HiRID Pre-processing

The steps to create the HiRID external validation datasets are outlined below. Tables 1-12 are included to aid explicability. These do not contain actual HiRID ICU data.

1) Found a broad range of synonyms for the six clinical variables (see Table 1). Our medical advisor and co-author (Prof Malcolm Sim) defined an acceptable range of values for each clinical variable, as shown in Table 2.

Variable name (Glasgow ICU format)	Synonyms
FiO_2	Fraction of Inspired Oxygen
SpO_2	Pulse Oximetry, Oxygen Saturation, Peripheral oxygen
Mean Arterial Pressure	MAP
Heart Rate	-
Adrenaline	Epinephrine
Noradrenaline	Norepinephrine

Table 1. Common synonyms for the six clinical variables.

Variable	Acceptable Range	Units
FiO_2	0.21 - 1.0	decimal fraction
SpO_2	0 - 100	%
Mean Arterial Pressure	0 - 200	mmHg
Heart Rate	0 - 300	/min
Adrenaline	0 - 10	mg/hr
Noradrenaline	0 - 10	mg/hr

Table 2. Units and acceptable ranges of variables present in the QEUH dataset.

2) Using HiRID documentation, found all relevant variables to include in data extraction. See Table 3.

Variable	HiRID name	HiRID units	HiRID ID
Adrenaline	Adrenalin	20 µg/ml Perfusor	1000649
	Adrenalin	100 µg/ml Perfusor	1000650
Noradrenaline	Noradrenalin	100 µg/ml Perfusor	1000656
	Noradrenalin	20 µg/ml Perfusor	1000657
FiO_2	Inspired oxygen concentration	%	2010
SpO_2	Peripheral oxygen saturation	%	4000
	Peripheral oxygen saturation	%	8280
Mean Arterial Pressure	Invasive mean arterial pressure	mmHg	110
	Non-invasive mean arterial pressure	mmHg	610
Heart Rate	Heart Rate	/min	200

Table 3. Description of variables used in creating the HiRID external validation datasets.

3)

i. Filtered 'Pharma_records' table for all records with readings for continuous Adrenaline and Noradrenaline (see Table 3 for variable IDs).

ii. Grouped each drug variable reading to the closest 15-minute timepoint.
- E.g., Adrenaline reading taken at '2100-01-06 08:34:15' was categorised as '2100-01-06 08:30:00'

iii. Calculated rate of drug administration for Adrenaline/Noradrenaline readings:
- Created 'min_time' field = minimum (earliest) drug administration time within given 15-minute period
- Created 'max_time' field = maximum (latest) drug administration time within given 15-minute period
- Created 'cumulativedose(min_time)' field = cumulative drug dose at 'min_time'
- Created 'cumulativedose(max_time)' field = cumulative drug dose at 'max_time'
- Drug amount administered within each 15-minute period:
  - cumulativedose(max_time) - cumulativedose(min_time) = cumulativedose_diff
- t(diff) = max_time – min_time
- Drug Rate = cumulativedose_diff/ t(diff) (units = micrograms/min)
- Converted rate to into mg/hr (as per training data drug units)

iv. Reformatted drugs table to one record per unique (patientid + timepoint), as per training data structure. See Table 4.

patientid	timepoint	Adrenaline	Noradrenaline
01	2198-01-03 18:15:00	0.0	0.2
01	2198-01-03 19:00:00	0.0	0.0
02	2198-01-03 19:45:00	0.1	0.0

Table 4. Example of table obtained after Step 3iv. Primary Key = (patientid + timepoint).

4)

i. Filtered 'Observations' table for readings of FiO_2/SpO_2/MAP/HR (see Table 3 for variable IDs).
- Note: as this 'Observations' table was too large to process using the facilities available, it was split into ten individual tables ('Observations_1' to 'Observations_10'), each table filtered accordingly, then the results concatenated.

ii. Grouped each physiological variable reading to the closest 15-minute timepoint.
- e.g., FiO_2 reading taken at '2100-01-06 09:56:15' was categorised as '2100-01-06 10:00:00'

iii. Filtered for all unique ICU admission timepoints, which contain readings for <u>all four</u> physiological parameters. This is the pool of relevant (patientid + timepoint) instances, where 'patientid' represents a unique ICU admission and 'timepoint' is the closest 15-minute timepoint to the parameter readings. See Table 5.

patientid	timepoint	variableid	pharmaid
01	2198-01-03 18:15:00	[2010 , 200, 4000, 110]	NaN
01	2198-01-03 19:00:00	[110, 4000, 2010, 200]	NaN
02	2198-01-03 19:45:00	[110, 8280, 2010, 200]	1000657

Table 5. Example of table obtained after Step 4iii. Primary Key = (patientid + timepoint)

iv. For the (patientid + timepoint) combinations found in Step 4iii, found all records of physiological parameter readings.
- Replaced variable ID numbers by name and listed variable values per unique (patientid + timepoint + variable). See Table 6.

patientid	timepoint	variable	value
01	2198-01-03 18:15:00	FiO2	[60.1, 59.5]
01	2198-01-03 18:15:00	HR	[97.0, 98.0, 97.0]
01	2198-01-03 18:15:00	MAP	[87.0, 82.0, 79.0, 85.0]

Table 6. Example of table obtained after Step 4iv. Primary Key = (patientid + timepoint + variable).

v. Removed all values outside acceptable range, as specified in Table 2.

vi. Extracted minimum, maximum and average parameter values per unique (patientid + timepoint+ variable).

vii. Removed instances where the % variation of values within each 15-minute period was greater than defined cut-off (see Tables 7 and 8). As advised by the study's medical advisor, a large variation should not be observed within a such a short time period and is indicative of a data error.
- % variation = ((max_val – min_val) / min_val) x 100

variable	% variation cut-off
FiO_2	5%
SpO_2	10%
MAP	25%
HR	25%

Table 7. Defined % variation cut-off

patientid	timepoint	variable	value	min_val	max_val	avg_val	variation(%)
01	2198-01-03 18:15:00	FiO2	[60.1, 59.5]	59.5	60.1	59.8	1.01%
01	2198-01-03 18:15:00	HR	[97.0, 98.0, 97.0]	97.0	98.0	97.3	1.03%
01	2198-01-03 18:15:00	MAP	[87.0, 82.0, 79.0, 85.0]	79.0	87.0	83.3	10.13%

Table 8. Example of table obtained after Step 4vii. Primary Key = (patientid + timepoint + variable).

viii. Converted FiO_2 readings from % to decimal fraction (as per training data units).

ix. Reformatted table to one record per unique (patientid + timepoint), as per training data structure, using average parameter values. See Table 9.

patientid	timepoint	FiO2	SpO2	MAP	HR
01	2198-01-03 18:15:00	0.60	100.0	83.3	97.3
01	2198-01-03 19:00:00	0.55	96.0	87.0	96.5
02	2198-01-03 19:45:00	0.60	90.0	78.0	98.0

Table 9. Example of table obtained after Step 4ix. Primary Key = (patientid + timepoint).

5) Created final HiRID validation dataset by merging the physiological parameter and drug rate tables. See Table 10.

patientid	timepoint	Adrenaline	Noradrenaline	FiO2	SpO2	MAP	HR
01	2198-01-03 18:15:00	0.0	0.2	0.60	100.0	83.3	97.3
01	2198-01-03 19:00:00	0.0	0.0	0.55	96.0	87.0	96.5
02	2198-01-03 19:45:00	0.1	0.0	0.60	90.0	78.0	98.0

Table 10. Example of final dataset obtained after Step 5. Primary Key = (patientid + timepoint).

6) Found discharge/death times for each unique ICU admission:

- HiRID does not store discharge/death times, therefore the time for the last record of Heart Rate was used as discharge time (if discharge status = alive) or death time (if discharge status = dead). This approach was recommended by the HiRID data owner, Martin Faltys. The discharge status, as well as age, sex and admission time were found in the 'General' table.
- Last recorded time for Heart Rate was found for each unique ICU admission across 'Observations_1' to 'Observations_10'.
- Columns for last hours before discharge/death time (1hr, 2hrs, 3hrs, 4hrs, 5hrs) were added (see Table 11). These columns were needed to create the HIRID external validation datasets.

patientid	admission time	discharge _status	d_time*	d_time – 1hr	d_time – 2hrs	d_time – 3hrs	d_time – 4hrs	d_time – 5hrs
01	2198-01-01 15:15:00	alive	2198-01-10 18:33:17	2198-01-10 17:33:17	2198-01-10 16:33:17	2198-01-10 15:33:17	2198-01-10 14:33:17	2198-01-10 13:33:17
02	2197-12-30 19:20:00	dead	2198-01-04 20:12:43	2198-01-04 19:12:43	2198-01-04 18:12:43	2198-01-04 17:12:43	2198-01-04 16:12:43	2198-01-04 15:12:43
03	2198-01-03 04:55:00	alive	2198-01-06 15:37:10	2198-01-06 14:37:10	2198-01-06 13:37:10	2198-01-06 12:37:10	2198-01-06 11:37:10	2198-01-06 10:37:10

Table 11. Example of table obtained after Step 6.
*d_time = discharge time (if discharge_status = alive) or death time (if discharge_status = dead)

7) Using the tables created in steps 5 and 6, created individual tables of readings for the following time periods before discharge/death: a) 1hr, b) 2hrs, c) 3hrs, d) 4hrs, e) 5hrs.

- E.g., Filtered Table 11 for all readings where (d_time – 1hr) < timepoint < (d_time) to create table of readings within 1hr before discharge/death.
-

patientid	timepoint	Adrenaline	Noradrenaline	FiO2	SpO2	MAP	HR	Discharge/death time	Discharge status
01	2198-01-03 18:15:00	0.0	0.2	0.60	100.0	83.3	97.3	2198-01-03 18:42:12	alive
01	2198-01-03 19:00:00	0.0	0.0	0.55	96.0	87.0	96.5	2198-01-03 23:54:10	dead
02	2198-01-03 19:45:00	0.1	0.0	0.60	90.0	78.0	98.0	2198-01-03 20:05:15	dead

Table 12. Example HiRID external validation (static) dataset.