

SHORT CONTRIBUTION

Open Access



Operationalizing deployment time in police calls for service

Samuel Langton¹, Tim Verlaan¹ and Stijn Ruiter^{1,2*} 

Abstract

Analyses of emergency calls for service data in the United States suggest that around 50% of dispatched police deployment time is spent on crime-related incidents. The remainder of time is spent in a social service capacity: attending well-being checks and resolving disturbances, for instance. These findings have made a considerable contribution to the discourse around public perceptions of the police and the distribution of public funds towards (or away) from law enforcement. Yet, an outstanding issue remains. No investigation has been undertaken into whether findings are robust to the different ways in which ‘time spent’ is operationalized in these studies. Using dispatch data for Amsterdam during 2019, this study compares three operationalizations of ‘time spent’. Additionally, in order to provide some context on the potential mechanisms through which these different operationalizations might yield different results, we report on dispatch numbers per incident category and provide an initial exploration into ‘multi-dispatch’ incident types. We find that general proportional breakdowns are fairly robust to the time measure used. However, for some incident categories (e.g. Health) and incident types (e.g. Shootings), analyzed in isolation, the results are not robust to the different operationalizations. We propose that the mechanism explaining this lack of robustness can be traced to the high dispatch numbers for specific incident categories and types, particularly those with an imminent threat to life.

Preregistration: This study has been preregistered under the title: *Scale and composition of emergency reactive police demand in Amsterdam, Netherlands* (<https://osf.io/qgww6/>).

Keywords Police, Calls for service, Demand

Background

Recent years have seen a number of studies investigate the scale (‘how much?’) and composition (‘what type?’) of demand for and supply of police services. In other words: what kinds of incidents do the public request police assistance for, and how much time is consumed by attending these incidents? Findings from the US have consistently demonstrated that a considerable amount of dispatched

police deployment time is consumed by incidents which do not involve crime, such as resolving community issues, attending mental health crises and dealing with traffic incidents (Langton et al., 2022; Lum et al., 2021; Ratcliffe, 2021). Amidst calls for a radical reform and defunding of police forces, these studies have raised an important discussion around the demand for and supply of policing services. With a considerable proportion (if not, the majority) of dispatched police deployment time consumed in response to requests for public assistance that do not (directly) involve crime, a rapid reduction in their capacity to meet this demand might have detrimental effects on public safety and well-being.

While findings have made a contribution to the discourse, a pressing issue remains: there has been no

*Correspondence:

Stijn Ruiter
s.ruiter@uu.nl

¹ Netherlands Institute for the Study of Crime and Law Enforcement, Amsterdam, Netherlands

² Department of Sociology, Utrecht University, Utrecht, Netherlands



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

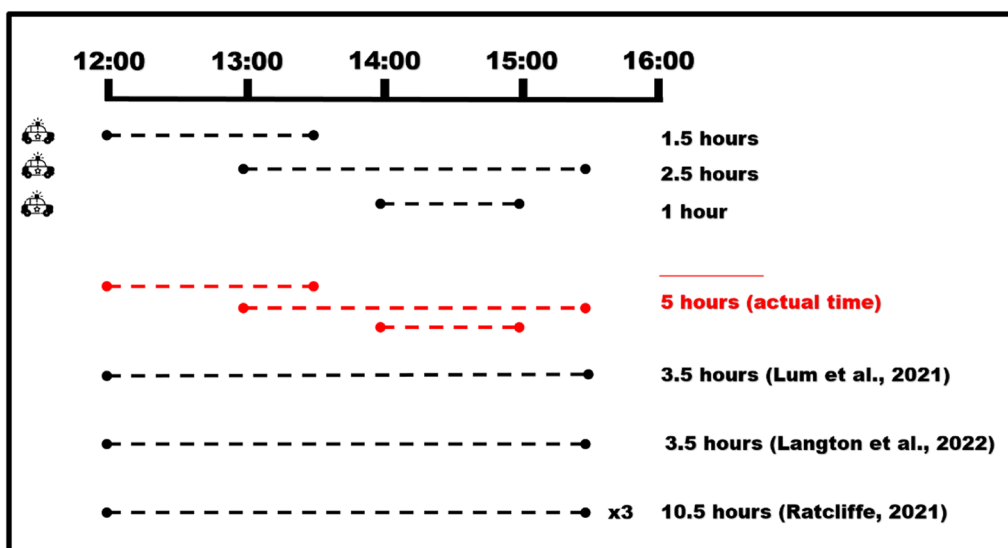


Fig. 1 Visualizing the different operationalizations of time spent. For the stylized example, we assume all operationalizations use the same start time

scrutiny of how ‘police dispatched deployment time’ is measured. We consider this issue to be particularly critical given the attention that some research has been given in mainstream media (Lum & Koper, 2021) and the impact that this might have on public sentiment and government policy. Lum et al. (2021), Ratcliffe (2021) and Langton et al. (2022) operationalized dispatched deployment time at the incident level: time measurements were taken from when the first police unit arrived on the scene, to when the last unit cleared the scene. We broadly define this measure as ‘first in last out’ (FILO). Ratcliffe (2021) provided a weighting to the operationalization by multiplying ‘FILO’ by the number officers that attended the incident. As visualized in Fig. 1, using a stylized example for a multi-dispatch incident, the operationalizations of Lum et al., (2021) and Langton et al. (2022) would underestimate the total dispatched deployment time, while Ratcliffe (2021) would overestimate it.

However, it is unclear how these different operationalizations might affect the proportional breakdowns of ‘time spent.’ If the underestimation or overestimation is uniformly distributed across all incident categories, the analytical outcomes will be robust to different operationalizations. As we have access to unit-level dispatch data, we can examine whether this is the case, or if any operationalizations introduce enough bias to compromise robustness. For clarity, we define ‘robust’ findings as those that consistently answer a research question, regardless of the time measurement used. This means that the level of robustness can vary depending on the specific research question.

With this research gap in mind, in this contribution we systematically compare different operationalizations of ‘dispatched deployment time’ using both incident and unit-level dispatch data in order to ascertain the impact on proportional breakdowns of time spent. We also report descriptive statistics on minutes consumed and the concentration of time according to incident classifications. This part of the study has been preregistered on an Open Science Framework repository (<https://osf.io/qgww6/>). To better grasp the prevalence of multi-dispatch incidents and their potential impact on time spent measures, we report descriptive statistics on dispatch numbers per incident class and visualize the distribution underlying the top 10 ‘multi-dispatch’ incident types in our data. Lastly, we share insights from our experience of double-coding an existing classification of police incidents (Ratcliffe, 2021).

Data & methods

Unit-level dispatch data

The data for this study are unit-level emergency dispatch records for the Amsterdam Police Unit during 2019. For this study, we only include incidents that occurred within the boundaries of the city itself and for which police units were dispatched. We refer readers to the preregistration document for further details on how the raw data were handled in preparation for analysis.

Table 1 Description of the different time operationalizations

Time measure	Description
Dispatched to clear (Cumulative)	Cumulative timespan between unit dispatched and unit cleared across all response units for an incident
Dispatched to clear (standard FILO)	Timespan between first unit dispatched and final unit cleared
Dispatched to clear (FILO*UNITS)	Timespan between first unit arrived and final unit cleared, multiplied by the number of units who have been present on scene

Operationalization of ‘time spent’

Besides the stated differences between existing studies in how time spent is operationalized, studies can also use different ‘start’ points. Ratcliffe (2021) used *dispatch* as the first timestamp, thereby including *travel time* in the operationalization of *time spent*. Lum and colleagues (2021) exclude *travel time* by using the *on-scene* timestamp as the start time. This is justified on the basis that it excludes potential *idle time* (“response times vary dramatically across different call types and are actually much longer than laypersons believe because officers do not immediately respond to many calls, given their low priority”, p. 267) and for what we would consider measurement error (“officers may be assigned to calls while on other calls, or can be reassigned to higher priority calls during a current call assignment”, footnote 14).

In the Netherlands, the dispatch system cannot link units to multiple calls, and the data are therefore not susceptible to that source of measurement error. Moreover, as agency over response prioritization mainly lies with the dispatch center, we can expect to see a much lower *idle time*. Nevertheless, the system does register an *on-way* time stamp for when the unit signals that they have begun their response to an incident. Indeed, the mean amount of time that elapses between *dispatch* and *on-way* is short (around 2 min) even across different incident types (see supplementary materials). We do not make any definitive claim as to what constitutes the ‘most appropriate’ start time. However, we do have the data available to test the impact of these different start times, in addition to the unit or dispatch-level operationalizations discussed above. Here, we would note that in our data 30% of responses for *on-way* are missing. This can occur due to technical malfunctions, time-pressure, forgetting and/or non-compliance.

We report on three operationalizations of time spent, as summarized in Table 1.¹ These three measures cover both existing FILO measures (weighted by the number of units and unweighted) in addition to a unit-level

cumulative sum. In all three operationalizations, we define the start point as *dispatch* although we report results for the same three operationalizations using both *on-way* and *on-scene* start points in the supplementary materials (nine in total). As such, we provide a comprehensive overview of the various different operationalizations available.²

Call classification

Call handlers classify calls according to a pre-specified list of incident types. We excluded those classifications which were clearly not initiated by the public. This left us with around 110,000 unique incidents during 2019 which required just over 220,000 dispatches. In the interests of consistency and international comparison, we then classified these existing incident types according to the demand classification first used by Ratcliffe (2021) and later Langton et al. (2022), namely: crime, health, traffic, community and quality of life. The classification procedure was undertaken by two raters independently, as outlined in the preregistration document. The degree of inter-rater reliability between the two raters was assessed using the Krippendorff’s alpha statistic (Krippendorff, 2011). The result (0.701, $p < 0.05$) exceeded the threshold stated in the preregistration document. As a result, the classification was considered appropriate and reliable enough for the Dutch context. Disagreements were resolved through group discussion together with the third author (author 3 initials). This highlighted a consistent issue, namely that of the demand classification ‘suspicious situation’. The raters could not definitively agree on whether this incident type could be classifiable under the Ratcliffe (2021) schema. A decision was made to create a new class for this incident type.

¹ Following a comment by an anonymous reviewer, we have changed the word ‘individual’ to ‘all’ in the description of cumulative measures for clarification, compared to the preregistration document.

² In the preregistration document, we stated that we would only report the various operationalizations with the *on-way* start point due to the concerns raised by Lum et al. (2021). At that stage, we did not realize that issues over idle time and measurement error were not a concern in the Dutch context. As mentioned, this calculation is still reported in the supplementary materials.

Table 2 Breakdown of counts and proportions of time spent

Demand class	Incident count	% incident count	% Cumul. dispatch to clear	% FILO dispatch to clear	% FILO*units dispatch to clear
Community	1166	1.06	0.54	0.66	0.38
Crime	34019	30.99	40.11	38.51	37.93
Health	6043	5.51	7.17	7.78	11.90
Quality of life	30484	27.77	19.43	21.59	17.56
Suspicious situation	22864	20.83	17.59	17.42	15.34
Traffic	15193	13.84	15.16	14.05	16.89

Methods

We refer readers to the preregistration document for a complete summary of the methods deployed to answer our main question, namely, how robust findings on proportional breakdowns of time are to different operationalizations of ‘dispatched deployment time’. We also report on dispatch numbers using descriptive statistics and an accompanying visualization on the distribution according to specific incident types. This second part of the analysis serves as a preliminary exploration into multi-dispatch incidents, given their potential role in determining dispatched deployment time calculations (see Fig. 1), and has not been preregistered.

Results

Demand breakdown

The overall breakdown of time according to each of the three operationalizations is summarized in Table 2. This highlights a number of initial insights. First, similar to findings in the US, in Amsterdam the public rely on the police for a diverse array of issues which are, more often than not, non-criminal. Taking the unit-level measure of ‘cumulative dispatched to clearance’ as an example, the 31% of crime incidents consumed 40% of total dispatched police deployment time during the year. Quality of life consumed 19% of time, suspicious situations 18%, traffic 15%, health 7% and community just 1%.

That said, we do observe some differences in the proportional breakdown according to the different operationalizations of time. For instance, using the cumulative measure, 7% of total dispatched deployment time is consumed by health incidents, but using the weighted FILO measure, this figure is 12%. Nevertheless, the impact of using different time measures is minimal in the sense that it does not change the finding that Amsterdam police spend the majority of their reactive dispatched deployment time resolving non-crime issues. The overall ranking of the demand classifications based on this relative proportion of ‘time spent’ is also the same, irrespective of the operationalization used, with the exception of the

weighted FILO measure. Here, traffic and suspicious situation swap place. We observe these findings irrespective of the start time used (see supplementary materials).

A proportional breakdown of time for the specific incident types nested within these broad demand classifications is visualized in Fig. 2. For brevity, we visualize the cumulative dispatched to scene clearance measure in isolation. Comparable graphics for all measures are available in the supplementary materials. Again, the picture is similar irrespective of the measure used. A considerable proportion of time spent attending crime incidents is consumed by violence and theft offences. The vast majority of time consumed by traffic-related duties involve accidents. Time spent on quality of life incidents is dominated by resolving nuisance calls (e.g., noisy neighbors) and conflicts (e.g., verbal confrontations). The largest health incident category is vague: unwell and sickness. Unfortunately, we cannot differentiate between physical ill-health (e.g., a fall) and mental ill-health, although we can see some distinctions with regards to accidents and suicide attempts. We know from observational work in the dispatch center itself that ‘suspicious situation—person’ is often used for confused or vulnerable people. In the Netherlands, dispatch centers assign nearby police, ambulance and fire service to respond to CPR by default, hence the substantial amount of time consumed by such incidents.

Time spent

In Table 3, descriptive statistics summarize the amount of time (in minutes) typically consumed by the different demand classifications. Given the skew in the data, we focus on the median. Incidents involving crime and health typically consume a comparable amount of time, with a median of 43 and 42 min respectively. Community issues, quality of life cases, and investigating suspicious situations consume less time (19, 22 and 25 min, respectively). Traffic incidents, on median, take a little longer to resolve (32 min). The equivalent statistics for all time measures are available in the supplementary materials.

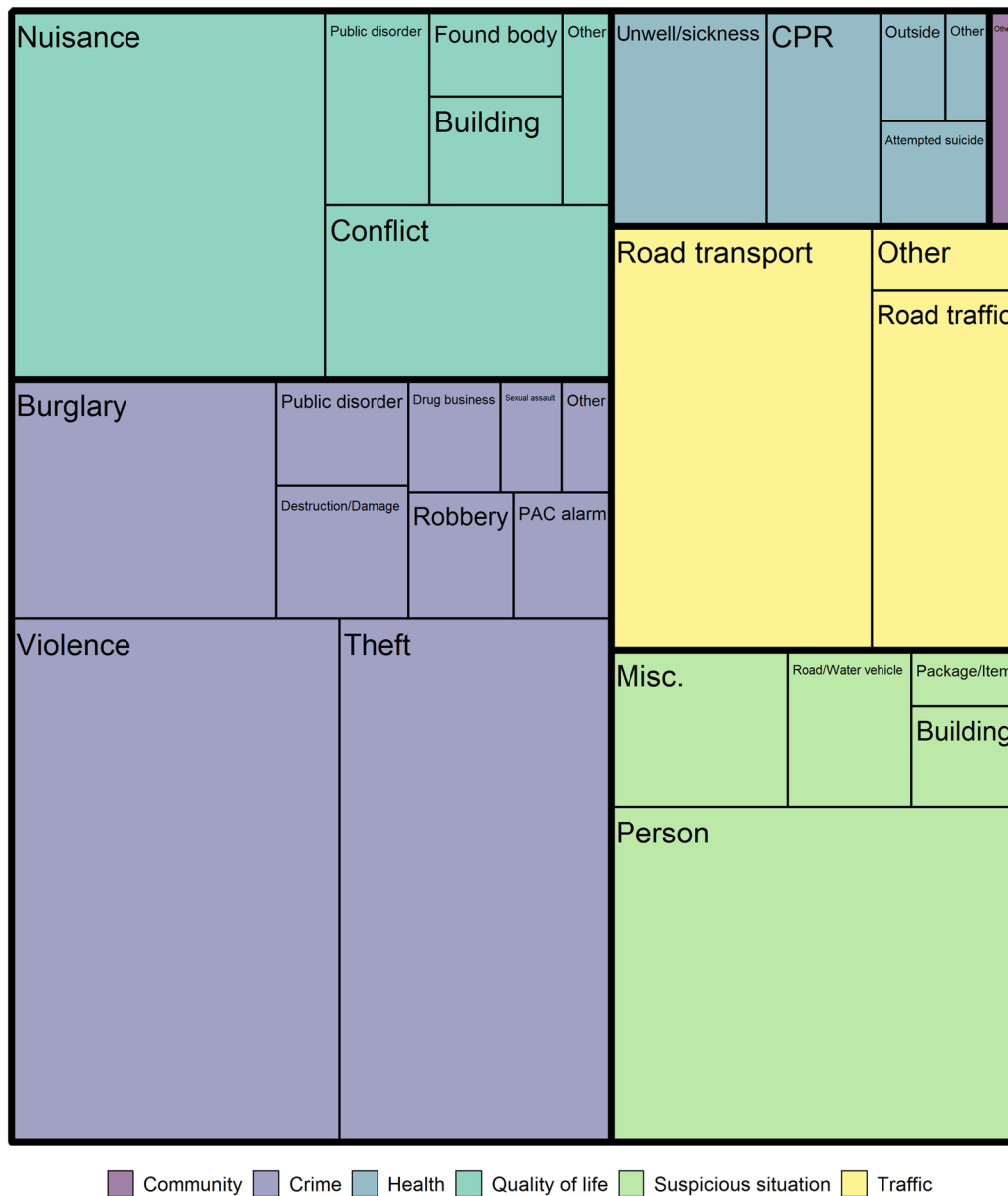


Fig. 2 Proportional breakdown of dispatched deployment time, operationalized using the ‘dispatched to clear’ measure

Table 3 Descriptive statistics about the amount of time (in minutes) consumed

Demand class	Min	Max	Mean	Median	SD
Community	0	571.60	34.33	19.03	52.50
Crime	0	6006.98	87.23	43.45	173.21
Health	0	3739.10	87.75	42.15	160.65
Quality of life	0	14638.42	47.16	22.12	154.96
Suspicious situation	0	10486.23	56.91	24.92	134.68
Traffic	0	15919.25	73.80	32.07	196.61

In Fig. 3, we use Lorenz curves to visualize the concentration of dispatched police deployment time according to each of the main three ‘time spent’ measures. Irrespective of the time measure used, we see that a considerable proportion of dispatched police deployment time in Amsterdam is consumed by a small number of incidents. Using the cumulative measure as an example (in blue), we can state that 25% of total dispatched deployment time is consumed by just 2% of incidents and 50% of time is consumed by 10% of incidents.

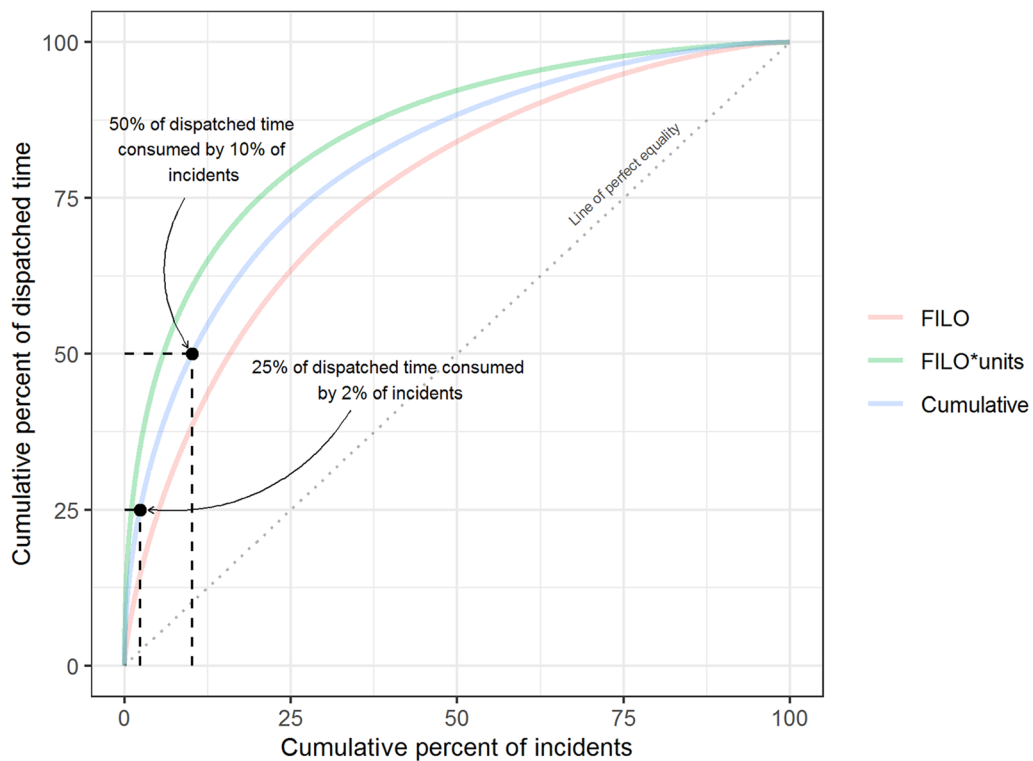


Fig. 3 Lorenz curves for each of the main time measures with an example of the thresholds highlighted for the ‘dispatched to clear’ measure

Table 4 Descriptive statistics about the number of dispatches by incident type

Demand class	Total dispatches	% dispatches	Min	Max	Mean	Median	SD
Community	1623	1	1	11	1.39	1	0.80
Crime	68,811	32	1	44	2.02	1	2.05
Health	21,587	10	1	32	3.57	3	2.37
Quality of life	47,455	22	1	62	1.56	1	1.55
Suspicious situation	40,499	19	1	32	1.77	1	1.47
Traffic	33,386	16	1	32	2.20	2	2.18

We observe different degrees of concentration depending on the method for operationalizing time spent: the standard FILO measure is *least* concentrated while the weighted FILO measure is *most* concentrated. This finding is consistent with the skew demonstrated in Fig. 1. Standard FILO does not account for the number of units: incidents for which there were multiple dispatches do not contribute any more to concentration than an incident for which only one unit attended. By contrast, the weighted measure accounts for units, but the multiplication with FILO (rather than unit-level time) contributes to a higher degree of concentration.

Multi-dispatch incident types

Descriptive statistics on the numbers of dispatches by demand classification are reported in Table 4. As expected, a large proportion of dispatches during the year were made in response to crime incidents (32%) although as reflected in the median, often only one unit is dispatched (noting that in the Dutch context this will often be a car containing two officers). The same is true for community, quality of life and suspicious situations. Traffic incidents have a median of two dispatches (this will often involve closing roads due to a traffic accident). Due to the CPR response procedure noted earlier, health

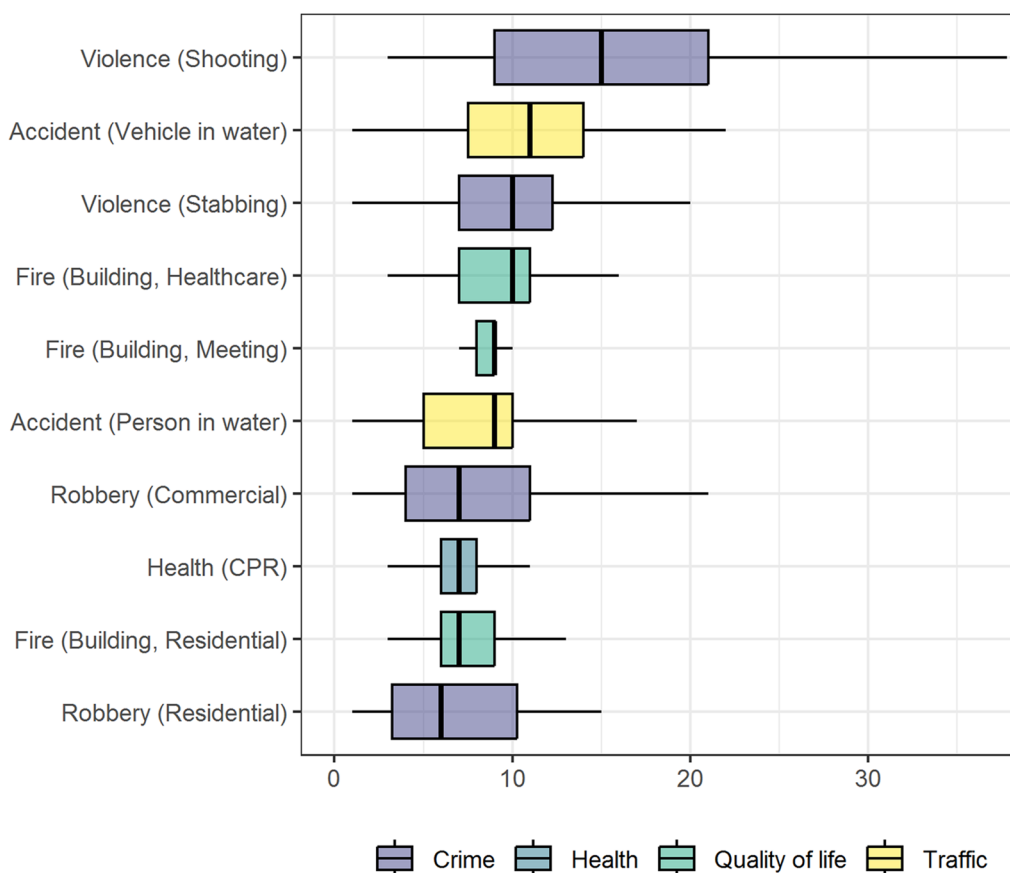


Fig. 4 Top-10 incident types by the median number of dispatches. Note that we focus on the interquartile range in isolation, given that the minimum and maximum values are already reported

incidents have the highest median number of dispatches with three. Perhaps most interestingly in terms of multi-dispatch incident types, the maximum number of dispatched units by demand classification is considerable. These range from community incidents involving eleven dispatches, to quality of life incidents for which 62 dispatches were made by emergency call handlers.

To further drill-down into multi-dispatch incident types, we visualize the distribution of dispatch counts for specific call types (see Fig. 4). These are the top-10 call types in terms of the median number of dispatches after having excluded those call types with less than twenty dispatches throughout the year. Eight out of these ten call types involve an imminent threat to life: serious violence, fires, CPR, and vehicle or people in water. The remaining two involve robbery (commercial and residential). In the Netherlands, incidents can be assigned three different priorities (1—most serious; 3—least serious). All ten of these incident types are priority-1.³

³ Priority levels 4 and 5 do exist in the system but are rarely used in practice. Such incidents account for 6% of incidents in our study period.

Discussion

In this short contribution, we set out to investigate to what extent proportional breakdowns of police ‘time spent’ are robust to the different operationalizations of *dispatched deployment time*. The proportion of total time spent on crime varies between 38 and 40% across the three measures. The ranking of different demand classifications by proportions of total time remains the same for cumulative (unit-level) and FILO (dispatch-level) operationalizations, irrespective of how we define the ‘start’ of deployment (dispatched, on way, or on scene). We observe only a minimally diverging ranking for the weighted FILO (Table 4).

Our descriptive exploration of dispatch numbers shows that there is variation between demand classifications in terms of the average (mean or median) number of responding units. Incidents involving health, for example, typically receive more dispatches compared to other incident types. In turn, we see a jump in the proportion of time consumed by Health when using the weighted FILO measure.

Only with unit-level data do we have the opportunity to calculate *actual* dispatched deployment time. But, many studies do not have access to the unit-level data required to calculate the cumulative operationalization. Based on our data in Amsterdam, if the goal of the analysis is to investigate the general breakdown of dispatched deployment time per demand class, the findings are fairly robust to the various operationalizations. Especially when research questions pertain to ranking of the categories based on relative time spent. Conversely, if the goal of the analysis is to investigate demand classes or incident types for their individual absolute and/or proportional size, not all demand classes and incident types will be robust to the various operationalizations. Namely, classes which receive systematically more dispatched units (e.g. Health) will be under- and overrepresented in absolute terms (like also in relative terms) by, respectively, the standard and weighted FILO measures used in existing US studies.

We make no claim of generalizability to the US, from which the research originated. Nevertheless, the underlying issues demonstrated here over the usage of CAD for the purposes of calculating dispatched deployment time are still highly relevant, and we certainly encourage future research to engage with the limitations and biases of the operationalization(s) used, ideally with reference to the prevalence of multi-dispatch incidents in the data, and in the context of the study region (e.g., dispatch practices).

Although we have shed new light on the various operationalizations of dispatched deployment time, some important factors remain unknown. We would propose using GPS tracking data of police units (e.g., used for different purposes in Dau et al., 2023) to capture unit movements towards an incident that were not initiated by or communicated to the dispatch center, such as those initiated after overhearing about an incident on the radio. These movements might constitute (unnecessary) flocking but are unmeasurable using the data used here, and difficult to determine using CAD data (Lum et al., 2021). Even then, neither data sources fully capture the context of multi-dispatch incidents. These investigations would also benefit from interviews and observations of dispatch center personnel and officers to better understand the conditions under which flocking might occur. We hope that these findings encourage further investigations in this area.

Lastly, as an additional point, we would like to highlight that—as our double-coding exercise demonstrated—recoding police incident types into broader a demand classification is no straightforward task. While we achieved an inter-coder reliability that surpassed the -preregistered threshold, the process was far from perfect. We suggest that all future research follow the

methodology we have described in the preregistration protocol to make this imperfect process at least as transparent as possible.

Acknowledgements

The authors would like to thank personnel from the Netherlands Police for providing advice on the data and/or facilitating access to the dispatch centre, particularly Ilse Klement and Lars van Jole. The authors would also like to thank Guido Legemaate (fire service) for his advice on GMS data more generally.

Author contributions

All authors contributed to the writing and conceptualization of the manuscript. All authors contributed to the initial data proposal request. Samuel Langton and Tim Verlaan conducted the data handling and descriptive analyses. All authors read and approved the final manuscript.

Funding

This research was carried out within the *What Works in Policing* research programme of the Netherlands Institute for the Study of Crime and Law Enforcement (NSCR) and funded by the Netherlands Police. The funder had no role in study design, data collection and analysis, decision to publish or preparation of the research.

Availability of data and materials

The data used for this study cannot be made publicly available for reasons of confidentiality. Additional materials and results, including the preregistration document, can be found on the Open Science Framework repository: <https://osf.io/qgww6/>.

Declarations

Competing interests

None.

Received: 5 July 2023 Accepted: 20 November 2023

Published online: 19 December 2023

References

- Dau, P. M., Dewinter, M., Vander Beken, T., Witlox, F., & Vandeviver, C. (2023). Police equilibrium: a global assessment of the demand and supply of police. *Police Soc*, 33(5), 537–554.
- Krippendorff, K. (2011). Computing krippendorff's alpha-reliability.
- Langton, S., Ruiters, S., & Verlaan, T. (2022). Describing the scale and composition of calls for police service: a replication and extension using open data. *Police Pract Res*. <https://doi.org/10.1080/15614263.2022.2102494>
- Lum, C., & Koper, C. S. (2021). Guest post: shifting mental health calls away from police will not greatly affect their workload. In *The Washington Post*. <https://www.washingtonpost.com/dc-md-va/2021/08/30/guest-post-911-analysis/>. Accessed 27 Nov 2023.
- Lum, C., Koper, C. S., & Wu, X. (2021). Can we really defund the police? A nine-agency study of police response to calls for service. *Police Q*. <https://doi.org/10.1177/10986111211035002>
- Ratcliffe, J. H. (2021). Policing and public health calls for service in philadelphia. *Crime Sci*, 10(1), 1–6.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.