Data and Code Disclosure

Belonging to: Modeling the Impact of Community First Responders

September 6, 2023

1 Data Sources

1.1 St John and GoodSAM Data

Hato Hone St John (in our paper: St John Ambulance Services, SJAS) is the leading provider of ambulance services in New Zealand. In addition to dispatching traditional ambulances, SJAS also operates a community first responder (CFR) system for out-of-hospital cardiac arrest (OHCA) patients through the GoodSAM app. The SJAS has kindly provided us with the time and location of all OHCAs in New Zealand from 2013 to 2020, as well as information on all CFR alerts sent out through the GoodSAM app since late 2017.

1.2 2013 Census Data

New Zealand collects census data every five years that counts the people and dwellings in New Zealand. The 2013 Census was held on March 5, 2013, and the data is publicly available on StatsNZ (`https://nzdotstat.stats.govt.nz/wbos/Index.aspx`). From this, we use population counts per area unit.

1.3 2013 Census Main Means of Travel to Work Data

The main means of travel to work data is part of the census data that collects information on the commute of workers aged 15 years and older between their home and their place of employment. This data is available upon request from StatsNZ. For convenience, we have also stored this raw data in a sheet titled `2013_residents_by_workplace` in the file `data/OutputFiles/MB AU data Auckland.xlsx`.

1.4 Ambulance Travel Time Data

The ambulance travel times data were obtained by Sam Ridler. Sam made estimates by using his ambulance simulation model JEMSS [1] and recommended multiplying them by 0.7 for lights and sirens. The result (after multiplication) is what is stored in `travelTimesAmbu-base2loc.txt`: it contains travel times from each St John base (in the order we sent them to Sam) to each area unit. This file is used as input to our optimization model.

[1] https://github.com/uoa-ems-research

2 Data Dictionary

2.1 Raw data: GoodSAM

GoodSAM data includes information on time and locations of all OHCAs from 2013 to 2020, and CFR alerts after the implementation of GoodSAM app in 2017. It has the following data variables:

- `responderId`: The ID for the CFR that is alerted.
- `Master_Incident_Number`: ID for an OHCA incidence.
- `alertId`: ID for the alert sent out.
- `MB_IncidentAdd`: Number of the meshblock within which the OHCA occurs.
- `CAD_Lat_with_fuzziness`: Latitude of the location of the ambulance responding.
- `CAD_Long_with_fuzziness`: Longtitude of the location of the ambulance responding.
- `respondersLatitude_with_fuzziness`: Latitude of the location of the CFR alerted.
- `respondersLongitude_with_fuzziness`: Longtitude of the location of the CFR alerted.
- `alertLatitude_with_fuzziness`: Latitude of the location of OHCA.
- `alertLongitude_with_fuzziness`: Longtitude of the location of OHCA.
- `responderCallStatus`: Status of the CFR alerted. The options are Accepted, Accepted-but-dropped-later, Rejected and Not Seen.
- `meterDistanceFromTheIncident`: The distance from the alerted CFR to patient in meter.
- `notificationTimeFormattedDate`: Time when alert is sent out to the CFR.
- `acceptanceTimeFormattedDate`: Time when the CFR presses the "Accept" button if he/she decides to accept the alert.
- `rejectionTimeFormattedDate`: Time when the CFR presses the "Reject" button if he/she decides to reject the alert.
- `droppedTimeFormattedDate`: If the CFR accepts the alert but drops it later, this records the time when he/she decides to drop the alert.
- `onSceneTimeFormattedDate`: Time when GPS detects that the CFR has arrived on-scene.
- `withPatientTimeFormattedDate`: Time when the CFR presses the button manually on the app to indicate that they are with the patient.
- `DHB`: District Health Board.
- `Time_PhonePickUp`: Time when the phone is picked up by emergency call center personnel.
- `Time_CallEnteredQueue`: Time when this OHCA call has entered the queue for ambulance dispatch.

- `Time_First_Unit_Assigned`: Time when an ambulance is assigned to respond to this OHCA call.
- `Time_First_Unit_Arrived`: Time when the ambulance has arrived on-scene.
- `defibObtainedTimeFormattedDate`: Time when defibrillation is administered to the patient.

2.2 Meshblock Data

The file `data/inputFiles/Meshblock data.txt` provides population and area information on all meshblocks in Auckland, and also the area unit that each meshblock belongs to. This data is also processed from the 2013 census day using Excel. We use this data to match each meshblock to the appropriate area unit.

- `Meshblock_(2013_areas)`: Meshblock code, i.e. 'MB_' + meshblock number.
- `MeshblockN`: Meshblock number.
- `TLA`: Territorial Local Authorities.
- `Local_board`: Local board of each meshblock.
- `AreaUnit`: The area unit that each meshblock belongs to.
- `Pop_usual`: Population of each meshblock during the day.
- `Pop_night`: Population of each meshblock during the night.

2.3 Population Data

We extracted the relevant parts of the 2013 Census Data to a file `data/inputFiles/population.txt` which provides the population and surface area of each area unit in Auckland. It has the following variables

- `Area Unit Code`: Code for each area unit.
- `Population day`: Population of each area unit during the day.
- `LandAreaSQKM`: Area of each area unit in square kilometers.

2.4 Variable names

Python	Paper	Explanation
BusyFraction	q	Fraction of time an ambulance is occupied (0.44)
nVolunteers	n	number of volunteers (raw count)
lambda	λ	demand (OHCA incidence estimates)
walkingSpeed	speed	6 km/h
x		vector that prescribes how much to recruitment from each area unit (sums to 1)
nu	ν	effective volunteer mass (after some recruited people have gone away to work)
threshold	τ	Response time threshold. Python counts from collapse; paper counts from EMS call arrival (which is one minute later)

3 Analysis process

3.1 Geographical region

The scope of our case study is defined via area units in our Local Board selection. This is found by matching the column `Local_board` (stored per meshblock) to the following list: Albert-Eden Local Board Area, Devonport-Takapuna Local Board Area, Henderson-Massey Local Board Area, Howick Local Board Area, Kaipatiki Local Board Area, Mangere-Otahuhu Local Board Area, Manurewa Local Board Area, Maungakiekie-Tamaki Local Board Area, Orakei Local Board Area, Otara-Papatoetoe Local Board Area, Papakura Local Board Area, Puketapapa Local Board Area, Waitemata Local Board Area, Whau Local Board Area. All area units that we included this way, are for convenience also indicated in column F of sheet *Summary Area Unit* of `data/OutputFiles/MB AU data Auckland.xlsx`

3.2 Data processing and cleaning

3.2.1 Estimating α

In this paper, we use GoodSAM data to estimate acceptance probability α. For this, we considered alerts sent on December 2017 or later, for which the `MB_IncidentAdd` falls within the Local Boards of our case study. We estimate α by counting all alerts with ResponderCallStatus Accepted or Accepted-but-dropped-later, and dividing by the total number of alerts.

3.2.2 Estimate OHCA Rates

To estimate incident rates according to the procedure in Dicker et al. (2019), we processed the 2013 census data `https://www3.stats.govt.nz/meshblock/2013/excel/2013_mb_dataset_Auckland_Region.zip?_ga=2.253476425.1370198847.1683544082-1530907325.1683544060)`. This url gives a folder with several files, the only one we used is called 2013-mb-dataset-Auckland-Region-individual-part-1.xlsm. There, the relevant sheet is: *2 Area unit* and the relevant columns are: *O* and *BK* and *FR* and *FS*. These provide raw counts of the number of males, 65+, Maori and Pacific. For each area unit that is in scope of our case study, we converted these to percentages by dividing the raw counts by the `Total people` for each area unit. The result was stored in a file `data/inputFiles/Input regression incident rate.xlsx`. This file is read in by `estimate_incident.py`.

- `Area Unit`: Area unit code.
- `Proportion Maori`: Percentage of the population within each area unit that is Maori.
- `Proportion Pacific`: Percentage of the population within each area unit that is Pacific.
- `Proportion 65`: Percentage of the population within each area unit that is aged 65 years or older.
- `Proportion Male`: Percentage of the population within each area unit that is male.

3.2.3 Empirical OHCA rates

The empirical incidence rates (visualized in Appendix 5) are obtained from the GoodSAM data. We did not restrict the dates to any specific time window. We compute empirical OHCA rates by counting the incidents per meshblock (using column MB_IncidentAdd) and aggregating over meshblocks that belong to the same area unit. To know which meshblock belongs to which area unit and which local board, see Section 3.1. We used these empirical counts of three area units, namely 514102 (CBD), 514103 (CBD) an 524200 (airport) to replace the model estimates from Dicker et al. (2019).

3.2.4 Population

We aggregated population counts Pop_usual on meshblock level from the 2013 census data. There are some area units that are only partially included in our defined Auckland region. For those, populations are calculated only using the meshblocks that are within our defined region (local boards). In addition, we have excluded all meshblocks with 0 population or 0 area. This data processing is visible in a sheet titled Area unit attributes in the file data/OutputFiles/MB AU data Auckland.xlsx. From this result, we copied the relevant columns (Area Unit Code, Population day and LandAreaSQKM) and pasted them to a file data/inputFiles/population.txt, which is read in by our optimization code.

3.2.5 Profile

We estimated a profile by processing the *2013 census* and *2013 census main means of travel to work* data. The workplaces in the raw data of *2013 Census Main Means of Travel to Work Data* contained two columns that required further attention: 'Auckland Unspecified' and 'New Zealand Unspecified'. We distributed these across the area units according to the same ratio as seen for people with a specified workplace. Per area unit, we calculated the part of the population that is 15 years or older in sheet Age by area unit column *C*. The profiles are then created by counting the part of the 15+ population for which a commute was provided in the *2013 census main means of travel to work* as workers, and counting the remainder of the 15+ population as non-workers. The result is a 287×288 profile matrix, stored in a sheet titled Profiles in the file data/OutputFiles/MB AU data Auckland.xlsx Each row, labeled with an area unit code, represents the profile of the corresponding area unit. Each column represents the percentage of people from each area unit that work in the labeled area unit. The last column, which is labeled as OutOfRegion, gives the percentage of people who work outside the region under study. The content of this previously mentioned sheet Profiles is copied into a new file titled data/inputFiles/Profile.xlsx, which is read in by the optimization code.

3.3 Using processed data to generate tables and figures

The code is written in Python3 and fully documented using Sphinx. The documentation is found in html files in the folder docs/build/html.

Our figures are created in LaTeX with tikz. The data behind these plots are generated as follows.

- The code for Figure 1 is estimate_incident.py. This Python script takes in population, meshblock, and input regression incident rate data. More details on estimating OHCA rates is included in Online Appendix 7. The result is stored in data/inputFiles/estimate_incident_rat.csv, and is the input file to optimization models.

- The code for Figure 2 is setup_model_from_file_evaluateOnly.py. This script evaluates both late arrival and survival without profiles, assuming a volunteer distribution that is proportional to the demand. It needs to take in population, ambulance travel time, and estimate incident rate data.

- The code for Figure 3 and Figure 4 is `setup_model_from_file.py`. This script uses the greedy algorithm to generate the optimal volunteer distributions without profiles and evaluates the performance of optimal solutions for both objectives. It needs to take in population, ambulance travel time, and estimate incident rate data. More details on the greedy algorithm are provided in the Online Appendix 4.

- The code for Figure 5 is `setup_model_from_file.py`. This script uses the Away-step Frank-Wolfe (FW) algorithm to generate the optimal volunteer recruitment distributions with time profiles and evaluate the performance of optimal solutions for both objectives. It needs to take in population, ambulance travel time, estimate incident rate, and profile data. More details on the FW algorithm are provided in our Online Appendix 6.

Finally, Table 1 was generated in an Excel file that we provide: `data/outputFiles/Table 1 Required volunteer density.xls`. It effectively takes Equation (2) without ambulances (or, equivalently $q = 1$) and target t, and sets this equal to the desired coverage. This equation is solved once for each (coverage, t)-combination.

4 Synthetic Data Set

We provide a synthetic data set of the GoodSAM data that includes all the variables that we used in this paper. The variables are left blank if they are not used in this paper to preserve the confidentiality of the data. We generate the synthetic data as follows:

- We randomly sample an incident from the GoodSAM data and record the meshblock number and the number of alerts sent out.
- We keep the meshblock number and randomly +1/-1 to the number of alerts.
- The response of each alert of this synthetic incident is then sampled individually from the dataset.
- Repeat this process for 50 times.

The code that was used to generate the synthetic data set is `create_mock_data.py` and the result is in `synthetic_goodsam_data.csv` in the folder data/Auckland data.

5 Github

Our code and data is available on `https://github.com/carolinetjes/CFRrecruitment`.