Details of data de-identification procedures

Variables

Only variables required to conduct the analyses included in the paper were kept.

Hospital study identification number

The number used to identify individual hospitals in the original dataset was replaced with a random number. No key has been retained to allow linkage of new and old hospital identification numbers.

Dates

Dates were shifted into the future using a random offset for each patient, preserving time intervals. To still allow for temporal splitting of the dataset a sequential number was generated based on the original date of arrival and stored as the variable seqn.

Times

Times were retained as they were in the original dataset.

Random deletion of 1% of observations

1% of the original observations were deleted. This was done to prevent identification of patients based on the seqn variable described under Dates above.

Age

The age variable was recoded from a continuous variable to an ordinal variable so that observations with an age > 89 years are grouped together.