European Transport Research Review

REVIEW **Open Access**

Advancing safety through connected and autonomous vehicles: a meta-analysis of market penetration and safety improvement rates from 2015 to 2024

Amirhossein Taheri[1*], Jing Yang[2], Steffen Müller[1], Dimitris Milakis[3] and Mohammed Quddus[4]

Abstract
The field of Connected and Autonomous Vehicles (CAVs) promises to improve road safety through advanced communication technologies and data sharing. This study addresses the need for a clearer understanding of the relationship between CAV Market Penetration Rates (MPRs) and Safety Improvement Rates (SIRs) by conducting an updated meta-analysis that includes 49 studies published between 2015 and 2024. Across these studies, the primary focus of safety surrogate measures (SSMs) was on Time-to-Collision (TTC) and Post Encroachment Time (PET), reflecting their dominant role in evaluating traffic safety impacts of CAVs. We applied sensitivity analysis to refine the data and mitigate the effects of variability across simulation parameters and scenarios. A power function was identified as the most accurate model to describe the MPR-SIR relationship, capturing how safety benefits scale with increasing adoption of CAVs. At low MPRs (10%–20%), safety gains are modest, with SIRs of 2.8% and 4.9%, respectively. Mid-range MPRs (30%–60%), while frequently studied, show overestimated safety benefits in raw findings; our corrected estimates yield SIRs of 8.0% at 30% MPR and 9.4% at 40% MPR, highlighting the complexity of mixed traffic environments. At high MPRs (70%–90%), safety benefits become substantial and more consistent, with an SIR of 39.4% at 90% MPR. These findings inform transportation planners and policymakers on the safety potential of CAVs, emphasizing the importance of high market penetration to realize meaningful safety improvements.

Keywords Meta-analysis, Intelligent connected vehicles, Autonomous vehicles, Connected vehicles, Traffic safety, Market penetration rate

*Correspondence:
Amirhossein Taheri
amirhossein.taheri@campus.tu-berlin.de
[1]Department of Motor Vehicles, Institute of Land and Maritime Transport, Faculty of Transport and Mechanical Systems, Technical University of Berlin (TU Berlin), TIB 13, Gustav-Meyer-Allee 25, 13355 Berlin, Germany
[2]Department of Medical Engineering, Institute for Machine Design and System Technology, Faculty of Transport and Mechanical Systems, Technical University of Berlin (TU Berlin), 10587 Berlin, Germany
[3]Institute of Transport Research, German Aerospace Center (DLR), Rudower Chaussee 7, 12489 Berlin, Germany
[4]Department of Civil and Environmental Engineering, Faculty of Engineering, Imperial College London, London SW7 2AZ, UK

1 Introduction

The transportation system is undergoing global changes driven by the rapid development of CAV technologies. These innovations aim to improve transportation efficiency and potentially enhance environmental sustainability [1]. However, CAVs also introduce new safety risks and challenges [2]. Studies show a complex relationship between deployment levels and safety benefits [3]. As the market penetration rate (MPR) of CAVs increases, safety improvements are expected due to enhanced vehicle communication, which can reduce crash risk. Yet, this relationship is multifaceted. Higher MPRs may improve road safety but can also introduce risks like overreliance on technology and unexpected traffic behaviors [4]. Understanding the interplay between MPR and safety is essential for policymakers to guide effective integration strategies.

Several key terms and metrics are central to understanding this study. Connected and Autonomous Vehicles (CAVs) refer to vehicles equipped with advanced communication technologies and varying levels of driving automation (Automation Levels 1–5) [5]. The Market Penetration Rate (MPR) indicates the proportion of vehicles on the road that are CAVs at a given time, reflecting the degree of technology adoption [6]. Surrogate safety measures are traffic-based indicators used to assess potential crash risk and safety improvements in simulation or observational studies. Common examples include Time-to-Collision (TTC), representing the time remaining before a collision would occur if current vehicle trajectories are maintained, Time Integrated Time-to-Collision (TIT), which integrates TTC over time to capture prolonged exposure to potential conflicts; Time Exposed Time-to-Collision (TET), which measures the duration a vehicle is exposed to critical TTC values; and Post-Encroachment Time (PET), which measures the time difference between one vehicle leaving a conflict point and another vehicle arriving at that point, indicating near-miss or crash risk [7]. Finally, the Safety Improvement Rate (SIR) quantifies the reduction in traffic conflicts or crashes due to increasing CAV adoption, standardizing outcomes across studies to allow comparison of safety benefits at different MPRs [6]. Safety evaluations rely heavily on modeling algorithms and traffic composition. Additionally, varying surrogate safety measures (e.g., Time-to-Collision (TTC), Post Encroachment Time (PET)) [8]. complicate comparisons across studies. To address this, the SIR was introduced as a standardized metric, quantifying safety benefits across different MPRs [6]. SIR compares traffic conflicts in non-CAV scenarios to those with varying CAV adoption levels, offering a clearer picture of how MPR influences road safety. Despite its potential, the relationship between MPR and SIR remains underexplored, particularly given rapid advancements and real-world CAV applications [9–11].

Building upon Xiao et al. (2021) [6], this study introduces an enhanced methodological framework designed to improve the robustness and generalizability of the Market Penetration Rate (MPR)–Safety Improvement Rate (SIR) relationship. While the earlier meta-analysis established the baseline relationship, the present study extends the dataset to include 49 studies published between 2015 and 2024, offering greater statistical power and a more balanced representation across low, mid, and high MPR levels. This study investigates the distribution and focus of safety surrogate measures used in previous research to clarify how variations in SSM selection may influence reported safety outcomes. Methodologically, we incorporate a structured sensitivity analysis alongside the traditional trim-and-fill approach to correct for potential publication bias and mitigate the influence of outliers and overrepresented mid-range studies. Furthermore, heterogeneity is managed through a variance-based weighting scheme within a random-effects model, improving the stability and reliability of aggregated estimates. These advancements result in smoother, bias-adjusted SIR trends and a more realistic understanding of how increasing CAV market penetration contributes to traffic safety improvements.

The structure of the paper is as follows: Sect. 2 describes the methodology of the meta-analysis used to assess the safety effectiveness of CAVs, including testing procedures, effect size combination, and evaluation of publication bias. Section 3 presents the results of the meta-analysis, followed by Sect. 4, which provides an in-depth discussion of the findings, methodological implications, and policy relevance. Finally, Sect. 5 offers the conclusions, summarizing the key insights and outlining directions for future research.

2 Methodology

Our methodology begins by developing a search strategy based on the keywords and exclusion criteria used in [6], yielding a preliminary set of research papers. The strategy is then expanded, incorporating additional exclusion criteria guided by the PRISMA guidelines [12] and the PICO model to refine the final study selection. After study selection, we extract data and assess heterogeneity. If heterogeneity is present, a random-effects model is used; otherwise, a fixed-effects model is applied. Sensitivity analyses help identify and remove outliers to improve data accuracy. To address potential publication bias, we use funnel plots and Begg & Egger's tests. If bias is found, the Trim-and-Fill method estimates and adjusts for missing studies, reducing bias. The corrected data are then used to establish the mathematical relationship between MPR and SIR.

2.1 Inclusion of studies: PRISMA and PICO models

We combined PRISMA's emphasis on transparent reporting with PICO's structured criteria to establish a rigorous and replicable meta-analysis process. The PRISMA 2020 [13] guidelines for systematic reviews and meta-analyses were followed (see Fig. 1). This framework, including a 27-item checklist and flow diagram, is widely recognized for enhancing the quality and transparency of systematic reviews. It standardizes reporting for intervention-based studies, ensuring reproducibility and comprehensiveness [12]. In our study, PRISMA guidelines structured key components of the meta-analysis. For instance, our title identifies the report as a meta-analysis, and the study selection, data collection, and eligibility criteria align with the PRISMA checklist. The results section also follows PRISMA by clearly detailing statistical methods and key findings, supporting transparent reporting.

The PICO framework helps formulate research questions and guide literature searches by focusing on four core elements: (a) Population – general transportation participants; like [6], we applied no specific filters; (b) Intervention – CAV adoption at various MPR levels; (c) Comparison – the baseline scenario without CAVs, used to assess intervention impact; and (d) Outcome – safety improvements measured by comparing different MPR scenarios to the baseline, as described in Eq. (1).

The initial step of our methodology involved identifying relevant studies. We used three bibliographic databases—Scopus, Web of Science, and Google Scholar—to conduct our meta-analysis, covering the period from 2020 to 2024. These databases were selected for their broad coverage of scientific, engineering, and technological research. The search, conducted on April 22, 2024, yielded 322 documents. We aimed to build on prior work by [6] by incorporating more recent studies on CAVs and their impact on traffic safety. To ensure inclusion of the latest developments, we considered studies published since 2020, including journal articles, conference papers, reports, book chapters, PhD dissertations, and master's theses.

Several factors led us not to directly replicate [6]'s search strategy. Initially, using their exact keywords produced only six relevant results from Google Scholar and none from the other two databases. To ensure a more comprehensive review, we expanded and refined the search string to also include studies from 2015 to 2020. This revised approach yielded results consistent with [6], validating our updated keyword strategy. Consequently, we applied this refined method to screen literature up to April 22, 2024, ensuring broader coverage and inclusion of the most recent research. The strategy used a combination of keywords and Boolean operators. The final search string was:

(autonom* OR automat* OR intelligent OR "self-driving" OR driverless OR connected) AND (vehicle OR mobility OR driv*) AND ("Conflicts Number" OR "Surrogate Measure" OR "Time To Collision" OR TTC OR "Post Encroachment Time" OR PET) AND ("Penetration Rate") AND ("Safety improve" OR "Safety enhance" OR "Safety increase").

The updated keywords focused on terminology related to CAVs, traffic conflicts assessment models, and safety metrics. The goal was to gather a diverse range of studies discussing the MPR of CAVs and improvement in traffic safety.

Following the search, the detailed literature selection process is depicted in Fig. 1, from initial retrieval to final inclusion, divided into two key phases. The first phase involved re-evaluating prior search, where 16 studies included in earlier reviews were reassessed for relevance and consistency. The second phase focused on identifying new studies. An initial search yielded 322 records from various databases. After removing 25 duplicate records, 297 remained for further screening. During this stage, 14 records were excluded for lacking direct relevance to CAVs, and 9 records were removed due to difficulties in retrieving full texts. Next, a comprehensive eligibility assessment was conducted on the remaining 274 studies, emphasizing English-language publications to maintain consistency. We excluded 65 studies that did not specifically address the impacts of CAVs on traffic safety, and 167 studies were excluded due to insufficient data on key conflict indicators, such as number of conflicts. Following this process, 33 new studies from the period 2020 to 2024 were considered suitable for inclusion. In the final step, these 33 recent studies were combined with the 16 previously included studies from [6], resulting in a total of 49 studies for our review. These additional studies represent a 206% increase over [6], and all of these data will be used in the new meta-analysis.

2.2 Data extraction

After completing the study selection based on the PRISMA flowchart and PICO model, we entered the data extraction, one of the key steps in the process. Data gathered from each study included details such as publication year and authors to establish the study's context and timeframe. The number of simulation runs was recorded to determine the sample size of each study and evaluate the robustness of the results. MPRs were highlighted as a key factor influencing the evaluation of CAV traffic safety. Various traffic safety metrics, such as number of conflicts, TTC, PET, were employed to measure the safety benefits of CAVs. To ensure data accuracy, the WebPlotDigitizer tool was used to extract data from charts, capturing data points when direct data was unavailable. This approach guarantees precise data collection [14].

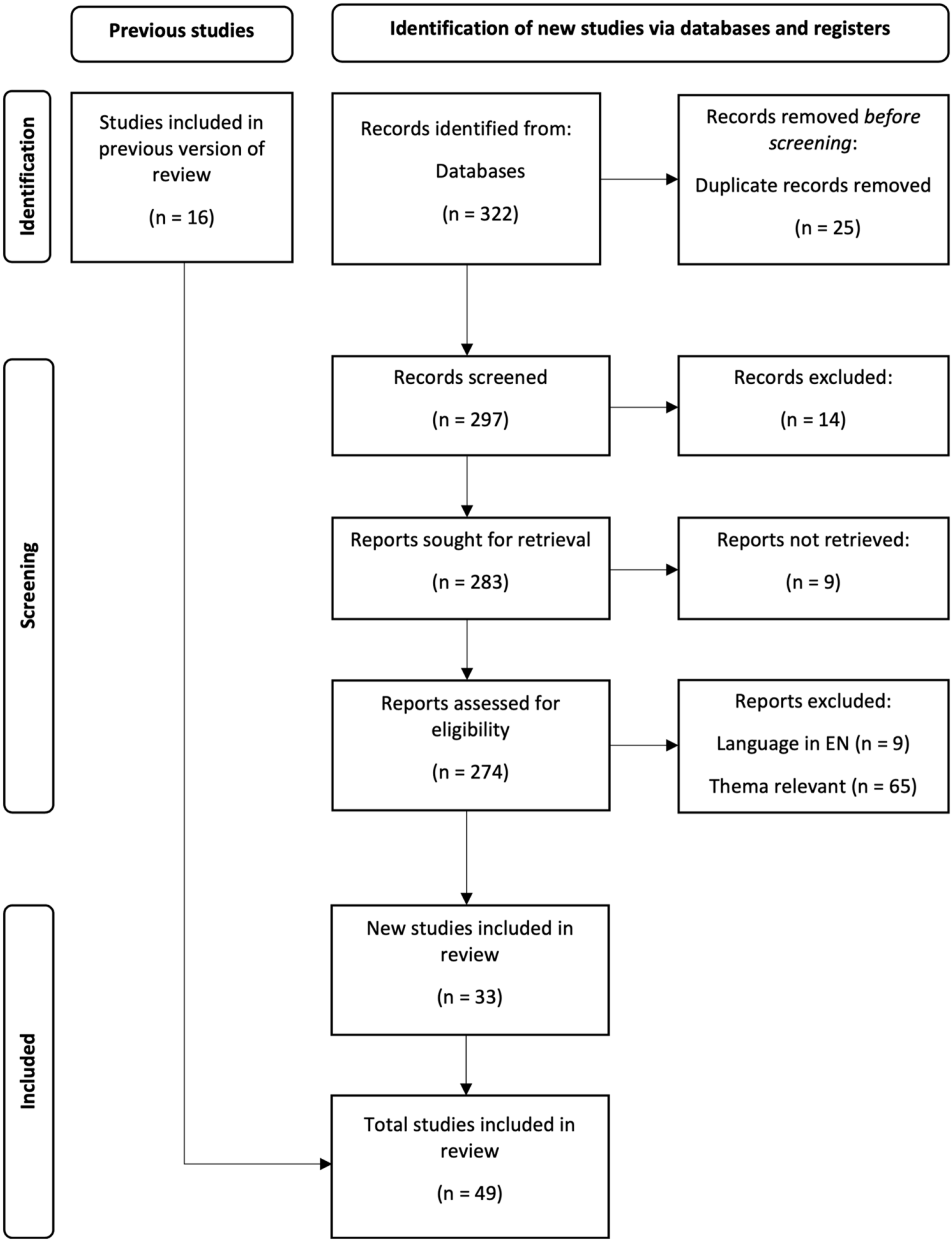


Fig. 1 PRISMA Flowchart of study selection and literature screening process

The dataset was systematically reviewed and recalculated, incorporating 16 papers that were utilized in [6]'s meta-analysis, which we reassessed and recalculated for this study, along with 33 newly selected studies from 2020 to 2024. These studies focused on simulation scenarios with various penetration rates of CAVs, evaluating their impact on safety using different safety metrics. To enable consistent comparison across studies, the SIR was used to standardize all safety metrics into a single, comparable factor. The selection process adhered strictly to predefined standards to ensure both accuracy and relevance, all detailed data were summarized in the supplementary fie. For each paper, the data extraction process involved gathering information on the authors, publication year, the number of simulations runs or observations (simulation times), and key metrics, including the MPR of CAVs, the SIR relative to baseline scenario (Eq. (1)), and the standard error (SE) calculated using Eq. (2). This comprehensive data collection enabled a detailed analysis of CAVs' impact on traffic safety across different studies, facilitating robust cross-study comparisons and improving the understanding of how varying CAV penetration levels affect safety outcomes.

2.3 Effect size

To standardize the evaluation of safety benefits across studies, the SIR was used as the primary metric. The SIR quantifies the impact of CAVs on safety by comparing the number of conflicts in scenarios without CAVs to those varying levels of CAV market penetration (see Eq. (1)). The formula is outlined below, where CN_0 represents the baseline number of conflicts without CAVs, indicating the conflict events in different market penetration rate. CN_x denotes the conflicts observed at CAV market penetrations from 10% to 90%, reflecting conflict events at different MPR scenarios [6].

$$\boldsymbol{SIR} = \frac{\boldsymbol{CN}_0 - \boldsymbol{CN_x}}{\boldsymbol{CN}_0} \times 100\% \tag{1}$$

Additionally, the standard error (SE), for SIR is computed as follows (see Eq. (2)):

$$\boldsymbol{se}\left(\boldsymbol{SIR}\right) = \sqrt{\frac{\boldsymbol{SIR}\left(1 - \boldsymbol{SIR}\right)}{\boldsymbol{n} - 1}} \tag{2}$$

Where n indicates the number of simulations running times in the experiments. The above formula is obtained from the Cochrane handbook of systematic evaluation and is used to assess the robustness of the effect values [6].

2.4 Inverse-variance estimation method

The next phase involved applying the inverse variance estimation method to calculate weights of effect values taken from studies. It works well for handling differences in effect values that occur across studies, and particularly fits for datasets showing considerable differences between them. The exact formula is presented as follows (Eq. (3)):

$$\boldsymbol{w_i} = \frac{1}{\boldsymbol{se_i^2}} \tag{3}$$

where w_i is the weight of study *i*; and se_i^2 stands for the variance of the effect size of study *i*.

Equation (4) gets applied when calculating a weighted mean effectiveness of the model is:

$$\bar{\boldsymbol{Y}} = \frac{\sum_{\boldsymbol{i}=1}^{\boldsymbol{n}} \left(\boldsymbol{w_i} \times \boldsymbol{Y_i}\right)}{\sum_{\boldsymbol{i}=1}^{\boldsymbol{n}} \boldsymbol{w_i}} \tag{4}$$

where $\bar{Y}$ is referring to as the weighted effect size and Y_i would signify the effect size of study i [15].

2.5 Assessment of heterogeneity

The measure of heterogeneity among the studies was examined through Cochran's Q statistic and I^2 index. The included studies differed in simulation scenarios, such as location, traffic volume, and vehicle types, which naturally introduce variability in effect sizes. Meta-analysis does not reduce this heterogeneity but quantifies it, allowing appropriate modeling and interpretation [16]. Cochran's Q statistic determines whether the observed variability in effect sizes across studies exceeds what would be expected by chance, guiding the choice between fixed-effects and random-effects models (see Eq. (5)).:

$$\boldsymbol{Q} = \sum \boldsymbol{w_i} \times \left(\boldsymbol{Y_i} - \bar{\boldsymbol{Y}}\right)^2 \tag{5}$$

As Cochran's Q is highly sensitive to the number of included studies and can be difficult to interpret on its own, we focus on I^2 and τ^2, which provide a clearer and more meaningful assessment of heterogeneity [16]. The I^2 homogeneity index was applied to find the variation in studies, measured with Eq. (6):

$$\boldsymbol{I}^2 = \boldsymbol{max}\left(0, \frac{\boldsymbol{Q} - (\boldsymbol{k} - 1)}{\boldsymbol{Q}}\right) \times 100\% \tag{6}$$

where *k* is representing the number of studies; and a greater I^2 value means that diversity among the studies is higher [17].

2.6 Random-effect models

Due to the presence of significant heterogeneity, as indicated by the I^2 index, a random effects model was utilized. This model is used in the analysis due to differences in study backgrounds and sample characteristics. Unlike the fixed effect model, which assumes that effect values remain constant across studies, the random-effect model accounts for variability in effect size between studies. It acknowledges that some variation in effect sizes is expected across different studies. This model focuses on the differences in effects between studies rather than within each individual study. By incorporating between-study variation, it provides a more comprehensive estimate of the overall effect (see Eq. (7)).

$$\boldsymbol{w_i^*} = \frac{1}{\boldsymbol{se_i^2} + \widehat{\boldsymbol{\tau}}^2} \tag{7}$$

where $\widehat{\tau}^2$ denotes the between-study variance (see Eq. (8)), estimated as:

$$\widehat{\boldsymbol{\tau}}^2 = \boldsymbol{max}(0, \frac{\boldsymbol{Q} - (\boldsymbol{k} - 1)}{\sum \boldsymbol{w_i} - \frac{\sum (\boldsymbol{w_i}^2)}{\sum \boldsymbol{w_i}}} \tag{8}$$

In this study, the between-study variance (τ^2) in the random-effects model was estimated using the DerSimonian–Laird (DL) method. The DL estimator was selected because our meta-analysis includes 49 independent studies, which is considered a large dataset for meta-analytic purposes. Prior research has shown that DL provides stable and reliable estimates of $\widehat{\tau}^2$ when the number of studies exceeds approximately 30 and heterogeneity is moderate [15, 18].

By incorporating $\widehat{\tau}^2$, the random effects model can more accurately capture the differences between studies, thereby enhancing the external validity for the findings. To assess the reliability of the effect sizes, 95% confidence intervals were calculated. The confidence interval was computed in Eq. (9):

$$95\%\mathbf{CI} = \bar{\boldsymbol{Y}} \pm 1.96 \times \sqrt{\frac{1}{\sum \boldsymbol{w_i^*}}} \tag{9}$$

where 1.96 signifies the value of the normal distribution. The utilization of 95% confidence intervals aids in estimating the range of effect values [18].

2.7 Sensitivity analyses

Following the selection of the random-effect model, sensitivity analyses are conducted. This step involves a rigorous examination of the data to identify and exclude outliers. By removing these outliers, we ensure that the meta-analysis results are not overly influenced by extreme values, thus enhancing the reliability and validity of the findings. Additionally, sensitivity analyses help refine the relationship between MPR and SIR, ensuring that any irregularities or biases in the data are addressed, providing a clearer understanding of the impact of CAVs on traffic safety [16].

2.8 Bias assessment

We assess potential publication bias using both funnel plots and Begg & Egger's tests. Funnel plots serve as visual tools to detect bias, where any asymmetrical distribution may indicate the presence of publication bias, particularly when smaller studies with less significant results are underreported. Begg & Egger's tests provide statistical evidence of this bias, and the results are judged based on the p-value. A p-value less than 0.05 suggest a statistically significant presence of bias. If bias is detected, further corrective measures, such as the Trim-and-Fill method, are implemented to adjust the analysis [19].

2.9 Trim-and-fill method

If publication bias is detected, the Trim-and-Fill method is employed to estimate and correct for the missing studies that may have caused the asymmetry in the funnel plot. This method hypothesizes the number of missing studies-often those with non-significant or negative results-and then imputes them into the analysis, aiming to create a more symmetrical funnel plot. By doing so, it recalculates the effect size, providing a less biased and more accurate estimate. This adjustment enhances the robustness of the meta-analysis by mitigating the impact of publication bias [20].

2.10 Modeling the MPR-SIR relationship: testing functional fits

In the final step of our analysis, we establish a relationship between MPR as the independent variable and SIR as the dependent variable. To determine the best fitting model for this relationship, we test four different functions: logarithmic, linear, exponential, and power. In the following result section, the explained methodology is applied to analyze the impact of CAVs on traffic safety.

3 Results

3.1 Relationship between MPR and SIR: initial analysis

The first results illustrate funnel plots to evaluate the relationship between SIR and MPR (see Fig. 2). In these funnel plots, the horizontal axis shows the SIR for each MPR, where positive values indicate a decrease in conflicts numbers. The vertical axis represents the standard error of each study, and weights are calculated from the standard error using Eq. (3) in the random effect model. Typically, funnel plots have an inverted funnel shape with

(A)

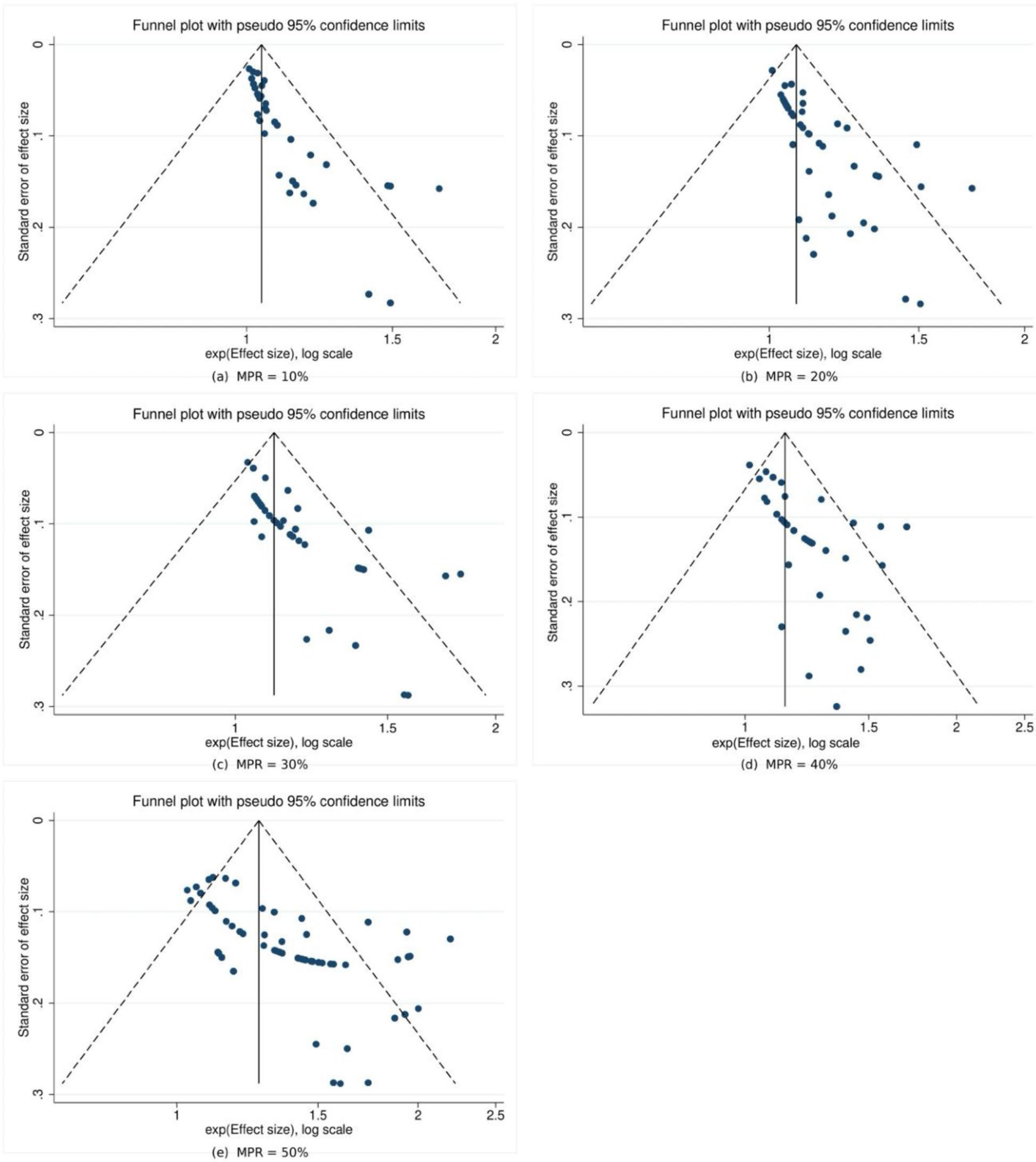


Fig. 2 Funnel plots of effect size (SIR) versus standard error under different MPR ranges. (**A**) Lower missing percentage rates (MPR = 10–50%). (**B**) Higher missing percentage rates (MPR = 60–90%). The solid line indicates the pooled effect, and diagonal lines indicate the 95% confidence limits

points spread around the estimated true values of independent study effects, suggesting no bias in the studies included. However, upon examining Fig. 2, it was evident that all funnel plots displayed a degree of asymmetry. For the 10%-40% MPR range (see Fig. 2(A)), the data points exhibited, asymmetry, suggesting potential publication bias. On the other hand, for 50%-90% MPR range (see Fig. 2(B)), the distribution appears more symmetric; however, some data points still fall outside the confidence

(B)

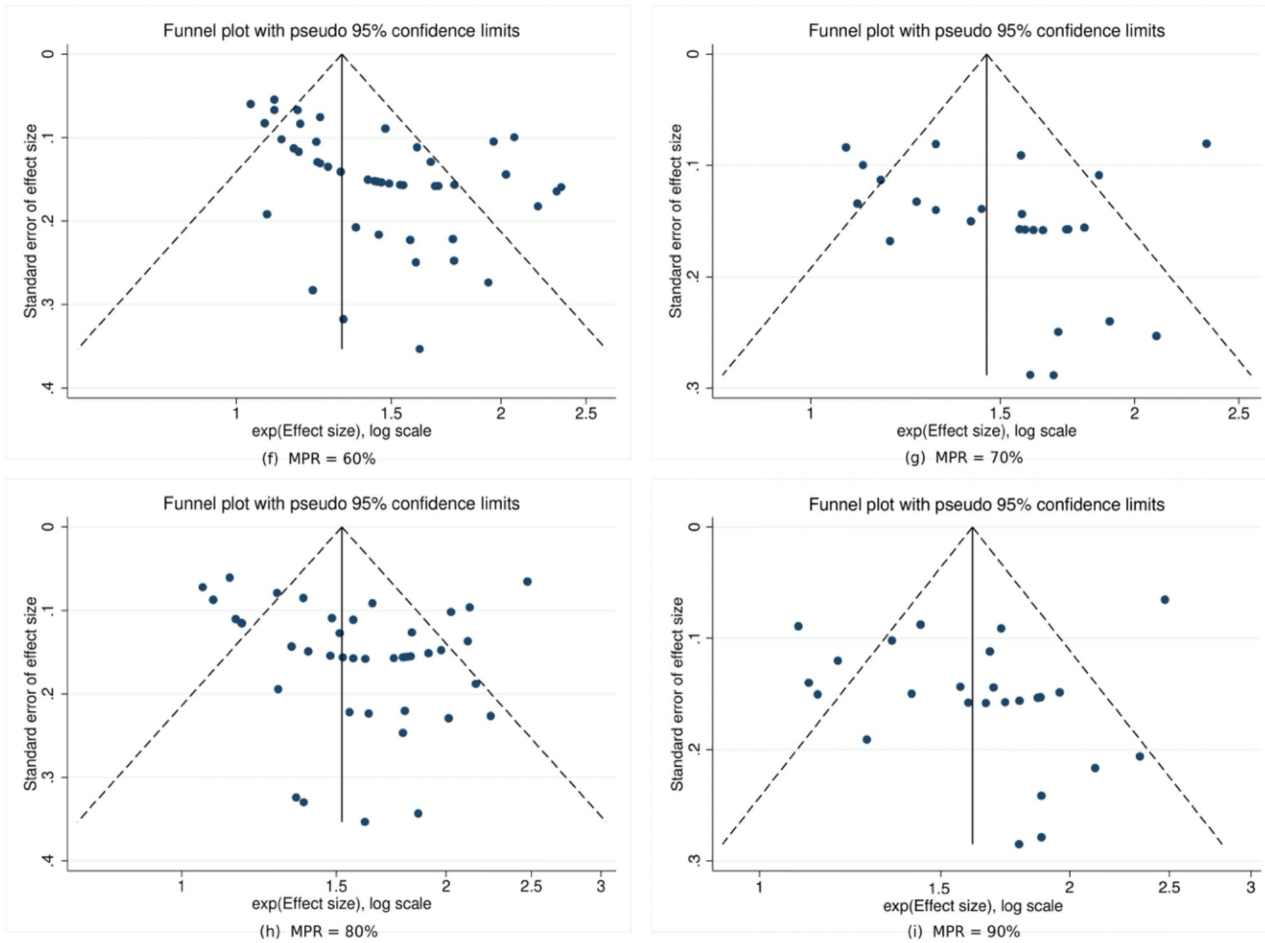


Fig. 2 (continued)

interval lines, indicating higher heterogeneity within studies.

Table 1 provides an analysis of publication bias across varying MPRs, considering the number of analyses and the results of Begg's and Egger's test. The results reveal a distribution of 79 analyses for low MPRs (10%-20%), 178 for mid-range MPRs (30%-60%), and 98 for high MPRs (70%-90%). The number of analyses varies significantly, peaking at 50% MPR with 56 analyses, while the lowest count is observed at 70% MPR with only 27 analyses. At lower to mid-range MPRs (10%-60%), both Begg's and Egger's test consistently yield low p-values (close to 0.000), indicating a significant presence of publication bias. However, at higher MPRs (70%-90%), the p-values from both tests rise substantially, reaching up to 0.315 and 0.665 for Begg's and Egger's tests, respectively, at 80% and 90% MPR levels. This reduction in publication bias at higher MPRs suggests a more balanced representation of study outcomes.

Before our first trim-and-fill adjustments, the SIR generally exhibited an upward trend across the MPRs, as shown in Table 1. Specifically, the SIR values increased from 4.1% at 10% MPR to the SIR value of 29.4% at 50% MPR, followed by a further rise to 49% at 90% MPR. After applying trim-and-fill method, the SIR values decreased across all MPRs, reflecting a correction for potential bias. For instance, the SIR reduced from 4.1% to 2.8% at 10% MPR and dropped significantly from 29.4% to 17.7% at 50% MPR. However, it is noteworthy that the SIR did not decrease for the MPRs of 30%, 40% and 90%, where values remained the same. Additionally, following the trim-and-fill adjustment, the SIR values no longer exhibited a constant upward trend; specifically, there was a decline from 21.7% at 40% MPR to 17.7% at 50% MPR.

The comparison between our meta-analysis and the findings by [6] reveals nuanced differences in SIR across varying MPRs. Both analyses confirm that the adoption of CAVs leads to noticeable safety improvements, but the effect and patterns of these improvements change at different penetration rates (see Fig. 4). At lower MPRs (10%-20%), the current study shows slightly lower SIRs compared to the previous analysis. For instance, at 10%

Table 1 Publication bias and Random Effects model analysis before and after sensitivity analysis

	Publication Bias				Random Effect Model							
		Begg's and Egg's test			Before the trim-and-fill				After the trim-and-fill			
	MPR	Number of analyses	P (Begg's test)	P (Egger's test)	Estimate SIR	95%CI	I^2	tau^2	SIR	95%CI	I^2	tau^2
Before sensitivity analysis	10%	35	0.000	0.000	4.1%	[0.021, 0.061]	0.00%	0.00%	2.8%	[0.009, 0.047]	27.30%	0.19%
	20%	44	0.000	0.000	7.9%	[0.055, 0.104]	11.20%	0.08%	4.9%	[0.027, 0.071]	38.70%	0.48%
	30%	37	0.000	0.014	18.5%	[0.122, 0.249]	78.50%	2.67%	18.5%	[0.122, 0.249]	78.50%	2.67%
	40%	39	0.000	0.000	21.7%	[0.148, 0.286]	88.00%	5.06%	21.7%	[0.148, 0.286]	88.00%	5.06%
	50%	56	0.000	0.000	29.4%	[0.241, 0.348]	60.80%	2.14%	17.7%	[0.113, 0.241]	74.10%	4.55%
	60%	46	0.001	0.000	35.6%	[0.288, 0.424]	71.10%	3.45%	21.4%	[0.134, 0.295]	80.30%	6.46%
	70%	27	0.076	0.526	38.7%	[0.300, 0.475]	70.20%	4.08%	28.5%	[0.191, 0.379]	80.60%	7.54%
	80%	43	0.315	0.124	45.8%	[0.383, 0.532]	78.50%	5.64%	33.8%	[0.259, 0.418]	85.10%	9.00%
	90%	28	0.139	0.665	49.0%	[0.391, 0.589]	87.60%	9.52%	49.0%	[0.391, 0.589]	87.60%	9.52%
After sensitivity analysis	10%	35	0.000	0.000	4.1%	[0.021, 0.061]	0.00%	0.00%	2.8%	[0.009, 0.047]	27.30%	0.19%
	20%	44	0.000	0.000	7.9%	[0.055, 0.104]	11.20%	0.08%	4.9%	[0.027, 0.071]	38.70%	0.48%
	30%	36	0.000	0.014	12.2%	[0.088, 0.156]	29.70%	0.31%	8.0%	[0.122, 0.249]	52.00%	0.99%
	40%	37	0.000	0.063	17.2%	[0.127, 0.217]	37.50%	0.54%	9.4%	[0.148, 0.286]	57.70%	1.50%
	50%	56	0.000	0.000	29.4%	[0.241, 0.348]	60.80%	2.14%	17.7%	[0.113, 0.241]	74.10%	4.55%
	60%	46	0.001	0.000	35.6%	[0.288, 0.424]	71.10%	3.45%	21.4%	[0.134, 0.295]	80.30%	6.46%
	70%	27	0.076	0.526	38.7%	[0.300, 0.475]	70.20%	4.08%	28.5%	[0.191, 0.379]	80.60%	7.54%
	80%	43	0.315	0.124	45.8%	[0.383, 0.532]	78.50%	5.64%	33.8%	[0.259, 0.418]	85.10%	9.00%
	90%	27	0.114	0.002	46.3%	[0.371, 0.555]	75.50%	5.29%	39.4%	[0.391, 0.589]	83.50%	8.72%

MPR, our study reports a 2.8% improvement in safety, which is lower than the 4.2% observed in the earlier study. Similarly, at 20% MPR, our SIR is 4.9%, compared to 8.6% previously. In contrast, at mid-range MPRs (30%-60%), our study highlights a significant deviation. Although, the SIR rises steadily from 18.5% at 30% MPR to 21.7% at 40%, a slight drop is observed at 50% MPR, where safety improvements fall to 17.7%. This decline is followed by a recovery to 21.4% at 60% MPR. This drop cannot be seen in the [6], where the SIR increases consistently across the same MPR range. At higher MPRs (70%-90%), both studies align more closely, though our findings reveal a slightly larger improvement. At 90% MPR, our analysis shows a 49% safety improvement, compared to 43.4% in the [6].

In addition to publication bias and safety improvement rates, Table 1 also reports the I^2 index and τ^2 values across different MPR levels to quantify heterogeneity among the included studies. The I^2 values range from 27.3% to 83.5%, indicating moderate to substantial heterogeneity, depending on the MPR level. This variation is expected given differences in simulation contexts, such as traffic volumes, study locations, and vehicle types. Higher τ^2 values were observed at certain MPR levels, suggesting greater dispersion in effect sizes. Together, these findings support the use of a random-effects model in our meta-analysis, as described in Sect. 2.5 and 2.6.

3.2 Refined relationship between MPR and SIR: Sensitivity analysis results

After applying the Trim-and-Fill method to correct for potential publication bias, a focused sensitivity analysis was performed to further refine the MPR–SIR relationship and assess the robustness of the pooled estimates. This additional step aimed to identify studies exerting disproportionate influence on the aggregated Safety Improvement Rate (SIR) values, particularly at specific Market Penetration Rate (MPR) levels where irregular patterns were observed. To achieve this, we targeted MPR levels of 30%, 40%, and 90%, where potential outliers were detected. For each study, the SIR and its standard error (SE) were calculated and summarized in Table 2, providing a transparent overview of individual effect sizes and their precision. Each study was sequentially removed, and the pooled SIR and corresponding confidence interval were recalculated to evaluate its influence on the overall estimate. Studies were identified as outliers when their exclusion caused a substantial shift in the pooled SIR or altered the confidence interval range.

At MPR = 30% [21], (*SIR = 0.887, SE = 0.073*), which examined freeway merging areas, exerted a strong influence on the pooled estimate. At MPR = 40%, two studies— [22] (*SIR = 0.970, SE = 0.057*), focusing on intersection coordination of connected vehicles, and [23] (*SIR = 0.131 and 0.014; SE = 0.077 and 0.027*),

Table 2 Studies Identified as Outliers through Sensitivity Analysis

Market Penetration Rate (MPR) Level	Study	Scenario Type	Safety Improvement Rate (SIR)	Standard Error (SE)
30%	Chen et al. (2024) [21]	Freeway merging area	0.887	0.073
40%	Olia et al. (2015) [22]	Intersection / connected vehicle coordination	0.970	0.057
40%	Karbasi et al. (2024) [23]	Low-speed urban network	0.131	0.077
40%	Karbasi et al. (2024) [23]	Low-speed urban network (secondary result)	0.014	0.027
90%	Arvin et al. (2020) [24]	Signalized intersection	0.450	0.166

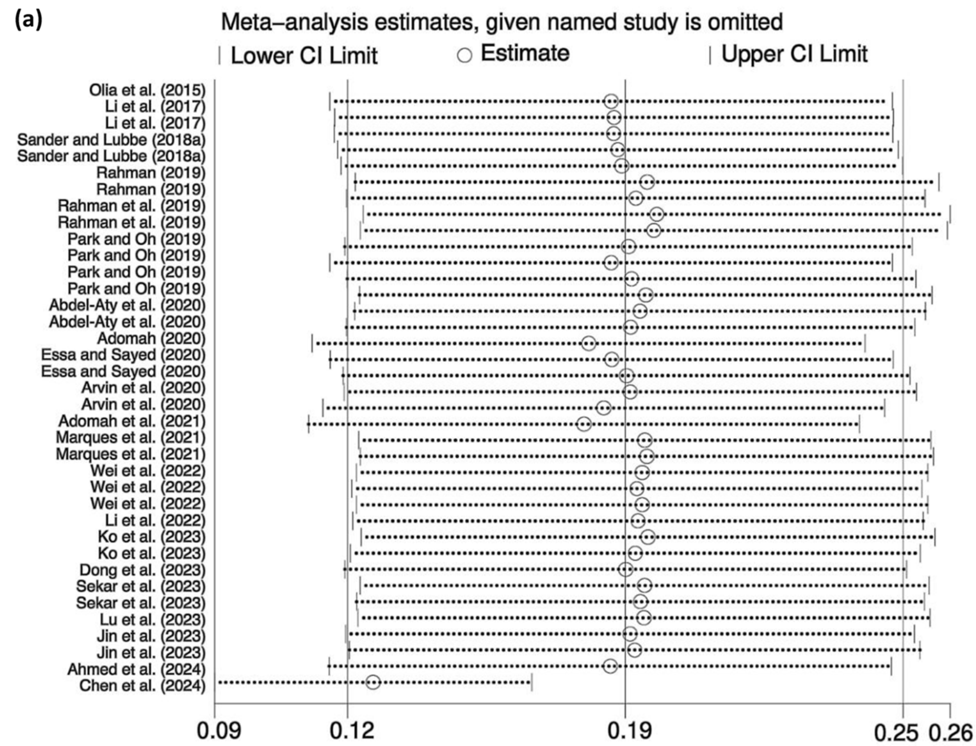
which analyzed low-speed urban networks—significantly affected both the effect size and confidence intervals. At MPR = 90% [24], (*SIR = 0.450, SE = 0.166*), evaluating automated vehicle performance at signalized intersections, also demonstrated a notable influence. These studies, primarily representing localized traffic environments (e.g., intersections, merging zones, or low-speed urban settings), tended to amplify safety effects at specific MPR levels due to limited generalizability of their simulation contexts. Consequently, they were excluded from the final pooled estimates to ensure a more stable and representative trend across MPR levels. Figure 3(a–c) illustrates the effect of removing these outliers, confirming the improved consistency, reliability, and robustness of the meta-analytic results.

After implementing the sensitivity analysis, the results of Begg's and Egger's test, along with the number of analyses have been updated to reflect the publication bias assessment across varying MPRs (see Table 1). The data now indicate a total of 79 analyses for low MPRs (10%-20%), 175 for mid-range MPRs (30%-60%), and 97 for high MPRs (70%-90%). The number of analyses varies significantly, peaking at 50% MPR with 56 analyses, while the lowest counts are observed at both 70% and 90% MPRs, each with 27 analyses. At the low MPRs (10%-20%), both Begg's and Egger's test consistently report low p-values (0.000), indicating a significant presence of publication bias. In the mid-range MPRs (30%-60%), there is a gradual increase in p-values; for example, Egger's test p-value rises to 0.014 at 30% MPR and 0.063 at 40% MPR. At higher MPRs (70%-90%), the p-values from both tests increase significantly, with Begg's test reaching 0.315 and Egger's test going up to 0.526 at 80% MPR, indicating a noticeable decline in publication bias and a more balanced representation of study outcomes in these ranges.

The following result derive from our second random-effect model, updated after implementing the sensitivity analysis. After the adjustment, the SIR exhibited a general decline across all MPRs compared to the initial estimates, although a consistent upward trend remained evident. Specifically, at 10% MPR, the SIR decreased from 4.1% to 2.8%, and at 20% MPR, it dropped from 7.9% to 4.9%. The same trend was observed for the MPRs at 50% and 60%, where the SIR decreased after trim and fill, yet the overall upward trend from 50% to 60% was maintained, with SIR values of 17.7% and 21.4%, respectively. In higher MPRs, the SIR was reduced from 38.7% to 28.5% at 70% MPR, but the upward trend resumed, reaching 33.8% at 80% and 39.4% at 90% MPR. These findings suggest that while the trim and fill led to lower SIR values across the spectrum, the overall trend of increasing safety benefits at higher penetration rates remains a significant observation in this analysis, as summarized in Table 1.

The comparison of SIR across different MPRs reveals some differences between the findings of [6], our current meta-analysis before sensitivity analysis, and the updated results after implementing sensitivity analysis. As shown in Table 1; Fig. 4, at lower MPRs (10%-20%), the SIR values remained consistent in our current meta-analysis, both before and after sensitivity analysis, at 2.8% and 4.9%, respectively. These values are lower than those reported by [6], which were 4.2% at 10% MPR and 8.6% at 20% MPR. In mid-range MPRs (30%-60%), notable changes are observed. At 30% MPR, the SIR in our analysis decreased significantly after sensitivity analysis, from 18.5% to 8.0%, aligning more closely with [6]'s reported SIR of 8.1%. Similarly, at 40% MPR, the SIR decreased from 21.7% before sensitivity analysis to 9.4% after, which is now lower than [6]'s value of 12.8%. At 50% and 60% MPRs, the SIR values remained unchanged after sensitivity analysis, holding steady at 17.7% and 21.4%, respectively, closing matching [6]'s results of 17.4% and 19.6%. For higher MPRs (70%-90%), the sensitivity analysis had minimal impact on the SIR values, with no changes observed at 70% and 80% MPRs, which remained at 28.5% and 33.8%, respectively. These results are comparable to [6]'s values of 31.0% and 35.2%. however, the highest MPR of 90%, the SIR dropped from 49.0% before sensitivity analysis to 39.4% after, which is slightly lower than [6]'s value of 43.4%.

Figure 5 summarizes the results of different modeling approaches used to examine the relationship between SIR and MPR. Among the models assessed, the power



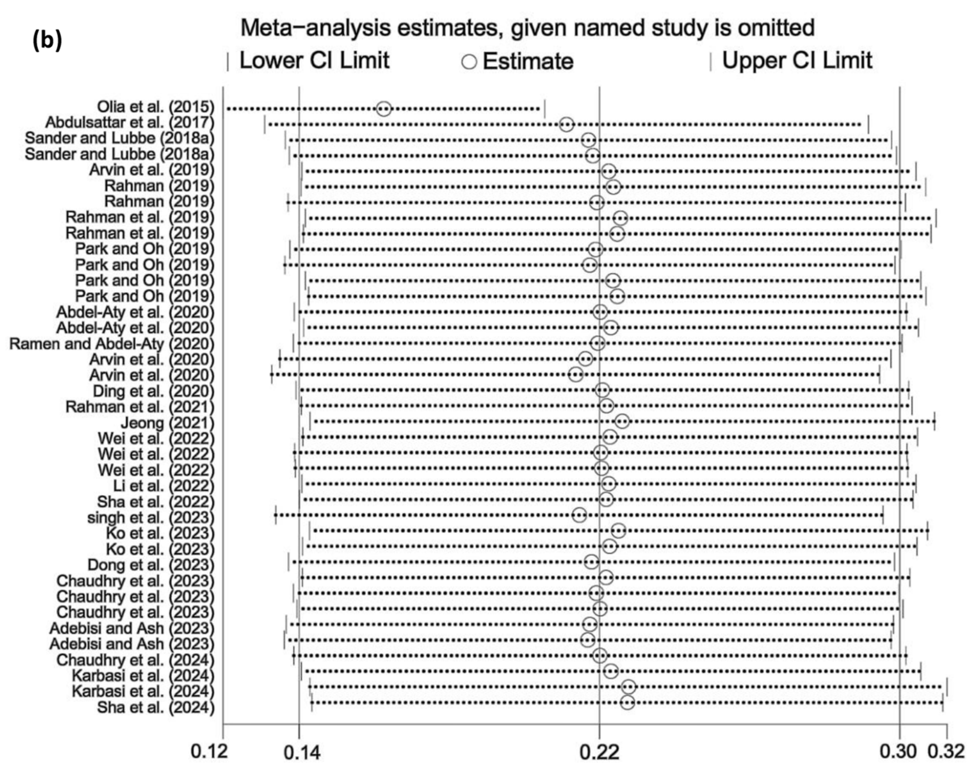


Fig. 3 Sensitivity analysis under different MPRs. (**a**) MPR = 30%. (**b**) MPR = 40%. (**c**) MPR = 90%

model stands out as the most effective, with the highest coefficient of determination ($R^2 = 0.9928$), indicating that it explains the majority of variance in SIR. The linear ($R^2 = 0.9714$) and exponential ($R^2 = 0.9663$) models also demonstrate good fit, but they do not capture the nonlinear relationship as effectively as the power

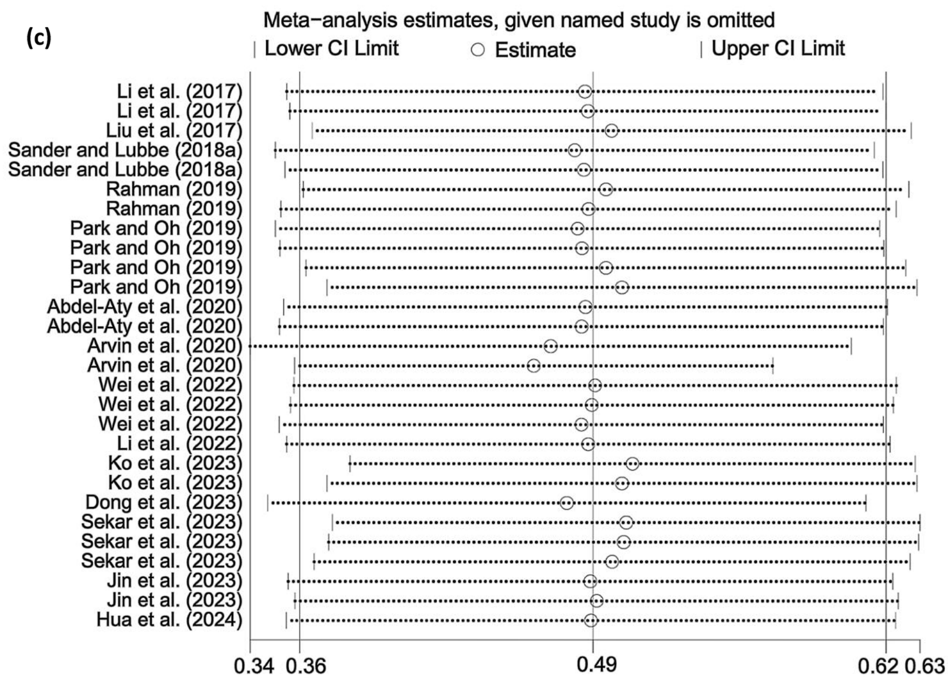


Fig. 3 (continued)

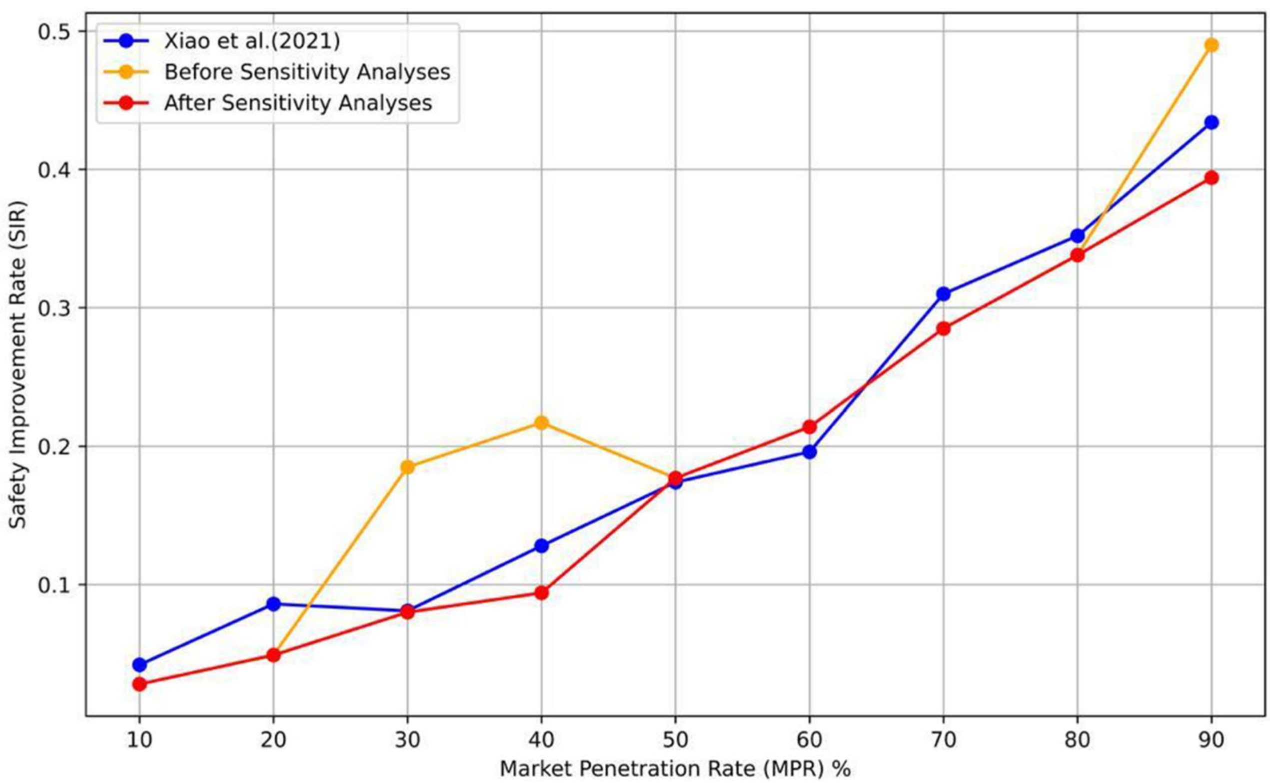


Fig. 4 Comparison of SIR across different MPR before and after sensitivity analysis with Xiao et al. (2021) [6]

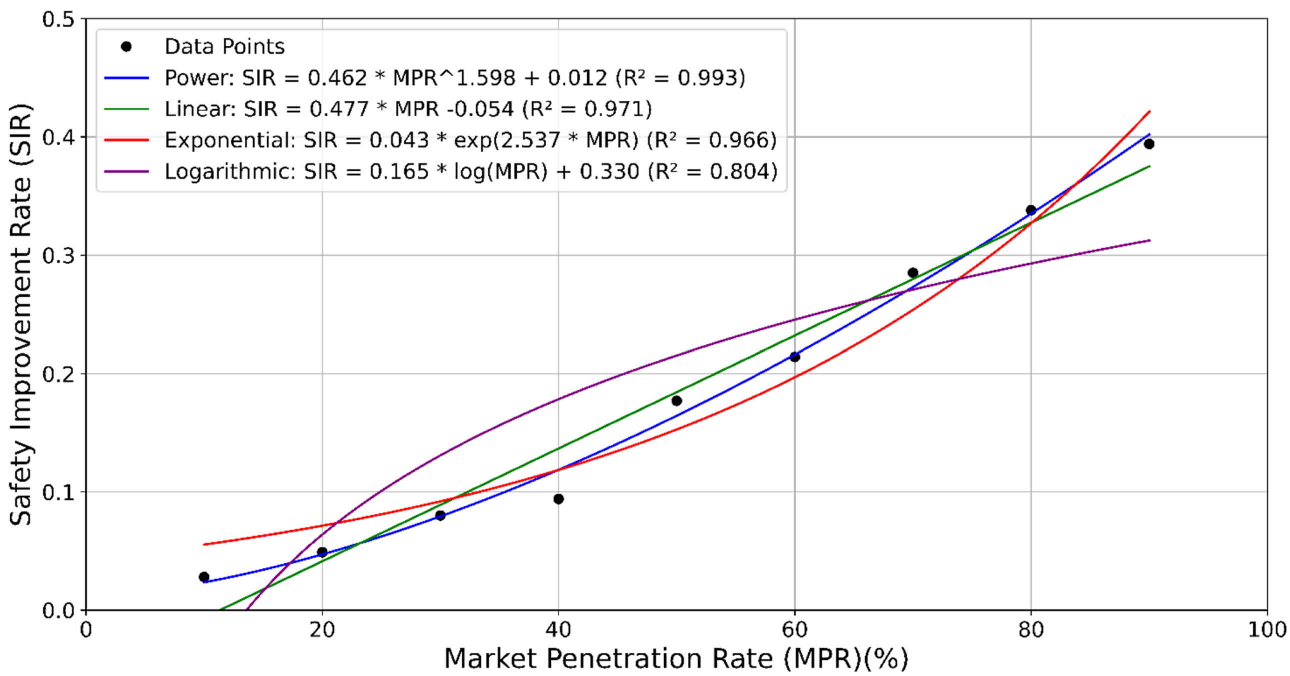


Fig. 5 Comparison of Different Fit Models for MPR and SIR

model. Furthermore, the power model's simplicity, strong explanatory power, and ease of interpretation make it a more suitable option for reporting. The goal is to balance model complexity with interpretability and generalizability to ensure that the results are practically useful. Figure 5 compares four candidate models for fitting the relationship between MPR and SIR, clearly showing that the power model provides the best fit.

4 Discussion

This study provides an in-depth analysis of how CAVs impacts road safety as MPR vary under different conditions, utilizing a meta-analysis. Building upon [6], this study extends the analysis timeline to include data from 2015 to 2024, adding 33 new studies to the original 16, analyzing 49 relevant studies in total. Additionally, the sensitivity analysis conducted in the current meta-analysis led to more reliable and refined results, enhancing the credibility of the study's outcomes. The current meta-analysis results align with [6] indicating that the widespread adoption of CAVs can lead to a significant reduction in traffic conflicts, particularly at higher MPRs. These outcomes have been consistently confirmed by various studies, with additional examples from different countries further demonstrating the positive impact of CAVs on road safety [6]. The findings also reveal that safety improvement rates for CAVs are relatively modest at lower MPRs (10%-20%), suggesting that managed lanes should be implemented during this phase [25]. As MPRs rise, allowing CAVs to use all lanes, they become beneficial, as interactions with conventional vehicles are reduced while communication among CAVs is enhanced. Higher MPRs are frequently associated with greater reductions in conflicts [6]. The results underscore the critical role of CAV technology in crash reduction, highlighting the importance of both technological advancement and policy development in realizing these safety benefits.

4.1 Frequency of analyses per MPR

Based on the number of analyses, mid-range MPRs (30%-60%) have the highest frequency, with 175 analyses, accounting for 49.9% of total analyses. Within this mid-range category, the MPR of 50% stood out with the highest share, representing 56 analyses, which constitutes approximately 16% of all total analyses. On the other hand, low MPRs (10–20%) have the least frequency, with a combined count of 79 analyses, representing of 22.5% of the total, suggesting relatively lower attention to these levels. High MPRs (70%-90%) fall in the middle with 97 analyses, making up 27.6% of the total, highlighting relatively less attention than the mid-range MPRs, however still higher than low MPRs. The mid-range analyses are approximately 2.2 times higher than those for low MPRs and about 1.8 times higher than those for high MPRs. The significant interest in the mid-range MPRs shows the effects of CAVs are likely more variable compared to both lower and higher MPR levels. Strong evidence of publication bias exists at lower to mid MPRs (10%-60%), indicated by consistently low p-values. In contrast, higher MPRs (70%-90%) show a lack of significant bias. The SIRs at lower to mid-range MPRs should be interpreted with caution, while findings at higher MPRs likely provide a more accurate depiction of the effects of CAV technology on traffic safety.

4.2 Relationship between SIR and MPR

The best functional form for the relationship between SIR and MPR is a power function form. This relationship indicates that both low and high MPRs contribute to improvement in conflict reduction and overall safety, with more pronounced benefits observed at higher MPRs. This trend aligns with the findings from previous studies, such as those by [26, 27], which suggest that higher MPRs generally facilitate more reliable traffic networks due to the increased volume of data shared among vehicles [28]. Our study reveals nuanced differences in SIRs for CAVs when compared to [6] across varying MPRs. At lower MPRs (10%-20%), our study shows improvements in SIR, albeit slightly lower than what reported in [6], indicating that safety improvements may be constrained when CAV technologies are not yet widely adopted. The observed improvements at these lower MPRs can be attributed to early benefits from the introduction of CAVs, such as enhanced driver awareness and automated safety features that provide some level of safety even when market penetration is limited. These moderate improvements at low MPRs may also result from limited data sharing and communication between vehicles, as well as heterogeneity of vehicle types present at these lower MPRs, as noted in the work of [24].

In the mid-range MPRs (30%-60%), sensitivity analysis plays a crucial role, particularly at 30% and 40%, where it corrects an initial overestimation of variability in the current meta-analysis. The irregularity in SIR at mid-range MPRs may be a transitional issue inherent to mixed traffic scenarios, where neither conventional vehicles nor CAVs dominate. As [25] demonstrated, even with the same CAV technology at a 50% MPR, crash reductions varied significantly between peak (34.9%) and off-peak (48.5%) traffic conditions. At higher MPRs (70% – 90%), our findings exhibit a stronger alignment with [6]; however, we note a more significant improvements in safety outcome in the current meta-analysis. The increased SIRs reflect the dominant presence of CAVs leading to enhanced data sharing and communication among vehicles. As the proportion of CAVs rises, traffic dynamics become more favorable, reducing the likelihood of crashes and improving safety outcomes overall. However, prior to sensitivity analysis, there was an exaggerated rise at 90% MPR. The sensitivity analysis effectively reduces this spike, bringing the results closer to [6]'s finding and ensuring a more balance trend. Overall, sensitivity analysis contributes to smooth out irregularities, particularly in the mid and high ranges, leading to more consistent and reliable results.

Achieving safety improvements with CAVs is not guaranteed at lower MPRs. Factors such as vehicle heterogeneity at signalized intersections and need to reach a certain MPR threshold to realize positive safety outcomes play a critical [24]. Simulation parameters and driving conditions significantly influence safety outcomes, particularly at low MPRs, where accuracy tends to decrease due to less data and variability in vehicular behavior. Several studies [21, 29] emphasize that the choice of evaluation methods adds complexity to interpreting results across different scenarios. For instance, several studies exhibited disproportionate influence on pooled results at specific MPR levels, as indicated by our sensitivity analysis [30]. These outliers— [21] at 30% MPR [22] and [23], at 40% MPR, and [24] at 90% MPR—shared a consistent pattern: each was based on localized or highly idealized simulation environments such as intersections, merging areas, or low-speed urban networks. These conditions tend to amplify safety improvements due to simplified interactions and assumptions of flawless communication or driver response. Similar tendencies have been noted in prior simulation-based studies [31–33], which caution that such idealized frameworks often overestimate real-world safety benefits. By identifying and excluding these outliers, the refined analysis aligns more closely with large-scale, mixed-traffic evaluations and provides a more balanced depiction of CAV safety performance across adoption scenarios. This outcome reinforces the methodological rigor of the current study and underscores the importance of context-aware synthesis when quantifying CAV safety impacts.

4.3 Distribution of safety surrogate measures in the meta-analysis

The distribution of safety surrogate measures (SSMs) among the analyses included in the meta-analysis (2015–2024) highlights the relative emphasis placed on different indicators of traffic risk. Measures based on Time-to-Collision (TTC)—including TTC itself, Time Exposed TTC (TET), and Time Integrated TTC (TIT)—collectively accounted for 64% of the analyses, reflecting their dominant role as primary indicators of potential collisions. Post Encroachment Time (PET) was used in 15% of analyses, while the remaining 21% employed a variety of other SSMs (e.g., Time Exposed Rear-End Crash Risk Index, Lane Changing Conflicts).

A temporal comparison with the period analyzed by [6] from 2015 to 2020 reveals notable shifts in focus over time (Fig. 6). During 2015–2020, TTC-based measures—including TTC, TET, and TIT—collectively accounted for approximately 60% of analyses, with TTC at 30%, TET at 15%, and TIT at 15%. PET represented only 5% and other SSMs made up 35%. In the 2020–2024 period, TTC-based measures slightly increased to 65%; this increase

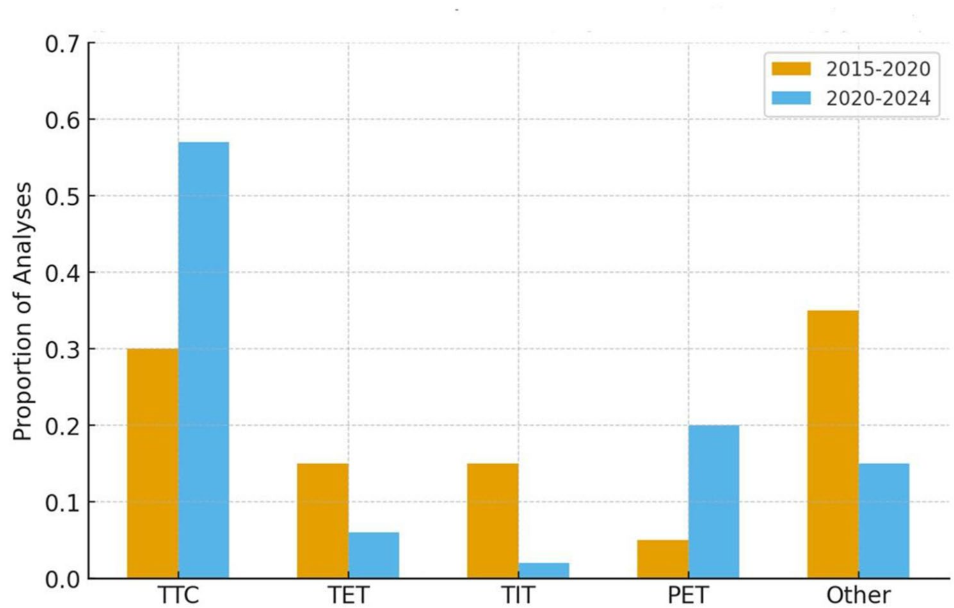


Fig. 6 Comparison of Safety Surrogate Measure Distribution Across Periods: TTC, TIT, TET, PET, and Other Measures

was primarily driven by a significant rise in TTC, from 35% to 57%, while TET and TIT decreased to 6% and 2%, respectively. PET usage rose sharply to 20%, reflecting a growing emphasis on near-miss and post-encroachment risk. Other SSMs also saw a substantial decline, dropping from 35% of analyses to approximately 15%. These trends indicate that, although TTC-based measures continue to dominate overall, the focus has shifted toward TTC itself, with PET increasingly adopted to capture additional dimensions of traffic safety, particularly in evaluating connected and automated vehicle interventions.

4.4 Advancements of this study compared to Xiao et al. (2021) [6]

A comparative summary between the present meta-analysis and [6] highlights several methodological and analytical advancements achieved in this study (Table 3). While [6] established the foundational relationship between Market Penetration Rate (MPR) and Safety Improvement Rate (SIR), the current analysis extends the study period from 2015 to 2024, encompassing 49 studies and 354 effect sizes—more than double the analytical scope of the earlier work. This broader dataset enhances statistical robustness and ensures a more balanced representation across low-, mid-, and high-MPR scenarios. A key advancement introduced in the current study is the explicit analysis of SSM distribution across all included studies—an aspect absents from the earlier meta-analysis—which enables the identification of methodological shifts and potential overreliance on specific safety indicators. Building on this, we applied sensitivity weighting and variance-based corrections to further reduce the influence of any single SSM on aggregated outcomes. Methodologically, the integration of a structured sensitivity analysis alongside the traditional trim-and-fill method provides stronger bias correction, mitigating the influence of outliers and overrepresented studies. Moreover, heterogeneity was more effectively managed using a variance-based weighting approach within the random-effects model, resulting in more stable and generalizable findings. The updated results reveal a smoother and more realistic SIR trend, increasing from 2.8% at 10% MPR to 39.4% at 90% MPR, compared to the steeper trajectory reported by [6]. Additionally, the power function (R^2 = 0.99) was identified as the best-fit model, outperforming the exponential form (R^2 = 0.97) previously observed.

4.5 Policy implications

This study provides actionable insights for policy and decision-making by establishing a bias-adjusted and methodologically robust relationship between Market Penetration Rate (MPR) and Safety Improvement Rate (SIR). The application of structured sensitivity analysis and the removal of outlier studies—mainly those derived from localized or idealized conditions such as intersections, merging zones, and low-speed urban networks—enhance the reliability of system-wide safety estimates. However, addressing these outliers also presented methodological challenges, as such studies often reflect critical yet narrowly scoped traffic environments where automation may perform exceptionally well or poorly. Recognizing and transparently managing these discrepancies is

Table 3 Comparison of key methodological and analytical features between Xiao et al. (2021) [6] and the current study

Aspect	Xiao et al. (2021)[6]	Current Study (2025)	Improvement / Advantage
Study period	2015–2020	2015–2024	Extended timeline includes latest CAV safety research and modeling advancements.
Number of studies included	16	49	Broader dataset enhances statistical power and representativeness.
Number of analyses (effect sizes)	139	354	Greater analytical depth across MPR levels ensures improved reliability.
Market Penetration Rate (MPR) range	10%–90%	10%–90%	Same range, but with more balanced study distribution across low, mid, and high MPRs.
Analysis of Safety Surrogate Measures (SSMs) distribution	Not analyzed	Comprehensively analyzed across all included studies	Identifies shifts in methodological focus and supports weighting corrections to reduce single-SSM influence.
Bias correction method	Trim-and-fill	Trim-and-fill + Structured Sensitivity Analysis	Additional bias control mitigates influence of outliers and overrepresented studies.
Handling of heterogeneity	Random-effect model	Random-effect model with variance-based weighting	Improved control of heterogeneity for more stable and generalizable results.
Key SIR findings	SIR increased from 4.2% (10% MPR) → 43.4% (90% MPR)	SIR increased from 2.8% (10% MPR) → 39.4% (90% MPR)	Smoothed and bias-corrected trend reflects more realistic safety improvements.
Observed publication bias	Moderate bias at low–mid MPRs	Significant bias corrected via sensitivity and trim-and-fill	Reduced asymmetry and improved reliability, especially at high MPRs.
Best-fit functional form	Exponential ($R^2 = 0.97$)	Power ($R^2 = 0.99=$)	Power model offers superior fit, interpretability, and predictive strength.
Overall contribution	Established the baseline MPR–SIR relationship	Updated, bias-adjusted, and methodologically advanced model for robust policy inference	Delivers updated and policy-relevant understanding of CAV safety effects.

essential to avoid overgeneralizing results beyond their valid operational contexts.

For policymakers, this distinction is particularly important. Overreliance on localized findings could lead to unrealistic expectations about early deployment outcomes or misaligned investment priorities. The refined analysis indicates that low MPRs (10%–20%) yield modest but consistent safety gains—suggesting that pilot programs and managed-lane strategies are more appropriate at early stages—while high MPRs (70%–90%) deliver substantial system-level improvements, supporting policies that promote large-scale adoption through infrastructure readiness, interoperability standards, and public education. By acknowledging the complexities associated with outlier studies and providing empirically grounded, bias-corrected estimates, this meta-analysis offers policymakers a transparent and realistic foundation for developing effective, evidence-based CAV deployment strategies.

4.6 Limitations and future research

Despite the comprehensive nature of this Meta-analysis, the findings remain subject to methodological variability and data limitations. The results are constrained by the heterogeneity of the included studies, particularly regarding simulation settings, surrogate safety measures, and definitions of connectivity and automation. Differences in experimental design, traffic composition, and assumed behavioral models may introduce biases that affect the generalizability of the aggregated results. Although sensitivity analyses and variance-based corrections were conducted to reduce the influence of individual study biases and methodological inconsistencies, residual uncertainty remains due to the inherent diversity of the underlying data. Therefore, while the meta-analytic synthesis provides valuable insights into overall safety trends, the results should be interpreted with caution and regarded as indicative rather than definitive evidence of CAV safety performance.

Future research should seek to disentangle the respective safety contributions of connectivity and automation by developing multi-dimensional meta-models that independently estimate the effects of (i) connectivity-only systems, (ii) automation-only systems, and (iii) fully integrated CAVs across different automation levels. It is important to note that the SIR, as defined in Eq. (3), consolidates outcomes from various surrogate safety indicators (e.g., TTC, PET) and assumes that, on average, these indicators exhibit comparable trends with respect to Market Penetration Rate (MPR). Although the meta-analysis applied sensitivity analyses and variance-based corrections to mitigate potential bias from any single surrogate measure, future meta-analytic frameworks could be further refined to calculate separate SIR values for individual surrogate safety measures, thereby offering a more nuanced and precise evaluation of how different technologies influence specific dimensions of road safety. Establishing standardized reporting protocols for surrogate measures, simulation parameters, and technology definitions would also enhance the comparability and reliability of results across studies. Expanding the geographical scope of CAV research to include a wider range of countries and regions is likewise recommended, as this would capture diverse cultural and transportation contexts and provide a clearer picture of CAV applicability worldwide. In addition, long-term investigations are needed to assess the safety performance of CAVs over extended periods, accounting for the influence of evolving technologies, infrastructure, and policy frameworks [34]. Such studies should also consider environmental and operational variables—such as weather, roadway characteristics, and traffic density—to develop more comprehensive safety assessment models. To bridge the gap between technology and policy, future studies could focus on maximizing the safety potential of CAVs through integrated interventions that combine technological advancements with supportive regulatory measures. These efforts may include enhancing automated driving algorithms, increasing public awareness and education, and promoting policies that encourage CAV adoption [35]. Moreover, future work should examine user acceptance and broader societal impacts—including employment and economic changes—as these factors are likely to influence MPR and, in turn, affect SIR outcomes [36].

5 Conclusions

This study set out to answer a central research question: How does the market penetration rate (MPR) of connected and automated vehicles (CAVs) influence overall traffic safety, as measured by the Safety Improvement Rate (SIR)? To address this, a comprehensive meta-analysis of 49 studies and 354 effect sizes published between 2015 and 2024 was conducted, extending and refining the work of [6].

The results demonstrate a clear positive relationship between MPR and SIR, best represented by a power function (R^2 = 0.99). Safety improvements increase progressively with higher CAV penetration, ranging from approximately 2.8% at 10% MPR to 39.4% at 90% MPR. These findings confirm that the widespread adoption of CAVs yields substantial safety gains due to enhanced communication, coordination, and automation among vehicles. However, at lower MPRs (10%–20%), improvements remain modest and more variable, reflecting the influence of mixed traffic conditions and limited data exchange between vehicles.

The study also highlights the importance of sensitivity analysis and bias correction. By removing outliers and adjusting for methodological inconsistencies, the refined model produces smoother and more realistic SIR trends, addressing prior overestimations at mid and high MPRs. This methodological advancement enhances the reliability of the aggregated results and provides a robust basis for future simulation and policy analyses.

Compared to [6], this study extends the dataset and introduces key methodological refinements—structured sensitivity analysis, variance-based weighting, and random-effects modeling—that enhance robustness and reduce bias from localized or idealized studies. These improvements yield a smoother, more realistic SIR trajectory and, by identifying a power function (R^2 = 0.99) as the best-fit model, provide a more accurate and predictive representation of the MPR–SIR relationship, offering deeper insights into CAV safety evolution across low-, mid-, and high-MPR scenarios.

Ultimately, this meta-analysis provides a transparent, bias-adjusted synthesis of the safety impacts of CAVs, reinforcing that higher market penetration levels correspond with substantial safety improvements. The findings contribute valuable empirical evidence for researchers and policymakers, emphasizing that effective CAV deployment strategies must consider not only technological readiness but also realistic adoption trajectories, infrastructure preparedness, and human factors. Continued research integrating real-world crash data, broader geographical coverage, and evolving vehicle technologies will be essential to fully capture the long-term safety potential of connected and automated mobility.

Abbreviations

CAVs	Connected and Autonomous Vehicles
FSL	Fixed Speed Limits
MPR	Market Penetration Rate
PICO	Population, Intervention, Comparison, Outcome
PET	Post Encroachment Time
SE	Standard Error
SIR	Safety Improvement Rate
SSM	Surrogate Safety Measures
TTC	Time-to-Collision
TIT	Time Integrated Time-to-Collision
TET	Time Exposed Time-to-Collision
VSL	Variable Speed Limits

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12544-026-00774-9.

Supplementary Material 1

Author contributions

Amirhossein Taheri: Conceptualization, Methodology, Data Curation, Formal analysis, Investigation, Visualization, Writing- original draft. Jing Yang: Methodology, Investigation, Data Curation, Visualization. Steffen Müller: Conceptualization, Supervision. Dimitris Milakis: Methodology, Writing – Original Draft. Mohammed Quddus: Writing – review & editing.

Funding

Open Access funding enabled and organized by Projekt DEAL. This research was supported by the Elsa-Neumann Scholarship for doctoral candidates, provided by the Dahlem Research School in the state of Berlin, Germany (Elsa Neumann Scholarship of the State of Berlin • Dahlem Research School • Freie Universität Berlin), under the grant number T77011.

Data availability
The dataset used for the meta-analysis, along with the full list of references, is provided as supplemental material accompanying this article.

Declarations

Generative artificial intelligence tools
To enhance the clarity and quality of English throughout the paper, we utilized GPT-4o mini, an advanced language model optimized for concise and natural language editing.

Competing interests
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Received: 1 July 2025 / Accepted: 13 February 2026
Published online: 04 March 2026

References
1. Nikitas, A., Thomopoulos, N., & Milakis, D. (2021). The environmental and resource dimensions of automated transport: A nexus for enabling vehicle automation to support sustainable urban mobility. *Annual Review of Environment and Resources*, *46*(1), 167–192. https://doi.org/10.1146/annurev-environ-012220-024657
2. Rana, M. M., & Hossain, K. (2023). Connected and autonomous vehicles and infrastructures: A literature review. *International Journal of Pavement Research and Technology*, *16*(2), 264–284. https://doi.org/10.1007/s42947-021-00130-1
3. Hadi, M., Amine, K., Hunsanon, T., & Arafat, M. (2021). Evaluating the benefits of red-light violation warning system in a connected vehicle simulation environment. *Transportation Research Record*, *2675*(11), 1372–1381. https://doi.org/10.1177/03611981211026662
4. Milakis, D., van Arem, B., & van Wee, B. (2017). Policy and society related implications of automated driving: A review of literature and directions for future research. *Journal of Intelligent Transportation Systems*, *21*(4), 324–348. https://doi.org/10.1080/15472450.2017.1291351
5. Wang, C., Xie, Y., Huang, H., & Liu, P. (2021). A review of surrogate safety measures and their applications in connected and automated vehicles safety modeling. *Accident Analysis and Prevention*, *157*, 106157. https://doi.org/10.1016/j.aap.2021.106157
6. Xiao, G., Lee, J., Jiang, Q., Huang, H., Abdel-Aty, M., & Wang, L. (2021). Safety improvements by intelligent connected vehicle technologies: A meta-analysis considering market penetration rates. *Accident Analysis and Prevention*, *159*, 106234. https://doi.org/10.1016/j.aap.2021.106234
7. Rahman, M. S. (2019). Assessing the safety and operational benefits of connected and automated vehicles: Application on different roadways, weather, and traffic conditions. *Doctoral dissertation, University of Central Florida*. https://stars.library.ucf.edu/etd/6561/. Accessed 23 February 2024.
8. Sha, D., Gao, J., Yang, D., Zuo, F., & Ozbay, K. (2023). Calibrating stochastic traffic simulation models for safety and operational measures based on vehicle conflict distributions obained from aerial and traffic camera videos. *Accident Analysis and Prevention*, *179*, 106878. https://doi.org/10.1016/j.aap.2022.106878
9. Xu, Z., Zhang, K., Min, H., Wang, Z., Zhao, X., & Liu, P. (2018). What drives people to accept automated vehicles? Findings from a field experiment. *Transportation Research Part C: Emerging Technologies*, *95*, 320–334. https://doi.org/10.1016/j.trc.2018.07.024
10. Nordhoff, S., Kyriakidis, M., van Arem, B., & Happee, R. (2019). A multi-level model on automated vehicle acceptance (MAVA): A review-based study. *Theoretical Issues in Ergonomics Science*, *20*(6), 682–710. https://doi.org/10.1080/1463922X.2019.1621406
11. Zhang, T., Tao, D., Qu, X., Zhang, X., Lin, R., & Zhang, W. (2019). The roles of initial trust and perceived risk in public's acceptance of automated vehicles. *Transportation Research Part C: Emerging Technologies*, *98*, 207–220. https://doi.org/10.1016/j.trc.2018.11.018
12. Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2010). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *International Journal of Surgery*, *8*(5), 336–341. https://doi.org/10.1016/j.ijsu.2010.02.007
13. Sohrabi, C., Franchi, T., Mathew, G., Kerwan, A., Nicola, M., & Griffin, M. … PRISMA 2020 statement: What's new and the importance of reporting guidelines. *International Journal of Surgery*, *88*, 105918. https://doi.org/10.1016/j.ijsu.2021.105918
14. Cramond, F., O'Mara-Eves, A., Doran-Constant, L., Rice, A. S. C., Macleod, M., & Thomas, J. (2018). The development and evaluation of an online application to assist in the extraction of data from graphs for use in systematic reviews. *Wellcome Open Research*, *3*, 157. https://doi.org/10.12688/wellcomeopenres.14738.2
15. Cooper, H. M., Hedges, L. V., & Valentine, J. C. (2009). *The handbook of research synthesis and meta-analysis*, 2nd Ed.,1-615. New York, NY: Russell Sage Foundation.
16. Huedo-Medina, T. B., Sánchez-Meca, J., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I2 index? *Psychological Methods*, *11*(2), 193–206. https://doi.org/10.1037/1082-989x.11.2.193
17. Veroniki, A. A., Jackson, D., Viechtbauer, W., Bender, R., Bowden, J., & Knapp, G. … Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Research Synthesis Methods*, *7*(1), 55–79. https://doi.org/10.1002/jrsm.1164
18. Guolo, A., & Varin, C. (2015). Random-effects meta-analysis: the number of studies matters. *Statistical Methods in Medical Research*, *26*(3), 1500–1518. https://doi.org/10.1177/0962280215583568
19. Lin, L., Chu, H., Murad, M. H., Hong, C., Qu, Z., Cole, S. R., & Chen, Y. (2018). Empirical comparison of publication bias tests in meta-analysis. *Journal of General Internal Medicine*, *33*(8), 1260–1267. https://doi.org/10.1007/s11606-018-4425-7
20. Duval, S., & Tweedie, R. (2000). A nonparametric Trim and Fill method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, *95*(449), 89–98. https://doi.org/10.1080/01621459.2000.10473905
21. Chen, P., Ni, H., Wang, L., Yu, G., & Sun, J. (2024). Safety performance evaluation of freeway merging areas under autonomous vehicles environment using a co-simulation platform. *Accident Analysis and Prevention*, *199*, 107530. https://doi.org/10.1016/j.aap.2024.107530
22. Olia, A., Abdelgawad, H., Abdulhai, B., & Razavi, S. N. (2015). Assessing the potential impacts of connected vehicles: Mobility, environmental, and safety perspectives. *Journal of Intelligent Transportation Systems*, *20*(3), 229–243. https://doi.org/10.1080/15472450.2015.1062728
23. Karbasi, A. H., Zheng, N., & O'Hern, S. (2024). Investigating the impact of automated vehicles on the safety and operation of low-speed urban networks. *Journal of Advanced Transportation*, *2024*(1), 1–21. https://doi.org/10.1155/2024/3835831
24. Arvin, R., Khattak, A. J., Kamrani, M., & Rio-Torres, J. (2020). Safety evaluation of connected and automated vehicles in mixed traffic with conventional vehicles at intersections. *Journal of Intelligent Transportation Systems*, *25*(2), 170–187. https://doi.org/10.1080/15472450.2020.1834392
25. Abdel-Aty, M., Wu, Y., Saad, M., & Rahman, M. S. (2020). Safety and operational impact of connected vehicles' lane configuration on freeway facilities with managed lanes. *Accident Analysis and Prevention*, *144*, 105616. https://doi.org/10.1016/j.aap.2020.105616
26. Yang, G., Ahmed, M., & Adomah, E. (2020). An integrated microsimulation approach for safety performance assessment of the Wyoming Connected Vehicle Pilot Deployment Program. *Accident Analysis and Prevention*, *146*, 105714. https://doi.org/10.1016/j.aap.2020.105714
27. Ko, J., Kim, H., Oh, C., & Kim, S. (2023). Impact of V2V warning information on traffic stream performance using microscopic simulation based on real-world connected vehicle driving behavior. *IEEE Transactions on Intelligent Transportation Systems*, *24*(11), 12241–12254. https://doi.org/10.1109/tits.2023.3287308
28. Rahman, M. H., Abdel-Aty, M., & Wu, Y. (2021). A multi-vehicle communication system to assess the safety and mobility of connected and automated vehicles. *Transportation Research Part C: Emerging Technologies*, *124*, 102887. https://doi.org/10.1016/j.trc.2020.102887
29. Adebisi, A., & Ash, J. E. (2023). Traffic safety performance evaluation in a connected vehicle environment with queue warning and speed harmonization applications. *Journal of Transportation Safety & Security*, *16*(8), 914–944. https://doi.org/10.1080/19439962.2023.2273545
30. Yue, L., Abdel-Aty, M., Wu, Y., & Wang, L. (2018). Assessment of the safety benefits of vehicles' advanced driver assistance, connectivity and low level automation systems. *Accident Analysis and Prevention*, *117*, 55–64. https://doi.org/10.1016/j.aap.2018.04.002

31. Mousavi, S. M., Osman, O. A., Lord, D., Dixon, K. K., & Dadashova, B. (2021). Investigating the safety and operational benefits of mixed traffic environments with different automated vehicle market penetration rates in the proximity of a driveway on an urban arterial. *Accident Analysis and Prevention*, *152*, 105982. https://doi.org/10.1016/j.aap.2021.105982
32. Marques, D., Bandeira, J., & Coelho, M. C. (2021). Traffic, emissions, and safety impacts of automated vehicles connecting a university and a science park. *2021 7th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, Heraklion, Greece, June 16–17, IEEE, 1–6. https://doi.org/10.1109/MT-ITS49943.2021.9529306
33. Jin, J., Huang, H., Li, Y., Zhang, G., Dong, Y., Zhou, B., & Xue, H. (2023). Variable speed limit modelling to improve traffic safety and efficiency of mixed traffic flow by a two-stage framework. *Transportmetrica A: Transport Science*, 1–25. https://doi.org/10.1080/23249935.2023.2253476
34. Matin, A., & Dia, H. (2023). Impacts of connected and automated vehicles on road safety and efficiency: A systematic literature review. *IEEE Transactions on Intelligent Transportation Systems*, *24*(3), 2705–2736. https://doi.org/10.1109/tits.2022.3227176
35. Pan, Y., Wu, Y., Xu, L., Xia, C., & Olson, D. L. (2024). The impacts of connected autonomous vehicles on mixed traffic flow: A comprehensive review. *Physica A: Statistical Mechanics and Its Applications*, *635*, 129454. https://doi.org/10.1016/j.physa.2023.129454
36. Matin, A., & Dia, H. (2024). Public perception of connected and automated vehicles: Benefits, concerns, and barriers from an Australian perspective. *Journal of Intelligent Connected Vehicles*, *7*(2), 108–128. https://doi.org/10.26599/jicv.2023.9210028

Publisher’s note