within each record, and the associated time stamps indicate the elapsed time since admission to the ICU. Measurements may be recorded at regular intervals ranging from hourly to daily, or at irregular intervals as required. Not all time series are available in all cases.

[#weight]

In a few cases, such as blood pressure, different measurements made using two or more methods or sensors may be recorded with the same or only slightly different time-stamps. Occasional outliers should be expected as well. [#weight]

* Note that *Weight* is both a general descriptor (recorded on admission) and a time series variable (often measured hourly, for estimating fluid balance).

OUTCOME-RELATED DESCRIPTORS

The outcome-related descriptors are kept in a separate CSV text file for each of the three record sets; as noted, only the file associated with training set A is available to participants. Each line of the outcomes file contains these descriptors:

- *RecordID* (defined as above)
- *SAPS-I score* (Le Gall et al., 1984 [http://www.ncbi.nlm.nih.gov/pubmed/6499483])
- *SOFA score* (Ferreira et al., 2001 [http://www.ncbi.nlm.nih.gov/pubmed/11594901])
- *Length of stay* (days)
- *Survival* (days)
- *In-hospital death* (0: survivor, or 1: died in-hospital)

The *Length of stay* is the number of days between the patient's admission to the ICU and the end of hospitalization (including any time spent in the hospital after discharge from the ICU). If the patient's death was recorded (in or out of hospital), then *Survival* is the number of days between ICU admission and death; otherwise, *Survival* is assigned the value -1. Since patients who spent less than 48 hours in the ICU have been excluded, *Length of stay* and *Survival* never have the values 0 or 1 in the challenge data sets. Given these definitions and constraints,

$$\begin{aligned} \text{Survival} > \text{Length of stay} &\Rightarrow \text{Survivor} \\ \text{Survival} = -1 &\Rightarrow \text{Survivor} \\ 2 \leq \text{Survival} \leq \text{Length of stay} &\Rightarrow \text{In-hospital death} \end{aligned}$$

ENTERING THE CHALLENGE

To begin, we recommend studying the training set as preparation for the Challenge itself. In particular, note that the SAPS-I score can be calculated readily from the time series, as the sample entries below do. To succeed in the Challenge, you should aim to outperform the sample entries (see Software below).

All entries in the Challenge must be in the form of source code that analyses a single Challenge record, producing a prediction (0: survival, or 1: in-hospital death) and an estimate of the risk of death (as a number between 0 and 1, where 0 is certain survival and 1 is certain death).

Your entry may be written in portable (ANSI/ISO) C or MATLAB/Octave m-code; other languages, such as Java, Perl, and R, may be acceptable (see [special requirements for entries in other languages](#) [\[#other-languages\]](#) below), but please ask us first, and do so no later than **7 April 2012**. Entries must accept properly-formatted input and produce properly-formatted output, either as [physionet2012.m](#) [\[physionet2012.m\]](#) does (if written in m-code), or as [physionet2012.c](#) [\[physionet2012.c\]](#) does (otherwise).

Acceptable entries are evaluated and scored by PhysioNet using an automated test framework, two versions of which are also available to participants for testing their entries unofficially prior to submitting them. The framework starts execution of an entry, supplies data from a single Challenge record, and collects the entry's analysis for that record; this process is a "run". The framework performs a separate run for each of the 4000 records in set B or set C.

Entries will be restarted for each run (each test record); they may not store information for use in later runs (for example, by writing files to be read later, or, in MATLAB entries, by setting global variables). Entries may include files that may be read but not modified during the test.

Awards will be presented to the most successful eligible participants during [Computing in Cardiology](#) [\[http://www.cinc.org/\]](#) (CinC) 2012. To be eligible for an award, you must:

1. Join [PhysioNetWorks](#) [\[https://physionet.org/users/\]](#) if you are not already a member, and follow the link from your PhysioNetWorks home page to "PhysioNet/CinC Challenge 2012" to register as a participant. Joining the project creates a Challenge Participant Page for you, where you will submit your entries and receive your scores.
2. Submit a preliminary Challenge entry via PhysioNetWorks no later than **25 April 2012**. (The period before this deadline is Phase 1.) You may submit **up to five Phase 1 entries** before this deadline. (Use them or lose them!) Each entry will receive scores for set B.
3. Submit an acceptable abstract on your work on the Challenge to [Computing in Cardiology](#) [\[http://www.cinc.org/\]](#) no later than **1 May 2012**. *Include a pair of set B scores for at least one preliminary entry in your abstract.* Please select "PhysioNet/CinC Challenge" as the topic of your abstract, so it can be identified easily by the abstract review committee.
4. Submit a final Challenge entry via PhysioNetWorks during Phase 2 (on or after **1 June** but no later than **25 August 2012**). You may submit **up to five Phase 2 entries** between 26 April and 25 August. Each entry will receive scores for set B.
5. Select one of your previously submitted Phase 1 or Phase 2 entries for testing using set C. If you have not made a choice before **26 August 2012**, we will test the entry that received the best event 1 score for set B. The set C scores will determine the final rankings of the entries, with the top-ranked entries in each event eligible for awards.
6. Submit a full (4-page) paper on your work on the Challenge to CinC no later than **1 September 2012**.
7. Attend CinC 2012 (**9-12 September 2012**, in Krakow, Poland) and present your work there.

An important goal of this Challenge, and of others in the annual series of PhysioNet/CinC Challenges, is to accelerate progress on the Challenge questions, not only during the limited period of the Challenge, but also afterward. In pursuit of this goal, we strongly encourage participants to submit open-source entries that will be made freely available after the conclusion of the Challenge via PhysioNet. If your entry is not intended as an open-source entry, please state this clearly within its first few lines.

Eligible authors of the entries that receive the best set C scores in each Challenge event will receive award certificates during the closing plenary session of CinC on 12 September 2012. In recognition of their contributions to further work on the Challenge problem, eligible authors of the open-source entries that receive the best set C scores will also receive monetary awards. No team or individual will receive more than one such monetary award.

SAMPLE ENTRIES, TEST FRAMEWORKS, AND SAMPLE RESULTS

We have provided sample entries written in MATLAB m-code and in C, test frameworks that can be used for batch-processing a set of Challenge data using a sample entry or your own entry, code for calculating unofficial scores, as well as the outputs of the sample entries for set A. Use this software to test your entry before submitting it, to verify that it can accept properly-formatted input and produce properly-formatted output. If you wish, you may incorporate code from the sample entries within your own entry, but you will have to add something of your own creation in order to succeed in the Challenge!

Participants may find the MATLAB and C functions for calculating SAPS-I scores to be useful. The calculated SAPS scores do not always match those given in the outcomes file, however.

For participants developing entries using MATLAB:

A valid entry written in m-code must be a function named `physionet2012`, with this signature:

```
[risk,prediction]=physionet2012(time,param,value)
```

The function must be able to run this way within the test framework, `genresults.m` (below), on a 64-bit GNU/Linux platform running MATLAB R2010b (or a later version). See `genresults.m` for definitions of the input and output variables. With prior approval, your entry may use most MATLAB toolboxes.

[physionet2012.m](#) [[physionet2012.m](#)]

Sample Challenge entry (requires `saps_score.m` below, and the MATLAB statistics toolbox)

[saps_score.m](#) [[saps_score.m](#)]

MATLAB function to calculate SAPS I score

[genresults.m](#) [[genresults.m](#)]

Test framework for MATLAB m-code entries. This code can also calculate (unofficial) scores for set A; to do so, you will need the known outcomes (`Outcomes-a.txt`, above) and `lemeshow.m` (below).

[lemeshow.m](#) [[lemeshow.m](#)]

MATLAB function for calculating unofficial event 2 scores

[Sample-SetA.txt](#) [[Sample-SetA.txt](#)]

Output of `physionet2012.m` for training set A, as collected by `genresults.m`

Scores calculated by `lemeshow.m` may differ slightly from the official scores (calculated using `score.c`, below) due to differences in rounding. Scores calculated by `score.c` will be used to determine the final rankings.

For participants developing entries using C or (with prior approval) another language:

A valid entry written in any language other than m-code must be provided in source form with instructions (a commented Makefile would be ideal) for producing an executable program named `physionet2012` from the source file(s). The executable program must be able to run in this way within the test framework on a 64-bit GNU/Linux platform:

```
physionet2012 <input-file >output-file
```

i.e., reading the contents of *input-file* (a Challenge data file such as [set-b/142675.txt](#) [[set-b/142675.txt](#)]) from its [standard input](#) [[faq.shtml#stdio](#)], and writing its analysis of the input to its [standard output](#) [[faq.shtml#stdio](#)], as a single newline-terminated line in this format:

```
142675,0,0.123
```

where the three fields are the RecordID, the binary prediction, and the risk estimate, as described below.

[physionet2012.c](#) [[physionet2012.c](#)]

C source for a sample Challenge entry (must be compiled together with `saps_score.c` below)

[saps_score.c](#) [[saps_score.c](#)]

C source for a function to calculate SAPS I score

[genresults.sh](#) [[genresults.sh](#)]

Test framework for entries not written in m-code; requires a shell (command-line interpreter) such as `sh` or `bash`. This code can also calculate (unofficial) scores for set A; to do so, you will need the known outcomes (`Outcomes-a.txt`, above) and a compiled version of `score.c` (below).

[score.c](#) [[score.c](#)]

C source for a program for calculating event 1 and event 2 scores

[plotresults.sh](#) [[plotresults.sh](#)]

Shell script for generating a risk decile plot as shown below (requires `plt` [[physiotools/plt](#)] and `ImageMagick` [<http://imagemagick.org/>])

[Outputs-a.txt](#) [[Outputs-a.txt](#)]

Output of `physionet2012.c` for training set A, as collected by `genresults.sh`

[results-a-c.png](#) [[results-a-c.png](#)]

Output of `plotresults.sh` for `physionet2012.c` evaluated on set A

For participants developing entries in R:

Download the sample R entry ([physionet2012.R](#) [[physionet2012.R](#)]) and the first data file from set A ([132539.txt](#) [[set-a/132539.txt](#)]) into your working directory, then test the sample R entry by running it using `Rscript`, like this:

```
Rscript physionet2012.R <132539.txt >output.txt
```

This creates an output file `output.txt`, containing one line:

```
132539,0,0.5
```

The sample R entry doesn't analyze the input; it simply reads it and produces a correctly-formatted output line. Use `physionet2012.R` as a model for your R-code. You can test your entry on set A using `genresults.sh` if you replace this line in it:

```
./physionet2012 <$R >>$OUT
```

with this one:

```
Rscript physionet2012.R <$R >>$OUT
```

CHALLENGE SCORING

As in previous challenges, participants may compete in multiple events:

Event 1

The goal is to predict in-hospital *mortality* with the greatest accuracy using a binary classifier. For this event, your entry must output a prediction (0: survival, or 1: in-hospital death) for each patient.

Event 2

The goal is to predict in-hospital *mortality percentage* (risk) within each decile range with the greatest accuracy. For this event, your entry must output a risk estimate between 0 and 1 for each patient.

Entries must output both predictions and risk estimates, but if you do not wish to compete in one of the two events your entry may output any acceptable values for that event.

Scoring for Event 1 is based on 2 metrics: Sensitivity (**Se**) and positive predictivity(**+P**). We define the numbers of true positives (**TP**), false positives (**FP**), false negatives (**FN**), and true negatives (**TN**) as below:

		Observed	
		Death	Survivor
Predicted	Death	TP	FP
	Survivor	FN	TN

Using these definitions, the 2 metrics and the scoring for Event 1 are given by:

Se = $TP / (TP + FN)$ [the fraction of in-hospital deaths that are predicted]

+P = $TP / (TP + FP)$ [the fraction of correct predictions of in-hospital deaths]

Score1 = $\min(\text{Se}, +P)$ [the minimum of Sensitivity and positive predictivity]

The sample MATLAB entry based on SAPS-1 earns an unofficial Event 1 score of 0.296 on set A, whereas random guessing yields a score of 0.139. A perfect (and almost certainly unattainable) Event 1 score is 1, so there is much room for improvement on the sample entry.

Scoring for Event 2 is based on the [Hosmer-Lemeshow H statistic](#)

(http://en.wikipedia.org/wiki/Hosmer%E2%80%93Lemeshow_test) (a common measure of model calibration), and on the spread, D, of risk estimates.

To calculate the H statistic for a given entry, the in-hospital mortality risks predicted by that entry are first sorted and the corresponding records are binned into deciles designated by $g = 1, 2, 3 \dots 10$. Thus the first

decile ($g = 1$) of the 4000-record set B contains the 400 records with the lowest predicted risk, the second decile contains the next 400 records, etc. The H statistic is then calculated as:

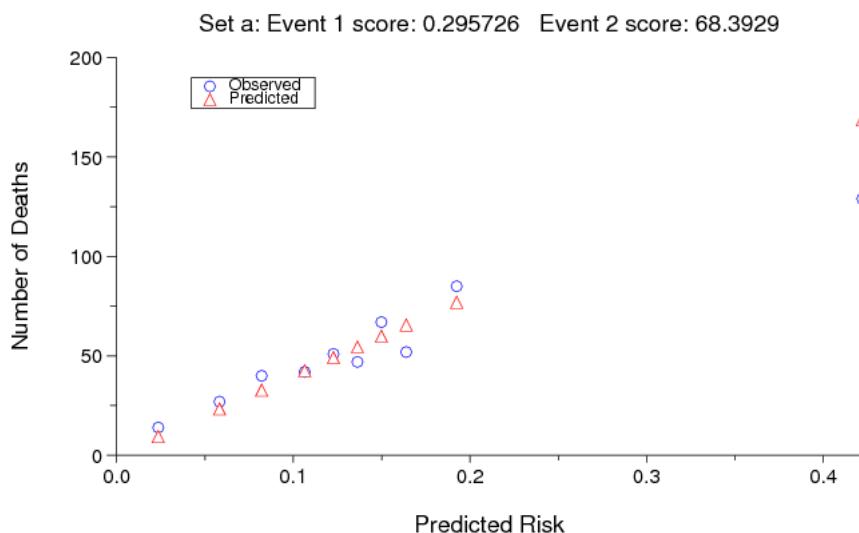
$$H = \sum_{g=1}^{10} \frac{(O_g - E_g)^2}{N_g \pi_g (1 - \pi_g) + 0.001}$$

where for each decile g , O_g is the observed number of in-hospital deaths, E_g is the predicted number of deaths, N_g is the number of records (400), and π_g is the mean estimated risk for records in the decile. This definition is modified as shown with respect to the standard definition of H by the addition of 0.001 in the denominator; we do this to avoid division by zero if π_g is zero or one. This can occur if an entry estimates a risk of zero (or one) for 400 or more records in a set; the effect of doing so, regardless of accuracy, will cause a substantial increase in H (lower values are better), so we strongly recommend *not* allowing your entry to output risk values below about 0.01 or above about 0.99.

To be useful as input to medical decisions for individual patients, risk estimates should accurately reflect individual patient risks, rather than simply the risk for the entire population of patients. For this reason, the event 2 score is also based on the range of risks, specifically on the difference, D, between the mean risk estimates in the top and bottom deciles from the H calculation ($\pi_{10} - \pi_1$). The event 2 score is thus the *range-normalized H statistic*, defined as **H/D**.

Files [lemeshow.m](#) [[lemeshow.m](#)] (in Octave/MATLAB m-code) and [score.c](#) [[score.c](#)] (in C) define functions that calculate **H/D** as defined above.

The sample MATLAB entry based on SAPS-1 achieves an unofficial Event 2 score of 68 on set A, and random guessing scores 9666. As for Event 1, an ideal score (0 in this case) is almost certainly unattainable, but it should be possible to improve on the sample entry. The figure below shows the observed and predicted numbers of deaths for each decile of risk assigned by the sample MATLAB entry for set A.



FREQUENTLY ASKED QUESTIONS ABOUT THE CHALLENGE

CAN I SUBMIT AN ENTRY WRITTEN IN ...?

We would really like to say yes! What will determine our answer is whether it is practical and reasonably efficient for us to evaluate your entry, and you can help us by testing and documenting the procedure needed to do so thoroughly.

If you can show us how to run your entry in an unmodified copy of our [test framework for entries written in C](#) [[genresults.sh](#)], using only compilers, interpreters, libraries, and other standard components that are freely available for GNU/Linux, we encourage you to submit code written in the language of your choice. We will accept entries written in Java, Perl, or R under these conditions, and (if you ask nicely) we will consider adding other languages to this short list.

Please give us as much extra time as possible to review entries not written in C or m-code.

WOULD IT STILL BE POSSIBLE TO PARTICIPATE IF WE SEND YOU CODE THAT IS COMPILED AND TESTED TO RUN ON THE DESTINATION HARDWARE AND OS?

Sorry, no. We do not require participants to submit open-source entries, but we will not test executables that we have not compiled ourselves from source code that we have inspected.

IF I DON'T SUBMIT ALL 5 PRELIMINARY (PHASE 1) ENTRIES, CAN I ADD THE UNUSED ONES TO MY QUOTA OF 5 PHASE 2 ENTRIES?

No. We are trying to encourage both experimentation with multiple approaches and sustained effort. In past Challenges some participants have used their entire allowance of entries before the first deadline, and others have saved their entries until hours before the final deadline. The most successful participants have usually reflected on each set of results, refining their ideas (and not merely their decision thresholds) before submitting the next entry. This approach yields better results, and it also allows us to review your entries and provide scores more rapidly than if we receive a large fraction of them just before the deadlines.

WILL IT BE POSSIBLE TO OUTPERFORM THE SAMPLE ENTRIES BASED ON SAPS-I?

Yes, certainly. SAPS-I uses fewer than half of the variables provided in the Challenge data sets, and only the first half (24 of 48 hours) of those. There is plenty of room for improvement in both events.

WHY ARE THE SCORES FOR THE C AND M-CODE SAMPLE ENTRIES DIFFERENT?

The scores are different because the two sample entries are not exactly the same algorithm. The SAPS calculations are very similar, but the C version makes use of data with time stamp of '24:00' (i.e., exactly 24 hours from the start) and the m-code version stops at '23:59'. This accounts for the very slight difference in the event 1 scores even though both entries use the same decision threshold. The risk estimates (which determine the event 2 scores) are completely different; the m-code entry uses a function that was fitted to the observed distribution of risk in set A, and the C entry uses a lookup table that is not optimized at all.

WHY DIDN'T YOU CHOOSE [MY FAVORITE STATISTIC] AS A SCORING METRIC?

Fund another award and we will consider adding another event!

No single measure can summarize all of the important aspects of performance on the Challenge problem in a single number; that's why we have multiple events. Since scoring metrics inevitably influence how participants design and refine their entries, we have chosen metrics that provide incentives to make *useful* predictions. Many alternatives would reward trivial and clinically irrelevant strategies; for example, one can obtain a high overall accuracy by predicting that all patients will survive, but such a predictor is of no value whatsoever.

WHAT IS THE PRECISE DEADLINE FOR PHASE 1? WHAT HAPPENS IF I CAN'T FIGURE OUT HOW TO SUBMIT A VALID ENTRY BY THEN?

On the off-chance that someone on Baker or Howard Islands in the mid-Pacific, or on a ship at sea between longitudes 172.5° W and 180°, submits an entry just before midnight local time on 25 April, and in order to be fair to everyone else, the deadline for phase 1 was noon GMT on 26 April; at that time, the software that collects entries stopped doing so.

The deadline is one of several that are important if you would like to be eligible for the Challenge awards (see [above](#) [\[#rules-and-dates\]](#)). Anyone who wishes to participate without eligibility for awards is welcome to join in unofficially during Phase 2 (1 June through 25 August).

SUGGESTED READINGS ON ACUITY SCORES AND MORTALITY PREDICTION

APACHE (1981)

Knaus WA, Zimmerman JE, Wagner DP, Draper EA, Lawrence DE. [APACHE - acute physiology and chronic health evaluation: A physiologically based classification system](#)

[\[http://www.ncbi.nlm.nih.gov/pubmed/7261642\]](http://www.ncbi.nlm.nih.gov/pubmed/7261642) . *Critical Care Medicine* 9(8):591-597 (1981 Aug).

SAPS-I (1984)

Le Gall JR, Loirat P, Alperovitch A, Glaser P, Granthil C, Mathieu D, Mercier P, Thomas R, Villers D. A simplified acute physiology score for ICU patients [<http://www.ncbi.nlm.nih.gov/pubmed/6499483>] . *Critical Care Medicine* **12**(11):975-977 (1984 Nov).

MPM-I (1985)

Lemeshow S, Teres D, Pastides H, Avrunin JS, Steingrub JS. A method for predicting survival and mortality of ICU patients using objectively derived weights [<http://www.ncbi.nlm.nih.gov/pubmed/4006490>] . *Critical Care Medicine* **13**(7):519-525 (1985 Jul).

APACHE II (1985)

Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system [<http://www.ncbi.nlm.nih.gov/pubmed/3928249>] . *Critical Care Medicine* **13**(10):818-29 (1985 Oct).

APACHE III (1991)

Knaus WA, Wagner DP, Draper EA, Zimmerman JE, Bergner M, Bastos PG, Sirio CA, Murphy DJ, Lotring T, Damiano A, et al. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults [<http://www.ncbi.nlm.nih.gov/pubmed/1959406>] . *Chest* **100**(6):1619-1636 (1991 Dec).

MPM II (1993)

Lemeshow S, Teres D, Klar J, Avrunin JS, Gehlbach SH, Rapoport J. Mortality Probability Models (MPM II) based on an international cohort of intensive care unit patients [<http://www.ncbi.nlm.nih.gov/pubmed/8230626>] . *JAMA* **270**(20):2478-86 (1993 Nov 24).

SAPS-II (1993)

Le Gall JR, Lemeshow S, Saulnier F. A new simplified acute physiology score (SAPS II) based on a European/North American multicenter study [<http://www.ncbi.nlm.nih.gov/pubmed/8254858>] . *JAMA* **270**(24):2957-2963 (1993 Dec 22-29).

Elixhauser Comorbidity Index (1998)

Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data [<http://www.jstor.org/pss/3766985>] . *Med Care* **36**(1):8-27 (1998 Jan).

SOFA (2001)

Ferreira FL, Bota DP, Bross A, Mélot C, Vincent JL. Serial evaluation of the SOFA score to predict outcome in critically ill patients [<http://www.ncbi.nlm.nih.gov/pubmed/11594901>] . *JAMA* **286**(14):1754-1758 (2001 Oct 10).

SAPS 3 (2005)

Metnitz PGH, Moreno RP, Almeida E, Jordan B, Bauer P, Abizanda Campos R, Iapichino G, Edbrooke D, Capuzzo M, Le Gall JR. SAPS 3—From evaluation of the patient to evaluation of the intensive care unit. Part 1: Objectives, methods and cohort description [<http://www.springerlink.com/content/t78746238246416/fulltext.pdf>] . *Intensive Care Med* **31**:1336-1344 (2005).
Moreno RP, Metnitz PGH, Almeida E, Jordan B, Bauer P, Abizanda Campos R, Iapichino G, Edbrooke D, Capuzzo M, Le Gall JR. SAPS 3—From evaluation of the patient to evaluation of the intensive care unit. Part 2: Development of a prognostic model for hospital mortality at ICU admission [<http://www.springerlink.com/content/k101329015377474/fulltext.pdf>] . *Intensive Care Med* **31**:1345-1355 (2005).

[Event-based method] (2006)

Silva A, Cortez P, Santos MF, Gomes L, Neves J. Mortality assessment in intensive care units via adverse events using artificial neural networks [<http://www.ncbi.nlm.nih.gov/pubmed/16213693>] . *Artificial Intelligence in Medicine* **36**(3):223-234 (2006 Mar).

APACHE IV (2006)

Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients [<http://www.ncbi.nlm.nih.gov/pubmed/16540951>] . *Critical Care Medicine* **34**(5):1297-310 (2006 May).

MPM0-III (2007)

Higgins TL, Teres D, Copes WS, Nathanson BH, Stark M, Kramer AA. Assessing contemporary intensive care unit outcome: an updated Mortality Probability Admission Model (MPM0-III) [<http://www.ncbi.nlm.nih.gov/pubmed/17255863>] . *Critical Care Medicine* **35**(3):827-835 (2007 Mar).

[Real-time method] (2009)

Hug CW, Szolovits P. ICU Acuity: Real-time models versus daily models [<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2815497/>] . *AMIA Annual Symposium Proceedings*, 260-264 (2009).

Questions and Comments [[/contact-us.shtml](#)]

If you would like help understanding, using, or downloading content, please see our Frequently Asked Questions [[/faq.shtml](#)].

If you have any comments, feedback, or particular questions regarding this page, please send

PhysioNet is supported by the National Institute of General Medical Sciences (NIGMS) [<http://www.nigms.nih.gov/>] and the National Institute of Biomedical Imaging and Bioengineering (NIBIB) [<https://www.nibib.nih.gov/>] under NIH grant number 2R01GM104987-09.

them to the webmaster [[contact-us.shtml](#)].

Comments and issues can also be raised on PhysioNet's [GitHub](#) [<https://github.com/MIT-LCP/physionet>] page.

Updated Wednesday, 31 August
2016 at 17:51 EDT