Analysis protocol

Title: Consequences of concurrency conflicts on delayed ambulance response and patient outcomes in Central Norway?

Roles:

Leader: Andreas Asheim
Analysis: Andreas Asheim, Lars Eide Næss
Clinical experts: Jostein Dale, Andreas Kruger, Jon-Ola Wattø, Oddvar Uleberg
Methodological experts: Sara M. Nilsen, Johan H. Bjørngaard
Proposed journals: Prehospital Emergency Care, International Journal of Epidemiology, The BMJ

Research question: How do delays from concurrency conflicts affect patient outcomes?

1. Background

Rapid emergency medical service response is important for the survival of patients having certain highly acute incidents.[1,2] Little conclusive is known about how response time affects outcome for other patient groups.[3]

Minimizing response time for all incidents has been a guiding principle for organizing emergency services, and response time has been widely used for assessing service quality. For any patient, a quicker response is preferable, although for some incidents, other factors may be more important.[4] Examples of this are the responders' competence and time to definite treatment. Shifting the use of limited resources to improve response time for less acute patients may improve overall response time measures while simultaneously increasing response time for the highly acute cases.

The nature of the emergency services makes it hard to conduct randomized controlled trials, including studies of response time and health outcomes. The complexity of medical emergencies means that assessing quality of emergency services through registered response times is not straightforward. Since the more severe cases likely get the quickest response, routinely collected data are likely to show that increased response time is negatively associated with patient survival. In addition, rural areas with longer distances, therefore longer response times, may have different demographic characteristics than densely populated areas. Routinely collected data are likely not sufficiently detailed to control such differences. Concluding on service quality by comparing response times may therefore be problematic.

2. Objectives

The aim of this project is to estimate the extent of concurrency conflicts in the emergency medical services (EMS) Central Norway from 2013 to 2022, and possible effects on patient mortality and hospitalisations. We will also estimate effects of concurrency conflicts on delayed response since an effect on patient outcomes would likely be mediated by delayed response.

3. Data

Data from these data sources will be used.

1. Data from the information system of the three emergency medical communication centrals (EMCCs) in Trondheim, Ålesund and Namsos
    a. Incidents (identifiers, classification, coordinates, time stamps)
    b. Patients (pseudonymous identifiers, patient characteristics)
    c. EMS missions (identifiers, time stamps)
2. Data on hospital activity from hospitals across the region
    a. Admissions (identifiers, acuity, time stamps, treatment and diagnoses)

Data on all individuals calling the emergency services within the area of the regional health trust, Helse Midt-Norge from 1 January 2013 to 31 December 2022 will be used. Only acute incidents are included in the main analysis.

3.1. Ethics and safety

The regional ethics committee (REK) has been informed about the study, and they evaluated the study not in need of a formal REK approval, reference number 283508. The use of the data has been approved by the regional network of EMS leaders in Central Norway (Regionalt fagledernettverk akuttmedisin).

The data will be obtained from the Helse Midt-Norge data warehouse, and all analyses will be performed on the secured computer systems of St. Olav's Hospital. The authors wil not have access to any data that directly identify any participant.

4. Analysis plan

This work will use data from the acute medical services that are collected for logistical purposes. In the data, there is no direct information on concurrency conflicts. To assess whether there could have been a concurrency conflict, we will use a machine-learning algorithm to predict up to five ambulances that are candidates to respond to an incident. The status of the candidate ambulances will then be checked, i.e., if they had a mission at the time of the incident. From this, an indicator of busy ambulance is devised. For a given incident, the indicator will be calculated as a probability that the preferred ambulance to respond to this incident is busy.

Consequences of a busy ambulances will naturally depend on both temporal and spatial variables. For example, incidents that occur where ambulances are few and far between may be more sensitive in the sense that a busy ambulance will result in a larger delay. Also, changes in organisation, for example adding or removing ambulances, may result in differences over time. There may be substantial variation in patient case mix across the region, and between different times of the year, days of the week, and times of the day. For example, one might expect other kinds of injuries to happen during a Friday night than a Thursday night. To handle differences in organisation, driving distances and case-mix over time and across the region, we will analyse only differences between incidents that happen close during comparable time periods. By analysing within groups of comparable incidents, the analysis addresses a hypothetical intervention: If two incidents happen at similar places at a similar time, what would be the difference if there was a difference in ambulance availability.

4.1. Eligibility criteria

All individuals in contact with the emergency services from 1 January 2013 to 31 December 2022, who were registered with a personal identification number in the databases, being involved in an acute incident with a known location.

4.2. Calculating the exposure

The exposure in this study is the probability of a concurrency conflict. This is calculated in several steps.

Calculating candidate ambulances

To calculate candidate ambulances, we perform a multinomial logistic regression analysis on a training set, which for these purposes is a collection of incidents where we know the first responding ambulance. From this we predict which ambulances are likely to respond to an incident, including a probability for each candidate ambulance. We will do this prediction separately for all incidents.

Training set

For a given incident (an index incident), a training set is selected as the 5000 closest incidents that happened less than 15 km away, within a year of the incident (up to one year before or after). These parameters are chosen to strike a balance between rural areas with few incidents and large distances, and urban areas with many incidents and shorter distances. All incidents that happened on the same day as the index incident will be removed from the training set, to avoid that the situation around the index incident affects the prediction.

Prediction

To account for differences in ambulance operating hours we will predict based on on day of the week and time of the day categorised as night (00:00-08:00), day (08:00-16:00) and evening (16:00-24:00). We will weigh responses to incidents close in time more, to lessen the influence of changes in organisation. Therefore, observations are weighted as $w=1/(1+dw^2)$, *dw* being the number of whole weeks from the incident. We use the multinomial logistic regression from the neural-net package *nnet* for *R* to do the calculation:

nnet:multinom(first_responder~TimeOfTheDay+DayOfTheWeek,data=trainingset,weights=w).

For each potential responder, this algorithm estimates the probability that this resource is the responding one. We will record the five most probable responders, $Ambulance_i,\ \ i = 1, \ldots, 5$, for each incident along with the probabilities, $p_i,\ \ i = 1, \ldots, 5$.

Calculating the indicator of busy ambulances

Using the list of all ambulance missions we assess the state of the five most probable responding ambulances at the time of the index incident, $Busy(Ambulance_i),\ \ i = 1, \ldots, 5$, being 0 if the ambulance was available, 1 if it was busy with any kind of mission. From the probabilities p_i and ambulance states, we compute the probability of a busy ambulance,

$$I \ = \sum_{i=1}^{5} Busy(Ambulance_i) \cdot p_i \approx E[Busy\ ambulace].$$

The value of *I* may be interpreted as an estimate of the probability that the index incident is subject to a concurrency conflict. For example, if ambulance A was busy at the time of an incident, and we have estimated that ambulance A was expected to respond to the incident with probability 50%. If, in addition, all other ambulances were available, the indicator value would be 50%.

The exposure in this study will be the variable *I*, which is takes a value between 0 (no concurrency conflict) to 1 (all candidate ambulances busy).

4.3. Outcomes

The two main outcomes in the study will be short-term and long-term mortality:

- Short-term: death within 2 or 7 days of the incident
- Long-term: Death within 60 and 365 days of the incident

In addition, we will investigate indicators of hospital use and hospital treatment costs:

- The probability of being admitted to the hospital within 12 hours, 2, 7, 60 and 365 days of the incident
- Number of days being hospitalised within 7, 60 and 365 days of the incident
- Number of hospital outpatient contacts within 60 and 365 days of the incident
- Hospital treatment costs within 60 and 365 days of the incident, measured by aggregated DRG-costs on all hospital stays and outpatient contacts

4.4. Matching of cases

We compare differences within groups of incidents that happened close during similar time periods. By studying only incidents that occurred close, we remove variability in response time from differences in driving distance and thus increase precision when estimating the association between busy ambulances and response time. We identify the neighborhood of the incident, *grunnkrets*, as defined by Statistics Norway,[5] and analyse within these. Neighbourhoods are the least aggregated geographical unit in Norway, defined as geographically coherent areas with up to 300-400 inhabitants in rural areas and 500-600 inhabitants in urban areas. In addition, we will compare within the same year, i.e., we analyse within incidents that occur in the same neighbourhood in the same year. This removes differences in response time that could be due to differences in organisation, how data are organised, and other gradual changes.

4.5. Adjustments

We will provide estimates adjusted by hour of the day (01:00, 02:00, 03:00, …, 24:00), day of the week and month. We will also adjust for the age of the patient with a 10-node spline curve, patient sex, and the number of days being hospitalised 90 days prior to the incident.

4.6. Cluster correction of standard errors

Data will be analysed on the patient level, while the exposure and matching are done at the incident-level. This means that one incident may have more than one patient. We will add a cluster correction to take correlation between patients within each incident into account.

4.7. Statistical methods

The analyses will be performed using the fixed-effects estimators from the *fixest*-package for *R*. We calculate estimates within groups of matched incidents. For continuous outcomes, we use ordinary least square (*feols*), for binary outcomes we will use logistic regression (via *feglm*), for count outcomes we will use Poisson regression (*fepois*). The code for an OLS analysis is,

```
model =  fixest::feols(outcome~Hour+Weekday+Month+sex+splines::ns(age,10) +
hospitalised_90d_prev + I | neighborhood_year_id, data=data).
```

We will also estimate effects of delay, as predicted by concurrency conflict, in an instrumental variable analysis. These analyses will only be estimated using 2-stage OLS (implemented in *feols)*,

```
model = fixest::feols(outcome~Hour+Weekday+Month+sex+splines::ns(age,10) +
hospitalised_90d_prev + I | neighborhood_year_id | Response~I, data=data).
```

For binary outcomes, this will provide estimates of risk-differences.

For all analyses, to account for clustering of observations, we obtain standard errors through the command,

```
summary(model,cluster=c(”incident_id”,”neighborhood_year_id”)).
```

4.8. Additional analyses

Balance tests

If, within the matched groups, patients involved in incidents with busy ambulances are systematically different from patients in incidents without busy ambulances, this could lead to imbalance in the design and biased results. We will investigate whether there are any signs of imbalance by considering associations between the indicator of a concurrency conflict and patient characteristics:

- age in year and in categories (0-20,20-40, …)
- sex
- number of days in the hospital the previous year
- acute hospital contact for cardiovascular disease previous year
- having an earlier EMS contact the previous 3 months)
- Being classified as having respiratory problems
- Missing personal identifier

Any substantial association may be an indication of an imbalance.

Stratified analyses

We will perform the analyses stratified on urban/rural/semi-urban municipalities to investigate whether rural municipalities, where consequences of concurrency conflicts likely are bigger, are more vulnerable.

We will investigate each year of the study separately, to investigate whether the results have been stable over time.

We will consider summer and winter, weekdays and weekends, and night, day, evenings to investigate whether specific time periods are vulnerable to ambulance availability.

We will provide separate estimates for incidents in neighbourhoods with different expected response times by stratifying the analyses on the median response time in the neighbourhood in the preceding year.

We also provide separate estimates for incidents in neighbourhoods with different travel time to hospital, computed by travel time along roads.

Analyses on specific patient groups

It may be expected that certain patient groups are particularly vulnerable to ambulance delays, for example patients with myocardial infarction and cardiac arrest. These patient groups may be hard to identify, since information on diagnosis is only available if the patient is admitted to hospital, i.e., we only have information on survivors. We will therefore use incident classification codes registered by the EMS Communication Center to select incidents that may involve vulnerable patients.

Sensitivity analyses

To investigate the impact of how incidents are grouped, we will perform the analyses with alternative conditions, in addition to being in the same neighbourhood:

- Same year, month, day of the week, shift
- Same year, month, day of the week
- Same year, month

Conventional analysis

The instrumental variable analysis may provide unbiased estimates of effects of ambulance delays. To assess the magnitude of this bias, we will also estimate the direct association between response time and outcomes, adjusted for all available pre-treatment variables.

References

1. Jena AB, Mann NC, Wedlund LN, Olenski A. Delays in emergency care and mortality during major US marathons. *New England Journal of Medicine* 2017; **376**(15): 1441-50.
2. Pons PT, Haukoos JS, Bludworth W, Cribley T, Pons KA, Markovchick VJ. Paramedic response time: does it affect patient survival? *Academic Emergency Medicine* 2005; **12**(7): 594-600.
3. Wilde ET. Do emergency medical system response times matter for health outcomes? *Health economics* 2013; **22**(7): 790-806.
4. Al-Shaqsi SZK. Response time as a sole performance indicator in EMS: Pitfalls and solutions. *Open access emergency medicine: OAEM* 2010; **2**: 1.
5. Norway S. Standard for delområde- og grunnkretsinndeling. 2023. https://www.ssb.no/klass/klassifikasjoner/12023).