A systematic review and meta-analysis on the effect social groups have on the egress times of pedestrian crowds

Yanghui Hu[1], Nikolai WF Bode[2]*

1 State Key Laboratory of Fire Science, University of Science and Technology of China, Hefei 230027, China

2 Department of Engineering Mathematics, University of Bristol, Bristol, BS8 1TW, UK

*corresponding author: Nikolai WF Bode. nikolai.bode@bristol.ac.uk

Word count: 6986 (not including references, appendix, tables, and figure captions)

A systematic review and meta-analysis on the effect social groups have on the egress times of pedestrian crowds

Conflicting results in the literature raise the question of how reliable findings from single experiments on pedestrian crowd movements are. An important example is the effect social groups have on crowd egress times from confined spaces where both increases and decreases were reported. We identify only six comparable studies and conduct the first analysis on this topic that integrates evidence from multiple studies quantitatively, accounting for different sample sizes. The aggregated findings suggest social groups increase average egress times but there is insufficient evidence to reject the null hypothesis of no effect. The conflicting results across published studies are thus likely to have arisen by chance, as experiments are statistically underpowered for determining a small effect. We find no evidence for publication bias in terms of findings or statistical power. Our work presents a quantitative basis for discussing the statistical reliability of experiments considering the high context-dependency of pedestrian dynamics.

Keywords: pedestrian dynamics; social groups; egress time; evacuation time; meta-analysis

Introduction

Research into pedestrian dynamics is concerned with the movement of walkers through the built environment. This field of investigation is highly interdisciplinary with methods including theory development, analysis of observational data and, increasingly, controlled experiments (Schadschneider and Seyfried 2011). Typical experiments are conducted with groups of volunteers, often undergraduate students, who are instructed to walk through an experimental setup that simulates a corridor, room, building, or similar (Haghani 2020). Experimental manipulations are aimed at investigating changes in the movement dynamics of pedestrian crowds and originally focused predominantly on the infrastructure, adjusting the width of exits or corridors, for example. Increasingly, experimenters have also turned to other aspects deemed to be important in pedestrian crowds, such as the physical or psychological characteristics of the individuals making up a crowd. Research in other fields has demonstrated potential problems with the evidence produced by such experimental studies (L'Abbé, Detsky, and O'Rourke 1987). Common issues relate to the repeatability of findings, consider the 'replication crisis' (Fidler and Wilcox 2018), and the extent to which single studies have sufficient statistical power to ensure they can be relied upon. Here we focus on the latter issue and investigate the combined evidence of different studies in pedestrian dynamics in an attempt to reconcile disagreements in findings. This approach, known as meta-analysis, is widely used in other research fields, exemplified for medical research by the Cochrane organisation, but currently underused in transportation studies in general and pedestrian dynamics in particular (for an example on the movement of individual pedestrians, see [Sharmin and Kamruzzaman 2018]).

We argue that the nature of experimental research on pedestrian dynamics makes it an interesting candidate for a meta-analysis, for the following reasons. First, the logistic difficulty and often cost involved in assembling large groups of participants in the same place at the same time mean that many studies only test one or a few groups of participants and only perform few experimental runs or replicates (typically no more than 20 [Haghani 2020]). Statistical theory dictates that for both larger natural variability in measurements and smaller magnitudes of the effect resulting from experimental manipulations, findings are less reliable, given such low numbers of replicates (studies have less statistical power [Cohn and Becker 2003]). Second, many studies consider similar measures of interest, often related to the flow or egress time of pedestrians, making comparisons across studies possible. Third, there are conflicting findings from different studies on the same topic, as discussed below, suggesting the need for a statistical investigation to reconcile these discrepancies. Fourth, the substantial increase in published experimental studies indicates there is a sufficiently broad evidence base for meta-analyses. In this contribution, we identify a topic in pedestrian dynamics suitable for and in need of a meta-analysis: the effect of social groups within crowds on the egress time, the time it takes to leave a room or building.

Instead of being homogeneous, pedestrian crowds are composed of individuals with different characteristics. One factor that distinguishes individuals is their membership in social groups, sets of pedestrians with a social relationship among them that is based on kinship or friendship, for example (Nicolas and Hassan 2021). Observations have shown that in daily life a high proportion of pedestrians walk in social groups. In addition, members of these social groups show distinctive behaviours, including trying to stay close to each other as much as possible, sharing the same destination, and even adopting formations to facilitate communication (Moussaïd et al. 2010; Federici et al. 2012; Singh et al. 2009; Gorrini, Bandini, and Sarvi 2014), so much so, that groups can be detected automatically from movement data (Li et al. 2021). Perhaps because of the distraction arising from talking to others or because they adjust to the slowest group member, the movement speed of social groups has been observed to be lower than that of individuals, on average (Moussaïd et al. 2010; Carey 2005; Schultz, Schulz, and Fricke 2010). During emergencies, social ties increase the likelihood that social groups evacuate together, help each other, wait for each other or search for each other (Drabek and Stephenson III 1971; Lindell, Kang, and Prater 2011; Yang et al. 2005; Thompson, Garfin, and Silver 2017).

These distinctive movement characteristics of social groups raise questions on the extent to which the presence of social groups within crowds affects the overall movement dynamics. Interestingly, different studies have found conflicting results in this regard. One of the most commonly studied aspects of crowd movement is the egress time. Some studies find that crowds with social groups show longer egress times compared with crowds composed entirely of unrelated individuals (Köster et al. 2011; Bode et al. 2015), suggesting that social groups could negatively affect evacuation processes. However, other work suggests that the presence of social groups reduces the egress times of crowds, perhaps because social groups can organize more effectively in space (von Krüchten and Schadschneider 2017). In addition, it has been suggested that the degree of familiarity between social group members and visibility levels are important for determining the effect of social groups on the movement dynamics of pedestrian crowds (Ma et al. 2017; Xie et al. 2018). This ambiguity in experimental findings is echoed by similarly varied findings from theoretical work using simulation models (e.g. reviewed in Bode et al. [2015]).

Here, we explore whether integrating the evidence from multiple studies helps to bring clarity about the effect of social groups on egress times. To the best of our knowledge, there is currently no systematic review and meta-analysis that attempts to comprehensively analyse and integrate these findings. We suggest this will be directly useful to understand the role of social groups and the extent to which they need to be considered in pedestrian crowd management, architectural design, and risk prevention. In addition, we suggest our work may be a useful case study to initiate a discussion on quantitative meta-research in pedestrian dynamics. Thus, we discuss the challenges, relevance, and importance of this type of research in detail throughout this contribution.

Before we present our methodological approach and findings, it is useful to discuss in general terms some of the challenges in conducting a meta-analysis on experimental work in pedestrian dynamics. It is uncommon that different studies investigate the same scenario. For example, egress is often studied from rooms of different dimensions and involving different sizes of social groups and different total numbers of participants. This means that measures may have to be standardised somehow before comparison across studies, to account for the inherent differences in experimental settings. It is important to ensure that this does not obfuscate the essence of what is measured. In addition, a key question from a statistical perspective is whether measurements taken from several experimental runs with the same group of participants can be considered independent data points or not. On the one hand, testing the same group of participants multiple times means individuals could habituate to the experimental setting or get tired, which may affect results. On the other hand, it could be argued that measurements on crowds are unlikely to be influenced by individuals, meaning that it would take a collective action of all participants to substantially skew results and such a collective action may be noticed by experimenters. The logistic difficulty of conducting experiments with crowds of volunteers, already discussed above, means we have to assume measurements from the same groups of participants can be treated as separate data points to ensure we have enough data for our meta-analysis.

While this discussion indicates that our meta-analysis should be interpreted carefully, we argue that the process of conducting a meta-analysis in this field of research itself is a useful starting point for a discussion on the repeatability, reliability, and interpretation of experimental findings in pedestrian dynamics.

Methods

Literature search strategy

Our systematic review was conducted according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Moher et al. 2009; Liberati et al. 2009). The purpose of our study was to investigate whether the presence of social groups has an impact on the time it takes pedestrian crowds to leave a confined space (egress time or evacuation time). We adopted two literature search strategies. First, we performed an electronic search on Scopus, Elsevier's abstract and citation database. We combined the keywords *(("small group" OR "social group" OR "social groups" OR "small groups") AND ("pedestrian" OR "crowd") AND ("experimental" OR "experiment") AND ("evacuation" OR "egress"))* and searched for them in titles, abstracts and keywords of publications. Search items were limited to journal articles and conference papers in English. Second, we conducted a manual search on the conference proceedings of the two conference series "Pedestrian and Evacuation Dynamics (PED)" and "Traffic and Granular Flow (TGF)" for the years 2012-2019 considering the same keywords as above (Weidmann, Kirsch, and Schreckenberg 2014; Chraibi et al. 2015; Daamen, Duives, and Hoogendoorn 2014; Knoop and Daamen 2016; Song, Ma, and Fu 2017; Hamdar 2019; Dederichs, Köster, and Schadschneider 2020; Zuriguel, Garcimartín, and Hidalgo 2020). In addition, we reviewed references cited by the retrieved articles for potentially relevant studies. The literature search was performed separately by two independent reviewers (the authors) on the 19th March 2021.

Selection criteria

The criteria for selecting studies to be included in this review were: 1) original studies on pedestrian dynamics, written in English and published as a journal or conference paper, 2) research presenting controlled experiments with volunteers which investigate the evacuation or egress time of pedestrian crowds (i.e. pedestrians leaving a confined space or passing through a bottleneck, such as a door), 3) studies that include two treatments contrasting pedestrians walking individually to walking in social groups (the size, definition, and simulation of social groups can vary across studies, as discussed below in section 'Data extraction and critical appraisal'), 4) egress must not include any changes in level using stairs or ramps. Changes in level via stairs or ramps cause movement features, such as, reduced speeds, the danger of tripping, and wall-following behaviour

(Zhang et al. 2018) that strongly influence crowd egress times (Frantzich 1996) and were not the subject of this review.

Data extraction and critical appraisal

The experiments considered in this review each included a number of experimental runs. In each experimental run, a crowd of volunteers simulates pedestrian egress or evacuation. Each experimental run thus yields one observation of the egress/evacuation time. Different experimental runs were conducted under two different experimental treatments that either simulated the presence of social groups or not.

Some studies investigate the effect the size of social groups has on crowd egress dynamics, such as comparing the effect when crowds are composed of social groups of size 2, 3, or 4 individuals. As there were too few studies testing different social group sizes, only the difference between the presence and absence of social groups was investigated. Data of different groups were amalgamated for the meta-analysis presented here. A possible justification for this approach is the observation that naturally occurring pedestrian crowds contain a mixture of social group sizes (Moussaïd et al. 2010).

For each selected study, we extracted the average egress time in seconds and its standard deviation from the article text, tables, or figures for each of the two treatments (averaging over different social group sizes, if applicable). In addition, we recorded the number of experimental runs for each treatment and the size of the pedestrian crowd.

Several studies tested further conditions in addition to the two treatments investigated here. For example, different exit widths were considered. As these studies only presented one experimental run for each treatment in each condition, data had to be amalgamated across these additional conditions. We achieved this by shifting the data to remove the mean differences across all conditions except for the two treatments investigated here. For example, in a study investigating two different exit widths, the difference in mean egress time across exit widths was subtracted from each egress time recorded for the narrower exit. Thus, the mean egress time was equalized for the two exit widths but the difference in egress time across the two treatments investigated here was not altered. Details of the data extraction for each selected study are included in the appendix.

Due to the experimental setups used, the recorded egress times differed substantially across studies. For example, consider the egress time of a crowd of almost 100 volunteers compared to the egress time of a dozen volunteers. Prior to statistical analysis, we thus standardized the egress times separately for each study by dividing the mean (and standard deviations) by the average egress time of the treatment in which pedestrians moved as individuals.

We performed a critical appraisal of the selected studies to provide an indication of the relative reliability, quality, and risk of bias across studies. In the Medical Sciences, widely accepted frameworks have been developed for this purpose (e.g. the Newcastle Ottawa Scale (Lo, Mertz, and Loeb 2014) or the Grades of Recommendation, Assessment, Development and Evaluation, GRADE [Canadian Task Force on Preventive Health Care 2011]). No such frameworks exist for research in pedestrian dynamics. Therefore, we recorded the five following indicators for study design and quality.

First, we reported the number of separate participant groups that completed the experiment. Some studies in pedestrian dynamics perform all experimental runs with one participant group. For experiments that rely on instructions, such as ones related to the simulation of social groups, it cannot be ruled out that differences in interpretation of instructions by different participant groups, which could lead to different behaviour during experiments. As such, testing multiple participant groups may generally be preferable.

Second, we determined if the two treatments under consideration were always completed in the same temporal order by participant groups (in all studies considered, all participant groups completed both treatments). A fixed temporal order could affect the behaviour of participant groups as they may habituate to the experimental context or get tired if many experimental runs were completed.

Third, we assessed the homogeneity of participant groups. For example, similar to research in other fields, experiments in pedestrian dynamics rely heavily on university students or similarly aged participants. Such participant groups are unlikely to be representative for the population in general and may therefore not show the correct population-level variability in measurements.

Fourth, we determined if the approach used for simulating social groups was applied consistently to all participant groups and experimental runs. For example, if a study used verbal instructions telling participants within social groups to remain close to each other during egress, the precise wording of these instructions had to be reported.

Fifth, the simulation of social groups may break up naturally occurring social groups within participant crowds. For example, individuals may be assigned randomly to social groups. Under such protocols, participants may additionally attend to their natural social preferences or even ignore instructions to do so.

The five aforementioned indicators are designed to give an indication of study design that can be easily summarized for many studies. As additional indicators may be relevant, this selection is intended as an indication and starting point for critical appraisal only.

All data extraction and processing were performed independently by the two authors using the R programming environment, version 4.0.2 (R Core Team 2020).

Statistical analysis

All statistical analyses were conducted in the R programming environment, version 4.0.2 (R Core Team 2020), using the packages 'esc' (Lüdecke 2019), 'meta' (Balduzzi, Rücker, and Schwarzer 2019), 'metafor' (Viechtbauer 2010), and 'dmetar' (Harrer et al. 2019a) . Throughout, we used the recommendations of a freely available guide that is also recommended as an introduction to meta-analysis to inform methodological choices (Harrer et al. 2019b).

The summary statistics used for measuring effect sizes were Hedges' g scores computed from the standardised mean egress times extracted from the literature, as described in section 'Data extraction and critical appraisal'. Hedges' g is a standardised measure of the mean difference across populations that corrects for biased effect size estimates and is particularly suited for meta-analyses including few studies (Hedges and Olkin 2014). Our meta-analysis used a random-effects model employing the DerSimonian-Laird estimator for the between-study variance to pool effect sizes across studies (Harrer et al. 2019b). In other words, the meta-analysis assimilated the information from all studies included and produced an estimate and confidence interval for the effect size based on this information. The statistical model used in our meta-analysis requires only two quantities to be estimated from the observed effect sizes from all included studies: the mean effect size across studies and the between-study variance (Kelley and Kelley 2012). While it is desirable to estimate these quantities based on many data points (i.e. many included studies), as only two quantities need to be estimated, it is plausible to estimate them based on a relatively low number of studies. Moreover, we found no qualitative difference in results when using a fixed-effects model instead of a random-effects model. A fixed-effects model only requires a single quantity, the mean effect size, to be estimated (Kelley and Kelley 2012).

We checked if the distributional assumptions of the meta-analysis hold qualitatively by plotting the quantiles for standardised effect estimates against the expected theoretical quantiles (following Wang and Lee [2020]).

It should be noted that there is a lively discussion on appropriate methods to be employed in meta-analysis within the Statistics community (e.g. [Jackson and White 2018]). Briefly, many methods make distributional assumptions, e.g. based on the Normal distribution, or require sufficient data to achieve accurate estimates. The number of studies included in the meta-analysis presented here is small which is likely to exacerbate some of these problems. Similarly, the data for some individual studies included in our meta-analysis only comprises a low number of experimental runs which also contributes to the uncertainty of meeting distributional assumptions in the calculation of confidence intervals, for example. We acknowledge this limitation and provide all data necessary for repeating this meta-analysis with alternative methods in the appendix.

We assessed the extent to which the effect sizes within our meta-analysis vary. It is important to check for heterogeneity to ensure the pooled effect size is a meaningful representation of the effect size distribution across studies and not dominated by outliers, for example. For completeness and because they have different properties, we report three measures for heterogeneity that are commonly used: *Cochran's Q, Higgin's & Thompson's* I^2, and *Tau-squared* (τ^2). *Cochran's Q* is a test statistic for the between-study variance. It increases with the number of studies and the number of replicates within studies and it therefore highly depends on the size of meta-analysis (Harrer et al. 2019b). I^2 is derived from Q and measures the proportion of the variation in effect sizes that is not caused by chance (Higgins and Thompson 2002). It is not sensitive to the number of studies, but it is still sensitive to the number of replicates within studies (Borenstein et al. 2017). τ^2 is insensitive to both aspects but often difficult to interpret (Harrer et al. 2019b).

Finally, we investigated the possibility of publication bias as part of our meta-analysis. Publication bias can occur in research fields where the outcome of studies influences the likelihood of them being published. For example, it is possible that experiments with inconclusive results or a very small number of replicates are less likely to be published (Rothstein, Sutton, and Borenstein 2005). Missing studies with low effect sizes could inflate the pooled effect size. We used a funnel plot to qualitatively explore the presence of publication bias (Borenstein et al. 2011). We plotted the effect size against the standard error for the effect in studies. Larger studies with more replicates have lower standard error and the premise of the funnel plot is that the scatter of effect sizes should increase symmetrically with increasing standard error (in a funnel shape). This would indicate that the reporting of effect sizes is not related to their statistical significance. An asymmetric funnel plot is indicative of publication bias (Borenstein et al. 2011).

Results

Literature search

Our electronic literature search retrieved 21 papers and we found three additional papers in conference proceedings. We removed duplicate publications and articles not published in English (n=4 papers). After excluding papers that did not meet our inclusion criteria (experiments involving stairs or other level changes, studies on observational data, and review articles; a total of n=12 papers), we screened eight potentially eligible publications. Two of these had to be removed from the study, because it was not possible to obtain standard deviations of egress times, either because they were not published (Lu et al. 2017) or because only a single replicate experimental run was conducted for the experimental condition without social groups (von Krüchten and Schadschneider 2017). Thus, we included six studies in our further review and meta-analysis (Bode et al. 2015;

Haghani et al. 2019; Köster et al. 2011; Xie et al. 2020; Bode 2016; Hu et al. 2021). Figure 1 summarises the process of our literature search.

Study characteristics

Table 1 shows a summary of the study characteristics. A description of how data was extracted from the studies and a summary of the data extracted can be found in table A1 in the appendix. The first included study was published in 2011 (Köster et al. 2011) and the most recent study was published in 2021 (Hu et al. 2021). We briefly review the key findings of the included studies reserving quantitative details for the meta-analysis in the next section.

Köster et al. (2011) developed a computational model for pedestrian crowd movement including social groups and conducted an experiment to validate their model. The experimental setting was egress from a classroom with students being positioned at their desks at the start. Ingress dynamics were also considered. Measurements from the experiments suggested an increase in egress times when social groups were present. Interestingly, the presence of social groups led to a reduction of ingress times but as this is not the topic of our work, we will not discuss this further.

Bode et al. (2015) conducted an experiment with a focus on disentangling the contribution of different elements of egress, namely pre-movement times and the movement towards and immediately in front of exits. In their experiment, participants were positioned at pre-defined starting positions and left a room that had six exits but only two were made available for egress. The experiment showed an increase in egress time when social groups were present, and the authors suggest this was mainly due to differences in pre-movement time and the time it took pedestrians to get close to exits rather than their movement dynamics in front of exits.

Bode (2016) conducted an experiment in which a crowd walked from an assembly area through a bottleneck at the end of a corridor. The author states that based on the data of this experiment alone, it is impossible to determine if social groups affect the egress time from the corridor.

Haghani et al. (2019) studied the egress of a crowd from a room with four exits, three of which were opened during experiments. Participants freely chose their initial positions inside an assembly area. In addition to the effect social groups have on the movement dynamics, this study also investigated the role of stress on pedestrian behaviour in a controlled way. Many additional aspects of the egress, such as decision-making, exit choice, and intra-group decision dynamics were studied. Most pertinent to this meta-analysis, Haghani et al. (2019) found a non-monotonic effect of social group size on the egress time: an increase for groups sizes of two and three was followed by a decrease for groups of size four.

Xie et al. (2020) investigated the effect of reduced visibility in addition to the effect of social groups being present. Egress took place from a room with four doors, only one of which was opened during the experiment and participants were placed onto pre-defined starting positions. The detailed analysis of the experimental data by Xie et al. (2020) also considered leadership, overtaking behaviour, and helping behaviour. The findings suggested an increase of egress times when social groups were present for good visibility, but the opposite effect for poor visibility.

Hu et al. (2021) tested the movement dynamics of pedestrian crowds that were composed entirely of independent individuals, entirely of social groups, or a mixture of the two. By adjusting the width of a bottleneck at the end of a corridor that participants had to walk through, they achieved a broad range of pedestrian densities. Egress times were not the main interest of this study and no specific findings on them were reported.

The main conclusion of the paper was that social groups did not affect the dynamics of the crowd at a macroscopic level but that they did affect microscopic movement dynamics.

This brief introduction to the included studies demonstrates that there is a substantial variety of experimental settings. Overall, findings seem to suggest that the presence of social groups increases egress times, but there are exceptions that may depend on the context. Next, we will integrate the evidence from these different studies quantitatively.

Meta-analysis

The forest plot in Fig. 2 shows that the effect sizes take a range of positive and negative values from the largest increase in egress times (Bode et al. 2015) to a decrease in egress times (Hu et al. 2021) when social groups are present. Based on data extracted from the six studies, our meta-analysis suggests that the presence of social groups appears to slightly increase egress times overall but that we cannot be confident in this effect (difference in effect size: 0.39, 95% CI=[-0.34,1.11], t=1.36, p=0.23; see Fig. 2). The 95% confidence interval for the difference across experimental settings includes zero (no effect) and we cannot reject the hypothesis that the overall effect is zero. Due to the standardization required to pool data from different experimental settings and because Hedges' g is used as the effect size, interpreting the estimated difference across experimental settings directly in terms of seconds per pedestrian is not possible (Hedges' g standardizes the mean difference across experimental settings using a pooled standard deviation). Seeing that we cannot be sure of the overall effect this limitation is not too problematic.

The confidence intervals for the effect sizes of the separate studies all overlap with the confidence interval of the pooled effect, suggesting there are no extreme outliers in our study (Fig. 2). The measures for heterogeneity are $Q=7.28$, $I^2=31\%$, and $\tau^2=0.17$. Following a previously suggested rule of thumb (Higgins et al. 2003; $I^2 = 25\%$: low heterogeneity; $I^2 = 50\%$: moderate heterogeneity; $I^2 = 75\%$: substantial heterogeneity), these measures indicate a low to moderate heterogeneity in our meta-analysis. We did not find any clear evidence for violations of the distributional assumptions of the meta-analysis (see Fig. A1 in the appendix). To further determine the sensitivity of our findings, we performed a leave-one-out analysis, recalculating the pooled effect with one study removed from the analysis at a time (see Fig. A2 in the appendix). While we found that removing studies with large positive or negative effects resulted in a shift of the pooled effect, our main findings of a positive effect size with a confidence interval including the no-effect line held qualitatively whichever study was removed. This indicates our results are not determined by individual studies.

Our qualitative review in section ‘Study characteristics’. revealed that many studies found an increase in egress times in the presence of social groups, yet our quantitative analysis suggests that we cannot be certain of this effect. How can this be explained? Fig. 2 shows that four out of six included studies indeed show an increase in egress times. However, two studies (Xie et al. 2020) and (Hu et al. 2021) find a decrease in egress times. In the case of Xie et al. (2020) this decrease arose when data from high and low visibility treatments were combined (mean differences between these treatments were accounted for). While it could be argued that data from low visibility treatments should be excluded, as it considers a fundamentally different context, the fact that (Hu et al. 2021) also find a negative effect under normal visibility conditions counters this argument. Moreover, another study conducted under normal visibility conditions that also indicated a decrease in egress times in the presence of social groups could not be included in this meta-analysis (von Krüchten and Schadschneider 2017). In our meta-analysis, the evidence provided by studies suggesting an increase in egress times is balanced by the evidence of studies suggesting a decrease in egress times resulting in a pooled effect that appears to be slightly positive but cannot be ruled out to be zero.

Evaluation of studies

Although the number of studies included in our meta-analysis is low, we suggest it is still interesting to explore the possibility of publication bias in findings and to attempt to summarize the risks to biased results inherent in the design of the included studies.

The funnel plot in Fig. 3 suggests there is no clear evidence for publication bias, as all studies lie symmetrically around the overall pooled effect size inside the funnel. The relatively even spread of standard errors (related to the reliability of study results based on the measurement variance and the number of replicates) suggests both small and larger studies are evenly represented in our meta-analysis.

Considering indicators for risks related to the design of studies, we found that all studies included homogeneous groups of participants meaning they may not be representative at the population level (Table 2). Reassuringly, all studies provided clear and consistent instructions to participants. Our remaining three indicators showed a mixed picture with most studies simulating social groups rather than making use of naturally occurring ones and some groups only testing one group of participants and/or fixing the temporal order of treatments. In principle, it would be possible to conduct an additional subgroup analysis (Harrer et al. 2019b) to determine quantitatively if any of the indicators for study design, or additional ones, such as expert ratings of study quality, influence the observed effect. However, given the number of studies available for our meta-analysis, this approach is not sensible here.

Discussion

In summary, we find that based on the pooled evidence across six studies we cannot reject the null hypothesis that social groups have no effect on the egress time of pedestrian crowds. Therefore, there is either insufficient evidence to clearly determine the effect or there is no effect. While our integration of evidence across studies has failed to provide clarity on the nature of the effect social groups have on pedestrian crowd egress times, we can provide an explanation for the conflicting results on this topic in the literature. The combination of high variability in measurements, the low number of replicates within studies, and a potentially low effect size lead to confidence intervals that contain the no-effect line. For several studies with these properties, we would expect to find results on either side of the no-effect line by chance. Seemingly conflicting results are thus likely to arise naturally from the properties of the effect and the design of studies.

If one experimental run with a crowd of participants is considered to contribute a single data point (as we have assumed here), it has to be concluded from a statistical perspective that many studies in the field of pedestrian dynamics are underpowered. In other words, they do not contain enough replicates to reliably detect effects (the largest number of replicates across the studies considered here was 18, see table A1 in the appendix). Recall that we had to make the additional assumption that replicates obtained with the same crowd of participants are statistically independent (they may not be). One solution to this issue would be a higher standardization of conducting experiments and reporting findings that would facilitate future meta-analyses. However, it should be noted that many studies in this field are not designed with such a crowd-level data analysis in mind. For example, some studies focus on smaller-scale dynamics related to the movement of individuals (e.g. leadership [Xie et al. 2020]) whilst others record more than one data point from an experimental run by sampling measurements from the crowd dynamics once it reaches a presumed steady-state (Hu et al. 2021). The extent to which such alternative approaches suffer from similar or different problems to the ones we discuss here is beyond the scope of this review (for a broader discussion, see [Bode and Ronchi 2019]).

For the type of analysis we present here, higher standardization and larger numbers of replicates are crucial. However, it is also important to critically reflect on the extent to which meta-analyses can be viewed as the gold-standard for pedestrian dynamics research. Several studies have pointed out that social groups may affect different elements of egress to a varying degree (e.g. pre-movement times and movement times [Bode et al. 2015; Haghani et al. 2019]). The variety of experimental settings across the studies included in our meta-analysis meant that for some studies, pre-movement times played a larger role (Bode et al. 2015; Haghani et al. 2019) than for others (Köster et al. 2011; Bode 2016; Xie et al. 2020; Hu et al. 2021). It could be suggested that a meta-analysis requires standardization to such an extent that important aspects of crowd egress dynamics may be excluded or may be dominated by other studies that do not capture such effects. For our specific example of pre-movement times, it should be noted that for one of the studies placing emphasis on this aspect of egress, the confidence interval for the effect still includes the no-effect line (Haghani et al. 2019). In addition, it is possible that different social group behaviours could have opposite effects on egress times. For example, social groups may take longer to make decisions about where to exit, but their decisions may be better, e.g. they may choose faster routes, or they may be able to cope better with congestion (see also the discussions in (Bode et al. 2015; Hu et al. 2021)).

More generally, it is acknowledged that pedestrian behaviour depends on the context (Moussaïd et al. 2010; Zanlungo et al. 2017) and the robust insights and standardization of meta-analyses therefore need to be balanced against a more exploratory line of investigation. For example, an undesirable outcome of a focus on meta-analyses would be a false sense of security about research insights based on findings from overly standardized or abstracted experiments that may elicit artificial behaviours or ignore processes that may become relevant in altered circumstances, such as emergency evacuations under time pressure. In this sense, meta-analyses could exacerbate issues with the ecological validity of controlled experiments. This discussion also makes it clear that it cannot be concluded that papers excluded from our meta-analysis are somehow inferior - they simply did not fit the format of our analysis but are still valid and important contributions to the literature.

Several of the studies included in our meta-analysis refer to each other or even build on each other methodologically. On the one hand, this can mean that it is more difficult to integrate their findings, as each study seeks to investigate a novel context rather than replicating published research. For example, very few of the studies included in our analysis use very similar physical settings. On the other hand, awareness of previous work can increase standardisation and therefore be beneficial for conducting meta-analysis. For example, Haghani et al. (2019) used the methods established in Bode et al. (2015) to create cohesion amongst simulated social groups by asking group members to talk to each other. How these opposing factors influence the performance of a meta-analysis depends on the extent to which findings can be meaningfully standardised across studies, as discussed above and in the following.

Fundamentally, our meta-analysis is a contribution to the discussion on the extent to which it is possible to generalise effects, such as the influence of social groups on the egress times of pedestrian crowds, given the importance of the physical setting, the different behaviours of social groups and the general context (e.g., emergency versus non-emergency context). One extreme would be to insist on investigating every context separately. The other extreme would be to assume pedestrian behaviour can always be averaged across contexts. It is likely that the most useful approach lies between these extremes and that some aspects of pedestrian behaviour be generalised across contexts, whilst others cannot. Our meta-analysis is closer to the second extreme and while this approach has limitations, which we discuss, we suggest it is important to quantitatively investigate the integration of evidence across contexts. Here we contribute one approach for how to do so and discuss it critically.

Study strengths and limitations

To the best of our knowledge, we have presented the first meta-analysis on the dynamics of pedestrian crowds. We have demonstrated how integrating evidence from several studies is useful for explaining seemingly conflicting findings in the literature. Nevertheless, our work has several limitations that we present here. First, the number of studies included in our meta-analysis is small meaning the available data for analysis is limited and the number of replicate runs for some individuals studies is also small. This may influence the validity of assumptions underlying our statistical methods. Second, we had to perform substantial data standardization prior to analysis. This hinders the interpretation of estimated effect sizes in our meta-analysis and could influence results, as discussed above. Third, to be able to perform a meta-analysis we had to focus on a narrow aspect of crowd dynamics. As discussed in the introduction, previous work has shown several ways in which social groups affect crowd dynamics and we could only consider one of these aspects here. Fourth, many of the studies included in our meta-analysis only tested one homogeneous participant group and they simulated social groups rather than making use of naturally occurring ones. Although this is not a limitation of our approach, it nevertheless raises questions of the extent to which our findings extend to the population level and to real-world contexts.

Conclusions

Based on our meta-analysis there is either insufficient evidence to clearly determine the effect social groups have on the egress time of pedestrian crowds or there is no effect. The low number of replicates in studies in combination with the variability in measurements and the possible size of any effect explain seemingly conflicting results across studies, as these are expected to arise by chance in such settings.

Our systematic review and meta-analysis highlight the usefulness of the following recommendations. These are mostly established principles of experimental design. a) Experiments should be repeated with different participant groups. b) Multiple experimental runs should be conducted for each experimental condition (the smaller the expected effect, the more replicates are needed). Considering points a) and b), experimenters may want to consider splitting the group of volunteers recruited to their study into several participant groups to allow for larger numbers of experimental runs with different participant groups if they are testing behavioural effects with high inherent variability, such as the effect of social groups. c) Experimental protocols, data measurement methods, and demographics of participant groups need to be communicated clearly to facilitate comparison across studies. The relevance of this recommendation is highlighted by the fact that we had to exclude one study because results were not reported, and we had to assume what results were reported in the figure of another study. d) Ideally, raw data and data processing code should be published alongside papers. At the least, it would be useful for studies to communicate data that can be compared to previous work, even if it is not the main focus of the study. In the context of our study, this would have helped with the standardisation of the effect size across studies.

We suggest that despite limitations meta-analyses are useful tools for research in pedestrian dynamics that help contextualize research findings by establishing the extent to which reported effects are statistically reliable. We also recommend that the outcomes of meta-analyses are not viewed as the final definitive evidence on crowd behaviour, as standardization may limit the extent to which findings apply in other contexts.

Acknowledgements

YH acknowledges the support from the National Natural Science Foundation of China [Grant No.71704168], from Anhui Provincial Natural Science Foundation, China [Grant No. 1808085MG217], and the financial support from the program of China Scholarships Council (No.201906340190).

Declaration of Interests Statement

The authors have no conflict of interests to declare.

References

Bode, Nikolai WF. 2016. The effect of social groups and gender on pedestrian behaviour immediately in front of bottlenecks. In 2016 8th International Conference on Pedestrian and Evacuation Dynamics, pp. 92-99. University of Science and Technology of China Press.

Bode, Nikolai WF, Stefan Holl, Wolfgang Mehner, and Armin Seyfried. 2015. "Disentangling the impact of social groups on response times and movement dynamics in evacuations." *PloS one* 10 (3):e0121227.

Bode, Nikolai WF, and Enrico Ronchi. 2019. "Statistical model fitting and model selection in pedestrian dynamics research." *Collective Dynamics* 4:1-32.

Borenstein, Michael, Larry V Hedges, Julian PT Higgins, and Hannah R Rothstein. 2011. *Introduction to meta-analysis*: John Wiley & Sons.

Borenstein, Michael, Julian PT Higgins, Larry V Hedges, and Hannah R Rothstein. 2017. "Basics of meta - analysis: I2 is not an absolute measure of heterogeneity." *Research synthesis methods* 8 (1):5-18.

Canadian Task Force on Preventive Health Care. 2011. "Grades of Recommendation, Assessment, Development, and Evaluation (GRADE)." URL https://www.gradeworkinggroup.org/ (accessed 3/6/2021).

Carey, Nick. 2005. "Establishing pedestrian walking speeds." *Karen Aspelin, Portland State University* 1 (01).

Chraibi, Mohcine, Maik Boltes, Andreas Schadschneider, and Armin Seyfried. 2015. *Traffic and Granular Flow'13*: Springer.

Cohn, Lawrence D, and Betsy J Becker. 2003. "How meta-analysis increases statistical power." *Psychological methods* 8 (3):243.

Daamen, Winnie, DC Duives, and SP Hoogendoorn. 2014. "The conference in pedestrian and evacuation dynamics (ped 2014)." *Transportation Research Procedia* 2:1-818.

Dederichs, Anne, Gerta Köster, and Andreas Schadschneider. 2020. "Proceedings of Pedestrian and Evacuation Dynamics 2018." *Collective Dynamics* 5:1-543.

Drabek, Thomas E., and John S. Stephenson III. 1971. "When Disaster Strikes." *Journal of Applied Social Psychology* 1 (2):187-203.

Federici, Mizar Luca, Andrea Gorrini, Lorenza Manenti, and Giuseppe Vizzari. 2012. Data collection for modeling and simulation: case study at the university of milan-bicocca. In *International Conference on Cellular Automata* (pp. 699-708). Springer, Berlin, Heidelberg..

Fidler, Fiona, and John Wilcox. 2018. "Reproducibility of scientific results." T*he Stanford Encyclopedia of Philosophy* (Winter 2019 Edition), Edward N. Zalta (ed.), URL https://plato.stanford.edu/archives/win2019/entries/scientific-reproducibility/ (accessed 3/6/2021).

Frantzich, Hakan. 1996. *Study of movement on stairs during evacuation using video analysing techniques.* Lund University, Institute of Technology, Department of Fire Safety Engineering.

Gorrini, Andrea, Stefania Bandini, and Majid Sarvi. 2014. "Group dynamics in pedestrian crowds: Estimating proxemic behavior." *Transportation research record* 2421 (1):51-6.

Haghani, Milad. 2020. "Empirical methods in pedestrian, crowd and evacuation dynamics: Part I. Experimental methods and emerging topics." *Safety Science* 129:104743.

Haghani, Milad, Majid Sarvi, Zahra Shahhoseini, and Maik Boltes. 2019. "Dynamics of social groups' decision-making in evacuations." *Transportation research part C: emerging technologies* 104:135-57.

Hamdar, Samer H. 2019. *Traffic and Granular Flow'17*: Springer.

Harrer, Mathias, Pim Cuijpers, Toshi Furukawa, and David Daniel Ebert. 2019a. "dmetar: Companion R Package For The Guide 'Doing Meta-Analysis in R'." URL http://dmetar.protectlab.org/ (accessed 3/6/2021).

Harrer, Mathias, Pim Cuijpers, Toshi Furukawa, and David Daniel Ebert.. 2019b. *Doing Meta-Analysis in R: A Hands-on Guide*. Chapman and Hall/CRC Press

Hedges, Larry V, and Ingram Olkin. 2014. *Statistical methods for meta-analysis*: Academic press.

Higgins, Julian PT, and Simon G Thompson. 2002. "Quantifying heterogeneity in a meta‐analysis." *Statistics in medicine* 21 (11):1539-58.

Hu, Yanghui, Jun Zhang, Weiguo Song, and Nikolai WF Bode. 2021. "Social groups barely change the speed-density relationship in unidirectional pedestrian flow, but affect operational behaviours." *Safety Science* 139:105259.

Jackson, D., and I. R. White. 2018. "When should meta-analysis avoid making hidden normality assumptions?" *Biom J* 60 (6):1040-58. doi: 10.1002/bimj.201800071.

Kelley, George A, and Kristi S Kelley. 2012. "Statistical models for meta-analysis: A brief tutorial." *World journal of methodology* 2 (4):27.

Knoop, Victor L, and Winnie Daamen. 2016. *Traffic and Granular Flow'15*: Springer.

Köster, Gerta, Michael Seitz, Franz Treml, Dirk Hartmann, and Wolfram Klein. 2011. "On modelling the influence of group formations in a crowd." *Contemporary Social Science* 6 (3):397-414.

L'Abbé, Kristian A, Allan S Detsky, and Keith O'Rourke. 1987. "Meta-analysis in clinical research." *Annals of internal medicine* 107 (2):224-33.

Li, Meng, Tao Chen, Hao Du, Na Ma, and Xinwei Xi. 2021. "Social group detection based on multi-level consistent behaviour characteristics." *Transportmetrica A: Transport Science* DOI: 10.1080/23249935.2021.1976877.

Lindell, M. K., J. E. Kang, and C. S. Prater. 2011. "The logistics of household hurricane evacuation." *Natural Hazards* 58 (3):1093-109.

Lo, Carson Ka-Lok, Dominik Mertz, and Mark Loeb. 2014. "Newcastle-Ottawa Scale: comparing reviewers' to authors' assessments." *BMC medical research methodology* 14 (1):45.

Lu, Lili, Ching-Yao Chan, Jian Wang, and Wei Wang. 2017. "A study of pedestrian group behaviors in crowd evacuation based on an extended floor field cellular automaton model." *Transportation research part C: emerging technologies* 81:317-29.

Lüdecke, Daniel. 2019. *esc: Effect Size Computation for Meta Analysis (Version 0.5.1)*.

Ma, Yaping, Lihua Li, Hui Zhang, and Tao Chen. 2017. "Experimental study on small group behavior and crowd dynamics in a tall office building evacuation." *Physica A: Statistical Mechanics its Applications* 473:488-500.

Moussaïd, Mehdi, Niriaska Perozo, Simon Garnier, Dirk Helbing, and Guy Theraulaz. 2010. "The walking behaviour of pedestrian social groups and its impact on crowd dynamics." *PloS one* 5 (4):e10047.

Nicolas, Alexandre, and Hassan, Fadratful Hafinaz. 2021. "Social groups in pedestrian crowds: Review of their influence on the dynamics and their modelling." *Transportmetrica A: Transport Science* DOI: 10.1080/23249935.2021.1970651.

R Core Team. 2020. "R: A Language and Environment for Statistical Computing." In.: R Foundation for Statistical Computing.

Rothstein, Hannah R, Alexander J Sutton, and Michael Borenstein. 2005. "Publication bias in meta-analysis." *Publication bias in meta-analysis: Prevention, assessment adjustments*:1-7.

Sara Balduzzi, Gerta Rücker, and Guido Schwarzer. 2019. "How to perform a meta-analysis with R: a practical tutorial." *Evidence-Based Mental Health*.

Schadschneider, Andreas, and Armin Seyfried. 2011. "Empirical results for pedestrian dynamics and their implications for modeling." *Networks & heterogeneous media* 6 (3):545.

Schultz, Michael, Christian Schulz, and Hartmut Fricke. 2010. "Passenger dynamics at airport terminal environment." In *Pedestrian and Evacuation Dynamics 2008*, 381-96. Springer.

Sharmin, Samia, and Md Kamruzzaman. 2018. "Meta-analysis of the relationships between space syntax measures and pedestrian movement." *Transport Reviews* 38 (4):524-50.

Singh, Harmeet, Robyn Arter, Louise Dodd, Paul Langston, Edward Lester, and John Drury. 2009. "Modelling subgroup behaviour in crowd dynamics DEM simulation." *Applied Mathematical Modelling* 33 (12):4408-23.

Song, Weiguo, Jian Ma, and Libi Fu. 2017. "Proceedings of Pedestrian and Evacuation Dynamics 2016." *Collective Dynamics* 1:618.

Thompson, Rebecca R, Dana Rose Garfin, and Roxane Cohen Silver. 2017. "Evacuation from natural disasters: a systematic review of the literature." *Risk analysis* 37 (4):812-39.

Viechtbauer, Wolfgang. 2010. "Conducting meta-analyses in {R} with the {metafor} package." *Journal of Statistical Software* 36:1-48.

von Krüchten, Cornelia, and Andreas Schadschneider. 2017. "Empirical study on social groups in pedestrian evacuation dynamics." *Physica a-Statistical Mechanics and Its Applications* 475:129-41. doi: 10.1016/j.physa.2017.02.004.

Wang, Chia-Chun, and Wen-Chung Lee. 2020. "Evaluation of the normality assumption in meta-analyses." *American journal of epidemiology* 189 (3):235-42.

Weidmann, Ulrich, Uwe Kirsch, and Michael Schreckenberg. 2014. *Pedestrian and evacuation dynamics 2012*: Springer Science & Business.

Xie, Wei, Eric Wai Ming Lee, Yanying Cheng, Meng Shi, Ruifeng Cao, and Yuchun Zhang. 2020. "Evacuation performance of individuals and social groups under different visibility conditions: experiments and surveys." *International journal of disaster risk reduction* 47:101527.
Xie, Wei, Yuchun Zhang, Yanying Cheng, Shiming Chen, Xuanwen Liang, and Wenbin Zhang. 2018. "Experimental study on movement speed and route choice of individuals and small groups under different visibility conditions." *Procedia engineering* 211:830-6.
Yang, LZ, DL Zhao, Jian Li, and TY Fang. 2005. "Simulation of the kin behavior in building occupant evacuation based on cellular automaton." *Building Environment* 40 (3):411-5.
Zanlungo, Francesco, Zeynep Yücel, Dražen Brščić, Takayuki Kanda, and Norihiro Hagita. 2017. "Intrinsic group behaviour: Dependence of pedestrian dyad dynamics on principal social and personal features." *PloS one* 12 (11):e0187253.
Zhang, Yuchun, Wei Xie, Shiming Chen, and Tao Li. 2018. "Experimental Study on Descent Speed on Stairs of Individuals and Small Groups Under Different Visibility Conditions." *Fire Technology* 54 (3):781-96.
Zuriguel, Iker, Angel Garcimartín, and Raúl Cruz Hidalgo. 2020. *Traffic and Granular Flow 2019*: Springer.

Appendices

Details of data extraction

Six studies were selected for inclusion in this meta-analysis and in the following we describe how data was extracted from each study and standardized.

Data from Bode et al. (2015) was obtained directly from the text. The relevant section in the text is cited verbatim here: '*The total time to empty the room, or last exit time, was 9.25±0.36s (mean±s.e.) in the group treatment and 7.85±0.27s (mean±s.e.) in the individual treatment (W = 36, P = 0.014).*'

Data from Haghani et al. (2019) were read off figure 13 in the published text. This was performed by hand and independently by the two authors whose readings were averaged. We assumed error bars in this figure represented one standard deviation, as information on this was not provided. The study presents egress times for different social group sizes and we averaged across these for both mean egress times and standard deviations (thus assuming independence of data for different group sizes). We also assumed data presented in figure 13 did not include data from other experimental conditions tested (e.g., elevated stress levels).

Data for Köster et al. (2011) was read off figure 4 in the published text. This was performed by hand and independently by the two authors whose readings were averaged.

Data for Xie et al. (2020) was obtained from figure 7 in the published text. Average egress times that were given in textual form in the figure were used. Different visibility conditions had to be accounted for as explained in section 'Data extraction and critical appraisal'. in the main text.

Raw data for Bode (2016) was available to the authors and the egress time for the first 30 participants to exit in each experimental run was obtained from this raw data. As the experiment was repeated at two locations, the procedure explained in section 'Data extraction and critical appraisal'. in the main text was used to account for the small difference in experimental settings.

Raw data for Hu et al. (2021) was available to the authors and the egress time for the first 80 participants to exit in each experimental run was obtained from this raw data. The procedure explained in section 'Data extraction and critical appraisal'. in the main text was used to account for differences across experimental runs with different bottleneck widths. In addition, we averaged egress times over crowds composed entirely of social groups and ones composed of a mixture of social groups and independent individuals.

Tables and figures

Table 1: Brief description of study characteristics. A social group of size one refers to independent individuals. This data is also available digitally as a supplementary file.

Study	Crowd size	Scenario description	Social group size	Group relationship
bode2015	12	10.0x10.0m room with 6 exits of fixed 1.0m width; two exits are opened in the experiment	1,3	Assigned by experimenters
haghani2019	Crowd size varies ≈74	10.6x10.6m room with a central obstacle and 4 exits of fixed 0.5m width; three exits are opened in the experiment	1,2,3,4	Assigned by experimenters
koester2011	30	Classroom with 1 fixed exit; dimensions were not recorded	1,2,3,4	Assigned by experimenters
xie2020	36	6.96x6.72m room with 4 exits of fixed 1.0m width; one exit is opened in the experiment	1,2,3,5	Natural
bode2016	Crowd size varies, egress of 30 pedestrians is considered here	2.0x3.0m corridor with a bottleneck of fixed 0.6m width and 1.5m length	1,3-4	Assigned by experimenters
hu2021	Crowd size varies, egress of 80 pedestrians is considered here	1.8x9.6m corridor with a bottleneck, with the following widths: 0.5m, 0.6m, 0.7m, 0.8m, 0.9m, 1.0m, 1.1m, 1.6m, 1.8m.	1,2	Assigned by experimenters

Table 2: Summary of risk of bias for studies included.

	bode2015	**haghani2019**	**koester2011**	**xie2020**	**bode2016**	**hu2021**
Same participant group(yes=1-no=0)	0	0	1	1	0	1
Treatment temporal order fixed(yes=1-no=0)	0	1	0	1	0	1
Homogeneous participant cohort(yes=1-no=0)	1	1	1	1	1	1
Nonconsistent social group instructions(yes=1-no=0)	0	0	0	0	0	0
Forced different groups to natural groups in cohort(yes=1-no=0)	1	1	1	0	1	1

Figure 1. Flow diagram of literature search strategy and outcomes.

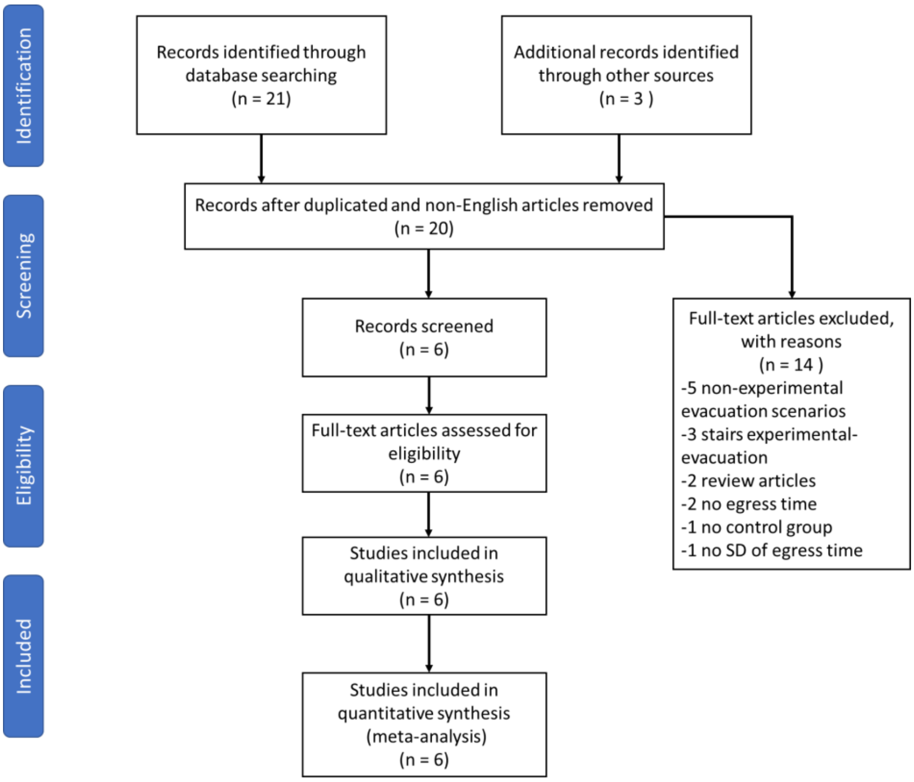


Figure 2. Forest plot comparing standardized egress times for pedestrian crowds without and with social groups. The first column shows the study ID. The second and third columns show the effect size and standard error of the effect size (Hedge's g computed from standardized egress times). The remaining three columns show again the effect sizes, the associated 95% confidence intervals and the weights of the study within the meta-analysis (weights depend on the variability of measurements and the number of replicates in studies). At the bottom of the figure, the overall estimated mean and prediction interval based on a random-effects model are shown. Marker sizes in the plot correspond to the weight of studies.

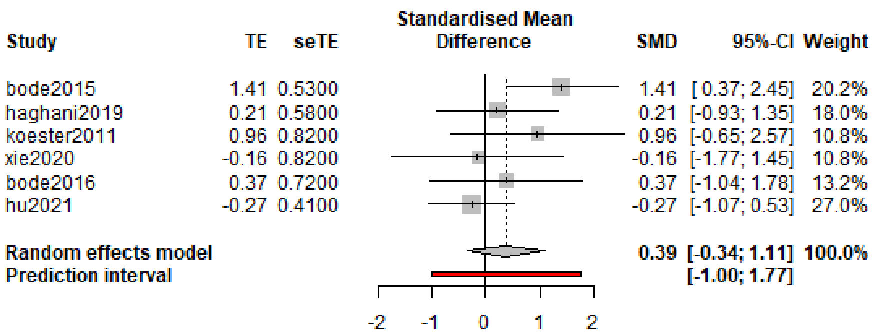
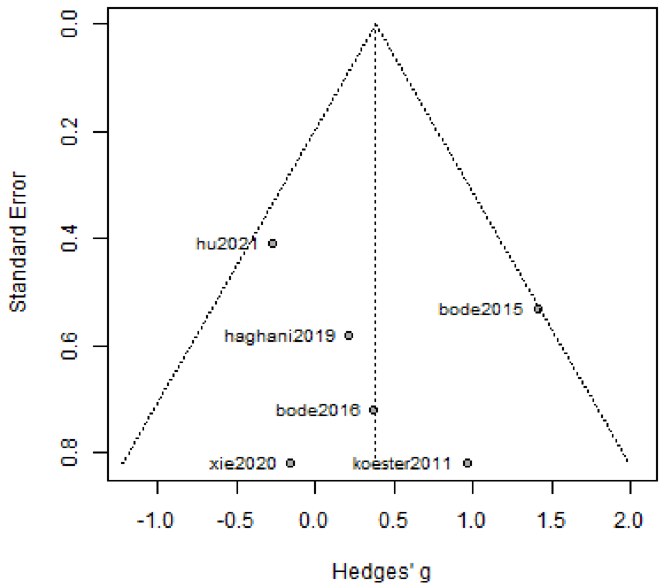


Figure 3. Funnel plot to visually assess publication bias. The x-axis shows the effect size of individual studies. The y-axis represents the standard error of each study. In general, larger studies have a smaller standard error and are thus plotted higher up the y-axis.



Appendix

Table A1. Raw data extracted from the studies included in the meta-analysis. This data is also available digitally as a supplementary file.

study parameters	bode2015	haghani2019	koester2011	xie2020	bode2016	hu2021
Mean egress time of individuals (s)	7.85	8.98	23.90	6.43	18.26	31.47
Mean egress time of social groups (s)	9.25	10.13	25.73	6.32	18.87	30.88
Standard deviation of egress time of individuals	0.81	5.16	1.39	0.51	1.20	2.01
Standard deviation of egress time of social groups	1.08	5.37	1.37	0.51	1.19	2.01
Mean egress time of individuals (standardized)	1	1	1	1	1	1
Mean egress time of social groups (standardized)	1.18	1.13	1.08	0.98	1.03	0.98
Standard deviation of egress time of individuals (standardized)	0.10	0.57	0.06	0.08	0.07	0.06
Standard deviation of egress time of social groups (standardized)	0.14	0.60	0.06	0.08	0.07	0.06
Number runs of social groups	9	6	4	3	4	18
Number runs of individuals	9	2	3	3	4	9
Crowd size	12	≈ 74	30	36	30	80

Figure A1. Quantile plot to check if the distributional assumptions of the meta-analysis hold. Observed quantiles for standardised effect estimates are plotted against the expected theoretical quantiles that follow a standard normal distribution (using the analysis code published in [Wang and Lee 2020]). A good match of data and distributional assumptions is demonstrated by a quantile plot clustered around the leading diagonal, shown as a solid line. The plot suggests good agreement with the exception of one outlier which corresponds to the study bode2015 (Bode et al. 2015). This may be explained by the fact that this study has the highest observed standardised effect size and is the only one for which the 95% confidence interval of the effect size does not include the zero-effect line. Another factor could be the importance of pre-movement times in this study that are also discussed in the main text. However, given the probabilistic nature of the assumptions and the amount of data, it is not possible to ascertain if this outlier is indicative of a systematic violation of the normality assumption. In other words, this outlier may have arisen by chance and may disappear with the inclusion of further studies.

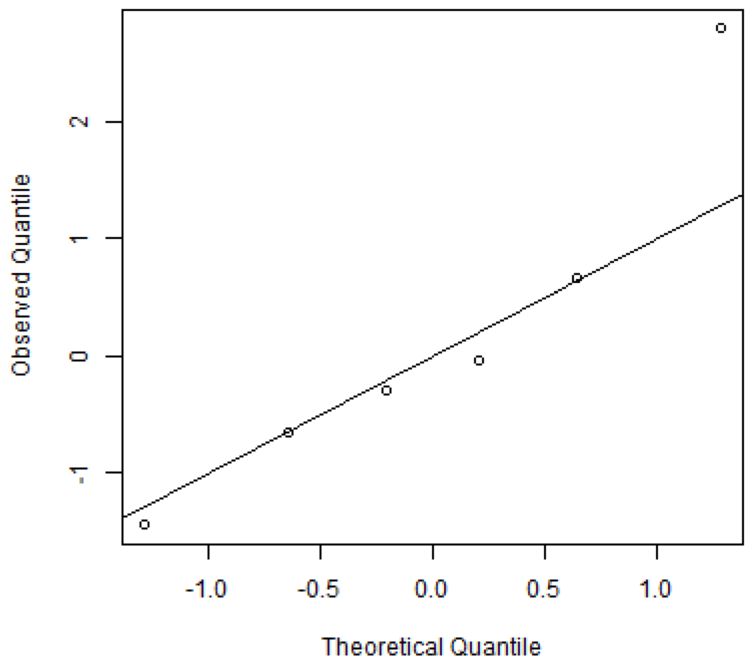


Figure A2. Forest plots showing recalculated pooled effects, removing the study indicated in the first column in each line. The plot and right-hand side show the pooled effect size and associated 95% confidence intervals calculated without the indicated study. The dashed line and shaded area show the pooled effect size and confidence interval when all studies are included (compare to Figure 2 in the main text).