Do community first responder networks have potential?

Nico Dellaert *
Eindhoven University of Technology, Department of Industrial Engineering and Innovation Sciences
n.p.dellaert@tue.nl

Loe Schlicher
Eindhoven University of Technology, Department of Industrial Engineering and Innovation Sciences
l.p.j.schlicher@tue.nl

Janneke Hillenaar
Eindhoven University of Technology, Department of Industrial Engineering and Innovation Sciences
janneke.hillenaar@hotmail.nl

Caroline J. Jagtenberg
Vrije Universiteit Amsterdam, School of Business and Economics, Department of Operations Analytics
c.j.jagtenberg@vu.nl

Abstract: In several countries, emergency medical services receive assistance from community first responders. In such a case, the dispatch center not only notifies an ambulance, but also one or more volunteers that are located near the incident. Due to their proximity to the patient, volunteers can often provide first aid or even life-saving help before the ambulance arrives. This paper is the first to mathematically model the stochastic response time of the first-arriving volunteer by considering uncertainties in dispatch time, volunteer acceptance delay and distance-dependent mode choice. By comparing the derived volunteer response-time distribution to the historical ambulance response-time distribution, it also quantifies the effect of a community first responder network. Besides that, several alert and retract rules are studied, in order to use the volunteers as effectively as possible. Alert rules set a different alert radius per district and retract rules stop the alert for late responding volunteers. Through a case study, this paper analyzes these rules and quantifies the impact of a potential community first responder network in the Netherlands and discusses managerial implications.

Key words: Community first responders, volunteer alert, dispatching, stochastic modeling

---

* corresponding author

1 Introduction

Emergency medical services in the United States need on average 7 minutes from the time of a 911 call to arrival on scene, with an increase to more than 14 minutes in rural settings (Mell et al. 2017). Similar response times can also be found in Europe, including the Netherlands and Belgium (Pauw 2021, Houweling 2021). Longer response times have been associated with worse outcomes in trauma patients, especially for emergent conditions such as cardiopulmonary arrest, severe bleeding, and airway occlusion. For a quicker response, Metelmann et al. (2021) suggests that *Community First Responders* can be of particular help. A community first responder (CFR) is a person who is trained and willing to voluntarily attend certain types of emergency calls in the area where they live or work. They can be members of the public who have received training in life-saving interventions such as defibrillation, or they can be off-duty paramedics, firefighters or other professionals (Jackson 2004). Their role is to provide life-saving treatment to the patient in the vital first few minutes until an ambulance arrives.

Since the introduction of smartphones, efficient CFR networks have emerged that alert registered CFRs based on their proximity to the incident. Such CFR networks are active in various countries including Canada, the United States, and the majority of Europe (Oving et al. 2019). The network is typically activated by the emergency call center, upon which a notification is sent to multiple nearby CFRs. Some networks have the option for CFRs to react with an accept or reject button. The most advanced networks can send cancellations, for example when sufficiently many CFRs have accepted or arrived.

Next to governments that deploy such CFR networks, also humanitarian network organizations consider it. For instance, in the Netherlands, we are in contact with the Red Cross (Hillenaar 2021), who are considering the deployment of such a CFR network. From several meetings with the Dutch Red Cross, we learned that at least two quantitative performance indicators are important in this consideration. The first one is the *response-time reduction* that a CFR network could realize, compared to an ambulance-only system. Clearly, this response-time reduction should be large enough to make the investment in a CFR network worthwhile. Second, the Red Cross sees *workload* of a CFR as an important performance indicator. Typically, when an alert is sent around, many CFRs respond and travel to the incident. This means that in the majority of the cases several CFRs arrive in vain. Because overburdening CFRs with requests can lead to a reduced willingness to accept future requests, the workload of a CFR should be limited.

In this paper we introduce a mathematical model that estimates the above-mentioned performance indicators for a currently non-existent CFR network. For the first performance indicator, we compare the response time of the already existing ambulance network with that of a potential CFR network. We obtain the response-time distribution of the ambulance network from historical data and we model the stochastic response time of the first-arriving CFR by considering uncertainties in dispatch time, CFR acceptance delay and distance-dependent mode choice. For the second performance indicator, we calculate the expected number of responding CFRs per incident, taking into account the availability of CFRs. To balance these two performance indicators, we also introduce several alert and retract rules to indicate the size of the area to alert, and whether outstanding alerts should be retracted once some people have responded. By means of a case study, we analyze these rules and quantify the potential –in terms of the two performance indicators– of a CFR network for a city in the Netherlands. For this case study, we make realistic estimates of acceptance probabilities, durations of delays, mode choice, and travel time. Finally, using Linear Programming, we also optimize which CFRs to alert: for each postal code, we prescribe an alerting radius that balances the city-wide number of alerts versus response-time improvement.

The rest of this paper is outlined as follows. In Section 2 we summarize literature on community first response networks, alerting strategies as well as related operations research models dealing with community first responder networks . Section 3 introduces a mathematical model to compute the response-time distribution of a given CFR base and discusses several alert and retract rules. Section 4 contains a case study of the city of Breda (the Netherlands) and we end with a conclusion and discussion in Section 5.

2 Literature review

This section describes how the performance of community first responder networks is measured in literature. It also captures alerting strategies: how many volunteers to dispatch for a single incident. Finally, it outlines the literature on analytics for responder networks, to which our paper contributes.

2.1 Optimization of ambulance logistics

Ambulance systems are often optimized in terms of response times, where fast responses are implicitly assumed to provide better patient outcomes. A common approach is to use Mixed Integer Programming to decide on the locations of bases - and potentially also the number of vehicles per base, for an overview see Li et al. (2011). The goal there is typically to maximize demand covered within a predefined time threshold. Besides a response-time threshold, a common alternative objective is maximizing the expected patient survival. This was first done by Erkut et al. (2008) and typically involves borrowing a so-called survival curve from the medical literature. Such curves are commonly known for out-of-hospital cardiac arrest, where the time to resuscitation clearly relates to survival.

The decisions on where to locate bases are made in a strategic phase, but optimization can also be helpful in real time. This is called ambulance redeployment: proactively moving idle vehicles around the region, to fill in gaps left by currently busy ambulances and hence improve the system preparedness. Mathematical techniques involve approximate dynamic programming (Maxwell et al. 2010, Schmid 2012) and heuristics (Jagtenberg et al. 2015). Some include workload of EMS personnel (Enayati et al. 2018). Maxwell et al. (2014) shows an upper bound for the performance that any redeployment policy may give.

It is common to always dispatch the closest idle ambulance, although a few papers investigated other options that deviate from that (Jagtenberg et al. 2017, Nasrollahzadeh et al. 2018).

Due to the randomness inherent to emergency incidents, a careful evaluation of logistics strategies often involves simulation. The literature includes open-source packages in Python (Dieleman and Jagtenberg 2024) and Julia (Ridler et al. 2022).

2.2 Performance of community first responder networks

CFRs are active in emergency healthcare systems of various countries, including the United States, Canada, United Kingdom, Ireland, Israel, and the Netherlands. A large number of retrospective data analyses have been done, and are summarized in Valeriano et al. (2021, Table 2). We next give an overview of studies that measured CFR performance in terms of response times.

Several studies took place in the United Kingdom. For example, Botan et al. (2022) did a retrospective study on a large part of the country in 2019. Using historical incident data, they showed that CFRs were most effective in rural areas, where they attended 6.2% of the calls. In 62.8% of those cases, the CFR arrived before the ambulance. In a smaller study, Campbell and Ellington (2016) measured the response-time benefit of a student CFR scheme, where twenty medical students were trained to be first responders to support ambulance services in an inner-city setting. Over 12 months, they attended 89 emergency calls. When CFRs arrived it was on average 3 minutes before the ambulance, with a difference varying throughout the day, peaking between 16:00 and 18:00.

A rural island in Denmark has had a CFR network for *all* medical emergencies in place since 2012. Sarkisian et al. (2020) found that in 85% of the cases, a CFR arrived before the ambulance. The median response time for CFRs was 5.5 minutes faster than the median response time for ambulances.

Just like on the Danish island, it turns out that in the Netherlands –a country with a long history of volunteer response– CFRs make a significant improvement in response times (Zijlstra et al. 2014). This network, however, is only activated for a very specific patient group (1-2%) of all emergency calls. The Dutch network is very dense and dates back to the pre-smartphone era, where CFRs were alerted by text message, based on their home or work address. Nowadays, both the text message service as well as an app are in place, and Slaa (2020) reports many empirical distributions for the system, including acceptance rates, delays and how these depend on CFR characteristics such as age, historical activity on the app, and distance to the patient.

In recent years, several studies have been done that do not look at volunteers in isolation, but combine their response with picking up up equipment (Zijlstra et al. 2014) or having AEDs delivered by drone (Matinrad and Reuter-Oppermann 2022). One example of a study that is not retrospective, is Barry et al. (2018). For Ireland, they identified the potential of their volunteer community by estimating the proportion of the Irish population that has the potential to receive a timely response. They used a predefined set of 536 volunteers, which they assumed are always available at their registered addresses, in addition to 105 ambulance base locations. Using a 10-minute response-time threshold they concluded that ambulances can reach around 62% of the population, while volunteers can increase this to over 91%. During off-peak hours the difference was estimated to be smaller (70% versus 92%), mainly due to an improved EMS speed.

To determine the added value of a CFR network, it is also important to consider the alerting strategy. We therefore next summarize the literature on this topic.

2.2.1 Alerting strategies

In recent years, several Dutch studies have reported on the operational alerting strategy. For example, Zijlstra et al. (2014) report that volunteers within a 1-km radius are dispatched. Moreover, according to van der Worp (2014), at most 10 volunteers were sent directly to the patient while up to 20 others were sent via an AED. More recently, HartslagNu (2021) investigated the radius within which they alert volunteers. They report that it was not possible to find one good single alerting radius for the entire country. Instead, they define an alerting radius based on (1) a maximum of 2 km, (2) the number of volunteers in the area (at most 100) and (3) the number of AEDs in the area (at most 3).

Limiting the number of alerts is necessary in light of volunteer fatigue: the notion that overburdening volunteers with requests leads to a reduced willingness to accept future requests. It is a known problem for volunteer firefighters (Dawson et al. 2015) and their retention is studied in Shrader (2012). For EMS volunteers, their motivation is studied in Israel (Khalemsky et al. 2020) and rural USA (Freeman et al. 2009). This subject is mainly researched qualitatively, by interviewing volunteers. Alerting many volunteers for one patient also increases the likelihood of volunteers arriving in vain. While a second responder on scene may still be helpful to assist the first, at some point a multitude of volunteers on scene is thought to discourage future responses.

Reducing the number of alerts while maintaining reasonable performance for the patient can for example be done through so-called phased alerting, where the idea is to send alerts in batches, with time lags in between, to see if previous ones have been accepted. Henderson et al. (2022) proposes to optimize phased alerting for cardiac arrest patients specifically, defining the objective in terms of a survival curve (the probability of survival depending on time to resuscitation).

In order to make a good choice as to which and how many CFRs should be alerted, it is important to have insight into how people travel to the incident. Studies such as Zijlstra et al. (2015) and Slaa (2020) could be useful for this purpose. For example, the first shows how for the Netherlands 54% of the CFRs went by car, 27% by bike and 15% on foot. In contrast, Slaa (2020) reports for the same system that 71.2% travel by bike whereas the rest mostly walk. A potential explanation for this difference is that the latter study was done five years later. During that time, the CFR density increased, which means volunteers are on average closer to the patient and thus, presumably, less likely to use a car. Neither of these studies reports on the relationship between distance and mode choice. Lastly, there is Jonsson et al. (2020), who studied Swedish community first responders and report a travel speed of 2.3 m/s on average and 1.8 m/s in densely populated areas. The latter group had the narrowest distribution, indicating homogeneity, which led the authors to conjecture that this group traveled on foot.

2.3 Operations Research models for community first responder networks and Ambulance Systems

In our paper, we make use of techniques from Operations Research to quantify the impact of a potential CFR network. Literature focusing on CFRs using OR techniques is, however, scarce. A first exception is the work of Paz et al. (2022). They study how to dispatch ambulances in the presence of CFRs. In their model, emergency incidents arise with different priorities, and the dispatch system can observe real-time location information on CFRs. The goal is to balance response for the current incident and response preparedness for future requests. They evaluate their solution with discrete event simulation and compare it against a procedure that does not incorporate CFR information. Another exception is the work of Van den Berg et al. (2024). They model volunteer presence as a Poisson Point process and bound the performance that can be expected from n volunteers by determining the optimal distribution of volunteer mass over a city. They assume all volunteers walk, leaving a mode choice –let alone a distance-dependent one– out of scope. A final exception is the work of Lancaster and Herrmann (2021). They use Monte Carlo simulation to compare the performance of a few hypothetical variants of CFR systems, with and without AEDs. They studied a region in the United States and used response times and cardiac arrest survival as their performance metrics.

Although OR literature focusing on CFRs is limited, there is a vast OR stream of literature on improving ambulance systems. These systems are often optimized in terms of response times, where fast responses are implicitly assumed to provide better patient outcomes. A common approach is to use Mixed Integer Programming to decide on the locations of bases –and potentially also the number of vehicles per base (for an overview, see Li et al. (2011)). The goal there is typically to maximize demand covered within a predefined time threshold. Besides a response-time threshold, a common alternative objective is maximizing the expected patient survival. This was first done by Erkut et al. (2008) and typically involves borrowing a so-called survival curve from the medical literature. Such curves are commonly known for out-of-hospital cardiac arrest, where the time to resuscitation clearly relates to survival.

The decisions on where to locate bases are made in a strategic phase, but optimization can also be helpful in real time. This is called ambulance redeployment: proactively moving idle vehicles around the region, to fill in gaps left by currently busy ambulances and hence improve the system preparedness. Mathematical techniques involve approximate dynamic programming (Maxwell et al. 2010, Schmid 2012) and heuristics (Jagtenberg et al. 2015). Recent works focussed on the inclusion of workload of EMS personnel (Enayati et al. 2018). Maxwell et al. (2014) performance guarantee bounds for large class of redeployment policies.

In practice, is it common to dispatch the closest idle ambulance, but this is not necessarily optimal. A few papers investigate other dispatching strategies that deviate from this closest idle one principle and demonstrate their benefits (see, e.g., (Jagtenberg et al. 2017, Nasrollahzadeh et al. 2018)). Methods that are used include Markov decision processes and approximate dynamic programming frameworks.

Due to the randomness inherent to emergency incidents, a careful evaluation of logistics strategies often involves simulation. Recently, several open-source packages are provided in literature to perform such simulations (see, e.g., (Dieleman and Jagtenberg 2024) for Python and (Ridler et al. 2022) for Julia).

3 Network first responder model

In this section, we develop our mathematical model, including the two performance indicators, and introduce relevant alert and retract rules, to analyze the performance of a potential CFR network. Moreover, we introduce an optimization problem that determines the maximal effectiveness of a potential CFR network.

3.1 Introduction

Our aim is to estimate the performance of an as-yet-unrealized CFR network, while taking into account the existence of an already-established professional responder network. As a first performance indicator, we compare the CFR response times with ambulance response times to determine the *expected response-time reduction*. Ambulance response times, defined as the duration between an emergency phone call and a responder arriving on scene, are often recorded with great accuracy. As such, we model the ambulance response time as a random variable with known distribution. When it comes to CFR response times, such records are either unavailable for networks in the development phase or are very limited for those that are already in existence. For that reason, we model the response time of a CFR as a sum of random variables, with each random variable characterizing a distinct task within the interval from receiving an emergency call to the CFR's arrival at the scene. Tasks that we explicitly model are (i) the *dispatch time*, which is the time between a 911 call and the moment the CFR system is activated, (ii) the *view delay*, which is the time for a CFR to check the message and respond, (iii) the *travel mode delay*, which is the time for a CFR between the acceptance of the call and the start of the trip, and (iv) the *travel time*, which is the time for a CFR to travel to the scene. These distinct tasks are also represented chronologically in Figure 3.1.

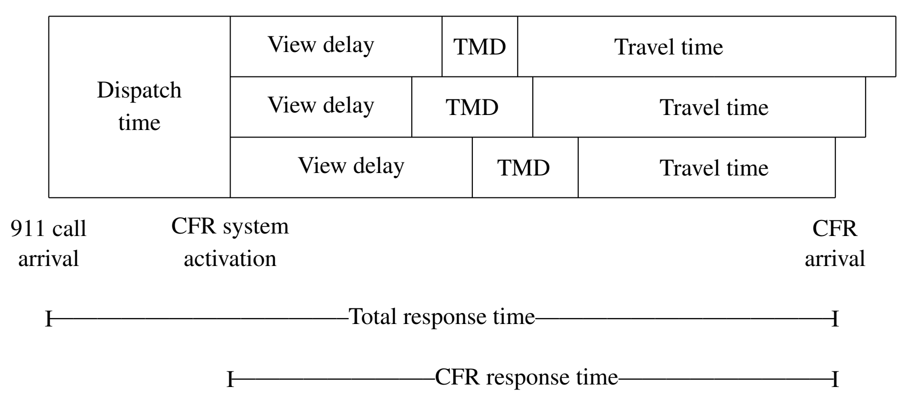

Figure 1 The duration of the four distinct tasks for three different CFRs responding to the same incident. Here, TMD stands for Travel Mode Delay.

Next to the expected response-time reduction, we also consider a second performance indicator: the *expected CFR workload*. We define this as the expected number of CFRs that travel to the scene per patient. In order to calculate this indicator, we introduce the concept of an alert area. Such an alert area is defined by a radius, representing a circle around the incident. For this area, we calculate the expected number of CFRs responding to an incident based on the number of *potential* CFRs present in the area and the probability that a CFR actually responds to the call. Later on in this paper, we also analyze how this metric depends on the chosen retract rule: such a rule may include informing CFRs that further assistance is not required anymore, once a sufficient number of CFRs have already accepted their alert.

In the following sections, we formally introduce the concept of an alert area, provide a detailed description of our CFR response-time modeling and describe the performance metrics and alert and retract rules.

3.2 Alert area

As already mentioned, we start by modeling the alert area around an incident. We specify such an area with an associated radius $r \in \mathbb{R}_{>0}$ and divide this area into concentric rings with $N = \{1, 2, \ldots, n\}$ the set of rings, $n \in \mathbb{N}_{>0}$ the number of rings, and $d_i \in \mathbb{R}_{\geq 0}$ the distance from the end of ring i to the incident location.[1] An example of such a setting, with $n = 5$ rings, is demonstrated in Figure 2.

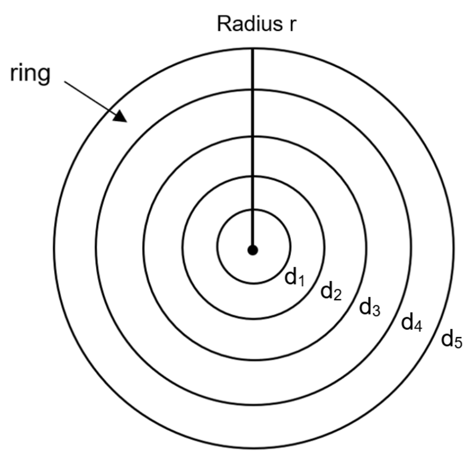

Figure 2 Visual representation of the situation with $n = 5$ rings.

Since our aim is to model the performance of a CFR network in a preparatory phase, we can not expect to know the real-time number and position of the CFRs. Instead, we follow Van den Berg et al. (2024) in modeling the presence of CFRs as a spatial Poisson process. The intensity of the Poisson process is defined by a rate $\rho \in \mathbb{R}_{>0}$ that can be interpreted as the average or expected number of *potential* CFRs per squared distance measure. This rate is subsequently multiplied by the probability that a CFR is *available* to respond (so-called Poisson thinning).[2] We assume that the probability of being available decreases per ring, which is more or less in line with the empirical findings in Slaa (2020, Fig 2.7). Combining these two assumptions

[1] Please, observe that $d_n = r$. For notational convenience, we also introduce $d_0 = 0$.

[2] For the difference between potential and available CFRs, see Table 1.

state	meaning
potential	potentially located in the alert area at the moment of CFR system activation
available	present in alert area and willing to respond [3]
responding	traveling to the patient

Table 1 States of CFRs.

leads to the following: for each ring $i \in N$, the number of *available* CFRs follows a Poisson distribution with mean $\lambda_i = (d_i^2 - d_{i-1}^2) \cdot \rho \cdot q_i \cdot \pi$, where $q_i = \alpha\beta^{i-1}$ with $\alpha, \beta \in [0,1]$. Factor q_i represents the probability of a CFR being available to respond and it models that CFRs far away from the incident are less inclined to respond than CFRs nearby. So, the probability that available CFR allocation $a = (a_i)_{i \in N}$ occurs, where $a_i \in \mathbb{N}_{\geq 0}$ describes the number of available CFRs of ring i, is

$$f_{AV}(a) = \prod_{i \in N} \left(\frac{\lambda_i^{a_i}}{a_i!} \cdot e^{-\lambda_i} \right). \tag{1}$$

For any allocation of a, the spatial Poisson process implies that CFRs are subsequently independently and *uniformly* distributed over the surface area of the circle, and thus each ring $i \in N$.

When CFRs travel to the incident, they use one type of transport mode. We let the set of transport modes be $M = \{foot, bike, car\}$ and the probability that a CFR in ring $i \in N$ uses mode $m \in M$ be $p_{im} \in [0,1]$.

3.3 CFR response time

In this subsection, we model CFR dynamics to estimate the CFR response time. As discussed in Subsection 3.1, the *total response time* is the sum of the duration of four distinct tasks. For each of them, we now introduce a discrete probability distribution.[4] First, we introduce $F_0 : \mathbb{N}_{\geq 0} \to [0,1]$ for the *dispatch time*. The dispatch time is the time between receiving a call and sending out messages to CFRs at the call center. Second, we introduce $F_1 : \mathbb{N}_{\geq 0} \to [0,1]$ for the *view delay*. The view delay refers to the time to check and answer the message. Third, we introduce $F_2 : M \times \mathbb{N}_{\geq 0} \to [0,1]$ for the *travel mode delay*. This delay may, for instance, consist of taking a coat and bag, getting the bike or getting into the car. Finally, we introduce $F_3 : N \times M \times \mathbb{N}_{\geq 0} \to [0,1]$ for the *travel time*. The travel time is the time spent traveling from the "home" location of the CFR to the incident location. To determine F_3, we use that CFRs are distributed uniformly over the area around the incident (see Section 3.2) and assume that velocity of transport mode m is uniformly distributed over the interval $[v_{m0}, v_{m1}]$ with $v_{m0}, v_{m1} \in \mathbb{R}_{>0}$ and $v_{m0} < v_{m1}$. Consequently, integrating over the alert area as well as over the velocity results in:[5]

[3] potential CFRs that react negatively on an alert are considered to be not available.

[4] Please, observe that we consider discrete time units. A typical discrete time unit could, for instance, be a second.

[5] A detailed derivation of F_3 is presented in Appendix 1.

$$F_3(i,m,t) = \begin{cases} 0 & \text{if } t \leq \frac{d_{i-1}}{v_{m1}} \\ \frac{\frac{1}{3}v_{m1}^3 t^2 - d_{i-1}^2 v_{m1} + \frac{2}{3}\frac{d_{i-1}^3}{t}}{(v_{m1}-v_{m0})(d_i^2-d_{i-1}^2)} & \text{if } t \leq \frac{d_{i-1}}{v_{m0}} \text{ and } \frac{d_{i-1}}{v_{m1}} \leq t \leq \frac{d_i}{v_{m1}} \\ \frac{\frac{2}{3t}\left(d_{i-1}^3-d_i^3\right)+v_{m1}(d_i^2-d_{i-1}^2)}{(v_{m1}-v_{m0})(d_i^2-d_{i-1}^2)} & \text{if } t \leq \frac{d_{i-1}}{v_{m0}} \text{ and } t \geq \frac{d_i}{v_{m1}} \\ \frac{\frac{1}{3}t^2(v_{m1}^3-v_{m0}^3)-d_{i-1}^2(v_{m1}-v_{m0})}{(v_{m1}-v_{m0})(d_i^2-d_{i-1}^2)} & \text{if } \frac{d_{i-1}}{v_{m0}} \leq t \leq \frac{d_i}{v_{m0}} \text{ and } \frac{d_{i-1}}{v_{m1}} \leq t \leq \frac{d_i}{v_{m1}} \\ \frac{\left(-\frac{2}{3}\frac{d_i^3}{t} - \frac{1}{3}v_{m0}^3 t^2 + d_{i-1}^2 v_{m0} + v_{m1}(d_i^2-d_{i-1}^2)\right)}{(v_{m1}-v_{m0})(d_i^2-d_{i-1}^2)} & \text{if } \frac{d_{i-1}}{v_{m0}} \leq t \leq \frac{d_i}{v_{m0}} \text{ and } t \geq \frac{d_i}{v_{m1}} \\ 1 & \text{if } t \geq \frac{d_i}{v_{m0}} \end{cases} \tag{2}$$

for all $i \in N, m \in M$ and $t \in \mathbb{N}_{\geq 0}$.

By taking the convolution over the three distributions (F_1, F_2 and F_3), we arrive at the CFR response-time distribution, which, for a given ring $i \in N$, is given by:

$$F_R(i,t) = \sum_{t'=0}^{t} \sum_{t_1=0}^{t'} \sum_{m=1}^{3} p_{im} \sum_{t_2=0}^{t'-t_1} f_1(t_1) \cdot f_2(m,t_2) \cdot f_3(i,m,t'-t_1-t_2) \text{ for all } t \in \mathbb{N}_{\geq 0}. \tag{3}$$

Subsequently, by convoluting over F_R and F_0, one can also derive the total response-time distribution.

3.4 Performance indicators

In this section, we derive our two key performance indicators, namely the expected response-time reduction and the expected CFR workload per incident. To derive the expected response-time reduction, we first introduce $F_{AM} : \mathbb{N}_{\geq 0} \to [0,1]$, which is the distribution of the arrival time of an already existing ambulance network. Second, we derive $F_{RFA} : \mathbb{N} \times \mathbb{N}^N \to [0,1]$, which is the distribution of the total response time of the *first* arriving CFR. This distribution F_{RFA} is derived from F_0 and F_R and reads as follows:

$$F_{RFA}(t) = \sum_{a \in \mathbb{N}_{\geq 0}^N} f_{AV}(a) \sum_{t_0=0}^{t} F_0(t_0) \left(1 - \prod_{i \in N} (1 - F_R(i, t-t_0))^{a_i}\right) \text{ for all } t \in \mathbb{N}_{\geq 0}. \tag{4}$$

Note that the term between brackets describes the probability that at least one CFR is earlier than $t - t_0$ time units. By convoluting this term with the distribution of the dispatch time, and subsequently with the probability of CFR allocation f_{AV}, we end up with our desired distribution F_{RFA}.

It is now possible to derive the first performance indicator, namely the expected reduction in response time:

$$\mathbb{E}[\text{response time reduction}] = \sum_{t \in \mathbb{N}_{\geq 0}} f_{AM}(t) \sum_{t' \in \mathbb{N}_{\geq 0}: t' < t} (t - t') F_{RFA}(t'). \tag{5}$$

To determine the second performance measure, the expected CFR workload, we calculate the expected number of *responding* CFRs per incident, which reads as follows:

$$\mathbb{E}[\text{CFR workload}] = \sum_{a \in \mathbb{N}_{\geq 0}^{N}} f_{AV}(a) \sum_{i=1}^{N} a_i = \sum_{i \in N} \lambda_i. \tag{6}$$

We conclude this section with Table 2, listing all parameters, variables, probabilities and cumulative density functions (cdf's).

Notation	Description
i	ring
N	set of rings
t	time, expressed in discrete time units
m	travel mode
F_0	cdf for the dispatch
F_1	cdf for the view delay
F_2	cdf for the travel mode delay
F_3	cdf for the travel time
F_R	cdf for the CFR's response time
a	vector of available CFRs, with a_i CFRs reacting in ring i
f_{AV}	probability to have allocation a of CFRs
F_{RFA}	cdf for the response time of the first arriving CFR
p_{im}	probability of using travel mode m in ring i
v_{m1}	maximum speed using travel mode m
v_{m0}	minimum speed using travel mode m
d_i	outer distance of ring i, with $d_0 = 0$
q_i	probability that a CFR in ring i is available
F_{AM}	cdf of the response time of ambulance

Table 2 Definitions of parameters, variables, and distributions.

3.5 Alert and retract rules

In the previous paragraphs, we implicitly assumed that all available CFRs see their alert and thus also respond to the call. Such a multitude of trips requires significant efforts on the CFRs' side (i.e., on the CFR workload). For that reason, this section introduces two rules to limit these efforts. The upcoming section subsequently investigates the impact of these rules on the CFR workload. The two rules read:

1. **Alert rule:** until now, we have assumed that all n rings are being alarmed upon a call. Instead of alerting all these rings, one could also decide to only alert the k out of n rings closest to the incident. Whenever we refer to such a rule, we call it the k-out-n alert rule.

2. **Retract rule:** Several existing CFR systems, such as the Dutch Hartslag Nu (HartslagNu 2021) system, receive a reply from their users via an app button or SMS response, indicating that they "accept" or "reject" the call. If a CFR system is able to receive and process such messages on the fly, they could potentially also retract some of the outstanding alerts when some other CFRs have already accepted. This may reduce the CFR workload as well. We consider the following retract rules:

   (a) Never retract the alert;

(b) Retract the alert in the rings outside i when a CFR accepts in ring i;

(c) Retract the alert in all rings when s CFRs have accepted.

For the setting with alert rule n-out-n and retract rule (a) we already derived the performance indicator formulas (see (5) and (6)), which we also call the *base setting*. In Appendix 2, we explain how to calculate the performance indicators for the setting with alert rule k-out-n and retract rules (b) and (c).

3.6 Maximizing effectiveness of potential CFR network

Instead of presenting a frontier of two performance indicators (e.g., the expected response time reduction *versus* the expected workload or the expected response time reduction of all incidents together *versus* the total number of trips of all CFRs) to support health organizations in defining what is the "best" configuration for a CFR network, one could also consider one of the performance indicators as given and see how to optimize the other one. In this section, we will exactly to that. More precisely, we formulate an optimization problem that identifies –per postal code– the number of rings to alert in order to maximize that expected response time reduction of all incidents in Breda together, given that a maximum number of expected yearly trips is available. In doing so, we will also investigate the impact of the various retract rules.

Let $B \in \mathbb{N}$ be the maximum number of expected yearly trips available. For this number B, we identify the number of alerted rings per postal code for the three retract rules, and for retract rule (c) we also identify the stopping number $s \in \mathbb{N}$, that maximizes the expected response-time reduction over all postal codes.

For retract rules (a) and (b), we introduce decision variable x_{di} for all $i \in N$ and all $d \in D$, with D the set of postal codes. Variable x_{di} takes a value of 1 if the first i rings are alerted in postal code $d \in D$ and 0 otherwise. We use $\mathbb{E}_{id}$[response time reduction] and $\mathbb{E}_{id}$[CFR workload] to denote the expected response-time reduction for postal code d and the expected CFR workload for postal code d respectively when i rings are alerted. Moreover, we denote by I_d the yearly number of incidents in postal code d.

For retract rule (a), we reuse the formulas (1), (4), (5) and (6), by limiting the set of rings to $\{1,..,i\}$ and using the density parameter and ambulance distribution per postal code to determine $\mathbb{E}_{id}$[response time reduction] and $\mathbb{E}_{id}$[CFR workload]. For retract rule (b), we adapted these formulas considerably. The results hereof are presented in Appendix 2.

For retract rules (a) and (b), we want to solve the following optimization problem:

$$\begin{aligned} &\text{Max} \sum_{d\in D}\sum_{i\in N} x_{id}\cdot \mathbb{E}_{id}[\text{response time reduction}]\cdot I_d \\ \text{s.t.} &\sum_{d\in D}\sum_{i\in N} x_{id}\cdot \mathbb{E}_{id}[\text{CFR workload}]\cdot I_d \le B \\ &\sum_{i\in N} x_{id} = 1 \text{ for all } d\in D \\ &\text{x}_{id} \in \{0,1\} \text{ for all } d\in D \text{ and all } i \in N. \end{aligned} \tag{7}$$

The model for retract rule (c) looks quite similar, but has an additional decision parameter $s \in S$, indicating the maximum number of CFRs that will be sent to an accident. Consequently, this time, we introduce decision variable x_{ids}, with a value of 1 if i rings are alerted in postal code $d \in D$ with a maximum of $s \in S$ CFRs, and 0 otherwise. Similar to the previous optimization problem, we use $\mathbb{E}_{ids}$[response time reduction] and $\mathbb{E}_{ids}$[CFR workload] to denote the expected response-time reduction for postal code d and the expected CFR workload for postal code d respectively when i rings are alerted and $s \in S$ CFRs are alerted at most. A detailed derivation of these expressions can be found in Appendix 2.

For retract rule (c), we want to solve the following optimization problem:

$$\begin{aligned} &\text{Max} \sum_{d \in D} \sum_{i \in N} \sum_{s \in S} x_{ids} \cdot \mathbb{E}_{ids}[\text{response time reduction}] \cdot I_d \\ \text{s.t.} &\sum_{d \in D} \sum_{i \in N} \sum_{s \in S} x_{ids} \cdot \mathbb{E}_{ids}[\text{CFR workload}] \cdot I_d \leq B \\ &\sum_{i \in N} \sum_{s \in S} x_{ids} = 1 \text{ for all } d \in D \\ &\text{x}_{ids} \in \{0,1\} \text{ for all } d \in D \text{ and all } i \in N. \end{aligned} \tag{8}$$

4 Case study

In this section, we investigate the potential of a CFR network in the city of Breda, the Netherlands. Breda is a city and municipality in the southern part of the Netherlands, located in the province of North Brabant. The municipality has approximately 185,000 inhabitants and it is divided into 25 postal codes (Figure 3).

For the case study, we will only consider the 17 postal codes that constitute the city (i.e., the dark yellow colored areas in Figure 3), leaving out 4 disconnected villages and the mostly industrial or agricultural districts. The characteristics of the 17 relevant postal codes, with their population, area in squared km, current number of CFRs, yearly number of incidents and the average ambulance response time (in minutes) are displayed in Table 3. The second and third column are retrieved from CBS (2022). The number of CFRs is provided by the Dutch Red Cross. The ambulance provider RAV Brabant Midden-West Noord (n.d.), responsible for the area including Breda, gave us access to the remaining columns. In particular, to derive the last column, they shared with us all ambulance response times for Breda in the year 2019. Based on these data points, we were also able to derive the empirical distribution F_{AM} per postal code.

To calculate the two performance indicators for each of the 17 postal codes, we also have to estimate some parameters. While in an existing network, most of the required parameters could be measured, we model a non-existent network and therefore use a mixture of intuition, experience and some references to estimate parameters. In line with already existing CFR networks (see, e.g., Hartslagnu (2024) who consider radii between 1-2 km) we decided to set a radius of $r = 1$ km for the alert area. In order to distinguish the travel modes and to have a reasonable number of alert options, we have chosen $n = 5$ rings with a bandwidth of 200 meters each. Moreover, based on a recent study by Slaa (2020), stating that on average 25% of the CFRs are available to respond, we decided to set $\alpha = 0.25$ and, in order to keep sufficient mass of CFRs responding in the last ring, we selected $\beta = 0.9$. With respect to the speed estimates for the

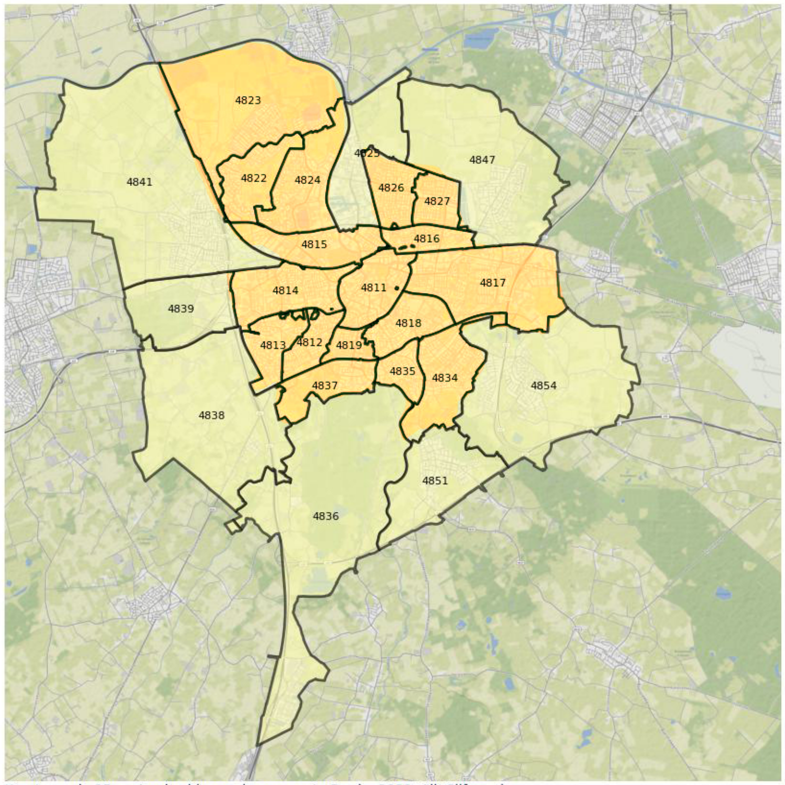

Figure 3 Map of all 25 postal codes of the municipality of Breda (2022), with the 17 selected codes in orange.

postal code	population	area	CFRs	incidents	av.amb.
4811	17,100	2.88	6	786	09:00
4812	10,210	1.97	1	329	09:35
4813	8385	2.01	4	260	09:07
4814	11,400	2.04	3	423	08:55
4815	4,355	1.10	1	193	08:13
4816	4,460	0.85	2	167	08:44
4817	15,425	3.92	2	469	09:54
4818	9,045	1.40	0	203	10:18
4819	3,630	0.58	1	101	09:02
4822	8,050	2.58	9	190	09:20
4823	8,875	3.06	5	238	10:05
4824	9,040	3.15	3	216	09:03
4826	9,335	1.94	6	420	09:17
4827	7,810	1.57	2	292	09:26
4834	11,905	2.59	2	407	09:33
4835	6,865	1.20	1	122	09:37
4837	2,255	0.60	0	98	09:42
total/average	148,145	33.44	48	4,914	09:20

Table 3 Characteristics of the 17 postal codes in focus of the city of Breda (2019)

different travel modes, we take into account that travel distances can be longer than 'as the crow flies' and assumed an interval of 3.6 km/h to 6 km/h for foot, 12 km/h to 20 km/h for bike and 20 km/h to 40 km/h for car. Next, to derive the probabilities to select a travel mode for a given ring, we started with the distance-related probabilities from Niemeijer and Buijs (2023), but as they considered non-urgent transportation, we increased the values for the faster options. The results are shown in Table 4.

Transport mode	foot	bike	car
ring 1	0.8	0.2	0
ring 2	0.4	0.5	0.1
ring 3	0.2	0.6	0.2
ring 4	0.05	0.6	0.35
ring 5	0	0.5	0.5

Table 4 Probabilities to choose transport mode per ring. These probabilities may differ for other cities or regions, depending on their population density. For instance, in rural areas, CFRs might respond with cars more often.

We also calculated the density parameter ρ per postal code. A natural way to derive this parameter is to divide the number of CFRs by the size of the area. However, because there exist some postal codes with no CFRs (leading to $\rho = 0$) we decided to calculate ρ as the average of the postal code density and the average neighboring postal code's density. These densities are listed in Section 4.1, Table 7. In addition, we have also selected distributions for the view delay, the travel mode delays and dispatch times (see Table 5). The range of the dispatch times is similar to the ones described in the study of (Hillenaar 2021).

	view delay	foot	bike	car	dispatch
distribution	NB(2,1/16)	NB(2,1/16)	NB(3,1/31)	NB(3,1/46)	U[60,120]
expectation (s)	30	30	90	135	90

Table 5 Probability distributions of view delay, the travel model delay (foot, bike and car) and dispatch times.

Using the selected velocities and probability distributions, we can calculate the expected total response times per ring and per transport mode for an arbitrary volunteer. The response times are reported in Table 6.

ring	1	2	3	4	5
foot	278	449	636	826	1016
bike	249	300	356	413	470
car	291	319	350	380	411

Table 6 Expected total response time (seconds) per ring and per transport mode for an arbitrary volunteer.

4.1 Evaluating the potential of a CFR network without retraction

We start with evaluating the performance of the CFR network in Breda without any alert retraction, that is, retract rule (a). The first performance indicator, i.e., the expected response-time reduction, when alerting $k = 1,2,3,4,5$ number of rings is represented in Table 7.

From Table 7 we learn that alerting one ring leads to a reduction of 11 seconds on average per occurring incident. This reduction increases with 20-30 seconds for every extra ring, leading to 112.9 seconds (almost two minutes) reduction for 5 rings. Hence, alerting all rings reduces the average response time from the

postal code	incidents	ρ	alert-1	alert-2	alert-3	alert-4	alert-5
4811	786	1.57	14.1	41.5	72.9	104.1	130.4
4812	329	0.98	9.9	29.9	54.0	79.8	103.5
4813	260	1.24	11.4	33.7	59.8	86.8	110.6
4814	423	1.25	11.0	32.8	58.7	85.2	108.6
4815	193	1.34	10.2	29.6	52.2	75.0	94.8
4816	167	1.74	14.7	42.4	73.4	103.5	128.1
4817	469	0.73	7.9	24.3	44.7	67.4	89.3
4818	203	0.49	5.7	17.9	33.8	51.9	70.1
4819	101	1.20	11.0	32.5	57.7	84.1	107.5
4822	190	2.27	20.8	59.6	101.8	140.2	169.6
4823	238	1.46	16.0	47.9	85.1	122.8	155.3
4824	216	1.23	11.2	33.1	59.2	86.0	109.8
4826	420	2.00	18.8	54.0	92.7	129.5	158.6
4827	292	1.32	12.9	38.2	67.8	98.2	124.8
4834	407	0.59	6.0	18.4	33.9	51.4	68.5
4835	122	0.63	6.4	19.8	36.4	55.0	73.0
4837	98	0.36	3.7	11.5	21.4	32.8	44.3
total/average	4914	1.44	11.8	34.8	61.6	89.0	112.9

Table 7 Expected response time reduction in seconds depending on number of alerted rings.

current average of 9:20 minutes (see Table 3) to 7:27 minutes. At the same time, averaged over all postal codes, alerting one ring leads to an expected CFR workload of 0.04, which increases to 0.74 when five rings are alerted. For retract rule (a) these quantities are linear to the density ρ.

For humanitarian network organizations such as the Red Cross, it may be hard to make a decision about the number of rings to alert, because the two performance indicators are in conflict with each other: alerting more rings reduces the expected response time, but at the same time also increases the total workload of CFRs. This becomes even more apparent when we also present the expected number of trips of all CFRs together and the expected response time reduction of all incidents together, which are visualized in Table 8. From this table, we learn that the expected response time reduction of all incidents together equals 9250 minutes, yearly. This sounds like an interesting potential. However, if we realize that 3620 CFR trips have to be performed by only 48 CFRs in Breda (see Table 3), then we may conclude that it is not always best to alert all rings. For instance, when we alert only 1 ring, a CFR responds more or less 4 times a year on average, with an average reduction of 5 minutes per incident per CFR.[6] In comparison, when 5 rings are alerted, the CFR responds 75 times with about a reduction of 2:30 minutes per response per CFR.

Number of alerted rings	1	2	3	4	5
Expected number of trips of all CFRs	193	714	1496	2481	3620
Expected response time reduction (min) of all incidents	963	2848	5048	7290	9250

Table 8 Results depending on number of alerted rings in Breda.

[6] The reduction of 5 minutes per incident per CFR is determined by dividing 963 by 193.

4.2 Results of maximizing effectiveness

In Table 9 we show the outcome of the three different optimization problems (i.e., for the three retract rules), in case of a yearly budget of 10, 20 and 30 trips per CFR. In Table 10, we also show these results in case all postal codes in Breda would have the national average density of 2.33 CFRs per km2, as well as with an average of 5, restricted to a budget of 10 and 20 only. In these tables, we refer to the different density scenarios as '(L)ow ($\rho = 1.44$)', '(M)edium ($\rho = 2.33$)' and '(H)igh ($\rho = 5$)'.

postal code	alert strategy								
	aL10	bL10	cL10	aL20	bL20	cL20	aL30	bL30	cL30
4811	1	1	1	2	2	2	2	2	3
4812	2	2	2	3	3	3	4	4	4
4813	2	2	1	2	2	2	3	3	3
4814	1	1	1	2	2	2	3	3	2
4815	1	1	1	1	1	2	2	1	2
4816	1	1	1	2	1	1	2	2	2
4817	3	3	3	4	4	4	5	5	5
4818	4	5	5	5	5	5	5	5	5
4819	1	2	2	1	2	2	3	3	3
4822	1	1	1	2	2	2	2	2	3
4823	3	2	4	3	3	5	4	4	5
4824	1	1	1	2	2	2	3	3	3
4826	1	1	1	2	2	2	2	2	3
4827	2	2	2	2	2	2	3	3	4
4834	3	2	2	3	4	4	4	5	4
4835	2	2	2	2	4	4	5	5	4
4837	2	4	1	5	2	4	4	5	3
reduction (min)	2092	2117	2207	3702	3790	4025	5048	5231	5656

Table 9 Number of alerted rings per rule and postal code and response time reduction over all postal codes for current number of CFRs, which we refer to as the (L)ow density scenario.

In Table 9 we notice that there are quite some differences in the optimal number of alerted rings over the postal codes, with usually more rings for low CFR density areas. Moreover, we see that the number of rings alerted does not decrease in the total amount of budget available and that retract rule (c) dominates retract rule (b), which dominates retract rule (a). This result is even stronger for the medium ($\rho = 2.33$) and high ($\rho = 5$) scenario, with 78 and 167 CFRs, respectively. For these scenarios, the effect of limiting the number of CFRs sent is stronger, as it happens more often that more than one CFR is available. Especially for the first two retract rules, the differentiation over the number of alerted rings per districts is now smaller, as all districts are assumed to have an identical CFR density. In the highest budget scenario with 3340 trips we have an average reduction of 2:48 minutes over all incidents and of 4:07 minutes for all incidents with a responding CFR. Compared to the results from Table 8 this is a considerable improvement: with almost 8 percent fewer trips, we reduce the response time with an additional 49 percent.

Although not explicitly reported in Table 8, we found that for the (L)ow $(\rho = 1.44)$ scenario, the choice for retract rule (c) was always to retract the alert after the first available CFR has accepted. Based on another numerical experiment (not reported here as well), it turns out that we need a budget of 50 trips per CFR or

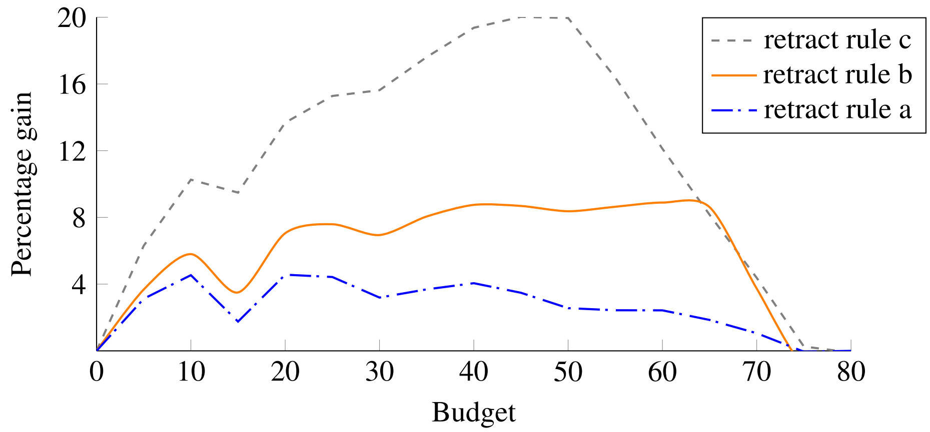

Figure 4 Response time that can be gained without increasing the budget, by differentiating the number of dispatched rings per district and potentially retract too. Gain is expressed against a benchmark policy that alerts the same number of rings everywhere

postal code	rule											
	aM10	bM10	cM10	aM20	bM20	cM20	aH10	bH10	cH10	aH20	bH20	cH20
4811	1	1	1	2	2	2	1	1	1	2	2	2
4812	2	2	2	2	2	2	2	2	2	2	2	4
4813	2	1	1	2	2	2	2	1	1	2	2	3
4814	1	1	1	2	2	2	1	1	1	2	2	3
4815	1	1	1	1	2	1	1	1	1	2	2	1
4816	1	0	1	1	2	1	1	1	1	2	2	2
4817	2	2	2	3	3	4	2	2	3	2	3	4
4818	2	2	3	3	3	5	2	2	5	3	3	5
4819	1	2	1	2	2	1	2	1	1	2	2	2
4822	1	1	2	2	2	2	1	1	2	3	2	3
4823	2	2	3	3	3	4	2	2	3	3	3	4
4824	1	1	1	2	2	2	1	1	1	2	2	2
4826	1	1	1	2	2	2	1	2	1	2	2	3
4827	1	2	2	2	2	2	1	1	2	2	2	3
4834	2	2	2	2	2	3	2	2	2	2	3	4
4835	2	2	2	2	2	3	1	2	2	3	2	4
4837	2	2	2	3	3	2	2	2	2	2	2	4
reduction (min)	3407	3451	3719	5870	6060	6744	6585	6827	7804	10790	11441	13768

Table 10 Number of alerted rings per rule (a/b/c) for a medium (M) and high (H) density scenario with 2 different budgets.

higher before to change this solution, i.e., to retract once two or more CFRs have accepted. In Figure 4 we show the relative improvement of the alert rules over a rule that uses equal x_{di}-values for all postal codes, without retracting, interpolating between the discrete values. We see that this gain is positive for all retract rules, but highest (in almost all cases) for retract rule (c). The highest gain is realized for a budget around 45-50, implying that differentiation in the number of rings to alert is maximized here.

5 Conclusion and Discussion

In this paper, we introduced a mathematical model that can estimate the expected response-time reduction and the expected CFR workload of an as-yet unrealized CFR network. To balance the two performance indicators, we introduced several alert and retract rules to indicate the size of the area to alert, and whether outstanding alerts should be retracted once some people have responded. By means of a case study, we analyzed these rules and quantify the potential –in terms of the two performance indicators– of a CFR network for the city Breda. For this case study, our model predicted that if CFRs respond around 20 times a year, which means around 20% of all ambulance trips, response times can be improved by 2 minutes on average compared to ambulance arrival. This reduction can even approach 4 minutes, when the number of CFRs will double more than twice (cf. the (M)edium and (H)igh scenarios). These are significant reductions, and could considerably change the health of hundreds of patients. On top of this, the case study also demonstrated the potential of retracting outstanding alerts after some CFRs have responded, especially at postal codes with a relative high number of CFRs. More specifically, retracting significantly reduces workload in the (H)igh scenario with 20 percent while only increasing the response time by a few seconds.

The numerical results in this paper are established under some assumptions. For instance, one is that CFRs are always deployed, once an ambulance is sent out. In practice, however, there might be considerations regarding CFRs. For instance, it is very well possible that for some urgent ambulance calls, CFRs are insufficiently trained. One way to deal with such a setting would be to compare the CFR response times with only those ambulance response times for which CFRs are of actual help. Figuring this out might, however, be quite complicated as it requires an analysis on many types of incidents, the capabilities/skills of CFRs as well as the subsequent question whether a CFR –with a certain set of capabilities– is able to handle a specific type of incident. As a proxy for this, one could perform a sensitivity analysis on the number of CFRs available –in the same spirit as we did in the numerical experiment with (L)ow, M(edium), and (H)igh CFR density, for the setting with all ambulances response times. The setting with a low density could then represent the case where only a subset of CFRs is trained to respond to specific types of incidents.

Closely related to this, one of our performance indicators is the response-time reduction. Although this performance indicator might be a good proxy of the potential benefit for society, ideally one would like to measure patient outcomes instead. However, this would require very detailed data, like 'What is the expected health gain of a patient with problem "X", when the first CFR arrives t_1 seconds after the emergency call and the ambulance arrives $t_2 > t_1$ seconds after the call?'. So far, this data is only available for very specific health issues, mostly expressed as survival data for cardiac arrests (which only represents 2% of the calls). Given the inability to translate response times to patient outcomes for the majority of the calls, we decided that the response-time reduction is the most appropriate performance indicator for now.

On a more technical note, we assumed in our numerical experiments that the yearly 'budget' of CFR trips is spent over all CFRs. This implies that on average, no CFR is too "busy". However, it might very well be that some CFRs are still very busy (e.g., a call every day) –and this seems to be undesirable from a central workload perspective. For that reason, we invested for our numerical experiment how busy the busiest CFR would be. For our 3H20-scenario, with a budget of 20 per volunteer, which is the most extreme scenario considered, we learned that the maximum average CFR use per district is 32. Based on a recent report by Slaa (2020), this seems to be an acceptable number per year for CFRs, supporting our assumption.

Future research directions include an investigation of the relative importance of response times and volunteer workload. Knowing this would inform the desired retract rule and consequently also the IT specifications or a CFR system. Moreover, one could make better-informed decisions if there was more detailed knowledge of how people move around their neighborhood. We are aware of at least one CFR system that tracks this; however, it is questionable whether that will ever be made public.

Finally, the developed model could also have emergency applications outside healthcare. As an example, consider off-duty police professionals that could be alarmed to assist in case of a terrorist attack (see, e.g., van Aken et al. (2024)). Alternatively, fire departments can use a CFR system to rapidly mobilize farmers equipped with large water-carrying trucks, typically used for crop irrigation, to preemptively saturate designated areas and slow the spread of wildfires.

Acknowledgments

We thank the regional ambulance provider Brabant Midden-West-Noord for allowing us to use their response-time data. Author CJJ thanks the SPRINTER consortium for fruitful discussions and the organizers of the Community First Responder Symposium 2024 for hosting their event.

Funding sources

This work was executed in part by author LS under NWO Grant VI.Veni.201E.017 and by author CJJ under TKI Dinalog grant number 2023-1-307TKI.

References

Barry, Tomás, Ainhoa González, Niall Conroy, Paddy Watters, Siobhán Masterson, Jan Rigby, Gerard Bury. 2018. Mapping the potential of community first responders to increase cardiac arrest survival. *Open Heart*, 5 (2), e000912.

Botan, Vanessa, Zahid Asghar, Elise Rowan, Murray D Smith, Gupteswar Patel, Viet-Hai Phung, Ian Trueman, Robert Spaight, Amanda Brewster, Pauline Mountain, et al. 2022. Community first responders' contribution to emergency medical service provision in the united kingdom. *Annals of Emergency Medicine*, .

Campbell, Alan, Matt Ellington. 2016. Reducing time to first on scene: an ambulance-community first responder scheme. *Emergency medicine international*, 2016.

CBS. 2022. Kerncijfers per postcode. `https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/geografische-data/gegevens-per-postcode`. Accessed: 2024-01-22.

Dawson, Drew, Katherine Mayger, Matthew JW Thomas, Kirrilly Thompson. 2015. Fatigue risk management by volunteer fire-fighters: Use of informal strategies to augment formal policy. *Accident Analysis & Prevention*, 84 92-98.

Dieleman, Nanne A., Caroline Jagtenberg. 2024. Electric ambulances: will the need for charging affect response times? *Available at SSRN: https://ssrn.com/abstract=4874479*, .

Enayati, Shakiba, Maria E Mayorga, Hari K Rajagopalan, Cem Saydam. 2018. Real-time ambulance redeployment approach to improve service coverage with fair and restricted workload for ems providers. *Omega*, 79 67-80.

Erkut, Erhan, Armann Ingolfsson, Güneş Erdoğan. 2008. Ambulance location for maximum survival. *Naval Research Logistics (NRL)*, 55 (1), 42-58.

Freeman, Victoria A, Rebecca T Slifkin, P Daniel Patterson. 2009. Recruitment and retention in rural and urban ems: results from a national survey of local ems directors. *Journal of Public Health Management and Practice*, 15 (3), 246-252.

HartslagNu. 2021. Let op: de alarmeringsstraal is gewijzigd. wat houdt dit in? URL `https://hartslagnu.nl/nieuws/let-op-de-alarmeringsstraal-is-gewijzigd-wat-houdt-dit-in/`. Accessed: 2023-08-15.

Hartslagnu. 2024. De alarmeringsstraal. `https://hartslagnu.nl/burgerhulpverlening/de-alarmeringsstraal/`. Accessed: 2024-01-22.

Henderson, Shane G, Pieter L van den Berg, Caroline J Jagtenberg, Hemeng Li. 2022. How should volunteers be dispatched to out-of-hospital cardiac arrest cases? *Queueing Systems*, 100 (3-4), 437-439.

Hillenaar, Janneke. 2021. The potential and feasibility of a community first responders network to reduce the response time for urgent cases in the dutch pre-hospital emergency system. Master's thesis Technische Universiteit Eindhoven.

Houweling, A. 2021. Waarom Nederlandse en Belgische ambulances nog steeds niet samenwerken. URL `https://www.zuidwestupdate.nl/nieuws/waarom-nederlandse-en-belgische-ambulances-nog-steeds-niet-samenwerken/`. Accessed: 2022-09-18.

Jackson, M. 2004. Volunteers boost ambulance crews. URL `http://news.bbc.co.uk/2/hi/health/3604976.stm`. Accessed: 2022-09-21.

Jagtenberg, Caroline J, Sandjai Bhulai, Robert D van der Mei. 2015. An efficient heuristic for real-time ambulance redeployment. *Operations Research for Health Care*, 4 27-35.

Jagtenberg, Caroline J, Sandjai Bhulai, Robert D van der Mei. 2017. Dynamic ambulance dispatching: is the closest-idle policy always optimal? *Health care management science*, 20 517-531.

Jonsson, Martin, Ellinor Berglund, Therese Djärv, Per Nordberg, Andreas Claesson, Sune Forsberg, Anette Nord, Hanno L Tan, Mattias Ringh. 2020. A brisk walk—real-life travelling speed of lay responders in out-of-hospital cardiac arrest. *Resuscitation*, 151 197-204.

Khalemsky, Michael, David G Schwartz, Raphael Herbst, Eli Jaffe. 2020. Motivation of emergency medical services volunteers: a study of organized good samaritans. *Israel Journal of Health Policy Research*, 9 (1), 1-12.

Lancaster, Greg, Jeffrey W Herrmann. 2021. Computer simulation of the effectiveness of novel cardiac arrest response systems. *Resuscitation Plus*, 7 100153.

Li, Xueping, Zhaoxia Zhao, Xiaoyan Zhu, Tami Wyatt. 2011. Covering models and optimization techniques for emergency response facility location and planning: a review. *Mathematical Methods of Operations Research*, 74 281-310.

Matinrad, Niki, Melanie Reuter-Oppermann. 2022. A review on initiatives for the management of daily medical emergencies prior to the arrival of emergency medical services. *Central European Journal of Operations Research*, 30 (1), 251-302.

Maxwell, Matthew S, Eric Cao Ni, Chaoxu Tong, Shane G Henderson, Huseyin Topaloglu, Susan R Hunter. 2014. A bound on the performance of an optimal ambulance redeployment policy. *Operations Research*, 62 (5), 1014-1027.

Maxwell, Matthew S, Mateo Restrepo, Shane G Henderson, Huseyin Topaloglu. 2010. Approximate dynamic programming for ambulance redeployment. *INFORMS Journal on Computing*, 22 (2), 266-281.

Mell, Howard K, Shannon N Mumma, Brian Hiestand, Brendan G Carr, Tara Holland, Jason Stopyra. 2017. Emergency medical services response times in rural, suburban, and urban areas. *JAMA surgery*, 152 (10), 983-984.

Metelmann, Camilla, Bibiana Metelmann, Dorothea Kohnen, Peter Brinkrolf, Linn Andelius, Bernd W Böttiger, Roman Burkart, Klaus Hahnenkamp, Mario Krammel, Tore Marks, et al. 2021. Smartphone-based dispatch of community first responders to out-of-hospital cardiac arrest-statements from an international consensus conference. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, 29 (1), 1-9.

Nasrollahzadeh, Amir Ali, Amin Khademi, Maria E Mayorga. 2018. Real-time ambulance dispatching and relocation. *Manufacturing & Service Operations Management*, 20 (3), 467-480.

Niemeijer, R, P Buijs. 2023. A greener last mile: Analyzing the carbon emission impact of pickup points in last-mile parcel delivery. *Renewable and Sustainable Energy Reviews*, 186. doi:https://doi.org/10.1016/j.rser.2023.113630.

Oving, Iris, Siobhan Masterson, Ingvild BM Tjelmeland, Martin Jonsson, Federico Semeraro, Mattias Ringh, Anatolij Truhlar, Diana Cimpoesu, Fredrik Folke, Stefanie G Beesems, et al. 2019. First-response treatment after out-of-hospital cardiac arrest: a survey of current practices across 29 countries in Europe. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, 27 (1), 112.

Pauw, S. 2021. Aanrijtijden ambulances verder verslechterd. URL `https://www.medischcontact.nl/nieuws/laatste-nieuws/nieuwsartikel/aanrijtijden-ambulances-verder-verslechterd.htm`. Accessed: 2022-09-21.

Paz, Juan Camilo, Nan Kong, Seokcheon Lee. 2022. Incorporating real-time citizen responder information to augment ems logistics operations: A simulation study. *Computers & Industrial Engineering*, 171 108399.

Ridler, Samuel, Andrew J Mason, Andrea Raith. 2022. A simulation and optimisation package for emergency medical services. *European Journal of Operational Research*, 298 (3), 1101-1113.

Sarkisian, Laura, Hans Mickley, Henrik Schakow, Oke Gerke, Gitte Jørgensen, Mogens Lytken Larsen, Finn Lund Henriksen. 2020. Global positioning system alerted volunteer first responders arrive before emergency medical services in more than four out of five emergency calls. *Resuscitation*, 152 170-176.

Schmid, Verena. 2012. Solving the dynamic ambulance relocation and dispatching problem using approximate dynamic programming. *European journal of operational research*, 219 (3), 611-621.

Shrader, Kristopher W. 2012. Improving the recruitment and retention of volunteer firefighters.

Slaa, Gunnar. 2020. Increasing cardiac arrest survival by improving the volunteer alerting algorithm. Master's thesis, University of Twente.

Valeriano, Ailish, Shyan Van Heer, François de Champlain, Steven C. Brooks. 2021. Crowdsourcing to save lives: A scoping review of bystander alert technologies for out-of-hospital cardiac arrest. *Resuscitation*, 158 94-121. doi:https://doi.org/10.1016/j.resuscitation.2020.10.035. URL `https://www.sciencedirect.com/science/article/pii/S030095722030544X`.

van Aken, Lotte, Loe Schlicher, Marco Slikker, Geert-Jan van Houtum. 2024. Fighting terrorism: How to position rapid response teams? *Naval Research Logistics (NRL)*, .

Van den Berg, Pieter, Shane G. Henderson, Caroline Jagtenberg, Hemeng Li. 2024. Modeling emergency medical service volunteer response. *Management Science*, To Appear. doi:10.2139/ssrn.3825060.

van der Worp, Wim. 2014. Resuscitation by text-message responders in The Netherlands. *Resuscitation*, 85 S55.

Zijlstra, Jolande A, Stefanie G Beesems, Rob J De Haan, Rudolph W Koster. 2015. Psychological impact on dispatched local lay rescuers performing bystander cardiopulmonary resuscitation. *Resuscitation*, 92 115-121.

Zijlstra, Jolande A, Remy Stieglis, Frank Riedijk, Martin Smeekes, Wim E van der Worp, Rudolph W Koster. 2014. Local lay rescuers with aeds, alerted by text messages, contribute to early defibrillation in a dutch out-of-hospital cardiac arrest dispatch system. *Resuscitation*, 85 (11), 1444-1449.

Appendix. 1

In this Appendix, we derive F_3, the distribution of travel time. In doing so, we first present the distribution of the travel time, given that velocity is fixed at $v \in \mathbb{R}_{>0}$. Because volunteers are uniformly distributed, this distribution, for a given $t \in \mathbb{N}_{\geq 0}$ coincides with the area under the ring with inner radius d_{i-1} and outer radius $\min\{d_{i-1}, v \cdot t\}$, divided by the area under the ring with inner radius d_{i-1} and outer radius d_i. Formally, for a given $t \in \mathbb{N}_{\geq 0}$ $i \in N$ and $m \in M$, we obtain the following distribution:

$$F_3(i,m,t,v) = \begin{cases} 0 & \text{if } t \leq \frac{d_{i-1}}{v} \\ \frac{(v \cdot t)^2 - d_{i-1}^2}{d_i^2 - d_{i-1}^2} & \text{if } \frac{d_{i-1}}{v} \leq t \leq \frac{d_i}{v} \\ 1 & \text{if } t \geq \frac{d_i}{v} \end{cases}.$$

In order to arrive at F_3, w need to differentiate between six different cases. In doing so, we first fix $m \in M$ and $i \in N$. Now, we distinguish between the six cases.

Case 1. $v_{m1} \cdot t \leq d_{i-1}$

In this case, even with the fastest velocity, volunteers cannot be on time at the incident. Hence,

$$F_3(i,m,t) = 0 \text{ for all } t \in \mathbb{N}_{\geq 0}.$$

Case 2. $v_{m0} \cdot t \leq d_{i-1}$ *and* $d_{i-1} \leq v_{m1} \cdot t \leq d_i$

Volunteers traveling at the lowest speed cannot be on time, but a fraction of the volunteers in the ring is on time when traveling at the highest velocity. Hence for any $t \in \mathbb{N}_{\geq 0}$, we have

$$\begin{aligned} F_3(i,m,t) &= \int_{\frac{d_{i-1}}{t}}^{v_{m1}} \frac{1}{v_{m1} - v_{m0}} F_3(i,m,t,v)dv \\ &= \frac{1}{v_{m1} - v_{m0}} \int_{\frac{d_{i-1}}{t}}^{v_{m1}} \frac{(v \cdot t)^2 - d_{i-1}^2}{d_i^2 - d_{i-1}^2} dv \\ &= \frac{\frac{1}{3} v_{m1}^3 t^2 - d_{i-1}^2 v_{m1} + \frac{2}{3} \frac{d_{i-1}^3}{t}}{(v_{m1} - v_{m0})(d_i^2 - d_{i-1}^2)}. \end{aligned}$$

Case 3. $v_{m0} \cdot t \leq d_{i-1}$ *and* $v_{m1} \cdot t \geq d_i$

Volunteers traveling at the lowest speed cannot be on time, while all volunteers in the ring are on time when traveling at the highest velocity. Hence for any $i \in N, m \in M$ and $t \in \mathbb{N}_{\geq 0}$, we have

$$\begin{aligned} F_3(i,m,t) &= \int_{\frac{d_{i-1}}{t}}^{\frac{d_i}{t}} \frac{1}{v_{m1} - v_{m0}} F_3(i,m,t,v)dv + \int_{\frac{d_i}{t}}^{v_{m1}} \frac{1}{v_{m1} - v_{m0}} F_3(i,m,t,v)dv \\ &= \frac{1}{v_{m1} - v_{m0}} \int_{\frac{d_{i-1}}{t}}^{\frac{d_i}{t}} \frac{(v \cdot t)^2 - d_{i-1}^2}{d_i^2 - d_{i-1}^2} dv + \frac{1}{v_{m1} - v_{m0}} \int_{\frac{d_i}{t}}^{v_{m1}} 1 dv \\ &= \frac{\frac{2}{3t}(d_{i-1}^3 - d_i^3) + v_{m1}(d_i^2 - d_{i-1}^2)}{(v_{m1} - v_{m0})(d_i^2 - d_{i-1}^2)}. \end{aligned}$$

Case 4. $d_{i-1} \leq v_{m0} \cdot t \leq d_i$ *and* $d_{i-1} \leq v_{m1} \cdot t \leq d_i$

A fraction of the volunteers traveling at the lowest speed is on time, and this holds as well for the volunteers traveling at the highest speed. Hence for any $i \in N, m \in M$ and $t \in \mathbb{N}_{\geq 0}$, we have

$$\begin{aligned} F_3(i,m,t) &= \int_{v_{m0}}^{v_{m1}} \frac{1}{v_{m1} - v_{m0}} F_3(i,m,t,v) dv \\ &= \frac{1}{v_{m1} - v_{m0}} \int_{\frac{d_{i-1}}{t}}^{\frac{d_i}{t}} \frac{(v \cdot t)^2 - d_{i-1}^2}{d_i^2 - d_{i-1}^2} dv \\ &= \frac{\frac{1}{3} t^2 (v_{m1}^3 - v_{m0}^3) - d_{i-1}^2 (v_{m1} - v_{m0})}{(v_{m1} - v_{m0})(d_i^2 - d_{i-1}^2)}. \end{aligned}$$

Case 5. $d_{i-1} \leq v_{m0} \cdot t \leq d_i$ *and* $v_{m1} \cdot t \geq d_i$

A fraction of the volunteers traveling at the lowest speed is on time, while all volunteers traveling at the highest speed are on time. Hence for any $i \in N, m \in M$ and $t \in \mathbb{N}_{\geq 0}$, we have

$$\begin{aligned} F_3(i,m,t) &= \int_{v_{m0}}^{\frac{d_i}{t}} \frac{1}{v_{m1} - v_{m0}} F_3(i,m,t,v) dv + \int_{\frac{d_i}{t}}^{v_1} \frac{1}{v_{m1} - v_{m0}} F_3(i,m,t,v) dv \\ &= \frac{1}{v_{m1} - v_{m0}} \int_{v_{m0}}^{\frac{d_i}{t}} \frac{(v \cdot t)^2 - d_{i-1}^2}{d_i^2 - d_{i-1}^2} dv + \frac{1}{v_{m1} - v_{m0}} \int_{\frac{d_i}{t}}^{v_1} 1 dv \\ &= \frac{\left(-\frac{2}{3} \frac{d_i^3}{t} - \frac{1}{3} v_{m0}^3 t^2 + d_{i-1}^2 v_{m0} + v_{m1} (d_i^2 - d_{i-1}^2) \right)}{(v_{m1} - v_{m0})(d_i^2 - d_{i-1}^2)}. \end{aligned}$$

Case 6. $v_{m0} \cdot t \geq d_i$

In this case, even with the slowest velocity, all volunteers are on time at the incident. Hence,

$$F_3(i,m,t) = 1 \text{ for all } t \in \mathbb{N}_{\geq 0}.$$

Appendix. 2

In this appendix, we consider the retract rules (b) and (c) from Section 3.5. In these rules the sequence of responding plays a large role, which makes the distribution of the home response delay no longer independent of the CFR, but dependent on this sequence. Therefore the order rank statistics for the home response delay will be used. Let $F_H^{jk}(t)$ be the view delay distribution for the CFR whose view delay rank is number j out of k available CFRs

$$F_H^{jk}(t) = \sum_{m=j}^{m} \binom{k}{m} F_1^m(t)(1-F_1(t))^{k-m} \text{ for } j \leq k \tag{9}$$

From this we can determine $F_R^{jk}(i,t)$, the CFR response-time distribution for the CFR in ring i whose view delay rank is number j out of k available CFRs, similar to equation (3):

$$F_R^{jk}(i,t) = \sum_{t'=0}^{t} \sum_{t_1=0}^{t'} \sum_{m=1}^{3} p_{im} \sum_{t_2=0}^{t'-t_1} f_H^{jk}(t_1) \cdot f_2(m,t_2) \cdot f_3(i,m,t'-t_1-t_2) \text{ for all } t \in \mathbb{N}_{\geq 0}. \tag{10}$$

As all CFRs have the same pdf for the view delay, all sequences of responding available CFRs have the same probability. We will now derive the expected number of CFR trips per incident and the expected response-time reduction. Denote the available CFR vector (CFR-allocation) by $\bar{a}$ and the responding CFR vector $\bar{r}$. The responding CFRs are the ones, whose alert has not been retracted. Obviously, $r_i \leq a_i$ for every ring i.

In retract rule (b), a CFR from an outer ring i will only respond if the view delay rank is lower than that of all $c_i := \sum_{j=1}^{i-1} a_j$ CFRs in its inner rings. The probability for the lower delay rank is $1/(1+c_i)$ and the corresponding reponse-time distribution for this CFR is $F_R^{1c_i}(i,t)$. For a district d with λ_{dj} expected available CFRs for ring j, we then first calculate the probability to find CFR-allocation a:

$$f_{AV}^i(a) = \prod_{j \leq i} \left(\frac{\lambda_{dj}^{a_j}}{a_j!} \cdot e^{-\lambda_{dj}} \right) \tag{11}$$

$$\mathbb{E}_{id}[\text{CFR workload}] = \sum_{a \in \mathbb{N}^i} f_{AV}^i(a) \left(a_1 + \sum_{j=2}^{i} \frac{a_j}{1+c_j} \right) \tag{12}$$

The probability that the vector of responding CFRs equals r, with $r_1 = a_1$ (and $r_j \leq a_j$ for $2 \leq j \leq i$), for a given CFR-allocation a, is given by

$$\mathbb{P}(\bar{r}|\bar{a}) = \prod_{j=2}^{i} \binom{a_j}{r_j} \left(\frac{a_j}{1+c_j}\right)^{r_j} \left(\frac{c_j}{1+c_j}\right)^{a_j-r_j} \tag{13}$$

$$F_{RFA}^{(b)}(i,t) = \sum_{t_0=0}^{t} F_0(t_0) \sum_{a \in \mathbb{N}^i} f_{AV}^i(a) \left(1 - \sum_{r} \mathbb{P}(\bar{r}|\bar{a}) \prod_{j \leq i} \left(1 - F_R^{1c_j}(j, t-t_0)\right)^{r_j} \right) \text{ for all } t \in \mathbb{N}_{\geq 0}. \tag{14}$$

The expected response-time reduction for retract rule (b) is then determined by:

$$\mathbb{E}_{id}[\text{response time reduction}] = \sum_{t\in\mathbb{N}_{\geq 0}} f_{AM}^{d}(t) \sum_{t'=0}^{t} (t-t')F_{RFA}^{(b)}(t'). \tag{15}$$

For retract rule (c), with a maximum of s CFRs per incident, the workload can be easily determined:

$$\mathbb{E}_{ids}[\text{CFR workload}] = \sum_{a\in\mathbb{N}^i} \min\left(s, \sum_{j=1}^{i} a_j\right) f_{AV}^{i}(a) \tag{16}$$

In the calculation of the response-time distribution, we have to distinguish two cases, depending on the total number of available CFRs within the i alerted rings, compared the the maximum number s. For low numbers of availabe CFRs, formula (4) can still be applied, but when more than s CFRs are available, the rank of the view delay becomes important. Let $g(k)$ denote the ring of the $k-$th ranked CFR, in terms of view delay, with $g(k) \in (1,..,i)$. Then we can write the response-time distribution as:

$$F_{RFA}^{(c)}(i,t) = \sum_{|a|<=s} f_{AV}^{i}(a) F_{RFA}(t) + \sum_{|a|>s} f_{AV}^{i}(a) \sum_{t_0=0}^{t} F_0(t_0) \left(1 - \sum_{r} \mathbb{P}(\bar{g}) \prod_{j=1}^{s} \left(1 - F_R^{j,|a|+1}(g(j), t-t_0)\right)\right) \text{ for all } t \geq 0. \tag{17}$$

Here,

$$\mathbb{P}\left(g(\bar{1}) = j\right) = \frac{a_j}{|a|} \tag{18}$$

$$\mathbb{P}\left(g(\bar{k}) = j\right) = \frac{a_j - \sum_{m=1}^{k-1} 1(g(m)=j)}{|a|+1-j} \tag{19}$$

Finally, the response-time reduction for retract rule (c) is thus determined by:

$$\mathbb{E}_{ids}[\text{response time reduction}] = \sum_{t\in\mathbb{N}_{\geq 0}} f_{AM}^{d}(t) \sum_{t'=0}^{t} (t-t')F_{RFA}^{(c)}(t'). \tag{20}$$