Predictors of police response time: a scoping review

Tim Verlaan[1,2]* and Stijn Ruiter[1,3]

Abstract

Background As rapid response has been a key policing strategy for police departments around the globe, so has police response time been a key performance indicator. This scoping review maps and assesses the variables that predict police response time.

Methods This review considers empirical studies, written in english, that include quantitative data from which an association between the outcome variable *police response time* and any predictor can be observed or derived. This review provides both a narrative synthesis as well as what we termed a *hybrid synthesis*, a novel way of synthesizing a large quantitative dataset which is considered too rich for a mere narrative synthesis and yet does not allow for meta-analysis.

Results The search, screening and selection process yielded 39 studies, which presented 630 associations between 122 unique predictor variables and police response time. In order to present the results in a digestible way, we classified these into categories and subcategories. All methodological steps and the findings are made public: https://github.com/timverlaan/prt.

Conclusions Most of the conclusion and discussion focuses on lessons learned and recommendations for future research, as it proved hard to draw any definitive conclusions on causal factors related to police response time. We recommend that future studies clearly describe mechanisms, focus on the components of police response time (reporting time, dispatch time, travel time—or a combination of these), attempt to standardize predictors and outcome variables, and we call for more research into *reporting time*. We conclude this review with a first attempt at deriving a causal model of police response time from the subcategories of predictor variables we observed in the empirical studies included in this review.

Trail Registration: https://osf.io/hu2e9.

Keywords Police response time, Scoping review, Predictors, Causal model

Introduction

For decades, *rapid response* has been a key strategy for police departments around the globe. There are two general reasons for this. First, to fulfill the police's life-line function, to be physically present when citizens are in direct need of help. Second, because faster response times to citizens' calls-for-service are commonly believed (and empirically suggested) to be associated with increased *witness availability* (e.g., Cordner et al., 1983), *on-scene arrests* (Cihan et al., 2012), *crime clearance rates*

*Correspondence:
Tim Verlaan
t.v.verlaan@vu.nl
[1] Netherlands Institute for the Study of Crime and Law Enforcement
(NSCR), Amsterdam, The Netherlands
[2] Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
[3] Utrecht University, Utrecht, The Netherlands

(e.g., Vidal & Kirchmaier, 2018), *crime deterrence* (e.g., Weisburd, 2021), *victim satisfaction* (e.g., Brandl & Horvath, 1991), *victim wellbeing* (e.g., Brown & Harris, 1989), and overall *citizen satisfaction* (e.g., Frank et al., 2005). As it is also an easy to track quantitative measure, *police response time* is one of the key performance indicators in policing. Police departments around the globe measure, assess, and publicly share their response times and aim to improve them. Therefore, knowledge of the factors that determine police response time is of great value to police in particular and society at large. Although there is a large body of research on predictors[1] of police response time, so far, there has been no research that gathered, mapped and synthesized the findings. This review aims to fill this gap by answering the following research questions:

1) What are the empirically tested predictor variables of police response time?
2) What conclusions can be drawn about the direction, size and statistical significance of the effects of these predictor variables on police response time?

The type of literature review carried out here is a scoping review, which is defined as "a form of knowledge synthesis that addresses an exploratory research question aimed at mapping key concepts, types of evidence, and gaps in research related to a defined area or field by systematically searching, selecting, and synthesizing existing knowledge." (Colquhoun et al., 2014). We used the PRISMA-ScR checklist (Tricco et al., 2018) and first developed a protocol, which we pre-registered before starting the research (https://osf.io/hu2e9). This review retrieves English written literature—scientific and grey— that presents quantitative, correlational data, between the outcome variable *police response time* and any empirically tested predictor variable. This study critically appraises the methodologies and evidence presented in the primary studies to assess if they support causal inference regarding the empirically tested predictor variables. We aim to quantitatively synthesize the effects of predictor variables on police response time and derive a comprehensive causal model.

Police response time
In this review, *police response time* is considered to consist of three components: *reporting time, dispatch time* and *travel time*. *Reporting time* starts the moment an incident takes place and ends when it is notified to the police.

Dispatch time starts when the police are notified of the incident and ends as soon as a patrol car is assigned to the call. *Dispatch time* can be subdivided into *queue time* and *handling time*. *Travel time* starts when the incident is assigned to a patrol officer or car(s) and ends when the first police officer arrives on scene. *Travel time* is determined by two factors: *travel distance* and *average travel speed*. Police response time can thus be expressed mathematically as in Eq. 1.

$$t_{response} \; = \; \Delta t_{incident\ reported} \; + \; \Delta t_{response\ initiated} \; + \; \frac{\bar{v}}{\Delta x} \quad (1)$$

Figure 1 is an extension of (Spelman & Brown, 1984) and displays the three stages of police response time schematically. Throughout this review, we refer to these three components of police response time. Any predictor variable with a causal link to police response time has to affect police response time through one of these components. The addition of this layer of complexity to police response time is needed to better understand the mechanisms through which predictor variables affect police response time and to consequently allow for better policies and interventions regarding police response time.

Methods
This methods section attempts to be as explicit as possible about the followed procedures without being excessively minute. Additional details (e.g., the exact queries used for each database) can be found in the referenced pre-registration protocol.

Screening criteria
This review only considers empirical studies, written in English, that include quantitative data from which an association (whether causal or not) between the outcome variable *police response time* and any predictor variable can be observed or derived. We made no further a priori selection based on research design, literature type (academic or grey), geographical location, or year of publication; all literature until January 2021 (which is when we carried out the literature search) was considered.

Information sources
The primary databases consulted were: Scopus, EBSCO/ Criminal Justice Abstracts, Clarivate Analytics/Web of Science Core Collection, ProQuest/International Bibliography of the Social Science (IBSS), EBSCO/APA PsycINFO, ProQuest/Criminal Justice Database, ScienceDirect, Wiley Online Library, and JSTOR. The secondary databases consulted were: Google Scholar, National Criminal Justice Reference Service—US, Global Policing Database.

---

[1] This study originally set out to identify and quantify the *determinants* of police response time. However, in order to be in line with the evidence base of the findings, the term *predictors* rather than *determinants* is used.
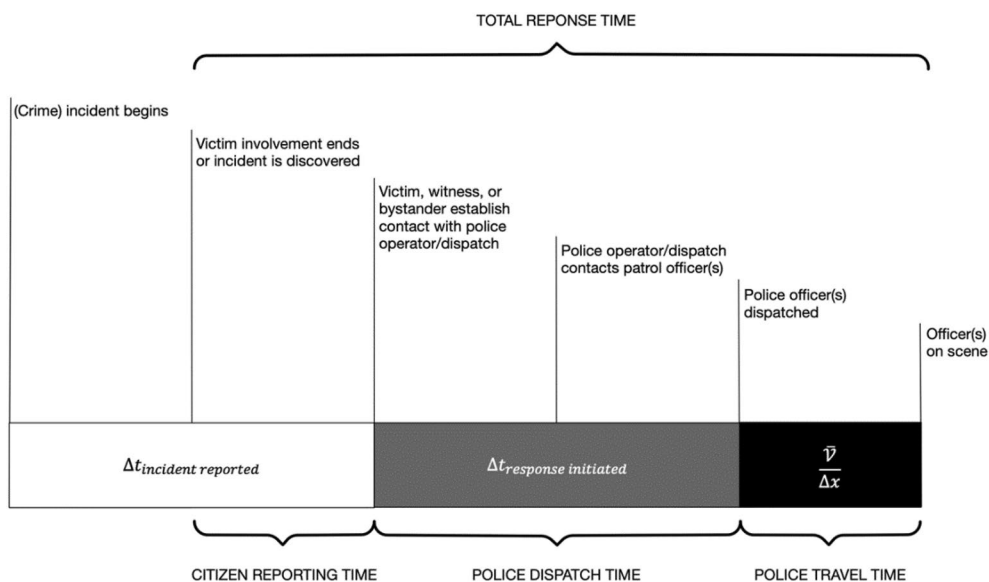
Fig. 1 The components of police response time—adjusted from Spelman and Brown (1984)

Table 1 Data retrieval per database

	Scopus	CJA	WoS	IBSS	PsychINFO	CJD	SD	Wiley	JSTOR	Total
Retrieval1	330	103	64	118	23	N/A	191	426	1559	2814
Retrieval2	89	29	24	102	6	1239	1593	152	769	4003

Search query

After a preliminary search, we decided on the broad search query (1) below, consisting of various instances of the concepts: *police* and *response time.* Both terms were separated by the operator W/25, meaning that the terms should occur within 25 words of each other in a text. The limit of 25 words was chosen as it resembles the upper bound of the mean word count of written sentences. Moreover, it allows to retrieve studies in which "response time" refers to "police response time" without the words having to be in that exact sequence and without including studies which include both terms, but in which these are unlinked. After the initial search phase, it was concluded the synonyms "rapid response" and "prompt response" should have also been included. Hence, a second search was carried out using query (2).

(1) ("police" OR "law enforc*") W/25 ("response tim*" OR "arrival tim*")
(2) ("police" OR "law enforc*") W/25 ("rapid response" OR "prompt response")
(3) ("police" OR "law enforc*") W/25 ("response tim*" OR "arrival tim*" OR "rapid response" OR "prompt response")

Titles and abstracts obtained from the queries of the primary databases were screened, supplemented with hand-searched articles from the secondary databases. The flow chart in Appendix A visualizes the entire data collection process.

Table 1 displays the number of studies each query yielded in each primary database. Note that the Criminal Justice Database was not available to the authors during the first round of queries. For that reason, we combined the first two queries into query (3), which was used for the second round CJD retrieval.

Screening phase

The retrieved dataset of studies from the primary databases was first cleaned and enriched with the study DOIs and using the update function of the Mendeley reference manager. This step greatly increased the number of abstracts that were retrieved. The set was subsequently deduplicated using Mendeley's *duplicate removal* function. Next, the remaining unique study titles and abstracts were imported into ASReview, an active learning software package created to screen literature for systematic and/or scoping reviews (Schoot et al., 2021). Using ASReview, the first author (TV) coded the

documents based on title and abstract as "maybe relevant" or "irrelevant". Documents were coded "maybe relevant" when the title and abstract suggested the study concerned empirical research linking one or more predictor variables to *police response time*. All other studies were marked as "irrelevant". In case the title or abstract contained insufficient information, the document was marked as "maybe relevant" in this part of the screening phase.

Using a machine learning algorithm on the user inputs, ASReview constantly presents what it predicts to be the most relevant document. This way, the user does not have to classify the entire set of documents, but can stop screening after a predefined set of consecutive documents are coded as "irrelevant". Any stopping criterion is arbitrary, but some are empirically shown to be sensical (e.g., Scherhag & Burgard, 2023). We settled on the stopping criterion of 100 consecutive "irrelevant" articles as advised to us in an informal meeting with the developers of the ASReview program. We used the following ASReview settings: classifier—"Naïve Bayes"; feature extraction—"tf-idf"; and query strategy—"Max", and we started the ASReview screening phase by providing the algorithm one relevant (Salimbene & Zhang, 2020) and one irrelevant (Zhu et al., 2020) study.

As the data retrieval was done in two rounds (adding the terms *rapid response*, *prompt response* and the CJD database in the second round), the ASReview screening phase was also done in two rounds. ASReview round one coded 111 studies (4.8%) as "maybe" relevant, and 597 (26%) as "irrelevant" before the stopping criterion was reached. ASReview round two coded an additional 11 studies as "maybe relevant".

Additionally, Google Scholar, National Criminal Justice Reference Service, Global Policing Database were hand searched. For all the studies that were retrieved in these secondary databases a judgement was made—in the browser, based on the title/abstract—if they were "maybe relevant" or "irrelevant".

Selection phase

All documents marked as "maybe relevant" during the screening phase progressed to the full-text selection phase in which eligibility was assessed. Studies were considered eligible if they concerned empirical research, had *police response time* as the outcome variable, and reported quantitative data from which an association between the outcome variable police response time and a predictor variable can be observed or derived. If a study failed on any of these criteria, it was excluded. The first violated criteria observed was noted for each discarded study. The first author (TV) did the full-text eligibility check for all literature that passed the screening test. The

selection phase resulted in 37 eligible studies. Next, forward and backward searches were carried out with these 37 studies as input, which resulted in 2 additional studies. Both these additional studies were not present in the original ASReview screening. Appendix B contains the list of included studies with some metadata. The "search overview" file on the public GitHub makes the selection process transparent (https://github.com/timverlaan/prt).

Data extraction

In order to extract the relevant information from the 39 studies, we created a data extraction form in Microsoft Excel. It was used to extract data specific on the outcome variable *police response time* and all observed predictor variables. Appendix C presents what data were extracted. In order to increase reproducibility of the data extraction, ATLAS.ti was used to code the *research location, time period, sampling strategy,* the presented *correlational & descriptive data* and the *conclusion* in the primary studies. The data extraction form is available on the GitHub page.

Data conversion and interpretation
Predictor variable categories

Because of the large number of unique predictor variables identified in the extracted data, we started with a data reduction step by classifying all predictor variables in categories and subcategories. This data reduction step helps with presenting the results in a digestible way. Table 2 specifies what additional information was coded per variable observed in the primary studies.

Strength of the associations

Some (models in the) primary studies did not directly present associations, or presented it in coefficients that do not directly allow for effect size calculation. In the first case, we attempted to extract descriptive statistics that allowed us to derive the size of the association ourselves. We converted multiple observed group means of population studies to percentage difference/change and sample means & standard deviations, t-statistics and z-statistics into Cohen's D where possible. In the second case, we converted the primary study's metric into a metric that allowed for effect size interpretation. We converted unstandardized B coefficients into a new metric "B*" by dividing it by the mean *response time* of the *calls for service* to make it analogous to the percentage change metric. If no mean was presented, we converted the unstandardized B coefficients into standardized Beta coefficients by extracting the standard deviation of both the outcome and the predictor variables. For the one study that presented results of an ordinal logit regression

Table 2 Variables added to the extracted information

Codename	Explanation
IV_label	Generalization of independent variable name
variable category	*Kinematics, exogenous, human involvement, incident characteristics, innovation, police strategies*
variable subcategory	*Day of the week, demand, road type, time of day, traffic, weather type, caller characteristics, community characteristics, victim characteristics, witness characteristics, incident characteristics, incident type, AGDS, AVL, emergency button, response unit type, travel distance, travel speed, call priority, department size, differential response system, jurisdiction area, no. officers per response unit, non-emergency number, on duty officers, patrol strategy, shift, workload*
Regression type	*Simple* or *multiple*
Effect direction	*+/-/none (for nominal and dichotomous predictor variables always none)*
Effect size interpretation	*Very small, small, medium, large (see Appendix D)*
Statistical significance	*/**/*** /non-applicable (i.e. population data)*
Internal validity	*Level I to V* (see section *Critical appraisal of research methodologies*)

Table 3 Maryland Scientific Methods Scale interpreted to suit the studies under review

Internal validity	Description
Level I	Correlation analysis, no control group nor attempt to establish counterfactual
Level II	'Before & after comparison' or comparison group without balancing covariates
Level III	Difference in Difference, attempt at balancing (OLS, matching) but uncontrolled differences likely remain
Level IV	Explicitly reasoned and proper balancing (OLS, matching), biases discussed but not addressed
Level V	RCTs, no selective sample attrition

(Kayes et al., 2019), we converted the unstandardized B coefficients into odds ratios.

In order to interpret the observed and converted Cohen's D, Pearson's R and odds ratios, we relied on Cohen's (2013) interpretation of effect sizes. In order to interpret the effect size of percentage difference/change measures, we posit our own interpretation and scale the interpretation of the unstandardized B* coefficient and the standardized beta coefficient analogous to this. We propose the following interpretation of the effect size of percentual change we believe to be reasonable for the context of police response time: 0–10% change, *very small*; 10–30% change, *small*; 30–50%, *medium*; >50%, *large*. All these data conversion and interpretation steps are detailed in Appendix D.

Direction of the associations
In our synthesis of the results, we discuss the direction of the association between scale/ordinal predictor variables and the outcome variable police response time. By a *positive* relation we mean that a unit/level increase of the scale/ordinal predictor variable corresponds with an increase in *police response time*, and thus police response is slower. By a *negative* relation we mean that a unit/level increase of the scale/ordinal predictor variable, *police*

response time decreases, and thus police response is faster. Interpreting *direction* of associations for nominal and dichotomous predictor variables is often meaningless. For example, different *crime types* (*property crime, violent crime,* etc.) might correspond with different police response times, but there is no logical direction of the association. For this reason, we only report whether there is very small/small/medium/large systematic variation in police response time between categories of those nominal variables, but more details can be found in the data extraction file published on GitHub. For some dichotomous variables though, especially those that capture pre/post interventions, the direction of the association has an intuitive interpretation. In such cases, this will be elaborated on.

Critical appraisal of research methodologies
In order to judge what conclusions can be credibly drawn regarding the presented associations between the outcome variable *police response time* and the predictors, the retrieved literature was critically appraised based on research design. We classified these using the Maryland Scientific Methods Scale (MSMS) (Farrington et al., 2003), as presented in Table 3. The higher the level of the research design, the more trust can be placed in

Table 4 Categorization of predictor variables

| Category	n	Subcategory	N	Unique variables	Measurement levels observed
Kinematics	9	Travel distance	8	3	Dichotomous, scale
		Travel speed	1	1	Scale
Exogenous	71	Day of the week	8	2	Dichotomous, scale
		Demand	5	1	Scale
		Road type	12	6	Dichotomous, nominal, ordinal
		Time of day	36	3	Dichotomous, nominal
		Traffic	4	2	Dichotomous
		Weather type	6	4	Dichotomous
Human involvement	150	Caller characteristics	39	16	Dichotomous, scale
		Community characteristics	87	23	Scale, dichotomous, nominal
		Victim characteristics	23	10	Dichotomous, nominal, ordinal, scale
		Witness characteristics	1	1	Nominal
Incident	276	Incident characteristics	27	9	Ordinal, dichotomous, nominal
		Incident type	249	8	Nominal, dichotomous
Innovation	19	AGDS	11	1	Dichotomous
		AVL	2	1	Dichotomous
		Emergency button	1	1	Dichotomous
		Response unit type	5	1	Nominal
Police strategies	93	Call priority	34	1	Ordinal, dichotomous, nominal
		Department size	7	3	Scale
		Differential response system	4	1	Dichotomous
		Jurisdiction area size	4	2	Scale
		No. officers per response unit	9	6	Dichotomous, nominal
		Non-emergency number	8	1	Dichotomous
		On duty officers	2	1	Scale
		Patrol strategy	16	11	Dichotomous, ordinal
		Shift	2	1	Nominal
		Workload	7	5	Scale

the internal validity of the observed associations. Strong causal claims are only warranted for research level V designs.

Results

Empirically tested predictors of police response time

Together, the 39 included studies yielded 630 empirically tested associations between predictor variables and (a component of) police response time. However, over two-thirds of these predictors originate from the three studies by Bieck (1978, 1980a, 1980b), which are based on relatively small sample sizes. The 630 associations relate to 122 unique predictor variables. Because it is impossible to present and discuss all these unique predictors in this article, we will present some descriptive statistics and refer the interested reader to the data extraction file available on GitHub. Of the 630 tested associations, 18% (n = 113) related to scale predictor variables, 6% (n = 40) to ordinal predictor variables, and 76% (n = 477) to

nominal predictor variables, the latter of which predominately represented *incident type*. The outcome variable *police response time* was operationalized in the following ways: 142 associations were tested for the component *reporting time* (23%), 81 for *dispatch time* (13%), 159 for *travel time* (25%), 237 for *dispatch time + travel time* (38%), and 11 for *reporting time + dispatch time + travel time* (2%). *Police response time* was virtually always measured as a scale variable (99%), which was operationalized as *minutes per call for service/incident* in most cases (95%). Police response time data mostly originated from US Police computer aided dispatch (CAD) systems, or predecessors of CAD. Only a few (older) studies used response times gathered by officers and observers (e.g., the Kansas City response time studies). Figure 2 shows a Sankey plot that presents the categories of predictors found in each study and against which component(s) of *police response time* the predictors variables within these categories were tested.
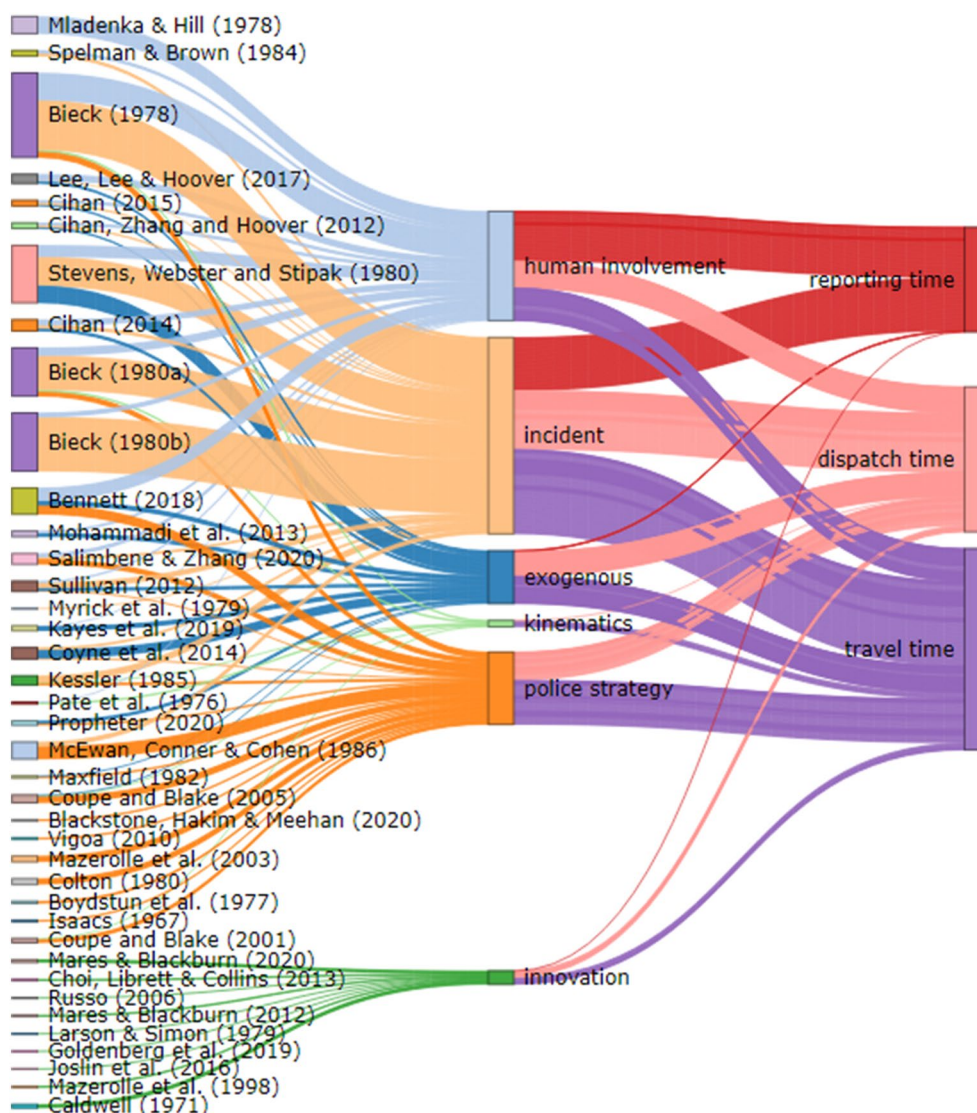
Fig. 2 Sankey plot of studies, categories and response time components. The Sankey diagram is available more clearly as an interactive visualization at: https://timverlaan.github.io/prt/sankey_prt.html

Categories of predictor variables

Capturing the 122 unique predictor variables in intuitive categories and subcategories—which make the analysis more tangible as well as do justice to the individual predictors—is no trivial task. We have based our categorization on the perceived (and sometimes described) mechanisms through which the predictor variables are assumed to affect (a component of) *police response time.* Table 4 presents the resulting categories and subcategories of this effort, next to the number of empirically studied associations within these (sub)categories and the measurement level of the predictor variables found within these (sub)categories.

Challenges to synthesizing the findings

After analyzing the included literature, we concluded that quantitative synthesis (in the form of meta-analysis) would not be possible. We identified five types of heterogeneity that hinder quantitative synthesis, here presented in no particular order:

First, large heterogeneity in the operationalization of predictor variables. For nominal and ordinal variables, this concerns heterogeneity in: the number of categories, the values these categories represent, coding system (e.g., dummy coding or deviation coding), and—in the case of dummy coding—the reference category selected. For scale variables, this concerns

heterogeneity in the unit of analysis. The following two examples illustrate this heterogeneity in operationalization:

> *Call priority* (ordinal) has been analyzed in many different operationalizations—e.g., *call priority ranging between 1 and 3* (Salimbene & Zhang, 2020) *or 1 and 5* (Colton, 1980)*; emergency/non-emergency* (Isaacs, 1967). While all clearly describe the same concept, they are also clearly measured using different categories. All priority levels correspond to varying target response times and even within categories, target response times vary (e.g., priority 1 can mean "within 10 min" or "within 15 min" in different jurisdictions). Lastly, the studies vary in the use of reference categories, it either being the lowest priority level (e.g., Salimbene & Zhang, 2020) or the highest priority level (Bennett, 2018).
>
> *Demand (scale)* is generally captured by *call volume.* This variable has been operationalized as *number of calls for service received in the same hour as the call* (Bennett, 2018); *the number of calls for service received in the same day as the call* (Sullivan, 2012); *or the average number of calls for service per hour* (Maxfield, 1982).

Second, large heterogeneity in the constitution of (regression) models. The majority of the predictor variables are tested in multiple (linear) regression models with varying sets of control and interest variables. Obviously, the inclusion and/or omission of relevant predictors variables (can) greatly influence the size, direction and statistical significance of predictor variables' parameter estimates.

Third, heterogeneity in the application of log-transformation of the outcome and (one) predictor variable. For 378 associations (60%), the outcome variable *police response time* (or one of its components) was log-transformed before it was analyzed; in all other cases, it was not. In four regression models, the predictor variable *call rate* was also log-transformed.

Fourth, large heterogeneity in the scope of the different models. Many of the models analyze *police response time* in relation to a subset of all calls for service—e.g., those related to *domestic violence* (i.e., Lee et al., 2017), *in-progress burglary* (e.g., Coupe & Blake, 2005), *in-progress assault* (i.e., Cihan, 2015), or *all part I crimes* (e.g., Kessler, 1985). Although we have not witnessed it being articulated in the studies, the underlying assumption seems to be that there are fundamental differences between these subsets in how the predictor variables relate to police response time. Unfortunately, none of the studies explicitly tested this assumption and the extracted data do not allow us to test it ourselves.

Fifth, large heterogeneity in the metrics in which research findings are presented, e.g., Pearson's R, Cohen's D, unstandardized B coefficients, Chi-squared, t-statistic, Z-statistic, F-statistic, mean difference, median difference, Kendall's tau and Beta coefficients. In many studies, a multitude of these metrics was presented without clear directions on what to base the conclusions on. While these metrics can mostly be converted into one another, the necessary information to do so was not always present or easily identifiable. Moreover, not all interpretations of the effect size of these metrics could be converted into one another. This is especially unfortunate, because (a component of) police response time is always expressed in unit of time (seconds, minutes, or hours, or a combination thereof). Standardized time differences/change (e.g., dividing the time difference by mean or median time) would therefore present the ideal metric for synthesis.

Strategy for synthesizing the findings

Given these obstacles, we deemed it impossible to quantitatively synthesize the results in the form of a meta-analysis. Instead, we present the results in a hybrid manner: we count the number of times parameter estimates for (sub)categories of predictor variables were in a certain direction and of a specific size. This hybrid synthesis, as presented in Appendix E, is our best attempt to present the wealth of data for 630 associations found in 39 studies in a way that is both digestible and all-encompassing. We refer anyone interested in the intricacies of the data to consult the full data extraction file on GitHub. In the next section, we give a narrative synthesis of the main findings presented in Appendix E.

Narrative synthesis
Kinematics
Average travel speed and *travel distance* can be reasoned to entirely explain police travel time. Any model that incorporates both accurately should thus—in theory—leave no room for other predictor variables. Only one study tested the effect of *travel speed* on *police response time* (Pate et al., 1976). Surprisingly, only a very small relation was found:

> *"… there was little correlation between driving speed and response time. However, the speed limits generally ranged only from 20 to 35 miles per hour. With so little variation among reported driving speeds, the low statistical correlations between driving speed and response time should not be construed to indicate that driving speeds do not affect response time. The method of estimating driving speeds may not be sufficiently reliable, and the actual range of speeds may be too narrow to permit the accuracy required for correlation analysis." (Pate et al., 1976, p. 28)*

The relation between *travel distance* and *travel time* was analyzed in three older studies, upon which five articles were based (Bieck, 1978, 1980a; Blake & Coupe, 2001; Coupe & Blake, 2005; Kessler, 1985; Pate et al., 1976). *Travel distance* was mostly measured in kilometers (scale variable). However, some studies also recoded it into binary variables *car in beat of incident* (Kessler, 1985) and *officers' location is in beat of incident* and included it—in tandem with the default scale variable—into a regression model (Bieck, 1978, 1980a). These recoded variables of essentially the same information generally had a smaller effect size and very likely interfered with the effect *of travel distance*. The studies that looked at *travel distance* found small to large positive effects on *travel time*.

Exogenous predictor variables

Exogenous variables all have—with the exception of *demand*—a very small association with police response time. *Demand*, although not tested using designs with a high internal validity, can easily be understood to have a large effect on *police response time.* A high demand for police response likely congests the dispatch center, which would increase *queuing times*. Besides, if a fixed number of available officers need to respond to a larger volume of calls-for-service, it will probably take them longer to respond because dealing with one call will increase the time before they can leave to respond to the next call. Between *police response time* and *time of day* and *day of the week* we observed some relations of a large size. However, these can be attributed to lack of control variables in the underlying models. For example, the one *large* relationship was found for 3.00–3.59 pm, which was the same time period the police unit under investigation had a *shift change*, but shift change itself was not accounted for in the model (Stevens et al., 1980).

Human involvement predictor variables

Human involvement variables show very small to small statistical relations to police response time. The direction also varies. One striking finding deserves further discussion. *Community characteristic* predictor variable *call rate*—operationalized as the *number of calls for service per 1,000 citizen per neighborhood per year*—shows a large relationship twice (Cihan, 2014, 2015). These parameter estimates seem anomalous in light of the other evidence that *call rate* shows two small negative and one very small positive relation to police response time. Because we also cannot think of a theoretical mechanism that would explain why call rate—in its current operationalization—would have such a large relationship to *police response time*, we urge the reader to interpret these large parameter estimates presented in Cihan (2014, 2015) with caution.

Incident predictor variables

Incident type is by far the most analyzed predictor variable. However, most of these analyses are based on observations from the Kansas City response time study (Bieck, 1978, 1980a, 1980b; Kessler, 1985), which was a relatively small study (sample sizes varied in the various presented regression models, max = 2,800, mean = 669.5, sd = 444.6) and the study has been criticized as a "highly non-random sample" (Vidal & Kirchmaier, 2018). Besides, *call priority level* has not been considered in the majority of the regression models based on the Kansas City response time study. The only evidence that sheds some light on the effect of *incident type* controlling for *call priority level* is found in Salimbene and Zhang (2020). They report only very small differences between the reference category *disorder* crime and incident categories *violent* crime, *property* crime and *other* crime types. *Incident characteristics* can roughly be broken down into four types: *(degree of) injury*, *weapon involvement*, *in-progress* and *source of the call for service* (911 or automated alarm system). Controlling for *call priority level*, *weapon involvement* and *source of the call for service* both have very small relations to *police response time*. Without controlling for *call priority level*, some (other) studies find the relationships for *(degree of) injury* to range from very small positive to large negative, and those for *in-progress* to range from very small to large negative. However, we do not place too much trust in these findings because *call priority level* was not controlled for.

Innovations

We labeled four predictor variables as *innovations*: *acoustic gunshot detection system (AGDS)*, *automatic vehicle location systems (AVL)*, *emergency button*, and *response unit type*. AGDSs are systems that should directly affect *reporting time*. Surprisingly, they have not yet been analyzed against reporting time. Instead, AGDSs show very small to small relations to the combined components dispatch time and travel time. In two studies by Mares and Blackburn (2012, 2021), this association also changes direction over time. At first, calls from the AGDS received slightly faster dispatch and travel times, while they later received slower dispatch and travel times. Dennis Mares argued (in personal communication) that this might be explained by *response fatigue*. AGDS-generated calls for service mostly turn out to lead to no suspects or arrests and are often believed to be false positives. *AVLs* were analyzed in two studies, using data gathered in 1974 (Larson & Simon, 1979) and 1999 (Russo, 2006). AVLs are meant to improve travel times by enabling dispatching of the closest car to an incident. Despite this clear mechanism, the systems had very small relations to *police response times*. Possibly, the technology at the time was not advanced enough to outperform status quo. It should be noted that the studies by Russo (2006) and Larson

and Simon (1979) have a low internal validity level and results should thus be interpreted with caution. The *emergency button* was only researched in one study with a relatively low internal validity score. Nevertheless, the results showed a large negative relation to *police response time* (reporting + dispatch + travel time), meaning the intervention lowered *police response time*. This shows the potential for innovations that are expected to affect *reporting time*, the component of *police response time* which Spelman and Brown (1984, p.55) suggested to be the largest contributor to *police response time*. *Response unit type* measures the difference in response time between two types of police cars in the late 1960s. Because the context of these findings is so outdated, we refrain from drawing any conclusions.

Police strategies

Police strategies is by far the broadest category we used. It includes: *call priority, workload, department size, differential response system, jurisdiction area size, number of officers per response unit, implementation of a non-emergency number, number of on-duty officers, patrol strategy,* and *shift*. The common denominator of all these predictors is that all are—in some way—measures that reflect decisions on how police operate. *Call priority* has a very small to large positive (positive due to inverse order of priority level, level 1 being highest) effect on *police response time*. There are some instances where the primary results suggested a negative relation, but this was entirely due to the use of a different reference category (referenced against highest instead of lowest priority level). *Call priority* mostly shows a large effect and from analyzing the results in context we deem this to be accurate. *Workload* is suggested to have varying relations with police response time for different priority levels. For high priority calls, it seems there is virtually always enough capacity, and *workload* is thus not a limiting factor. In contrast, variation in *police response time* for low priority calls seems to be explained for a large part by *workload*. *Department size* (−), *jurisdiction area size* (+), *number of officers per response unit* (−), *number of units responding* (−) and *number of on-duty officers* (−) all show very small to small relations with *police response time*, with their respective direction reported between brackets. Within subcategory *department size*, the predictor variable *number of officers employed* had a negative direction while the predictor variable *number of citizens employed* showed a negative direction once and a positive direction twice. *Shift* captures the difference between the morning, night, and evening shift (while controlling for other variables), and these showed very small to small relations to response times. Coupe and Blake (2005) report that *shift change* has a significant effect on response times, as was also hinted at in Stevens et al., (1980). Unfortunately, the

effect size and direction of *shift change* is not mentioned in the study, only its p-value (Coupe & Blake, 2005).

Discussion

First, we discuss the general characteristics of the reviewed literature. Second, we discuss the results as presented in Appendix E and the narrative synthesis. Third, we discuss the limitations of this review. Finally, we present recommendation for future research into police response time and our attempt towards building a causal model.

Discussion of the literature

Looking at the general characteristics of the included studies, a few things immediately stand out. First, the vast majority of studies (90%) originate from the United States (and most from large urban areas), with only three studies from the United Kingdom and one from Iran. Without sufficient research from other countries, it remains unclear how the urban and US-centric findings generalize. In countries where predictor variables have a greater range or variance, they might also differently affect *police response times*. For example, while *road type* did not prove to be a limiting factor to *police response time* in the large cities studied in the United States, it is easily imagined that mountainous dirt roads in rural parts of the world will constitute a limiting factor to *police response time*.

Second, research interest into police response time seems to have come in two waves, with a quiet period between 1985 and 1995. Studies from the first wave mainly focus on *incident* predictor variables. Studies from the second wave, especially more recently, mainly focus on whether and how *police response times* vary with *human involvement* variables. Remarkably, some important predictor variables like *travel distance* and *(average) travel speed* were controlled for in the first wave—when measuring them was very difficult and less reliable—more often than in the second (current) wave when measuring them—and thus controlling for them—is far easier and more reliable.

Third, since 1977, all data used in the included studies were extracted from computer aided dispatch (CAD) systems from (local) police departments or other such large-scale databases. These CAD data should be considered population data for the police department they are extracted from. With the exception of Bennett (2018), none of the studies included in this review choose the local police departments from which they extract data through a sampling procedure. Instead, police departments are selected based on convenience and/or other

criteria. In the absence of a predefined sample space and explicit sampling procedure, the results of these studies should not be interpreted as generalizing to any larger spatial unit such as state, country or continent. As the studies are based on population data that themselves do not generalize to a larger superpopulation, there is actually no need for the use of p-values or inferential statistics in general: all the variation and effects observed are, by definition, the real variation and effects. In other words, these studies provide stronger evidence than they give themselves credit for, although it is evidence that does not present a clear basis for generalization beyond the study area.

Fourth, the literature fails to build towards the common goal of gaining a comprehensive understanding of determinants of (and their effects on) police response time. Perceived mechanisms and causal models are scarcely communicated, and if they are, not in an explicit manner (i.e. in non-specific text rather than in causal diagrams). The models seem not to get gradually improved based on insights from prior research. Overall, the literature seems to be stuck in a paradigm of correlation rather than trying to work towards a causal model.

Discussion of the results

In total, the relationship between 122 unique predictor variables and (a component of) police response time was analyzed. Given all the obstacles on synthesizing the findings mentioned before, it is not possible to draw any definitive conclusions on the effects of predictor variable (sub)categories. Nevertheless, we will discuss some observations and lessons learned from the synthesis of each variable (sub)category.

Kinematics directly determines travel time and therefore have a causal link with police response time. More effort should be directed to accurately measuring and including *travel distance* or *(average) travel speed* in studies of police response time. Note though that including both in a regression model will leave no room for other predictors. Having those measures might actually shift the research interest from the explanation of police response time as such to explaining why police officers drive at different speeds or are positioned at different distances from incident locations.

Exogenous variables (except *demand*) have negligible associations with police response time in the contexts where they have been studied. Only in very extreme cases can variables like *weather* or *traffic* be expected to have a noteworthy effect on police response time. For this reason, we consider the inclusion of these *exogenous variables* (except *demand*) in a causal model as optional in urban areas of most developed countries.

The association between *human involvement variables* and police response time is potentially contentious and easily linked to societal debates around 'defund the police' and 'black lives matter'. It is, therefore, vital that the research into this is rigorous in its design, cautious in its claims and clear on the distinction between correlation and causation. Nevertheless, most studies that have looked into this make bold claims we feel are not warranted by the research presented. The most authoritative article on *human involvement variables* (Bennett, 2018) retrieved CAD data from 40 police agencies and measured the effect of many *human involvement variables* on dispatch and travel time. Effect sizes and direction varied greatly, which caused Bennett (2018, p. 30) to write the following:

> *"Notice that in every case the standard deviation is much larger than the mean, in most cases by a factor of three or more. While the sign of the means mostly correspond to the estimates from the combined regression, in every case there are estimates on both sides of zero."*

From this we conclude that the associations of human involvement variables with police response time are far from homogenous or clear, and scholars should be careful to draw any conclusions. More (rigorous) research is needed.

Incident variables show varying effect sizes, from very small to large. However, in most of the models in which these varying findings are observed *call priority level* is not controlled for. *Call priority level* is determined exclusively by incident variables and itself has a direct causal mechanism to police response time. Unsurprisingly, *incident variables* were found to have only a very small effect on police response time once *call priority level* was controlled for. We believe that the variation in police response time for different incident variables—not explained by *call priority level*—is indeed very small. Nevertheless, the mechanisms behind these small differences present an interesting avenue for future research.

All current *innovations* aim to shorten reporting time by either bypassing citizen action or facilitating it. Even though the evidence base is either currently lacking or very small, we believe these innovations can actually have

large effects on shortening reporting time and, consequently, police response time.

Regarding the subcategory *police strategies*, we limit our discussion to a few predictor variables that are suggested to—empirically and/or theoretically—have a large effect on police response time: *workload* (consisting of *demand* and *police capacity), call priority level*, and *police discretion*. Of these variables, *call priority level* has the largest effect as it directly dictates police actions. Any variation among incidents of the same priority level can thus be attributed to actions not dictated. In part it will be related to *police discretion*, which can influence police response time in three ways: police dispatchers handling calls for service faster/slower, patrol officers traveling to incidents faster/slower, or patrol officers positioning themselves so they are closer to/further away from (future) incidents. *Workload* mainly determines when a response is initiated and therefore has a greater effect on calls with lower priority.

Limitations
Limitations of data gathering
The first limitation of this scoping review is that the search terms did not include the separate components of police response time (i.e., reporting time, dispatch time and travel time)—the importance of which was only realized during the analysis of the retrieved data. Nevertheless, we doubt that any relevant study only uses the component terms without any of the various search terms we included for police response time (i.e., "rapid response", "prompt response", "arrival time", and "response time"). A second limitation of this literature review is that no double collecting, screening, selecting and coding of the literature was done. While this undeniably lowers the reliability of the findings, we are confident that—largely thanks to our open and systematic way of analyzing and interpreting the data—the results meet the standards of reproducibility and reliability. We believe saturation was reached in searching for relevant literature as both the forward and backward searches only yielded a single extra study—neither of which was found to be present in the literature retrieved from primary databases. Third, the disproportionate concentration of data emanating from the United States, coupled with an inadequate representation of data from other geographical origins, likely has implications for generalizability of this review's findings. While articulating these specific implications is challenging, our preceding discussion attempts to illuminate the potential ramifications arising from the limited diversity observed in the study regions.

Limitations of data analysis
In order to present the results in a digestible manner, we presented the observed statistical relations on a four-point scale (very small/small/medium/large) in Appendix E. Assigning those values for Cohen's D, Pearson's R and odds ratios was straightforward using Cohen's (2013) classification or extrapolations of this work (see Appendix D). However, for unstandardized B coefficients, standardized Beta coefficients and percentual change, we could not rely on existing rules of thumb and had to decide on what constitutes a very small/small/medium/large statistical relation. Such judgment calls are inherently arbitrary; using different cutoff values might lead to different results. We also wanted to provide some indication of the (average) internal validity of the scientific evidence per subcategory by using the Maryland Scientific Methods Scale. All these steps should be taken as a first attempt to present the large data collection effort in a way that is still digestible by a reader. We encourage the reader to critically assess these choices or to even suggest other ways of synthesizing large quantitative data collections without solely relying on a narrative synthesis. As systematic reviews are periodically updated, we hope constructive criticism is able to improve the validity of our interpretations for updated versions.

Limitations of the findings
In many ways, this study suffers from attempting to standardize the analysis of data that proved to be too heterogeneous to be captured in such a rigorous manner. As a result, the article might now read as a critique of prior work in this field rather than having an exclusive focus on uncovering the empirically tested determinants of police response time and their effect size, direction and significance. Although we are aware of this limitation, we believe this is a necessary assessment of the research evidence and the article presents a better way forward than any (less critical) alternative would offer.

Recommendations for future research into police response time
In order to help police response time research progress—and in response to the obstacles to quantitative synthesis we had to deal with—we have four general recommendations for future research.

First, we argue that the research into police response time should shift from the current correlational paradigm to a causational paradigm. Many exploratory studies have now been done and the field would benefit from research that works towards a causal model of police response

time. Moreover, this shift would likely result in higher MSMS scores than are currently witnessed. Obtaining all the data required to properly test such a causal model can be challenging and might well be beyond the possibilities of many research projects. Nevertheless, clearly defining and communicating the perceived causal model and pitfalls in testing it should be possible for all research projects and would already be a big step forward for this line of research.

Second, we argue that each *call priority level* corresponds to a distinct *police response time* distribution. Therefore, we recommend to analyze the different call priority levels separately. Moreover, in case the goal is to investigate how specific determinants limit faster police response time, we would even argue that researchers should only analyze variables if they can be assumed to present a limiting factor. Only when the police are aimed at realizing the shortest possible response time, all variables can be assumed to present a limiting factor. Therefore, we recommend—if the goal is to investigate how specific determinants limit faster police response times— to only analyze the highest priority levels calls.

Third, research into police response time would benefit greatly from more standardization in operationalization of both predictor and outcome variables. Such standardization is preferably reached through consensus among experts. In order to start this conversation, we present some suggestions on four matters we deem most in need of standardization in the current literature: transformations, measure of central tendency, coding, standardization for effect size interpretation:

1. The majority of the 630 associations were tested against log-transformed components of the outcome variable *police response time*. In most of the studies that used log-transformation, it was argued useful for making the outcome variable less skewed, or, to be beneficial for regression analysis as it improves normality of the error distribution (e.g., Cihan et al., 2012, p. 316; Lee et al., 2017, p. 67). We argue against log-transformation for these reasons and believe skewedness of the data should be investigated rather than treated a nuisance. We judge much of the skewedness of police response time data to be a result of measurement error – i.e., by police officers forgetting to log their *on-scene arrival*, resulting in a number of very large *police response times*. A more appropriate response to the observed skewedness would thus be *outlier detection*, as employed by Choi et al. (2014).

2. If outlier detection does not suffice to curb large skewedness of the data, we argue that the *median* would be the better measure of central tendency over the *mean*, which is now more commonly used. To illustrate: in the study by Bieck (1978, p. 140), the *mean* police response time (reporting + dispatch + travel) is reported to be 3 h and 57 min, while the *median* response time is reported to be 18 min and 50 s; we argue in such cases the median better reflects the central tendency of police response times.

3. With the exception of one study (i.e., Stevens et al., 1980), which uses deviation coding, all studies that included nominal variables in multiple regression models use dummy coding. This can only be done with an arbitrary reference category, and we observed over twenty different reference categories used for the predictor *incident type* alone. We argue that deviation coding is preferable as it would improve comparability and facilitate synthesis of findings.

4. In order to make normative judgements on the size of observed difference or associations, many studies standardize the associations by using the standard deviation of the predictor and/or outcome variable distribution. For example, Coupe and Blake (2005, p. 249) report: "As expected, response times were strongly associated with response distances (Pearson's r = +0.47, p < 0.01)". We argue, instead of presenting (on standard deviation) standardized correlation measures and its standardized interpretation (e.g., "*strong association*"), that it is generally more meaningful to present the associations in the intuitive units it is measured in and interpret effect sizes in context, for example, by comparing the unstandardized differences/effects to the median response time.

Fourth, more research should be conducted into *reporting time*. Several of the earliest studies (1980a; Bieck, 1978, 1980b; Spelman & Brown, 1984) show that *reporting time* constitutes the largest share of total *police response time* and claim it should thus receive most focus when aiming to improve response times. To illustrate, a quote from Spelman and Brown (1984):

> "*Most of the time, the [reporting] delays are so substantial that even our fastest response to the crime will be ineffective in producing arrests. In short, we have focused on using high technology dispatching equipment and sophisticated deployment schemes*

to reduce police response time, when we should have focused on reducing citizen delays." (Spelman & Brown, 1984, pp. xi–xii)

While the evidence base upon which these claims are made is rather slim, may be outdated and originates from a very specific setting (five cities in the USA), we judge *reporting time* to be a very promising and often neglected research topic, especially when the aim of the research is to contribute to improving *police response time*. However, it is difficult to accurately measure reporting time. We believe more research into AGDSs would provide a good start, as AGDSs automatically record timestamps for each incident (e.g., "shots fired" or "shooting"), which could then be compared to the reporting timestamps of calls-for-service in order to obtain an accurate reporting time. Furthermore, the only intervention tested in relation to *reporting time* (i.e., *emergency button*) showed a very large effect of on shortening overall police response time (Joslin et al., 2016).

Towards a causal model of police response time

Although the methodologies of the reviewed studies do not allow for causal inference, the predictor variables studied could well be determinants of police response time. In fact, for many of the predictor variables, a mechanism by which they influence (a component of) police response time can easily be reasoned. This is most evident for *travel distance* and *travel speed* as these fully determine travel time, but it also applies to other predictor variables.

Because most studies neglect to communicate the (perceived) mechanisms, much remains unknown about the causal model of *police response time*. In order to work towards a causal model of police response time, we present a first conceptualization of a causal model that consists of the subcategories identified from the empirical literature included in this review. Appendix F depicts this first conceptualization in the form of a causal directed acyclic graph (Textor et al., 2016). The different categories are depicted using different colors. Next to the subcategories we used to synthesize the literature, we

included two mediating variables (*coverage* and *willingness to help*) and one moderating variable (*personal bias/preference*) in the model in order to complete the perceived mechanisms in the model.

The causal model presented in Appendix F is by no means a finished product; it is likely underdetermined (e.g., regarding *reporting time*) and at the same time oversaturated (e.g., because of the inclusion of very specific subcategories). Nevertheless, it provides a first mapping of the current state of the literature and could be used to guide future scientific inquiries into police response time and show how these will be related. It can be helpful in identifying the variables that should be controlled for when trying to estimate specific total, direct, or indirect effects of a certain (set of) predictor variable(s) on a component of police response time. Based on this first causal model, we would like to make one final recommendation:

Instead of investigating the effects of variables on overall police response time, we argue it is better to focus on a certain component of *police response time*. For example, studies that investigated the effects of *human involvement* variables on overall *police response time* often find very small to small relationships: certain communities would receive slightly slower or faster response times. Conclusions are then drawn—though never very explicitly—that these findings suggest racial biases among police personnel. However, with the empirical designs used, it is often unclear whether the observed disparities are explained by travel speed, travel distance or handling time. Disparities in travel distance would have to be attributed to a focus on specific locations within a police jurisdiction and such decisions are generally not made at the level of individual police officers. Studies would present stronger evidence for racial bias among police officers if they could show disparities in the (mean, median or max) *travel speed* of patrol officers or *call handling time* of dispatchers. In other words, more specific causal models and analyses are easier to communicate, control for confounding, guard for omitted variable bias and generally less susceptible to noise and therefore, provide a stronger evidence base and contribute more to the development of the field.

Appendix
Appendix A: flow chart
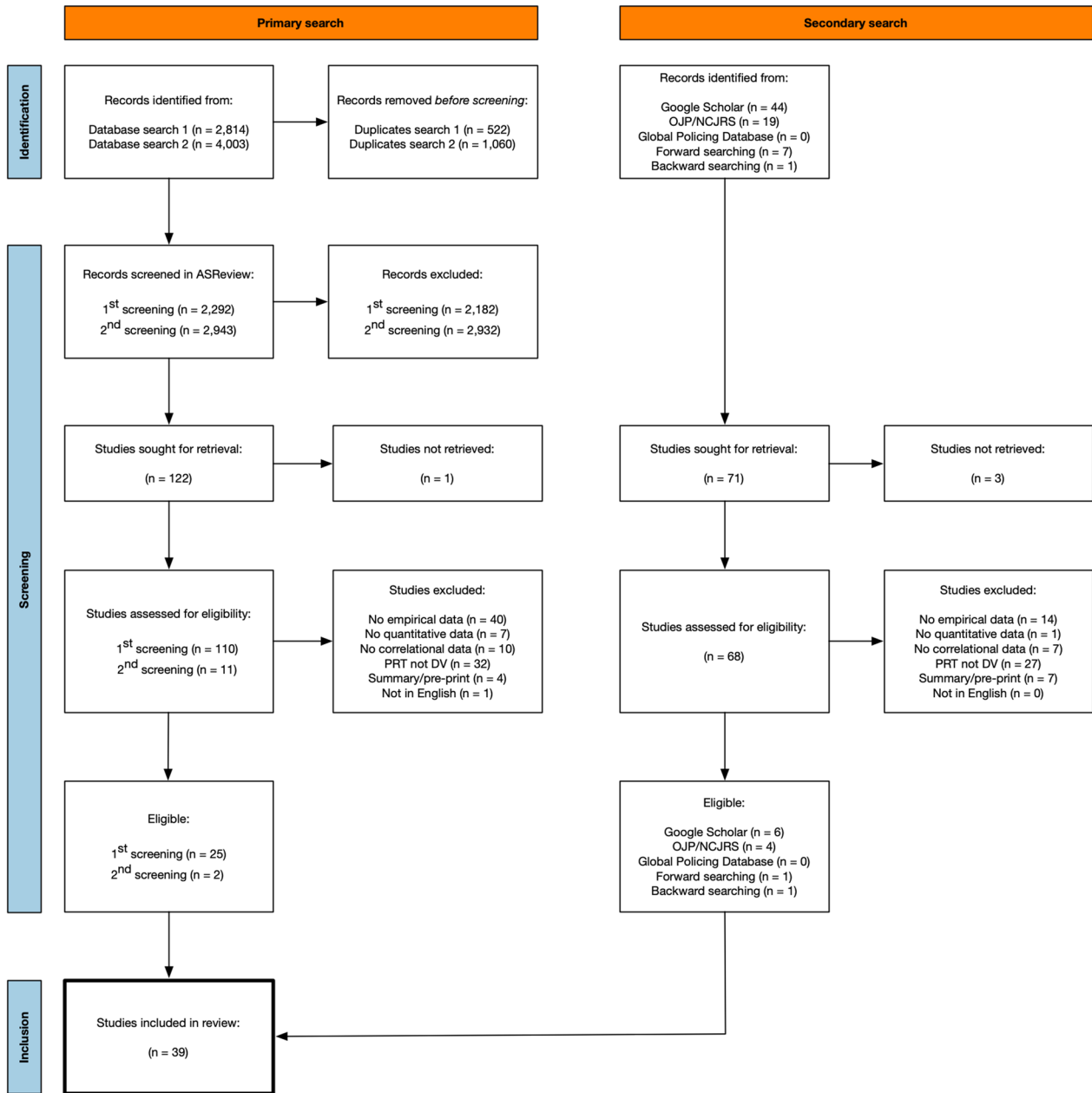
Appendix B: included studies

Authors	Pub. Year	Research location	Design	Eval?	Evaluated	Study type	Sampling technique
Bennett	(2018)	40 jurisdictions in USA	Observational	No	–	Sample	Stratified random sampling
Bieck	(1978)	Kansas City (USA)	Observational	No	–	Sample	Non-random sampling
Bieck	(1980a)	Kansas City (USA)	Observational	No	–	Sample	Non-random sampling
Bieck	(1980b)	Kansas City (USA)	Observational	No	–	Sample	Non-random sampling
Blackstone, Hakim & Meehan	(2020)	Salt Lake City (USA)	Observational	Yes	Verified Response	Population	–
Blake & Coupe	(2001)	West Midlands (UK)	Observational	Yes	One or two-officer car	Sample	Disproportionate stratified sampling
Boydstun et al	(1977)	San Diego (USA)	Experimental	Yes	One or two-officer car	Sample	Random sampling
Caldwell	(1971)	Durham (UK)	Observational	No	–	Population	–
Choi, Librett & Collins	(2014)	Brockton (USA)	Quasi-experimental	Yes	AGDS	Population	–
Cihan	(2015)	Houston (USA)	Observational	No	–	Population	–
Cihan	(2014)	Houston, Dallas (USA)	Observational	No	–	Population	–
Cihan, Zhang & Hoover	(2012)	Houston (USA)	Observational	No	–	Population	–
Colton	(1980)	San Diego (USA)	Observational	No	–	Population	–
Coupe & Blake	(2005)	West Midlands (UK)	Observational	No	–	Sample	Disproportionate stratified sampling
Coyne et al	(2014)	Country-wide USA	Observational	Yes	Mobilizing reserve troops	Population	–
Goldenberg et al	(2019)	Camden (USA)	Observational	Yes	AGDS	Population	–
Isaacs	(1967)	Los Angeles (USA)	Observational	No	–	Population	–
Joslin et al	(2016)	Syracuse (USA)	Observational	Yes	Vocera badge	Population	–
Kayes et al	(2019)	Florida (USA)	Observational	No	–	Population	–
Kessler	(1985)	Kansas City (USA)	Observational	Yes	One or two-officer car	Sample	non-random sampling
Larson & Simon	(1979)	St Louis (USA)	Experimental	Yes	AVM	Population	–
Lee, Lee & Hoover	(2017)	Houston (USA)	Observational	No	–	Population	–
Mares & Blackburn	(2021)	St Louis (USA)	Quasi-experimental	Yes	AGDS	Population	–
Mares & Blackburn	(2012)	St Louis (USA)	Quasi-experimental	Yes	AGDS	Population	–
Maxfield	(1982)	San Francisco (USA)	Observational	No	–	Population	–
Mazerolle et al	(2002)	Dallas (USA)	Observational	Yes	Non-emergence call system	Population	–
Mazerolle et al	(1998)	Dallas (USA)	Experimental	Yes	AGDS	Population	–
McEwan, Conner & Cohen	(1986)	Garden Grove; Greensboro; Toledo (USA)	Experimental	Yes	Differential Police response	Population	–
Mladenka & Hill	(1978)	Houston (USA)	Observational	No	-	Sample	Random sampling
Mohammadi et al	(2013)	Sistan and Baluchistan (IRAN)	Observational	No	–	Population	–
Myrick et al	(1979)	Tucker (USA)	Observational	No	–	Population	–
Pate et al	(1976)	Kansas City (USA)	Observational	No	-	Sample	Non-random sampling
Propheter	(2020)	Sacramento (USA)	Observational	Yes	Presence of Urban sport facilities	Population	–

Authors	Pub. Year	Research location	Design	Eval?	Evaluated	Study type	Sampling technique
Russo	(2006)	Altamonte Springs (USA)	Observational	Yes	AVL	Population	–
Salimbene & Zhang	(2020)	Northeast Texas (USA)	Observational	No	–	Population	–
Spelman & Brown	(1984)	Jacksonville, Peoria, Rochester, San Diego (USA)	Observational	No	–	Sample	Stratified random sampling
Stevens, Webster and Stipak	(1980)	York (USA)	Observational	No	–	Sample	Disproportionate stratified sampling
Sullivan	(2012)	Southwest Sky (USA)	Observational	No	–	Population	–
Vigoa	(2010)	Doral (USA)	Observational	No	–	Population	–

Appendix C: data extraction form

Column name	Explanation
Independent Variable (IV)	Predictor variable name as presented in the primary study
IV dummy	
IV label	
IV type	Predictor variable type: *nominal, dichotomous, ordinal* or *scale*
IV description	A description of what the predictor variable entails
IV reference category	If *nominal* or *ordinal*: IV reference category is noted here
IV unit	If *scale*: IV unit of analysis is noted here
IV transformation	*None, Ln*
DV component	*Reporting, dispatch* or *travel time;* or any combination
DV type	The outcome variable type: *nominal, ordinal* or *scale*
DV unit	Unit of analysis of police response time (usually *min/incident*)
DV transformation	*None, Ln*
Mechanism described	
Counterfactual/balancing	
IV function	
Study reference	Authors (year published)
Source	Table or page where extracted data was found
Correlation test	Statistical test used to gage the correlation between IV and DV
Conclusion study	The conclusion the study draws based on the correlational data

Column name	Explanation
Sample/population	Does the study concern population or sample data?
Sample space	To what geographic unit does this data generalize?
Research design	*Observational/quasi-experimental/experimental*
Intervention research	*Yes/no*
Evaluated	Description of what (IV) is evaluated
CAD data?	*Yes/no*
Alternative dataset	If not CAD, what then?
Sampling strategy	Sampling strategy for analysis unit of analysis (usually *CFSs*)
Sample info	Additional information on sampling choices specified
Sample priority level 1	
N (sample)	Sample size
Sample unit	Unit of analysis
Unit scope	What subset of CFSs are analyzed?
Metric	Metric in which the association is expressed
Value	Value of the metric
Standard error	Standard error of the value of the metric
Significance	p-value of the value of the metric
Date of data extraction	The starting and ending day, month and year of data extraction
Descriptives	Descriptive statistics IV such as mean, standard deviation, median, degrees of freedom are collected for studies using population data

Appendix D: data conversion and interpretation form
See Tables 5, 6

Table 5 Metric conversion table

Input metric(s)	Condition	Conversion	Output metric
Population Means	Population study	$\frac{(\mu_{post}-\mu_{pre})}{\mu_{pre}}$	% Change
Sample Means, sample standard deviations, sample size	Sample study	$\frac{(\overline{X}_{post}-\overline{X}_{pre})}{\left(\frac{(N-1)*SD^2_{post}+(N-1)*SD^2_{pre}}{(N+1)+(N-1)}\right)}$	Cohen's D
Z-statistic, sample size	–	$2*\frac{z}{\sqrt{N}}$	Cohen's D
t-statistic	–	$2*\frac{t}{\sqrt{N-1}}$	Cohen's D
Unstandardized B, standard deviations	Standard deviations provided	$B*\frac{SD_{predictor\ variable}}{SD_{outcome\ variable}}$	Beta
Unstandardized B, Grand mean	Grand mean provided	$\frac{B}{Mean}$	B*
Unstandardized B	Logit regression	e^B	Odds ratio

Table 6 Metric interpretation table

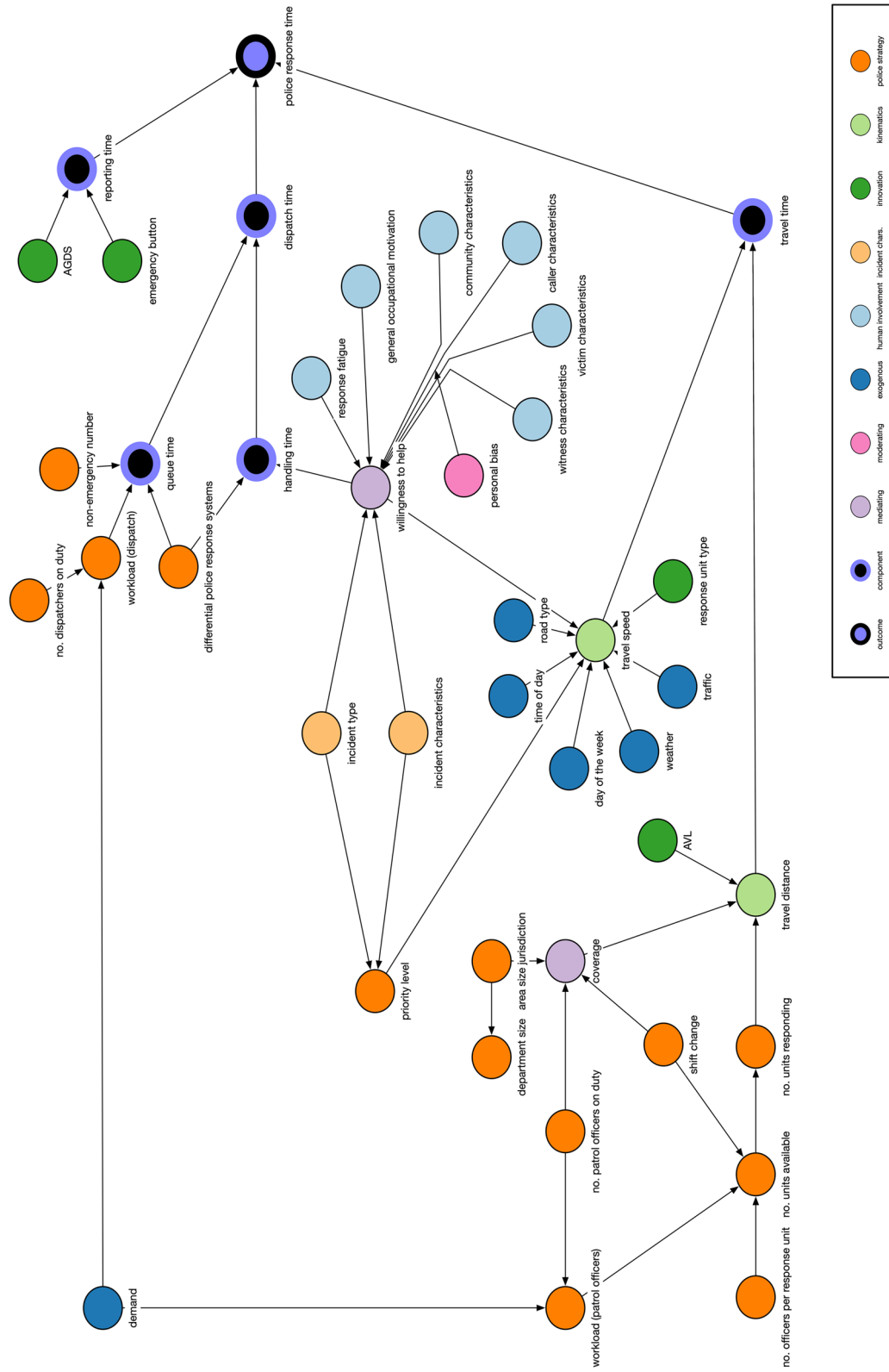| Input metric	Very small	Small	Medium	Large	Interpretation based on:
Cohen's D	<, ±0.2	[±0.2,±0.5]	[±0.5,±0.8]	>±0.8	(Cohen, 2013)
Pearson's R	<, 0.1]	[0.1, 0.3]	[0.3, 0.5]	>0.5	(Cohen, 2013)
Odds Ratio	<, 1.44]	[1.44, 2.48]	[2.48, 4.27]	>4.27	Conversion from (Cohen, 2013)
% Change	<, 0.1]	[0.1, 0.3]	[0.3, 0.5]	>0.5	We posit these as reasonable interpretations for % changes in *police response times*
B*	<, 0.1]	[0.1, 0.3]	[0.3, 0.5]	>0.5	Analogous to % Change
Beta	<, 0.1]	[0.1, 0.3]	[0.3, 0.5]	>0.5	Analogous to B*, under the assumption of equal standard deviations between the predictor and outcome variable

Appendix E: Hybrid synthesis of predictor variable subcategory results

Category	Subcategory	Effect size interpretation	Direction	N	Mean internal validity
Exogenous	Day of the week	Large	None	1	3
	Day of the week	Very small	None	7	3
	Demand	Large	+	1	1
	Demand	Medium	+	1	3
	Demand	Very small	+	3	3
	Road type	Very small	–	3	3
	Road type	Very small	None	9	3
	Time of day	Large	None	1	1
	Time of day	Medium	None	6	1
	Time of day	Small	None	11	1.18
	Time of day	Very small	None	18	1.89
	Traffic	Very small	None	4	3
	Weather type	Very small	+	2	3
	Weather type	Very small	None	4	3

Category	Subcategory	Effect size interpretation	Direction	N	Mean internal validity
Human involvement	Caller characteristics	Large	None	1	1
	Caller characteristics	Small	None	18	3
	Caller characteristics	Very small	None	20	3
	Community characteristics	Large	−	1	3
	Community characteristics	Large	+	2	3
	Community characteristics	medium	−	3	3
	Community characteristics	Medium	+	1	3
	Community characteristics	Small	−	14	2
	Community characteristics	Small	+	12	1.67
	Community characteristics	Small	None	5	1
	Community characteristics	Very small	−	20	2.7
	Community characteristics	Very small	+	14	2.43
	Community characteristics	Very small	None	15	1.27
	Victim characteristics	Large	None	1	1
	Victim characteristics	Medium	None	1	1
	Victim characteristics	Small	−	3	3
	Victim characteristics	Small	None	3	3
	Victim characteristics	Very small	−	3	3
	Victim characteristics	Very small	+	6	3
	Victim characteristics	Very small	None	6	2.33
	Witness characteristics	Small	None	1	1
Incident	Incident characteristics	Large	None	4	2
	Incident characteristics	Medium	−	3	1
	Incident characteristics	Medium	None	4	1.5
	Incident characteristics	Small	−	3	1.67
	Incident characteristics	Small	+	1	3
	Incident characteristics	Small	None	2	2.5
	Incident characteristics	Very small	−	1	3
	Incident characteristics	Very small	+	2	2
	Incident characteristics	Very small	None	7	2.71
	Incident type	Large	None	10	1.6
	Incident type	Medium	None	16	1.25
	Incident type	Small	None	100	1.52
	Incident type	Very small	None	123	1.54
Innovation	AGDS	Medium	None	4	2
	AGDS	Small	None	2	2
	AGDS	Very small	None	5	2
	AVL	Very small	None	2	2
	Emergency button	Large	None	1	2
	Response unit type	Medium	None	2	2
	Response unit type	Small	None	2	2
	Response unit type	Very small	None	1	2
Kinematics	Travel distance	Large	+	2	2
	Travel distance	Medium	+	2	1
	Travel distance	Small	+	1	3
	Travel distance	Small	None	3	3
	Travel speed	Very small	+	1	1

Category	Subcategory	Effect size interpretation	Direction	N	Mean internal validity
Police strategy	Call priority	Large	−	2	3
	Call priority	Large	+	8	1.25
	Call priority	Large	None	11	1
	Call priority	Medium	−	1	1
	Call priority	Medium	+	1	1
	Call priority	Medium	None	2	1
	Call priority	Small	−	2	2
	Call priority	Small	+	1	1
	Call priority	Small	None	3	1
	Call priority	Very small	−	2	1
	Call priority	Very small	+	1	1
	Department size	Small	−	2	3
	Department size	Very small	−	2	3
	Department size	Very small	+	2	3
	Department size	Very small	None	1	3
	Differential response system	Large	None	1	1
	Differential response system	Very small	None	3	1
	Jurisdiction area size	Small	+	1	1
	Jurisdiction area size	Very small	None	3	3
	No. officers per response unit	Small	None	4	2
	No. officers per response unit	Very small	None	5	2
	Non-emergency number	Medium	None	2	2
	Non-emergency number	Small	None	1	2
	Non-emergency number	Very small	None	5	2
	On duty officers	Very small	None	2	3
	Patrol strategy	Medium	−	1	1
	Patrol strategy	Small	−	1	3
	Patrol strategy	Small	None	5	3
	Patrol strategy	Very small	None	9	3
	Shift	Small	None	1	3
	Shift	Very small	None	1	3
	Workload	Large	+	2	1.5
	Workload	Medium	+	1	1
	Workload	Small	+	2	1.5
	Workload	Very small	+	1	3
	Workload	Very small	None	1	3

Appendix F: causal model of police response time
See Fig. 3

Fig. 3 Proposed causal model of the observed and discussed subcategories of predictor variables

Declarations

Competing interests
The authors declare that they have no competing interests.

References
Bennett, D. S. (2018). *Police response times to calls for service.* Stanford.

Bieck, W. H. (1978). Response time analysis volume II: Analysis: Vol. II.

Bieck, W. H. (1980a). Response time analysis volume III: Part II crime analysis.

Bieck, W. H. (1980b). Response time analysis volume IV: Noncrime call analysis: Vol. IV.

Blackstone, E. A., Hakim, S., & Meehan, B. (2020). Burglary reduction and improved police performance through private alarm response. *International Review of Law and Economics, 63*, 105930. https://doi.org/10.1016/j.irle.2020.105930

Blake, L., & Coupe, R. T. (2001). The impact of single and two-officer patrols on catching burglars in the act: A critique of the audit commission's reports on youth justice. *British Journal of Criminology, 41*(2), 381–396. https://doi.org/10.1093/bjc/41.2.381

Boydstun, J. E., Sherry, M. E., & Moelter, N. P. (1977). Patrol staffing in San Diego One- Or Two- Officer Units.

Brandl, S. G., & Horvath, F. (1991). Crime-victim evaluation of police investigative performance. *Journal of Criminal Justice, 19*(3), 293–305. https://doi.org/10.1016/0047-2352(91)90008-J

Brown, B. B., & Harris, P. B. (1989). Residential burglary victimization: Reactions to the invasion of a primary territory. *Journal of Environmental Psychology, 9*(2), 119–132. https://doi.org/10.1016/S0272-4944(89)80003-9

Caldwell, R. (1971). *Optimal distribution policy for mobile police patrols.* Durham University.

Choi, K.-S., Librett, M., & Collins, T. J. (2014). An empirical evaluation: Gunshot detection system and its effectiveness on police practices. *Police Practice and Research, 15*(1), 48–61. https://doi.org/10.1080/15614263.2013.800671

Cihan, A. (2014). Social disorganization and police performance to burglary calls: a tale of two cities. *Policing an International Journal of Police Strategies & Management, 37*(2), 340–354. https://doi.org/10.1108/PIJPSM-06-2013-0058

Cihan, A. (2015). Examining the neighborhood effects on police performance to assault calls. *Police Practice and Research, 16*(5), 391–401. https://doi.org/10.1080/15614263.2014.927765

Cihan, A., Zhang, Y., & Hoover, L. (2012). Police response time to in-progress burglary: A multilevel analysis. *Police Quarterly, 15*(3), 308–327. https://doi.org/10.1177/1098611112447753

Cohen, J. (2013). Statistical power analysis for the behavioral sciences. *Routledge.* https://doi.org/10.4324/9780203771587

Colquhoun, H. L., Levac, D., O'Brien, K. K., Straus, S., Tricco, A. C., Perrier, L., Kastner, M., & Moher, D. (2014). Scoping reviews: Time for clarity in definition, methods, and reporting. *Journal of Clinical Epidemiology, 67*(12), 1291–1294. https://doi.org/10.1016/j.jclinepi.2014.03.013

Colton, K. W. (1980). Police and computer technology: The case of the San Diego computer-aided dispatch system. *Public Productivity Review, 4*(1), 21. https://doi.org/10.2307/3380055

Cordner, G. W., Greene, J. R., & Bynum, T. S. (1983). *The sooner the better: Some effects of police response time. Police at work: Policy issues and analysis* (pp. 145–164). Sage Publications.

Coupe, R. T., & Blake, L. (2005). The effects of patrol workloads and response strength on arrests at burglary emergencies. *Journal of Criminal Justice, 33*(3), 239–255. https://doi.org/10.1016/j.jcrimjus.2005.02.004

Coyne, C. J., Hall, A. R., McLaughlin, P. A., & Zerkle, A. (2014). A hidden cost of war: The impact of mobilizing reserve troops on emergency response times. *Public Choice, 161*(3–4), 289–303. https://doi.org/10.1007/s11127-014-0200-4

Farrington, D. P., Gottfredson, D. C., Sherman, L. W., & Welsh, B. C. (2003). The maryland scientific methods scale. In D. P. Farrington, D. L. MacKenzie, L. W. Sherman, & B. C. Welsh (Eds.), *Evidence-based crime prevention* (pp. 27–35). Routledge.

Frank, J., Smith, B. W., & Novak, K. J. (2005). Exploring the basis of citizens' attitudes toward the police. *Police Quarterly, 8*(2), 206–228. https://doi.org/10.1177/1098611103258955

Goldenberg, A., Rattigan, D., Dalton, M., Gaughan, J. P., Thomson, J. S., Remick, K., Butts, C., & Hazelton, J. P. (2019). Use of ShotSpotter detection technology decreases prehospital time for patients sustaining gunshot wounds. *Journal of Trauma and Acute Care Surgery, 87*(6), 1253–1259. https://doi.org/10.1097/TA.0000000000002483

Isaacs, H. (1967). A study of communcations, crimes, and arrests in a metropolitan police department. In Task Force Report: science and technology.

Joslin, J. D., Goldberger, D., Johnson, L., & Waltz, D. P. (2016). Use of the vocera communications badge improves public safety response times. *Emergency Medicine International, 2016*(2014), 1–3. https://doi.org/10.1155/2016/7158268

Kayes, M. I., Al-Deek, H., Sandt, A., Carrick, G., Staves, C., Gamaleldin, G., Faruk, O., & Al-Sahili, O. (2019). Characteristics of law enforcement response to wrong-way driving events in Florida. *Transportation Research Record: Journal of the Transportation Research Board, 2673*(9), 514–524. https://doi.org/10.1177/0361198119846456

Kessler, D. A. (1985). One-or two-officer cars? A perspective from Kansas City. *Journal of Criminal Justice, 13*(1), 49–64. https://doi.org/10.1016/0047-2352(85)90025-X

Larson, G. C., & Simon, J. W. (1979). *Evaluation of a police automatic vehicle monitoring (AVM) System: A study of the St. Louis experience 1976–1977.* U.S. Department of Justice, Law Enforcement Assistance Administration, National Institute of Law Enforcement and Criminal Justice.

Lee, J.-S., Lee, J., & Hoover, L. T. (2017). What conditions affect police response time? Examining situational and neighborhood factors. *Police Quarterly, 20*(1), 61–80. https://doi.org/10.1177/1098611116657327

Mares, D., & Blackburn, E. (2012). Evaluating the effectiveness of an acoustic gunshot location system in St. Louis. *MO. Policing., 6*(1), 26–42. https://doi.org/10.1093/police/par056

Mares, D., & Blackburn, E. (2021). Acoustic gunshot detection systems: a quasi-experimental evaluation in St. Louis. *MO. Journal of Experimental Criminology, 17*(2), 193–215. https://doi.org/10.1007/s11292-019-09405-x

Maxfield, M. G. (1982). Service Time, dispatch time, and demand for police services: Helping more by serving less. *Public Administration Review, 42*(3), 252. https://doi.org/10.2307/976012

Mazerolle, L. G., Rogan, D., Famega, C., & Eck, J. E. (2002). Managing citizen calls to the police: The impact of Baltimore's 3–1–1 call system. *Criminology & Public Policy, 2*(1), 97–124. https://doi.org/10.1111/j.1745-9133.2002.tb00110.x

Mazerolle, L. G., Watkins, C., Rogan, D., & Frank, J. (1998). Using gunshot detection systems in police departments: The impact on police response times and officer workloads. *Police Quarterly, 1*(2), 21–49. https://doi.org/10.1177/109861119800100202

McEwen, J. T., Connors, E. F., & Cohen, M. I. (1986). *Evaluation of the differential police response field test*. Department of Justice, National Institute of Justice.

Mladenka, K. R., & Hill, K. Q. (1978). The distribution of urban police services. *The Journal of Politics, 40*(1), 112–133. https://doi.org/10.2307/2129978

Mohammadi, M., Martiniuk, A. L. C., Ansari-Moghaddam, A., Rad, M., Rashedi, F., Ghjasemi, A., & Ahangari, H. (2013). Police response time to road crashes in south-east of Iran. *Journal of the Pakistan Medical Association, 63*(12), 1523–1527.

Myrick, J. A., Pittman, J. V., Sayford, N. F., & Thomason, C. Y. (1979). *Evaluation of the role of police in the EMS system. Final report*.

Pate, T., Ferrara, A., Bowers, R. A., & Lorence, J. (1976). *Police response time: Its determinants and effects*. Washington: Police Foundation.

Propheter, G. (2020). Do urban sports facilities have unique social costs? An analysis of event-related congestion on police response time. *International Journal of Urban Sciences, 24*(2), 271–281. https://doi.org/10.1080/12265934.2019.1625805

Russo, C. W. (2006). *AVL and response time reduction: image and reality*. University of Central Florida.

Salimbene, N. A., & Zhang, Y. (2020). An examination of organizational and community effects on police response time. *Policing: an International Journal, 43*(6), 935–946. https://doi.org/10.1108/PIJPSM-04-2020-0063

Scherhag, J., & Burgard, T. (2023). Performance of semi-automated screening using Rayyan and ASReview: A retrospective analysis of potential work reduction and different stopping rules. *Big Data & Research Syntheses 2023, Frankfurt, Germany*. https://doi.org/10.23668/psycharchives.12843

Spelman, W., & Brown, D. K. (1984). *Calling the police: Citizen reporting of serious crime*. US Department of Justice, National Institute of Justice.

Stevens, J. M., Webster, T. C., & Stipak, B. (1980). Response time: Role in assessing police performance. *Public Productivity Review, 4*(3), 210. https://doi.org/10.2307/3379854

Sullivan, E. (2012). *The Effects of Crime Incident Characteristics and Neighborhood Structure on Police Response Time*. Arizona State University.

Textor, J., van der Zander, B., Gilthorpe, M. S., Liśkiewicz, M., & Ellison, G. T. H. (2016). Robust causal inference using directed acyclic graphs: The R package 'dagitty.' *International Journal of Epidemiology, 45*(6), 1887–1894.

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., Moher, D., Peters, M. D. J., Horsley, T., Weeks, L., Hempel, S., Akl, E. A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M. G., Garritty, C., & Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine, 169*(7), 467–473. https://doi.org/10.7326/M18-0850

van de Schoot, R., de Bruin, J., Schram, R., Zahedi, P., de Boer, J., Weijdema, F., Kramer, B., Huijts, M., Hoogerwerf, M., Ferdinands, G., Harkema, A., Willemsen, J., Ma, Y., Fang, Q., Hindriks, S., Tummers, L., & Oberski, D. L. (2021). An open source machine learning framework for efficient and transparent systematic reviews. *Nature Machine Intelligence, 3*(2), 125–133. https://doi.org/10.1038/s42256-020-00287-7

Vidal, J. B. I., & Kirchmaier, T. (2018). The effect of police response time on crime clearance rates. *Review of Economic Studies, 85*(2), 855–891. https://doi.org/10.1093/restud/rdx044

Vigoa, O. (2010). Comparative analysis between routine preventative uniformed police patrols and crime reduction and response to calls for service. In *ProQuest Dissertations and Theses*. Lynn University.

Weisburd, S. (2021). Police presence, rapid response rates, and crime prevention. *The Review of Economics and Statistics, 103*(2), 280–293. https://doi.org/10.1162/rest_a_00889

Zhu, R., Hu, X., Li, X., Ye, H., & Jia, N. (2020). Modeling and risk analysis of chemical terrorist attacks: A Bayesian network method. *International Journal of Environmental Research and Public Health, 17*(6), 2051. https://doi.org/10.3390/ijerph17062051

Publisher's Note