Handover document to DHSC – Urgent and Emergency Care Modelling

Overview of modelling:

The simulation models the flow of patients over a 15-day period through an emergency department. This 15-day period can be changed, although we do recommend running for a reasonable length of time to ensure sufficient counts in the model output metrics.

The model is a discrete-event simulation that models each patient moving and interacting with different aspects of care. The simulation is built with SimPy, which is a process-based discrete-event simulation framework based on Python.

The level of detail includes:

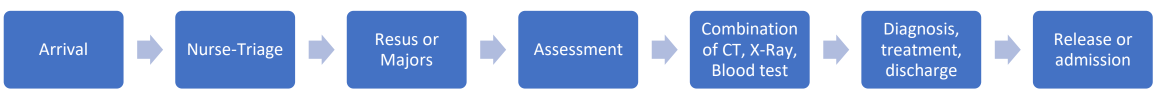
- Events happening in different ED rooms (Resus, Majors, Minors, Waiting room, SDEC, etc.). These are modelled as *containers* in SimPy, which help model the production and consumption of a homogeneous resource.
- Events occurring with different clinicians ('Generic doctors', triage nurses, RATs, SDEC clinicians). Clinicians are modelled as *preemptive resources* in SimPy. This means that some entities (= patients) will prevent other entities with lower priority from using the resource and kick them out. For example, if a higher acuity patient needs a clinician that is busy with a lower acuity patient, there may be an interruption.
- Events happening across the patient care process (arrival, registration, assessment, diagnostic tests, treatment, discharge, admission, etc.)

Patients are modelled differently depending upon their acuity, since it is known that different acuity patients pass through different ED areas and undergo different frequencies of activities, and that processes take a different amount of time depending on acuity. The model incorporates the idea that doctors can be interrupted (depending on patient acuity), beds and seats can be occupied, and clinicians can be unavailable because they are busy elsewhere.

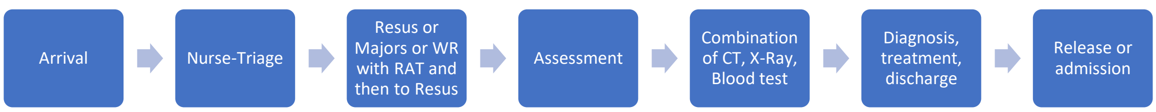
Approximate process maps:

Below are the high-level process maps modelled across different acuities.

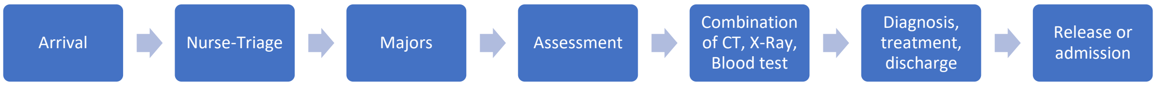
Acuity 1-2 - Ambulance:



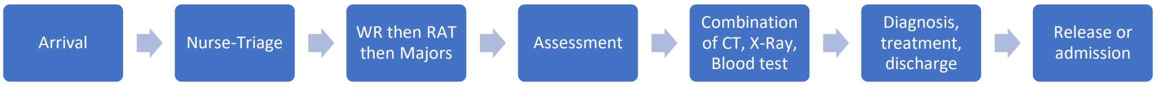
Acuity 1-2 – Non-Ambulance:



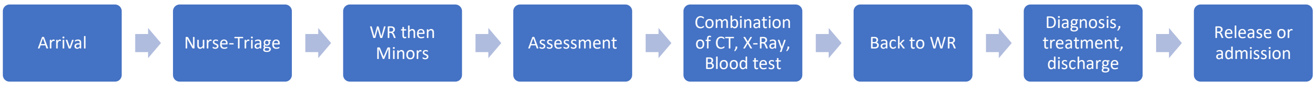
Acuity 3 - Ambulance:



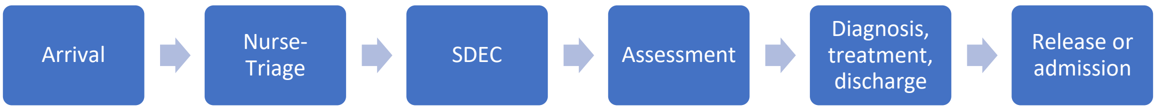
Acuity 3 – Non-Ambulance:



Acuity 4 -5 (Non-SDEC):



Acuity 4 only (SDEC):



Main assumptions:

- We have modelled a pre-defined process map relevant to PAH, as highlighted above. This will be different between hospitals (urban vs rural, small vs large, etc.).
- The parameters used in populating the model have been informed, based on the data availability and where possible stakeholder informed opinion. Inevitably, we have had to make a number of judgement calls for creating the baseline model. For a full list of the parameter values and the underlying assumptions, please see the 'UECsim_parameters_assumptions_v1.xlsx' spreadsheet, found within the GitHub 'docs' folder.
- Acuity not only determines the routes that patients go through in ED but also the time of many of the processes: time for assessment, treatment and discharge vary depending on acuity. In general, lower acuities have shorter times with clinicians.
- The default model currently includes an SDEC – to turn the SDEC off completely, please set the "sdec_Percentage" parameter to 0 in the config file. All the remaining sdec-related parameters will then be ignored.
- The model does not include a number of elements within UEC that may influence patient flow. These include:

    - Human efficiency factors, such as, rerouting patients from the conventional pathway when the ED is full or creating additional capacity when urgently needed.
    - Prioritisation and care of specific patients within each acuity, depending on their needs.
    - We are assuming fast flow through an SDEC i.e. with high number of clinicians, beds, and with no diagnostics. SDEC patients are assumed currently to be acuity 4 only. Otherwise, the times to assess/treat/discharge patients in the SDEC follow the same values as in the main ED (for acuity 4).
    - The model considers different routes within the ED for ambulance and non-ambulance patients. Nothing prior to arrival is currently accounted for in the modelling (e.g. ambulance handover times). The model does not differentiate ambulance vs non-ambulance patients in the model outputs.

Model parameters, including inputs and outputs:

See 'Assumptions and data file.xlsx' file in GitHub.

Code information:

- How is the code organised?:
    - All of the code is on GitHub. The main folder contains:
        - Main scripts (scenario_tool.py, simulation_flask.py, simulationFlaskED.py, Monte-Carlo.py). The main modelling script is simulationFlaskED.py.
        - Input files (config.json, config_EDsizes.json, config_Edtypes.json). Please note that config_EDsizes.json and config_EDtypes.json are not currently part of the default run with the main scripts.
        - Process_Mining folder – Contains various process mining example HTML files, and process mining R code. Process mining is a data-driven method of visualising processes, in this case, the flow of events across patients in our ED model. This can be useful when validating the modelled pathway to ensure that events are happening in the intended order.
        - Outputs folder – Contains simulation outputs.
        - Utils folder – Contains helper functions.
        - Legacy_Model_Documents – Shows the original process map and config text file.
- How to run the sim?
    - There is a requirements.txt in the root folder that lists all the packages and versions needed to run the project. Please follow standard instructions to install these (e.g., pip install -r requirements.txt) – it is recommended to use a virtual environment.
    - The simulation can be run using either scenario_tool.py (running without the flask user interface, e.g., python scenario_tool.py) or simulation_flask.py (running with the flask user interface e.g., python simulation_flask.py and then clicking on the http link to open the UI. Once the UI is open, please click submit). Please note, the simulation will take ca 5 minutes to run in the current settings and for a simulation time of 15 days.

Recommendations and future improvements:

- Refine model parameters, particularly stakeholder informed estimates, and add in parameter uncertainty.
- Build into the model the effect of human factors.
- Validate the model for other hospitals to prevent overfitting.
- Validate the modelled process map.
- Monte-Carlo the simulation – there is a substantial degree of variability when running simulations with the same parameters more than once so reflecting this uncertainty is essential. We have made a basic Monte-Carlo script which runs 'scenario_tool.py' over a set number of iterations, outputting statistics on 4 hr admit-discharge percentages. We would suggest using this as a basis for Monte-Carlo simulations.
- Refine the definition of 'generic doctor' used in the model. Currently, a 'generic doctor' is defined as a clinician who performs an assessment or diagnosis, treatment or discharge. Generic doctors are not SDEC, nurse-triage or RAT clinicians. Regardless, we note that the number of generic doctors parameter value has uncertainty, and future modelling should refine the 'generic doctor' definition.
- Incorporate the effect of work shifts into the model. The current setup assumes that the same number of staff members are present at all times. To calculate that number for PAH we are simply using the maximum number of clinicians working at any one time of the day. Adding different numbers of resources depending on the time of the day would make a more realistic framework.
- Perform scenario modelling recognising that changing one parameter can impact a number of other model input parameters.

Some notes/signposts

1. Time assess/treat/discharge is the time of the process itself, rather than the time from arrival to having finished an event (which would include waiting time, etc.). It is challenging to get times of processes from PAH/ECDS etc; time data is generally available between events, for example, the time between being seen by a triage nurse and being treated but not from the beginning to end of treatment.

2. Uncertainty (standard deviation parameters) – Currently we have not implemented this for many of the time parameters, due to the data quality or lack of data. In many cases, uncertainty will not be particularly important (e.g., for time assess) because the average times are very short. However, we note that uncertainty would be important for some of the larger value time parameters, including, time to admit/get a bed, treatment time, etc.

3. Currently time to get a bed isn't influenced by bed occupancy. It would be important to get more granular data on time to get a bed (or maybe just correct the current median time by a factor depending on occupancy?).

4. If the simulation fails, the most common error shows (corridor_json not found) – this is more or less a generic message (not necessarily anything wrong with corridor_json). The error occurs as a result of the large 'try' statement within the run function in simulationFlaskED.py failing. First port of call – remove the current simulation_flask.js, shut down Python and rerun. Then try debugging.

5. Independence of diagnostic tests. Currently patients can undergo any of CT, Xray and bloods tests. For each test, a yes/no answer for each patient is drawn depending on the probability of each acuity getting that particular test. So, a patient can get all three tests, or none, or any combination. This is a big assumption, considering that diagnostics may be correlated, for example, people who have CT scans may not have X-rays. Data is required to explore this independence assumption.

6. See UECsim_parameters_assumptions for logic behind parameter assumptions. Note there is variability on how informed the parameters are, from some being directly taken from hospital-level data, to some being informed by PAH clinicians, to others being guided by intuition.

7. Generally – real-world data does not necessarily follow a normal distribution, so asking for means/stdevs when requesting data from trusts may lead to biased estimates. At a minimum, ask for medians as well – ideally get distribution data from ECDS or the trust. Preferably, future modelling should setup IG agreements early on in order to work with patient-level data so distributions for all relevant parameters can be created. When fitting distributions, there will be a risk of overfitting the model. To overcome this, the model needs to be validated across hospitals.

8. Note that the PAH process map as it stands does not fully reflect the reality and what exactly is coded in the model. Major differences include the representation of mental health patients (PAH has no MH department), increased diagnostic options not limited by acuity, neglect of modelling the FAU for complexity, acuity 1-2 patients going to both majors and resus and not solely resus, acuity 4 and not also acuity 5 patients going to SDEC, requirement that all acuity 4-5 patients receive diagnosis, treatment and discharge unless they leave due to patience being exceeded, etc.

9. Generally, specific acuity patients undergo processes in different ways. For example, acuity 1-2 patients use clinicians who are not interrupted to the same extent, as higher acuity 4-5 patients.

10. The model considers four types of clinician: ‘Generic’, ‘SDEC’, ‘Nurse triage’, and ‘RAT’. The model outputs shown on the pdf exclude SDEC patients in reporting (so reported times are for the main ED only, not SDEC).

11. Diagnosis, treatment, and discharge (DTD) are treated as one combined event.

12. For assessment by any clinician and DTD a doctor is requested, the event is performed, and the doctor is released. In this way, the availability of a clinician can be a limiting factor in flow which reflects reality.

13. When a patient moves to a room, a bed has to be available else there is a time delay. In this way, flow is limited by bed availability which reflects reality.

14. ED flow depends strongly on numbers of triage nurses and the associated assessment time, since all patients undergo nurse-triaging and PAH has low numbers of NTs. This is an example where the effect of human factors could be important, for example, in generic clinicians potentially offering additional triaging capability, or modelling nurse-triage by-passing etc.

15. Mental health patients are modelled for acuity 3 only. Mental health patients follow the same patient flow process as for non-mental health patients, except that they are treated as having increased waiting time between assessment and DTD.

16. A small script to run multiple replicates of the same scenario can be found in Monte_Carlo.py. This can be edited to increase the number of replicates and will plot the 4hr metric across the replicates as it they are running (note that you will need to close the plot window for the next run to continue). This plot can be used to diagnose how stable the metric is and whether fewer/more replicates are needed.

17. Process_mining.R can be useful for validating the model pathway. For example, by:

    a. Visualising the model pathway in terms of patient numbers who follow different routes in ED and who undergo different activities.
    b. Visualising the model pathway in terms of average time spent undergoing different activities. This is potentially useful for identifying time-limiting bottlenecks.
    c. Visualising the distribution of different activities, for example, assessment time, admission time, etc.
    d. Examining corelations between activities, as illustrated by process matrices.
    e. Identifying inconsistencies/errors in modelling, for example, from specific acuity patients undergoing unexpected activities or from activities occurring in an unplanned order.

Suggestions for additional modelling:

<u>Parametrisation of mean admission times (= time between decision to admit and admission)</u>

The current implementation has a median time to admission over a year for PAH. This is a fixed parameter. In reality, the time could vary substantially over one year.

The previous model implementation had these values (units in minutes):

```
"Target_Parameters_Default_Values": {
    "W": {
        "mean_target": 120,
        "std_target": 60.0
    },
    "S": {      josephlillington, 3 weeks ago
        "mean_target": 30,
        "std_target": 15.0
    }
},
```

These are based on whether the hospital has a strong/weak culture i.e. their propensity to admit patients: this determines whether a hospital can clear out beds quickly and improve flow.

In order to look at the impact of variation in mean_target on flow, we propose:

- Get data (directly from PAH on the level of variation or alternatively ECDS has the time between the decision to time to admit and arrival, and so we could approximate mean_target from this) for a year and analyse the variation.
- This might include determining the shape of the relationship with real data and looking at the spread of variation (e.g. by making the parameter follow a normal distribution with mean and stdev).

<u>Parametrisation of mean bed times (= times to get a bed).</u>

The current implementation has a median time to get a bed over a year for PAH. This time is additional to the admit time mentioned above. Mean_bed is a fixed parameter. However, we assume in reality that bed occupancy will have an effect on this time to get a bed.

The previous implementation of the model used these values:

```
"Bed_Occupancy_Default_Values": {
    "80": {
        "mean_bed": 30,
        "std_bed": 30.0
    },
    "85": {
        "mean_bed": 120,
        "std_bed": 60.0
    },
    "90": {
        "mean_bed": 480,
        "std_bed": 120.0
    },
    "95": {
        "mean_bed": 600,
        "std_bed": 240.0
    }
},
```

We propose, either:

- Get data (from ECDS if there is G&A occupancy data and time to get a bed; or directly from PAH) for a year and analyse the variation
- confirm if there is a correlation between occupancy and time to get a bed.
- add an occupancy parameter – depending on the shape of the relationship, it could be similar to the previous implementation above (but with PAH-adjusted parameters depending on different levels of occupancy)

Or:

- If there not correlation between occupancy and time to get a bed, that we analyse the variation in time to get a bed, only.

Mental health patient effect on flow:

Mental health patients are modelled as acuity 3 patients, only.

For MH patients who arrive by ambulance, these patients go to majors and have assessment, diagnosis, treatment, and discharge (DTD), as other patients. They are modelled as having no diagnostics. Between assessment and DTD, they have a long waiting time, currently set to be 2 hours.

For MH patients who do not arrive by ambulance, they do not go to SDEC. Instead, they go the waiting room and then to majors. Again, these patients who go to majors have assessment, diagnosis, treatment, and discharge (DTD), as other patients. They are again modelled as having no diagnostics. Between assessment and DTD, they have a long waiting time, currently set to be 2 hours.

We propose:

1) Validating the pathway with PAH:
    a. Is flow only affected for acuity 3 MH patients?
    b. Do MH patients normally not receive diagnostics (CT, X-Ray, Bloods).
    c. When is the additional delay time present – is it between assessment and DTD as modelled, or is the time-limiting event when waiting to be discharged from ED or during admission (either time to admit or when waiting for time to get a bed).
2) Better estimating the wait time modelled (either from PAH opinion or their data; or by estimating the delay from ECDS with the available data fields). We might then introduce uncertainty on the parameter value, as well, as validate our average time modelled.