


Size matters

The use and misuse of statistical significance in discrete choice models in the transportation academic literature

Journal Article**Author(s):**

Parady, Giancarlos; Axhausen, Kay W. 

Publication date:

2024-12

Permanent link:

<https://doi.org/10.3929/ethz-b-000636311>

Rights / license:

[Creative Commons Attribution 4.0 International](#)

Originally published in:

Transportation 51(6), <https://doi.org/10.1007/s11116-023-10423-y>



Size matters: the use and misuse of statistical significance in discrete choice models in the transportation academic literature

Giancarlo Parady¹ · Kay W. Axhausen²

Accepted: 2 September 2023 / Published online: 28 September 2023
© The Author(s) 2023

Abstract

In this paper we review the academic transportation literature published between 2014 and 2018 to evaluate where the field stands regarding the use and misuse of statistical significance in empirical analysis, with a focus on discrete choice models. Our results show that 39% of studies explained model results exclusively based on the sign of the coefficient, 67% of studies did not distinguish statistical significance from economic, policy or scientific significance in their conclusions, and none of the reviewed studies considered the statistical power of the tests. Based on these results we put forth a set of recommendations aimed at shifting the focus away from statistical significance towards proper and comprehensive assessment of effect magnitudes and other policy relevant quantities.

Keywords Discrete choice models · Effect size · Statistical significance · Statistical power · Policy implications · Practice implications

Introduction

Generally speaking, the purpose of academic transportation research is to better understand transport-related human behavior to better inform transportation policy design and implementation (Parady et al. 2021), and with the advent of cheap computing power, statistical models have become a key tool to help explain transport-related phenomena. However, along with the widespread use of statistical models across many fields relying on quantitative analysis, came its widespread “cookbook” use, where statistical significance is used to define the importance, or lack thereof, of any given variable, regardless of its practical importance. The seminal work of McCloskey and Ziliak (1996) shed some light on

✉ Giancarlo Parady
gtrncoso@ut.t.u-tokyo.ac.jp
Kay W. Axhausen
axhausen@ivt.baug.ethz.ch

¹ Department of Urban Engineering, The University of Tokyo, Bunkyo-Ku, Tokyo-to 113-8656, Japan

² IVT, ETH Zürich, 8093 Zurich, Switzerland

consistent errors in the use of “statistical significance” in the field of economics. But this is not by any means exclusive to the economics field, and while we suspect the transportation field is no stranger to such errors, this has yet to be systematically evaluated, which motivates this work. In this paper we adapt McCloskey and Ziliak (1996)’s nineteen questions to fifteen questions relevant to the academic transportation literature to evaluate where the field stands regarding the use and misuse of statistical significance in empirical analyses, with a focus on discrete choice models. This study complements the study of Parady et al. (2021) on model validation in that it helps provide a cross-sectional overview on how statistical models are used in the field, ultimately aiming to promote better modeling practices.

McCloskey and Ziliak’s key findings

As an unbiased selection of best practice in economics, McCloskey and Ziliak (1996) reviewed all full-length papers published in the *American Economic Review* in the 1980s that used regression analysis. For the 182 papers selected, they asked nineteen questions about the use of statistical significance, and recorded answers as “yes” for sound statistical practice, “no” for unsound practice, or “not applicable.” Their key findings are summarized below:

1. 70% of reviewed studies did not distinguish statistical significance from economic, policy or scientific significance.
2. One third of studies used only the *t*- and *F*-statistics as criteria for variable inclusion in the analysis of the paper.
3. 72% of studies did not discuss the scientific conversation within which the magnitude of a coefficient can be judged to be “large” or “small.”
4. 59% of studies ambiguously used the word “significant” to mean statistically different from the null sometimes and to mean practically important at other times.
5. 32% explicitly stated to have used statistical significance as an exclusive criterion to drop variables from a model.
6. Only 4% of studies considered the statistical power of their tests.
7. 69% of studies did not report descriptive statistics of variables used in the model.

Evaluating the use of statistical significance in the transportation academic literature

Article selection criteria

To get a more comprehensive view of the current state of affairs, we extended the scope of the review to the whole field, rather than a “selection of best practices” in the field. We based our selection criteria on Parady et al. (2021) so that the results presented here can complement their findings on validation practices and provide a cross-sectional overview on how statistical models are used in the field. A key difference, however, is that we did not exclude papers using stated preference surveys from the review. Using the Web of Science Core Collection maintained by Clarivate Analytics we reviewed discrete choice

model reporting practices in the academic transportation literature published between 2014 and 2018.¹ Articles were selected based on the following criteria:

1. Peer-reviewed journal articles published between 2014 and 2018.
2. Analysis uses discrete choice models.
3. Target choice dimensions are:
 - (a) Destination choice
 - (b) Mode choice
 - (c) Route choice
4. Articles that analyze other choice dimensions are considered if and only if the article includes at least one of the three target choice dimensions defined in 3.
5. Web of Science Database search keywords are:
 - (a) Destination choice model
 - (b) Mode choice model
 - (c) Route choice model
6. Web of Science Database fields are:
 - (a) Transportation
 - (b) Transportation science and technology
 - (c) Economics
 - (d) Civil engineering
7. Research scope is limited to human land transport and daily travel behavior (tourism, evacuation behavior, and freight transport articles were excluded).
8. Studies using numerical simulation only were excluded.
9. Methodological papers were only included if they used empirical data. In addition, as a new criterion for this study, papers must have a clear focus on policy or policy related variables (papers whose stated contribution is exclusively methodological were excluded).
10. For route choice models, Stochastic user equilibrium (SUE) models are excluded.

The final number of articles reviewed was 95 articles,² selected randomly out of a total of 283 articles. This is equivalent to 34% of all articles matching the inclusion criteria.

The fifteen questions for the transportation field

While this study is based on McCloskey and Ziliak (1996), we have adapted the questions to reflect the idiosyncrasies of the transportation field. In addition, for some questions, we expanded the yes–no dichotomy to include partially satisfactory practices. While whenever

¹ Data was collected in early 2019, hence the review scope spanned from 2014 to 2018.

² See Appendix 3 for an explanation on why this number of papers was used.

possible we tried to align our scoring criteria with their criteria, the way some questions were operationalized was not clearly described in their paper, so it is possible that there are differences in scoring criteria and methods. We advise the reader to keep this in mind as we compare both studies in the next subsections.

To adapt the questions to the field, we started with an a-priori set of questions adapted based on the authors experience, and pre-tested them on a random subset of the eligible papers ($N=29$, not included in main review), iteratively modifying the questionnaire until reaching a satisfactory set of fifteen questions, that is, a set of questions that do not overlap too much with other questions, questions that are not ambiguous in meaning and questions that are appropriate to the context of transportation research. Moving forward, when referring to McCloskey and Ziliak's nineteen questions we use the nomenclature MZ-1, ..., MZ-19.

The fifteen questions for the transportation field are:

Q1. Does the paper report descriptive statistics and units for model variables?

This question is equivalent to MZ-2, and points to the fact that knowledge of variable units and basic descriptive statistics (at least, means for continuous variables and relative frequency for categorical variables) is crucial to properly interpret model results. We have extended the answers to "yes, largely," "yes, partially" and "not at all."

Q2. Are estimated coefficients used to calculate elasticities, marginal effects or some other quantity of interest that addresses the question of "how large is large"?

This is equivalent to MZ-3, but in discrete choice modeling this is even more critical given that, as opposed to the linear regression case, coefficient estimates are not directly interpretable. In transportation, elasticities, marginal effects, and to a lesser extent odd ratios are usually reported. Other quantities of interest usually reported are marginal rates of substitution, in particular, the value of travel time. While such quantities are not strictly speaking measures of effect size, they are quantities of policy importance, hence within the scope of this and subsequent questions.

Note that amount of explained variance is often referred to as a measure of effect size. While we agree that this measure does convey an idea of a set of variables or a model's explanatory power, such measure does not have a direct policy interpretation in terms of magnitude in the way an elasticity or a marginal effect does. In that regard it is not relevant here or in subsequent questions.

We classified papers into three categories: "yes, in a comprehensive manner," for such cases where effects sizes or other quantities of interest are reported for (i) most variables in the paper or (ii) those variables the authors have explicitly identified as important in the objectives or hypotheses statements of the paper; "yes, partially," for the cases when these are only reported for one or a few variables, but not necessarily covering all the key variables; and "not at all."

Since reporting of coefficients is a convention of the field, this question ignores whether or not coefficients are reported in the first place. That is, a paper that reports measures of effect magnitudes for all variables in addition to the model coefficients, and a paper that exclusively reports measures of magnitude for all variables, will both be classified as "yes, in a comprehensive manner."

Q3. Does the paper report all standard errors, t-statistics, and goodness of fit statistics like the likelihood ratio test, and rho-square?

This question is an alternative to MZ-6, which asks if "the paper eschews reporting all *t*- or *F*-statistics or standard errors, regardless of whether a significance test is appropriate." Since it is a convention in the field that null hypotheses are implied based on the objectives of the study, full model results including coefficients, *t*-statistics,

and goodness of fit statistics are usually reported, and in our experience demanded by reviewers if not. We reformulated the question to evaluate to what extent the model reporting convention is met (coefficients, significance statistics, and goodness of fit), with the explicit understanding that coefficients and significance test statistics should not be given a primary position in a paper over the results directly related to effect magnitudes and policy discussions.

Q4. Does the paper mention the power of the tests?

This is equivalent to MZ-8, referring to whether the authors make any reference to the statistical power of a test.

Q5. If so, does it evaluate the power function?

This is equivalent to MZ-9 and asks whether or not the authors evaluated the power of any test in the paper.

Q6. Does the paper eschew "asterisk econometrics," the ranking of coefficients according to the absolute size of the test statistic?

This is equivalent to MZ-10. This practice is not conventional in the field, so we would expect a priori to not happen very often.

Q7. In the model results section, does the paper eschew "sign econometrics," remarking on the sign but not the magnitude of the effect?

This is equivalent to MZ-11. It refers to the practice of describing models based on the sign of the coefficient (usually in addition to the size of the t-statistic) without considering the magnitude of the effect in question, and whether such effect is large enough to matter in practical terms. We explicitly limited the scope of this question to the section where the model results are first introduced to account for the fact that it is plausible a researcher completely discusses a model in terms of sign econometrics (a practice we discourage while acknowledging the role conventions play in perpetuating this) but then proceed to conduct some separate analysis that does give some idea of magnitude. This includes remarking on quantities derived from coefficients such as the value of travel time.

We classified papers into three categories: "yes, comprehensively" for cases where "sign econometrics" are eschewed for most variables in the paper, "yes, partially" when this is the case for one or a few variables only, but not necessarily covering all the key variables; and "not at all."

Q8. Does the paper discuss the magnitude of estimated effects or other quantities of interest?

Q9. Does the paper make a judgement on magnitudes, making the point that some effects or quantities of interest are practically influential or important and some are not?

These questions relate to MZ-12 and focus on whether the paper makes the point of the practical importance (as opposed to statistical significance) of observed effects. While in their original questionnaire they refer specifically to coefficients, as we noted earlier, coefficients are not directly interpretable in discrete choice models, so we expanded the scope of this question to cover any form of analysis that focuses on magnitude, including simulations. Furthermore, in the pre-test phase we decided to split this question in two parts. Q8 refers to whether or not there was a discussion of magnitude, either by discussing elasticities, marginal effects, marginal rates of substitution, etc., or by conducting simulation analyses. Such discussions include the interpretation of estimated magnitudes and their relative comparisons. Q9, on the other hand, explicitly asks whether the authors make a judgement on the magnitudes observed, that is, whether they explicitly judge an effect or quantity of interest to be "large," "medium," "small" or "practically important" or "practically negligible" based on some criteria. This distinction is important, because the largest effect among a group of effects might still be small in practical terms. And while it might seem trivial at

first, such judgement of magnitude is an important part of a quantitative study, and it is not necessarily that straightforward a task.

For these two questions, we classified papers into three categories: "yes, comprehensively" for such ideal cases where they do so for (i) most variables in the paper or (ii) those variables the authors have explicitly identified as important in the objectives or hypotheses statements of the paper; "yes, in a limited manner" when they do so for one or a few variables only, but not necessarily covering all the key variables, and "not at all."

Q10. Does the paper discuss the scientific conversation within which an effect or other quantity of interest would be judged large or small?

This is equivalent to MZ-13. It asks whether the author compares her own findings against previous studies in the literature or commonly accepted values in the field. We marked this question as "yes" if at least one quantity of interest is compared against the literature.

Q11. Does the paper avoid choosing variables for inclusion solely on the basis of statistical significance?

This is equivalent to MZ-14. It asks whether authors explicitly state dropping variables from a model based exclusively on statistical significance, disregarding the magnitude of the effect. Papers using stepwise variable selection methods are also marked as "no." As in McCloskey and Ziliak (1996), papers are only marked "no" when authors state so explicitly, so this can be thought of as a lower bound for this criterion.

Q12. Does the paper do a simulation to determine whether the estimated effects or other quantities of interest are reasonable and/or to better illustrate the magnitude of estimated effects?

This is equivalent to MZ-17. It also includes policy simulations. Note that this is different from the case where simulation is necessary to estimate magnitudes, for example marginal effect of dummy variables, or elasticities in open-form models, although we recognize in some instances this difference might be blurry.

Q13. In the conclusion and implications sections, is statistical significance kept separate from economic policy and scientific significance?

This is equivalent to MZ-18. For example, papers that conclude summarizing variables that were found to be statistically significant and even proceed to infer policy suggestions from these "findings" are marked as "no" as they are mixing statistical significance with practical importance. The scope of this question is limited to the part of the conclusion and implications sections that refer directly to model results.

Q14. In the estimation, conclusion and implications sections, does the paper avoid using the word "significance" in ambiguous ways, meaning "statistically significant" in one sentence and "large enough to matter for policy or science" in another?

This is equivalent to MZ-19. We limited the scope of this question to the estimation, conclusion, and implications sections.

Q15. Does the article report confidence intervals of effect sizes, using them to interpret practical importance and not merely as a replacement for pointwise statistical significance?

This is technically not part of the nineteen original questions, but an additional question McCloskey and Ziliak wished they could have added (Ziliak and McCloskey 2007). Note that in this question we refer to the confidence interval of a measure of effect magnitude such as elasticities or marginal effects, not of coefficients. We classified papers into three categories: "yes, in a comprehensive manner," "yes, in a limited manner," and "not reported for any variable." In this case, "yes, in a limited manner" refers to the case where confidence intervals are reported but not used in the discussion.

Questions eliminated from the questionnaire.

Readers will note that we have indeed eliminated some questions from the original questionnaire. MZ-1 refers to the use of a small number of observations such that statistically significant differences are not just a result of large sample size. To avoid the issue of what a “small number” of observations is, we excluded this question. Instead, we report the distribution of the minimum sample sizes used in the reviewed studies (see Sect. “The issue of power”).

MZ-4 asks if the proper null hypotheses are specified, arguing that the most common mistake is testing against a null of zero when another null is of interest. However, in transportation, it is common that null hypotheses are implied based on the objectives of the study and are not stated explicitly. For example, Khan et al. (2014) state as their objective “to evaluate the effects of built environment variables on the use of non-motorized travel modes.” As such, a priori hypotheses of parameter sizes, which are not directly interpretable, are very rarely specified, with the clear exception of theoretically defined parameters such as the scale parameter in the nested logit model. As such, such a question would not be very informative and hence removed.

MZ-5 on whether coefficients are carefully interpreted, was removed since it largely overlaps with Q2 and Q8. MZ-7 on using statistical significance as the only criteria of importance at its first use, and MZ-15 on avoiding using statistical significance as the only important criteria after the “crescendo” were removed because such issues could be covered to a large extent by Q7~Q10. Finally, MZ-16 on whether statistical significance was decisive, and conveyed the sense of an ending, was deleted because it was rather ambiguous, hard to operationalize, and its content could also be covered with Q7~Q10.

Main findings

The results of the review are summarized in Table 1. The key findings are listed below, with the values in parenthesis summarizing the values reported by McCloskey and Ziliak (1996) for reference purposes.

1. 67% (MZ:70%) of reviewed studies did not distinguish statistical significance from economic, policy or scientific significance.
2. 86% (MZ: 72%) of studies did not discuss the scientific conversation within which the magnitude of a coefficient can be judged to be “large” or “small.”
3. 62% (MZ: 59%) of studies ambiguously used the word “significant” to mean statistically different from the null sometimes and to mean practically important at other times.
4. 39% (MZ: 53%) explained model results exclusively based on the sign of the coefficient.
5. 24% (MZ: 32%) explicitly stated to have used statistical significance as an exclusive criterion to drop variables from a model.
6. 0% (MZ: 4%) of the reviewed studies considered the statistical power of the tests.
7. 0% of the reviewed studies reported confidence intervals (of effect magnitudes) and used them to interpret economic or policy significance. 7% did however report these intervals but did not explicitly use them in the discussion.

Table 1 Answer to the fifteen questions in the transportation field measured in percentage

Does the article	Yes	Yes, comprehensively /largely	Yes, limitedly/ partially	Applicable articles
Q4: Mention the power of the tests?	0.00%	–	–	94
Q5: Examine the power function?	–	–	–	0
Q15: Report effect confidence intervals, using them to interpret economic significance not merely as a replacement for pointwise statistical significance?	7.37%	0%	7.37%	95
Q10: Discuss the scientific conversation within which an effect or other quantity of interest would be judged large or small?	13.68%	–	–	95
Q12: Do a simulation to determine whether the estimated effects or other quantities of interest are reasonable and/or to better illustrate the magnitude of estimated effects?	29.47%	–	–	95
Q13: In the conclusions and implications sections, keep statistical significance separate from economic policy and scientific significance?	32.63%	–	–	95
Q9: Make a judgement on effect magnitudes?	36.84%	13.68%	23.16%	95
Q14: In the estimation, conclusions, and implications sections, avoid using the word "significance" in ambiguous ways?	37.63%	–	–	93
Q7: In the model results section, eschew "sign econometrics"?	60.64%	27.66%	32.98%	94
Q8: Discuss the magnitude of estimated effects or other quantities of interest?	64.21%	33.68%	30.53%	95
Q2: Use coefficients to calculate elasticities, or some other quantity that addresses the question of "how large is large"?	65.26%	45.26%	20.00%	95
Q11: Avoid choosing variables for inclusion solely on the basis of statistical significance?	75.53%	–	–	94
Q3: Report all traditionally reported statistics?	76.84%	–	–	95
Q1: Report descriptive statistics for model variables?	78.95%	65.26%	13.68%	95
Q6: Eschew "asterisk econometrics"?	100.00%	–	–	94

For Q14, papers that do not mention the word "significant" are marked as NA

For Q4, Q6, Q7 and Q11 papers that did not directly show model results (i.e., reported in a separate article) are marked as NA

Size matters: on the reporting and discussing of effect magnitude.

Focusing specifically on the questions concerning effect size, first we want to highlight two very positive findings. The first is the complete eschewing of “asterisk econometrics,” that is, the ranking of coefficients according to the absolute size of the test statistic. The second is the high levels of reporting of descriptive statistics. 79% (MZ: 31%) of studies did so, 65% did so extensively, and 14% did so only partially. As we discussed earlier, such information is necessary for properly interpreting model results. 77% of studies also reported all traditionally reported model statistics, that is, coefficients, measures of significance, and goodness of fit. However, while we do not oppose such reporting per se, we want to underscore this should not be the highpoint of the paper, as it often is. If anything, it should be supplementary material the reader can consult if required, thus moving the paper towards matters of practical importance. In that regard, 65% (MZ: 67%) of studies used coefficients to calculate elasticities, or some other quantity of interest that conveys magnitude. 45% did so extensively (either for the majority of the model variables, or the variables the authors explicitly stated being key to the study) while 20% did so only partially. In line with this, 64% (MZ: 80.2%) of studies explicitly discussed the magnitude of estimated effects or other quantities of interest. 34% did so extensively, 31% did so only partially (difference due to rounding error). That is, in addition to reporting magnitudes, the authors make the decision to discuss the model results in terms of such magnitudes. For example, exploring the impacts of walk and bike infrastructure on mode choice, Aziz et al. (2018) report in a very orthodox manner that “*the direct elasticity value indicates that 1% increase in the total bike lane proportion (normalized by area) in the home and work census tracts will increase the probability to choose bike by 1.13%.*” Heinen and Ogilvie (2016) also discuss effect magnitudes in the context of the impacts of the introduction of a new guided busway in Cambridge, UK, stating that “*the results correspond with individuals living, for example, 4 km from the busway being from 60 to 70% more likely (depending on the indicator of variability) to have increased their active travel share [more than 20%] than those living 9 km away.*”

It is important to note that among papers classified as “yes, partially” for Q2, a large share simply reported one or several measures of value of travel time, and then reverted back to reporting only coefficient signs for all the other variables. In fact, 39% (MZ: 53%) of studies explained model results exclusively based on the sign of the coefficient, with no reference to effect magnitudes or other quantities of interest. While convention plays a role here, it is worth noting that some authors did in fact, after reporting models exclusively sign-wise, go on to conduct simulations to determine whether the estimated effects or quantities of interest are reasonable, or to evaluate policy effects. That being said, 22% of papers did not report or discuss any magnitude whatsoever. That is, the discussion was exclusively limited to sign direction and statistical significance.

Regarding the question of “how large is large?” it is worth pointing that 63% of studies failed to make a clear judgement of magnitude. We make the distinction here from the question regarding explicitly discussing estimated magnitudes (Q8) because such magnitudes were frequently discussed in relative terms, and in many cases the authors fell short of a judgement on whether the quantity of interest was “large” or “small” or “policy-relevant” or “not relevant” based on any criteria. In this regard, M. Khan et al. (2014) make clear judgements of magnitude when they state that “*network connectivity (measured as 4-way intersections within 0.5 mile) plays a major role: a single*

standard deviation change in this variable is estimated to increase walking probability by 34%” and go on to state that “parking prices and free-parking availability variables were not found to have much of an effect.” It is clear from these statements that based on the authors’ judgement, network connectivity is a variable of practical importance. Here we also want to highlight that they evaluated magnitudes by estimating percentage change in dependent variables given one standard-deviation changes in the independent variables. Such an approach is designed to overcome one key limitation of traditionally reported elasticities (defined as the percentage change in the dependent variable given one-percent change in the independent variable,) which is that a 1% change might be easier to achieve for some variables than others.

de Luca and Di Pace (2015) also make clear judgments of magnitude when they discuss value of travel time estimates and state that “aside from being similar to those estimated in different Italian case studies (Cantarella and de Luca 2005), [the magnitude] indicates the extreme importance of parking location. Assuming that the average one-way travel monetary cost is equal to 3 €, 10 min walking time (about 700 m at 4 km/h) is more than half of the whole travel monetary cost.” This is an ideal form, as it gives a clear economic interpretation of the quantity in question and a clear judgment of its magnitude. In addition, they compare estimated magnitudes to similar studies in the literature.

The questions of “how large is large?” is, however, a difficult question with no easy immediate answer, if anything, underscoring the importance of addressing it. The very concepts of “small” or “large” are difficult to characterize and might require some degree of convention. In his seminal work on power analysis, Cohen (1988) argued that “all conventions are arbitrary, one can only demand they not be unreasonable.” And while he noted that it was desirable to have, and actually proposed and characterized universal effect size measures, free of unit variability and applicable to various research issues and statistical models, he warned that “the meaning of any effect size is, in the final analysis, a function of the context in which is embedded.” Thus, addressing the question of how large is large requires a clear understanding of the scientific context of the study. On this point, 86% (MZ: 72%) of studies failed to discuss the scientific conversation within which the magnitude of an effect or other quantity of interest can be judged to be “large” or “small” by referencing to values reported in the literature of at least one variable. Allard and Moura (2018) provide this scientific context by reporting a table comparing several values of time and willingness to pay for long distance intermodal service characteristics, the object of their study.

For some variables, judgement of magnitude is not that straightforward, and even impossible to discuss in economic terms. This is especially so for latent constructs, where the meaning of unit changes or percentage changes are not clear-cut. Hess et al. (2018) address this issue in the context of latent attitude constructs, and propose that instead of arbitrarily looking at percentage changes in attitudes, pointless due to the scale of such constructs, it would be more meaningful to test what would happen if everyone’s attitudes were like those of a particular segment of the population.

The issue of power

Another issue worth highlighting is that none of the reviewed studies considered the statistical power of the tests (MZ: 4%). While a comprehensive exposition of statistical power

		REALITY	
		There is no effect (null is true)	There is an effect (null is false)
DECISION BY RESEARCHER (based on model results)	There is no effect (null not rejected)	True negative probability = $1 - \alpha$	False negative · Type II error probability = β
	There is an effect (null rejected)	False positive · Type I error probability = α (Statistical significance)	True positive probability = $1 - \beta$ (Statistical power)

Fig. 1 Outcomes of a statistical test and its relation to statistical significance and statistical power

is beyond the scope of this paper, we will briefly discuss the main issues with the goal of sparking a well-deserved debate on the matter.

As illustrated in Fig. 1, the statistical power of a test gives the probability that a statistical test will correctly identify an effect that exists in reality (that is, a *true positive*). In statistical jargon, it is the conditional probability that the test will reject the null hypothesis when the null is *actually false*. It can also be defined as the probability of avoiding Type II errors (*false negatives*).

Statistical power is a function of sample size, statistical significance and more importantly, effect size. That is, given a statistical significance level α , the smaller the effect, the larger the sample required to detect it. Its most common uses are to evaluate the power a statistical test had on a completed study, and to calculate necessary sample sizes given anticipated effect sizes and power (Cohen 1988). In other words, it is used to answer two questions: (a) assuming that the effect we are looking for actually exists and has magnitude m , for sample size n , what is the probability we will detect such effect (i.e., correctly reject the null) at significance level α ? And (b) what sample size do we need in our study to identify an effect of magnitude m , at significance level α with power level $1 - \beta$?

To give a concrete example, a statistical test with power of 0.2, means that the researchers will accept the null hypothesis, when the null is false, four out of five times. In the words of Ziliak and McCloskey (2007), “power puts a check on the naïveté of the gullible.” This is particularly critical for “small” effects which will require larger sample sizes to be detected.

In the case of multivariate modeling, where multiple statistical tests are conducted, Maxwell (2004) showed in his analysis of underpowered studies in psychology, that required samples sizes to detect at any given power level (a) any single prespecified effect, (b) at least one effect, or (c) all effects, differ. For example, he showed that for a multivariate linear regression with five predictors³ while the power of correctly detecting at least one effect (that is, the probability of a true positive) was > 0.99 , the power for correctly detecting all effects was only 0.22. Maxwell argued that given

³ Assuming correlation between each predictor with other predictors and with the outcome variable = 0.3, $n = 400$ and $\alpha = 0.05$.

that review studies continue to show lack of power in the literature, tests of individual hypotheses often lack sufficient power, even when adequate power exists for detecting at least one effect somewhere in the collection of tests. In that regard, while in the transportation field, at the usual sample sizes of large-scale household travel surveys, power issues might not be much of a problem, in the studies we reviewed, the median sample size (when several samples are reported, the minimum was used) was 1,404 (choice situations) and the 20th percentile was 527. Although we do not make claims of the applicability of Maxwell's findings to discrete choice models, his results do underscore the need to address such concerns in the transportation field. While many power studies have been conducted in fields like psychology (Cohen 1962; Rossi 1990) and education (Brewer 1972), to the best of our knowledge, such analyses have not been conducted in the transportation field, so the state of affairs is not known. It is important, however, to note that the research in the psychology field rests on the extensive work on statistical power by Cohen (1962, 1988) who defined "scale-free" conventions to characterize "small," "medium," and "large" effects, and while Cohen largely focused on t-test for means, correlation coefficients, proportional differences, and linear regression, the literature on statistical power for discrete choice models is scarce (Chen et al. 2010; de Bekker-Grob et al. 2015). Finally, regarding sample size determination, it must be pointed out that while there is a comprehensive literature dealing with sample size for discrete choice experiments (Rose et al. 2008; Rose and Bliemer 2013) existing theory largely ignores the issue of minimum sample size requirements in terms of power (de Bekker-Grob et al. 2015).

Conflating statistical significance with practical importance

Another issue worth pointing out is the conflation of statistical significance and practical importance. This is evident in the above stated fact that 22% of papers did not report or discuss any magnitude whatsoever. This means that one in five reviewed papers completely defined the importance of their findings based on statistical significance. 24% (MZ: 32%) of studies explicitly dropped variables solely based on statistical significance, and 62% (MZ: 59%) of studies used the term "significant" either conflating it with practical importance or in a way the reader cannot discern which meaning the author is pointing to. In some cases, the interpretation of the size of the t-statistic was misinterpreted as a measure of effect size. In a study of social interaction effects on the decision-making process, the study's authors state of a latent construct of walking preference that "*this component is the most statistically significant variable...indicating the strong influence that parents have on the development of their children's attitudes towards walking,*" misinterpreting a large t-statistic with a strong influence on outcome. Similarly, in a study of mode-shifting behavior, the study's authors argue that "*bus service level has the most significant positive t-value, which indicates that improving the bus service level can increase the shifting proportion of car travelers to bus significantly.*" Here it is also unclear whether or not the "significant" increase in modal shift proportion is meant to signify a considerable increase in practical terms, or just a difference statistically different from zero.

Finally, 67% of studies mixed statistical significance and practical importance in the discussion and conclusion sections, the most widespread practice being reporting a set of variables as important, based on statistical significance, without having properly discussed these in terms of their practical importance, that is, their effect magnitudes.

Do top journals do any better?

To answer this question, we developed a simple score to get an idea of the overall performance of the reviewed papers. We scored each “Yes, comprehensively/largely” as 1, each “Yes, limitedly/partially” as 0.5 and each “No” as 0, averaged it over the total number of valid questions for each paper and normalized it to 100. Q3 was excluded from this score as we do not believe that mechanically reporting all traditionally reported statistics necessarily implies good practice.

After calculating the mean and standard deviation of scores for articles in top journals (defined as first quartile journals in the Scimago Journal Ranking in the field of transportation) we found that articles in top journals scored marginally better (mean score: 44.4, standard deviation: 40, N: 69) than non-top journals (mean score: 42.6, standard deviation: 39.3, N: 26) but the difference was surprisingly small in practical terms and, for what is worth, not statistically significant.

Recommendations to the field

Based on the discussion above, we put forth a set of recommendations aimed at shifting the focus away from statistical significance towards proper and comprehensive assessment of effect magnitudes and policy relevant matters.

Make reporting of effect magnitudes and their confidence intervals mandatory.

Statistical significance should not be more than one of many criteria of evaluation, but it should certainly not be the most important one. The discussion of statistical models should focus on effect magnitude and other policy relevant quantities. In that regard, confidence intervals of effect sizes (not of coefficients) give a clearer image of effect magnitude and the levels of uncertainty surrounding the estimates, and does without what Maxwell (2004) calls that “*air of finality*” that the presence or not of asterisks next to coefficient estimates tend to convey. For elasticities and marginal effects, confidence intervals are usually estimated via bootstrapping (Parady et al. 2023), or via the Delta-method. It is important to note however, that reporting just average effects might obscure differences in effects among different subpopulations. As such, authors should also strive to report effect magnitudes for the sub-populations of interest in their study, as defined by the research questions they seek to answer. Note that popular discrete choice modeling tools such as Apollo for R, Biogeme for Python, and NLOGIT already provide tools to easily estimate marginal effects, elasticities, and conduct simulations.

Also note that we do not oppose reporting model coefficients, and acknowledge their usefulness should a researcher want, for example, to use a particular model for simulation of individual outcomes (i.e., agent-based simulations). However, notwithstanding this merit, due to their lack of direct interpretability, these should be relegated to secondary position in the paper, or even an appendix, where interested readers can consult if necessary.

Provide to the extent possible judgements of magnitude that convey what the authors consider are “small,” “medium,” or “large” effects (or other quantities of interest) and the basis for such judgement.

While we acknowledge this is certainly not an easy task, there is a discussion to be had regarding what effects or quantities are policy relevant and how to assess such relevance. Furthermore, such discussions should ideally be accompanied by a discussion on the cost implications of changing the policy variables in question. As discussed earlier in the context of elasticities reporting, while the effect of a 1% increase in a policy variable might be practically larger for some variables than others, the costs associated with that 1% increase might be higher as well. As such, an explicit discussion of the cost implications of such increases, while rarely conducted, is of high importance to policy making and should be actively encouraged.

Compare, whenever possible, effect magnitudes or other quantities of interest to existing studies.

For the most regularly reported values, such as value of travel time, there is a myriad of studies reporting such values for many contexts (Abrantes and Wardman 2011; Axhausen and Fröhlich 2012; Kato et al. 2011), so there are no reasons why such comparisons cannot be made. For less often reported values, given the irregularities in reporting discussed above and differences in variable definition and measurement there will be certainly times when such a task will be difficult, but should magnitude reporting become mandatory and authors strive to provide judgements on such magnitudes, in time, proper discussion of scientific context should be widespread, thus catalyzing a virtuous cycle of proper reporting practices and discussion of magnitudes.

For new studies, take statistical power into consideration when defining sample size to guarantee the effects the researcher wants to detect can in fact be detected with enough power. For studies using secondary data (i.e., national household survey data, etc.) report post-hoc power levels of tests reported in the study.

Certainly, the literature on this issue is rather scarce, but the work of de Bekker-Grob et al. (2015) should be a starting point.

In short, we recommend researchers to submit their articles to the fifteen questions we formulate here, in order to identify potential areas where a particular draft can be improved. In addition, a cross-sectional review and evaluation as the one presented in this article should be periodically conducted to evaluate how well the field is progressing over time.

Conclusion

In this study we reviewed the academic transportation literature published between 2014 and 2018 to evaluate where the field stands regarding the use and misuse of statistical significance in empirical analysis, with a focus on discrete choice models. Our results showed repeated errors in the use of statistical significance and a lack of a clear focus on effect magnitudes for a considerable share of studies.

We want to reiterate that the ultimate objective of transportation academic research is to better inform transportation policy design and implementation, which requires proper discussion of effect magnitudes and their practical implications. In that sense, note that the purpose of this study is not to criticize the field but to point out ways it can realign itself better with this ultimate objective.

Appendix 1: Reviewed articles and scores (in the order in which they were reviewed)

No	Reference	Journal	Top journal	N	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Score	
1	Khan et al. (2014)	Transport Policy	Yes	4,741	1	1	0	0	NA	1	1	1	1	1	1	1	1	1	0	0	78.57
2	Standen et al. (2017)	Journal of Transport and Health	No	229	1	1	0	0	NA	1	1	0	0	0	0	0	0	1	0.5	0	39.29
3	Anderson et al. (2014)	EURO J Transp Logist	Yes	2,952	0	1	1	0	NA	1	0.5	1	0.5	1	1	0	1	0	0	0	57.14
4	Hess et al. (2018)	TR part A	Yes	5,413	0	1	0	0	NA	1	0.5	1	0.5	0	1	1	0	NA	0	0	46.15
5	Mehdizadeh et al. (2018)	Transportation	Yes	735	1	0	1	0	NA	1	0	0	0	0	1	0	0	1	0	0	35.71
6	Lee, (2015)	Journal of Urban Planning and Development	No	1,149	0.5	0	1	0	NA	1	0	0	0	0	1	0	0	1	0	0	32.14
7	Zhang et al. (2014)	Transportation	Yes	716	1	0	1	0	NA	1	0	0	0	0	1	0	0	0	0	0	28.57
8	Ding et al. (2014)	Journal of Urban Planning and Development	No	18,510	1	0.5	1	0	NA	1	0	0.5	0.5	1	1	1	0	0	0	0	53.57

No	Reference	Journal	Top journal	N	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Score
9	Jánošíkova et al. (2014)	Transportation Planning and Technology	No	23,808	1	1	1	0	NA	1	1	1	0.5	1	1	0	0	1	0	67.86
10	Clark et al. (2014)	Transportation	Yes	1,855	1	1	0	0	NA	1	0	0	0	0	1	0	0	0	0	28.57
11	Orozco-Fon- talvo et al. (2018)	Journal of Transport and Health	No	420	0.5	0	1	0	NA	1	0	0	0	0	1	0	0	1	0	32.14
12	Aziz et al. (2018)	Transportation	Yes	3,357	1	1	1	0	NA	1	1	0.5	0.5	0	1	0	0	0	0	50.00
13	Gokasar and Gunay, (2017)	Journal of Air Transport Management	No	410	0	0.5	1	0	NA	1	0.5	0.5	0	0	0	0	0	0	0	25.00
14	Gerber et al. (2017)	TR part A	Yes	2,167	1	0	1	0	NA	1	0	0	0	0	1	0	0	0	0	28.57
15	Yamamoto et al. (2018)	Travel behavior and society	Yes	91	0.5	0	1	0	NA	1	0	0	0	0	1	0	0	0	0	25.00
16	Hasnine et al. (2018)	J. of Transport Geography	Yes	3,208	1	0.5	1	0	NA	1	0.5	0.5	0.5	1	1	1	1	0	0	64.29
17	Paleti et al. (2017)	Journal of Choice Modelling	No	3,000	0.5	0	0	0	NA	1	0	0	0	0	0	1	0	1	0	25.00
18	Habib, (2014)	Transportation	Yes	3,003	1	0.5	1	0	NA	1	0.5	0	0	0	1	0	0	0	0	35.71

No	Reference	Journal	Top journal	N	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Score
19	Collins and MacFarlane, (2018)	TR part A	Yes	906	1	1	0	0	NA	1	1	0	0	0	0	0	0	1	0	35.71
20	Rahman and Baker, (2018)	Transport Policy	Yes	1,060	1	1	1	0	NA	1	0.5	0	0	0	1	0	0	0	0	39.29
21	Zhang et al. (2018)	Journal of Advanced Transportation	No	319	0	0	1	0	NA	1	0	0.5	0	0	0	1	0	0	0	25.00
22	Qin et al. (2017)	Transportation Planning and Technology	No	NA	1	0	1	0	NA	1	0	1	0.5	0	1	1	1	0	0	53.57
23	Qin et al. (2018)	TR part F	Yes	NA	1	0	1	0	NA	1	0	0.5	0	0	1	1	0	0	0	39.29
24	Habib and Sasic, (2014)	Journal of Choice Modelling	No	264,023	0	0.5	1	0	NA	1	0.5	0	0	0	0	0	0	0	0	21.43
25	Khoo and Asitha, (2016)	Travel behavior and society	Yes	NA	1	0.5	1	0	NA	1	0.5	0	0	0	1	0	0	0	0	35.71
26	Zhang et al. (2015)	KSCE Journal of Civil Engineering	No	313	0.5	0	0	0	NA	1	0	0	0	0	0	0	0	0	0	10.71
27	Heinen and Ogilvie, (2016)	Journal of Transport and Health	No	450	1	1	0	0	NA	1	1	1	0	0	1	0	0	1	0.5	53.57

No	Reference	Journal	Top journal	N	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Score
28	Allard and Moura, (2018)	TR part A	Yes	9,976	0	1	1	0	NA	1	1	1	0.5	1	1	0	1	0	0	60.71
29	Hasnine and Habib, (2018)	Transport Policy	Yes	1,555	1	0	1	0	NA	1	0	0.5	0	0	1	1	0	0	0	39.29
30	Assi et al. (2018)	Case Studies on Transport Policy	No	597	0	1	1	0	NA	1	1	0.5	0	0	0	0	0	1	0	39.29
31	Lee et al. (2014)	International Journal of Sustainable Transportation	Yes	6,246	1	1	1	0	NA	1	0.5	1	0	0	1	0	0	0	0	46.43
32	Habib et al. (2014)	TR part A	Yes	1,069,252	0	0.5	0	0	NA	1	0.5	0.5	0	0	1	0	0	0	0	25.00
33	Keyes and Crawford-Brown, (2018)	TR part F	Yes	1,615	0	1	0	0	NA	1	0	0	0	0	0	0	0	1	0.5	25.00
34	Rotaris and Danielis, (2015)	Transport Policy	Yes	NA	0	0	0	NA	NA	NA	NA	1	1	0	NA	1	1	0	0	40.00
35	Paulssen et al. (2014)	Transportation	Yes	519	0.5	0.5	0	0	NA	1	0.5	0.5	0.5	0	1	0	1	1	0	46.43
36	Satiennam et al. (2016)	IATSS Research	No	2,400	1	0.5	1	0	NA	1	0.5	1	0.5	0	1	1	1	0	0	60.71

No	Reference	Journal	Top journal	N	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Score
37	Hensher and Ho, (2016)	TR part E	Yes	301	1	1	1	0	NA	1	1	1	1	0	0	1	0	0	0	57.14
38	Pnevmatikou et al. (2015)	Transportation	Yes	1,038	1	0.5	1	0	NA	1	0.5	0.5	0	0	1	0	0	1	0	46.43
39	Carteni et al. (2016)	Transport Policy	Yes	4,888	0.5	1	1	0	NA	1	1	1	1	1	1	0	1	0	0	67.86
40	de Luca and Di Pace, (2015)	TR part A	Yes	NA	1	1	1	0	NA	1	1	1	1	1	1	0	1	0	0	71.43
41	Sarkar and Mal-likarjuna, (2018)	Travel behavior and society	Yes	561	1	0.5	1	0	NA	1	0	0	0	0	1	0	0	1	0	39.29
42	Khan et al. (2016)	International Journal of Sustainable Transportation	Yes	230	1	1	1	0	NA	1	0.5	1	1	0	1	1	1	0	0	67.86
43	Vij et al. (2017)	TR part A	Yes	7,860	0.5	1	0	0	NA	1	0.5	1	1	0	1	1	1	0	0	57.14
44	Ho and Hensher, (2016)	J. of Transport Geography	Yes	1,965	0	0	0	0	NA	1	0	0	0	0	0	0	0	1	0	14.29
45	Sun et al. (2018)	Journal of Transport and Health	No	764	1	1	0	0	NA	1	1	0.5	0.5	0	1	0	0	0	0	42.86

No	Reference	Journal	Top journal	N	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Score	
46	Gan and Ye, (2014)	Journal of Modern Transportation	No	1,120	1	0	1	0	NA	1	0	0	0	0	0	0	0	0	0	0	21.43
47	Mahpour et al. (2018)	TR part F	Yes	399	1	0	1	0	NA	1	0	0	0	0	1	0	0	1	0	0	35.71
48	Danapour et al. (2018)	Case Studies on Transport Policy	No	437	1	0	1	0	NA	1	0.5	0.5	0	1	0	0	0	0	0	0	35.71
49	Heinen, (2016)	TR part F	Yes	564	1	1	0	0	NA	1	1	1	0	0	1	0	1	1	1	0.5	60.71
50	Ermagun and Samimi, (2015)	Transport Policy	Yes	2,653	1	1	1	0	NA	1	1	1	0.5	1	1	0	0	1	0	0	67.86
51	Mitra and Buliung, (2014)	J. of Transport Geography	Yes	945	1	0	1	0	NA	1	0	0	0	0	1	0	0	1	0	0	35.71
52	Vij and Walker, (2014)	TR part B	Yes	60,000	0	1	1	0	NA	1	0.5	1	1	0	1	1	1	0	0	0	60.71
53	Halldórsdóttir et al. (2017)	International Journal of Sustainable Transportation	Yes	11,656	1	1	1	0	NA	1	1	1	1	1	1	0	1	0	0	0	71.43
54	Thigpen et al. (2015)	TR part D	Yes	1,480	1	0	1	0	NA	1	0	1	1	0	1	1	0	1	0	0	57.14

No	Reference	Journal	Top journal	N	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Score
55	Schoner et al. (2015)	J. of Transport Geography	Yes	614	1	0	1	0	NA	1	0	0	0	0	0	0	0	0	0	21.43
56	Liu et al. (2018)	Transport Policy	Yes	752	1	0.5	1	0	NA	1	0.5	0.5	0.5	0	0	0	1	0	0	42.86
57	Cole-Hunter et al. (2015)	TR part D	Yes	769	1	1	0	0	NA	1	0	0	0	0	1	0	0	1	0.5	39.29
58	Anta et al. (2016)	Transportation	Yes	891	0	0	1	0	NA	1	0	0.5	0	0	1	1	0	1	0	39.29
59	Hyland et al. (2018)	Travel behavior and society	Yes	NA	1	1	1	0	NA	1	1	1	0.5	0	1	1	1	0	0	67.86
60	Bueno et al. (2017)	TR part A	Yes	21,771	1	1	1	0	NA	1	1	1	1	0	0	1	0	0	0	57.14
61	Ahmad Ter-mida et al. (2016)	TR part A	Yes	2,045	1	0	1	0	NA	1	0	0	0	0	1	0	0	1	0	35.71
62	Hsu and Saphores, (2014)	Transportation	Yes	1,362	1	1	0	0	NA	1	1	0.5	0.5	0	1	0	1	0	0	50.00
63	Motoaki and Daziano, (2015)	TR part A	Yes	NA	1	1	1	0	NA	1	0.5	1	0.5	0	1	1	1	1	0.5	75.00
64	Zaidan and Abulibdeh, (2018)	Journal of Air Transport Management	No	434	0	1	1	0	NA	1	0.5	0.5	0	0	1	0	0	0	0.5	39.29

No	Reference	Journal	Top journal	N	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Score
65	Mattson et al. (2018)	Research in Transportation Economics	No	4,724	0.5	1	1	0	NA	1	0.5	0.5	0	0	1	0	0	1	0	46.43
66	Yang et al. (2016a, b)	TR part A	Yes	1,733	1	0.5	1	0	NA	1	1	1	0.5	0	1	1	1	0	0	64.29
67	Lin et al. (2018)	TR part D	Yes	304	0	1	1	0	NA	1	0.5	0.5	0	0	0	0	0	0	0	28.57
68	Efthymiou and Antoniou, (2017)	Transport Policy	Yes	600	0	0	0	0	NA	1	0	0	0	0	1	0	0	0	0	14.29
69	Di Ciommo et al. (2014)	TR part A	Yes	974	1	0	0	0	NA	1	0	0	0	0	1	0	0	0	0	21.43
70	Yang et al. (2015)	Journal of Advanced Transportation	No	1,574	1	0.5	1	0	NA	1	0.5	0.5	0.5	0	1	0	1	0	0	50.00
71	Stone et al. (2014)	Transport Policy	Yes	359	1	1	1	0	NA	1	1	0	0	0	1	0	0	1	0	50.00
72	Kamargianni et al. (2014)	Transportation	Yes	9,714	1	0	0	0	NA	1	0	0	0	0	1	0	0	0	0	21.43
73	Fernández-Antolín et al. (2016)	Journal of Choice Modelling	No	1,686	0.5	1	1	0	NA	1	1	1	0	1	1	0	1	1	0	67.86
74	Yang et al. (2016a, b)	TR part C	Yes	1,404	0	0.5	1	0	NA	1	0.5	0.5	0	0	0	1	0	0	0	32.14

No	Reference	Journal	Top journal	N	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Score
75	Mitra and Buiung, (2015)	Transport Policy	Yes	945	1	0.5	1	0	NA	1	0.5	0.5	0	0	1	0	0	0	0	39.29
76	Wang et al. (2017)	J. of Transport Geography	Yes	2,127	0	0	1	0	NA	1	0	0	0	0	0	0	0	1	0	21.43
77	Rose and Hensher, (2014)	Transportation	Yes	5,556	1	1	1	0	NA	1	0.5	1	0	1	1	0	0	1	0	60.71
78	Ji et al. (2017)	International Journal of Sustainable Transportation	Yes	709	1	1	1	0	NA	1	0	0.5	0	0	0	0	1	0	0	39.29
79	He and Giuliano, (2017)	Transportation	Yes	799	1	1	1	0	NA	1	0	0	0	0	1	0	0	1	0	42.86
80	Liu et al. (2015)	Transport Policy	Yes	147,715	1	1	1	0	NA	1	0.5	0.5	0	0	1	0	0	0	0	42.86
81	Arman et al. (2018)	Transportation Planning and Technology	No	22,212	1	0.5	1	0	NA	1	0.5	0.5	0	0	1	0	0	0	0	39.29
82	Guan and Xu, (2018)	TR part A	Yes	4,769	1	0.5	1	0	NA	1	0.5	0.5	0	0	1	0	0	1	0	46.43
83	Sarkar and Chunchu, (2016)	Journal of Urban Planning and Development	No	567	1	1	1	0	NA	1	1	1	0.5	0	1	1	1	0	0	67.86

No	Reference	Journal	Top journal	N	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Score
84	Kristoffersson et al. (2018)	Transport Policy	Yes	1,131	0	0	1	0	NA	1	1	1	1	0	1	1	1	1	0	64.29
85	Lin et al. (2017)	International Journal of Sustainable Transportation	Yes	1,488	1	1	1	0	NA	1	0.5	1	0.5	0	0	1	1	0	0	57.14
86	Manoj and Verma, (2015)	J. of Transport Geography	Yes	386	1	0	1	0	NA	1	0	0	0	0	1	0	0	0	0	28.57
87	Kunhikrishnan and Srinivasan, (2018)	Transport Policy	Yes	872	1	0	1	0	NA	1	0	0.5	0	0	1	1	0	0	0	39.29
88	Dong et al. (2016)	International Journal of Sustainable Transportation	Yes	3,805	0.5	0	1	0	NA	1	0	0	0	0	1	0	0	1	0	32.14
89	Bridgell, (2014)	Transportation Planning and Technology	No	NA	0.5	1	0	0	NA	1	1	1	0.5	0	1	1	1	NA	0	61.54
90	Moniruzzaman and Farber, (2018)	International Journal of Sustainable Transportation	Yes	8,903	1	0	1	0	NA	1	0	0.5	0	0	1	1	0	0	0	39.29

No	Reference	Journal	Top journal	N	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	Score
91	Helbich, Bocker and Djijst, (2014)	J. of Transport Geography	Yes	4,317	1	0	1	0	NA	1	0	0	0	0	0	0	0	1	0	28.57
92	Basheer et al. (2018)	Transportation in Developing Economies	No	402	1	0	1	0	NA	1	0	0	0	0	1	0	0	0	0	28.57
93	Bhat et al. (2015)	TR part B	Yes	5,716	0	0.5	1	0	NA	1	0.5	0.5	0	0	1	0	1	1	0	46.43
94	Tilahun et al. (2016)	J. of Transport Geography	Yes	1,984	1	1	1	0	NA	1	1	1	0.5	0	0	0	1	0	0	53.57
95	Irfan et al. (2018)	Journal of Transportation Engineering, Part A: Systems	No	402	0.5	1	1	0	NA	1	1	1	1	0	1	1	1	0	0	67.86

Appendix 2: Original questions by McCloskey and Ziliak (1996)

Does the paper...

1. Use a small number of observations, such that statistically significant differences are not found at the conventional levels merely by choosing a large sample?
2. Report units and descriptive statistics for regression variables?
3. Report coefficients in elasticity form, or in some useful form that addresses the question of “how large is large”?
4. Test the null hypotheses that the authors said were the ones of interest?
5. Carefully interpret coefficients?
6. Eschew reporting all *t*- or *F*-statistics or standard errors, regardless of whether a significance test is appropriate?
7. At its first use, consider statistical significance to be one among other criteria of importance?
8. Mention the power of the tests?
9. Examine the power function?
10. Eschew “asterisk econometrics,” that is, ranking the coefficients according to the absolute size of the *t*-statistic?
11. Eschew “sign econometrics,” that is, remarking on the sign but not the size of the coefficients?
12. Discuss the size of the coefficients?
13. Discuss the scientific conversation within which a coefficient would be judged “large” or “small”?
14. Avoid choosing variables for inclusion solely on the basis of statistical significance?
15. Use other criteria of importance besides statistical significance after the crescendo?
16. Consider more than statistical significance decisive in an empirical argument?
17. Do a simulation to determine whether the coefficients are reasonable?
18. In the “conclusions” and “implications” sections, distinguish between statistical and substantive significance?
19. Avoid using the word “significance” in ambiguous ways, meaning “statistically significant” in one sentence and “large enough to matter for policy or science” in another?

Appendix 3: Regarding the number of reviewed articles

The final number of articles reviewed was 95 articles, selected randomly out of a total of 283 articles. Using the score system explained in Sect. “Do top journals do any better?”, Fig. 2 plots the cumulative average score of reviewed papers for each question. It can be seen that there is great variability in papers scores for each question at the beginning of the review, but the scores stabilize as the number of papers increase. The review was concluded at 95 articles. This is equivalent to 34% of all articles matching the inclusion criteria.

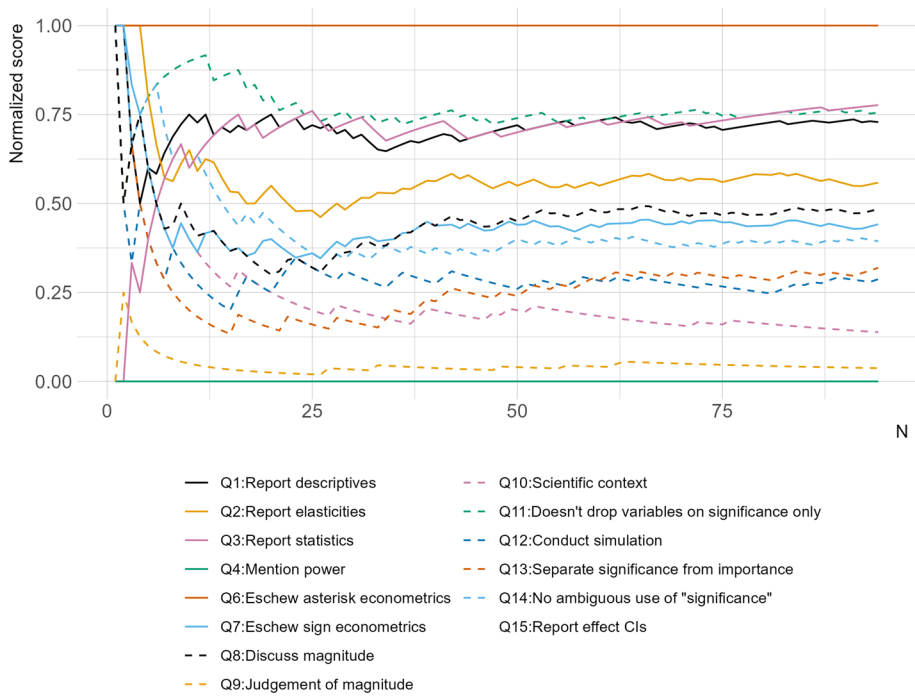


Fig. 2 Cumulative average score of papers for each of the fifteen questions

Acknowledgments This work was supported by Japan Society for the Promotion of Science KAKENHI Grant Number 20H02266.

Author contributions The authors confirm contribution to the paper as follows: study conception: Giancarlo Parady, Kay W. Axhausen. Study design, literature review and draft writing: Giancarlo Parady. Draft revision and editing: Giancarlo Parady, Kay W. Axhausen.

Funding Open access funding provided by The University of Tokyo.

Declarations

Competing interests The authors declare no competing interests.

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

References

- Abrantes, P.A.L., Wardman, M.R.: Meta-analysis of UK values of travel time: an update. *Transp. Res. Part A Policy Pract* **45**(1), 1–17 (2011). <https://doi.org/10.1016/j.tra.2010.08.003>
- Ahmad Termida, N., Susilo, Y.O., Franklin, J.P.: Observing dynamic behavioural responses due to the extension of a tram line by using panel survey. *Transp. Res. Part a: Policy Pract* **86**, 78–95 (2016). <https://doi.org/10.1016/j.tra.2016.02.005>
- Allard, R.F., Moura, F.: Effect of transport transfer quality on intercity passenger mode choice. *Transp. Res. Part a: Policy Pract* **109**, 89–107 (2018). <https://doi.org/10.1016/j.tra.2018.01.018>
- Anderson, M.K., Nielsen, O.A., Prato, C.G.: Multimodal route choice models of public transport passengers in the Greater Copenhagen Area. *EURO J. Transp. Logist.* **6**(3), 221–245 (2014). <https://doi.org/10.1007/s13676-014-0063-3>
- Anta, J., Pérez-López, J.B., Martínez-Pardo, A., Novales, M., Orro, A.: Influence of the weather on mode choice in corridors with time-varying congestion: a mixed data study. *Transportation* **43**(2), 337–355 (2016). <https://doi.org/10.1007/s11116-015-9578-1>
- Arman, M.A., Khademi, N., de Lapparent, M.: Women’s mode and trip structure choices in daily activity-travel: a developing country perspective. *Transp. Plan. Technol.* **41**(8), 845–877 (2018). <https://doi.org/10.1080/03081060.2018.1526931>
- Assi, K.J., Nahiduzzaman, K.M., Ratrout, N.T., Aldosary, A.S.: Mode choice behavior of high school goers: evaluating logistic regression and MLP neural networks. *Case Stud. Trans. Policy* **6**(2), 225–230 (2018). <https://doi.org/10.1016/j.cstp.2018.04.006>
- Axhausen, K.W., Fröhlich, P.: Übersicht zu Stated Preference- Studien in der Schweiz und Abschätzung von Gesamtelastizitäten, Statusbericht 2012. (2012)
- Aziz, H.M.A., Nagle, N.N., Morton, A.M., Hilliard, M.R., White, D.A., Stewart, R.N.: Exploring the impact of walk–bike infrastructure, safety perception, and built-environment on active transportation mode choice: a random parameter model using New York City commuter data. *Transportation* **45**(5), 1207–1229 (2018). <https://doi.org/10.1007/s11116-017-9760-8>
- Basheer, S., Srinivasan, K.K., Sivanandan, R.: Investigation of information quality and user response to real-time traffic information under heterogeneous traffic conditions. *Transp. Dev. Econ.* **4**(2), 1–11 (2018). <https://doi.org/10.1007/s40890-018-0061-5>
- Bhat, C.R., Dubey, S.K., Nagel, K.: Introducing non-normality of latent psychological constructs in choice modeling with an application to bicyclist route choice. *Transp. Res. Part B: Methodol.* **78**, 341–363 (2015). <https://doi.org/10.1016/j.trb.2015.04.005>
- Brewer, J.K.: On the power of statistical tests in the “American Educational Research Journal.” *J. Res. Sci. Teach.* **9**(3), 391–401 (1972). <https://doi.org/10.1002/tea.3660090410>
- Bridgelall, R.: Campus parking supply impacts on transportation mode choice. *Transp. Plan. Technol.* **37**(8), 711–737 (2014). <https://doi.org/10.1080/03081060.2014.959354>
- Bueno, P.C., Gomez, J., Peters, J.R., Vassallo, J.M.: Understanding the effects of transit benefits on employees’ travel behavior: evidence from the New York-New Jersey region. *Transp. Res. Part a: Policy Pract* **99**, 1–13 (2017). <https://doi.org/10.1016/j.tra.2017.02.009>
- Cantarella, G.E., de Luca, S.: Multilayer feedforward networks for transportation mode choice analysis: An analysis and a comparison with random utility models. *Transp. Res. Part C: Emerg. Technol.* **13**(2), 121–155 (2005). <https://doi.org/10.1016/j.trc.2005.04.002>
- Carteni, A., Cascetta, E., de Luca, S.: A random utility model for park & carsharing services and the pure preference for electric vehicles. *Transp. Policy* **48**, 49–59 (2016). <https://doi.org/10.1016/j.tranpol.2016.02.012>
- Chen, H., Cohen, P., Chen, S.: How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Commun. Stat. Simul. Comput.* **39**(4), 860–864 (2010). <https://doi.org/10.1080/03610911003650383>
- Clark, A.F., Scott, D.M., Yiannakoulis, N.: Examining the relationship between active travel, weather, and the built environment: a multilevel approach using a GPS-enhanced dataset. *Transportation* **41**(2), 325–338 (2014). <https://doi.org/10.1007/s11116-013-9476-3>
- Cohen, J.: The statistical power of abnormal-social psychological research: a review. *J. Abnorm. Soc. Psychol.* **65**(3), 145–153 (1962). <https://doi.org/10.1037/h0045186>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. In *Statistical Power Analysis for the Behavioral Sciences*. Psychology Press
- Cole-Hunter, T., Donaire-Gonzalez, D., Curto, A., Ambros, A., Valentin, A., Garcia-Aymerich, J., Martínez, D., Braun, L.M., Mendez, M., Jerrett, M., Rodriguez, D., de Nazelle, A., Nieuwenhuijsen, M.: Objective correlates and determinants of bicycle commuting propensity in an urban environment. *Transp. Res. Part D: Transp. Environ.* **40**(2), 132–143 (2015). <https://doi.org/10.1016/j.trd.2015.07.004>

- Collins, P.A., MacFarlane, R.: Evaluating the determinants of switching to public transit in an automobile-oriented mid-sized Canadian city: a longitudinal analysis. *Transp. Res. Part a: Policy Pract* **118**, 682–695 (2018). <https://doi.org/10.1016/j.tra.2018.10.014>
- Danapour, M., Nickkar, A., Jeihani, M., Khaksar, H.: Competition between high-speed rail and air transport in Iran: the case of Tehran-Isfahan. *Case Stud. Transp. Policy* **6**(4), 456–461 (2018). <https://doi.org/10.1016/j.cstp.2018.05.006>
- de Luca, S., Di Pace, R.: Modelling users' behaviour in inter-urban carsharing program: a stated preference approach. *Transp. Res. Part a: Policy Pract* **71**, 59–76 (2015). <https://doi.org/10.1016/j.tra.2014.11.001>
- de Bekker-Grob, E.W., Donkers, B., Jonker, M.F., Stolk, E.A.: Sample size requirements for discrete-choice experiments in healthcare: a practical guide. *Patient* **8**(5), 373–384 (2015). <https://doi.org/10.1007/s40271-015-0118-z>
- Di Ciommo, F., Comendador, J., López-Lambas, M.E., Cherchi, E., Ortúzar, J.D.: Exploring the role of social capital influence variables on travel behaviour. *Transp. Res. Part a: Policy Pract* **68**, 46–55 (2014). <https://doi.org/10.1016/j.tra.2014.08.018>
- Ding, C., Mishra, S., Lin, Y., Xie, B.: Cross-nested joint model of travel mode and departure time choice for urban commuting trips: case study in Maryland-Washington, DC Region. *J. Urban Plann. Dev.* **141**(4), 04014036 (2014). [https://doi.org/10.1061/\(asce\)up.1943-5444.0000238](https://doi.org/10.1061/(asce)up.1943-5444.0000238)
- Dong, H., Ma, L., Broach, J.: Promoting sustainable travel modes for commute tours: a comparison of the effects of home and work locations and employer-provided incentives. *Int. J. Sustain. Transp.* **10**(6), 485–494 (2016). <https://doi.org/10.1080/15568318.2014.1002027>
- Efthymiou, D., Antoniou, C.: Understanding the effects of economic crisis on public transport users' satisfaction and demand. *Transp. Policy* **53**, 89–97 (2017). <https://doi.org/10.1016/j.tranpol.2016.09.007>
- Ermagun, A., Samimi, A.: Promoting active transportation modes in school trips. *Transp. Policy* **37**, 203–211 (2015). <https://doi.org/10.1016/j.tranpol.2014.10.013>
- Fernández-Antolí, A., Guevara-Cue, A., de Lapparent, M., Bierlaire, M.: Correcting for endogeneity due to omitted attitudes: empirical assessment of a modified MIS method using RP mode choice data. *J. Choice Modell.* **20**, 1–15 (2016). <https://doi.org/10.1016/j.jocm.2016.09.001>
- Gan, H., Ye, X.: Leave the expressway or not? Impact of dynamic information. *J. Modern Transp.* **22**(2), 96–103 (2014). <https://doi.org/10.1007/s40534-014-0043-1>
- Gerber, P., Ma, T.Y., Klein, O., Schiebel, J., Carpentier-Postel, S.: Cross-border residential mobility, quality of life and modal shift: a Luxembourg case study. *Transp. Res. Part a: Policy Pract* **104**, 238–254 (2017). <https://doi.org/10.1016/j.tra.2017.06.015>
- Gokasar, I., Gunay, G.: Mode choice behavior modeling of ground access to airports: a case study in Istanbul, Turkey. *J. Air Transp. Manag.* **59**, 1–7 (2017). <https://doi.org/10.1016/j.jairtraman.2016.11.003>
- Guan, J., Xu, C.: Are relocatees different from others? Relocatee's travel mode choice and travel equity analysis in large-scale residential areas on the periphery of megacity Shanghai, China. *Transp. Res. Part a: Policy Pract* **111**, 162–173 (2018). <https://doi.org/10.1016/j.tra.2018.03.011>
- Habib, K.N.: Household-level commuting mode choices, car allocation and car ownership level choices of two-worker households: The case of the city of Toronto. *Transportation* **41**(3), 651–672 (2014). <https://doi.org/10.1007/s11116-014-9518-5>
- Habib, K.M.N., Sasic, A.: A GEV model with scale heterogeneity for investigating the role of mobility tool ownership in peak period non-work travel mode choices. *J. Choice Model.* **10**(1), 46–59 (2014). <https://doi.org/10.1016/j.jocm.2014.01.003>
- Habib, K.M.N., Swait, J., Salem, S.: Using repeated cross-sectional travel surveys to enhance forecasting robustness: accounting for changing mode preferences. *Transp. Res. Part A Policy Pract* **67**, 110–126 (2014). <https://doi.org/10.1016/j.tra.2014.06.004>
- Halldórsson, K., Nielsen, O.A., Prato, C.G.: Home-end and activity-end preferences for access to and egress from train stations in the Copenhagen region. *Int. J. Sustain. Transp.* **11**(10), 776–786 (2017). <https://doi.org/10.1080/15568318.2017.1317888>
- Hasnine, M.S., Habib, K.N.: What about the dynamics in daily travel mode choices? A dynamic discrete choice approach for tour-based mode choice modelling. *Transp. Policy* **71**, 70–80 (2018). <https://doi.org/10.1016/j.tranpol.2018.07.011>
- Hasnine, M.S., Lin, T.Y., Weiss, A., Habib, K.N.: Determinants of travel mode choices of post-secondary students in a large metropolitan area: the case of the city of Toronto. *J. Transp. Geogr.* **70**(June), 161–171 (2018). <https://doi.org/10.1016/j.jtrangeo.2018.06.003>
- He, S.Y., Giuliano, G.: Factors affecting children's journeys to school: a joint escort-mode choice model. *Transportation* **44**(1), 199–224 (2017). <https://doi.org/10.1007/s11116-015-9634-x>

- Heinen, E.: Identity and travel behaviour: a cross-sectional study on commute mode choice and intention to change. *Transport. Res. F: Traffic Psychol. Behav.* **43**, 238–253 (2016). <https://doi.org/10.1016/j.trf.2016.10.016>
- Heinen, E., Ogilvie, D.: Variability in baseline travel behaviour as a predictor of changes in commuting by active travel, car and public transport: a natural experimental study. *J. Transp. Health* **3**(1), 77–85 (2016). <https://doi.org/10.1016/j.jth.2015.11.002>
- Hensher, D.A., Ho, C.Q.: Experience conditioning in commuter modal choice modelling—Does it make a difference? *Transp. Res. Part E Logist. Trans. Rev.* **95**, 164–176 (2016). <https://doi.org/10.1016/j.tre.2016.09.010>
- Hess, S., Spitz, G., Bradley, M., Coogan, M.: Analysis of mode choice for intercity travel: application of a hybrid choice model to two distinct US corridors. *Transp. Res. Part a: Policy Pract* **116**, 547–567 (2018). <https://doi.org/10.1016/j.tra.2018.05.019>
- Ho, C.Q., Hensher, D.A.: A workplace choice model accounting for spatial competition and agglomeration effects. *J. Transp. Geogr.* **51**, 193–203 (2016). <https://doi.org/10.1016/j.jtrangeo.2016.01.005>
- Hsu, H.P., Saphores, J.D.: Impacts of parental gender and attitudes on children's school travel mode and parental chauffeuring behavior: results for California based on the 2009 National Household Travel Survey. *Transportation* **41**(3), 543–565 (2014). <https://doi.org/10.1007/s11116-013-9500-7>
- Hyland, M., Frei, C., Frei, A., Mahmassani, H.S.: Riders on the storm: exploring weather and seasonality effects on commute mode choice in Chicago. *Travel Behav. Soc.* **13**, 44–60 (2018). <https://doi.org/10.1016/j.tbs.2018.05.001>
- Irfan, M., Khurshid, A.N., Khurshid, M.B., Ali, Y., Khattak, A.: Policy implications of work-trip mode choice using econometric modeling. *J. Trans. Eng. A Syst.* **144**(8), 04018035 (2018). <https://doi.org/10.1061/jtepbs.0000158>
- Jánošíkova, L., Slavík, J., Koháni, M.: Estimation of a route choice model for urban public transport using smart card data. *Transp. Plan. Technol.* **37**(7), 638–648 (2014). <https://doi.org/10.1080/03081060.2014.935570>
- Ji, Y., Fan, Y., Ermagun, A., Cao, X., Wang, W., Das, K.: Public bicycle as a feeder mode to rail transit in China: the role of gender, age, income, trip purpose, and bicycle theft experience. *Int. J. Sustain. Transp.* **11**(4), 308–317 (2017). <https://doi.org/10.1080/15568318.2016.1253802>
- Kamargianni, M., Ben-Akiva, M., Polydoropoulou, A.: Incorporating social interaction into hybrid choice models. *Transportation* **41**(6), 1263–1285 (2014). <https://doi.org/10.1007/s11116-014-9550-5>
- Kato, H., Sakashita, A., Tsuchiya, T., Oda, T., Tanishita, M.: Estimating value of travel time savings by using large-scale household survey data from Japan. *Transp. Res. Rec.* **2231**, 85–92 (2011). <https://doi.org/10.3141/2231-11>
- Keyes, A.K.M., Crawford-Brown, D.: The changing influences on commuting mode choice in urban England under Peak Car: a discrete choice modelling approach. *Transp. Res. F: Traffic Psychol. Behav.* **58**, 167–176 (2018). <https://doi.org/10.1016/j.trf.2018.06.010>
- Khan, M., Kockelman, K.M., Xiong, X.: Models for anticipating non-motorized travel choices, and the role of the built environment. *Transp. Policy* **35**, 117–126 (2014). <https://doi.org/10.1016/j.tranpol.2014.05.008>
- Khan, S., Maoh, H., Lee, C., Anderson, W.: Toward sustainable urban mobility: investigating nonwork travel behavior in a sprawling Canadian city. *Int. J. Sustain. Transp.* **10**(4), 321–331 (2016). <https://doi.org/10.1080/15568318.2014.928838>
- Khoo, H.L., Asitha, K.S.: User requirements and route choice response to smart phone traffic applications (apps). *Travel Behav. Soc.* **3**, 59–70 (2016). <https://doi.org/10.1016/j.tbs.2015.08.004>
- Kristoffersson, I., Daly, A., Algiers, S.: Modelling the attraction of travel to shopping destinations in large-scale modelling. *Transp. Policy* **68**, 52–62 (2018). <https://doi.org/10.1016/j.tranpol.2018.04.013>
- Kunhikrishnan, P., Srinivasan, K.K.: Investigating behavioral differences in the choice of distinct Intermediate Public Transport (IPT) modes for work trips in Chennai city. *Transp. Policy* **61**, 111–122 (2018). <https://doi.org/10.1016/j.tranpol.2017.10.006>
- Lee, J.: Impact of neighborhood walkability on trip generation and trip chaining: case of Los Angeles. *J. Urban Plan. Dev.* **142**(3), 05015013 (2015). [https://doi.org/10.1061/\(asce\)up.1943-5444.0000312](https://doi.org/10.1061/(asce)up.1943-5444.0000312)
- Lee, J.S., Nam, J., Lee, S.S.: Built environment impacts on individual mode choice: an empirical study of the Houston-Galveston Metropolitan Area. *Int. J. Sustain. Transp.* **8**(6), 447–470 (2014). <https://doi.org/10.1080/15568318.2012.716142>
- Lin, J.J., Wang, N.L., Feng, C.M.: Public bike system pricing and usage in Taipei. *Int. J. Sustain. Transp.* **11**(9), 633–641 (2017). <https://doi.org/10.1080/15568318.2017.1301601>
- Lin, J.J., Zhao, P., Takada, K., Li, S., Yai, T., Chen, C.H.: Built environment and public bike usage for metro access: a comparison of neighborhoods in Beijing, Taipei, and Tokyo. *Transp. Res. Part D: Transp. Environ.* **63**(1), 209–221 (2018). <https://doi.org/10.1016/j.trd.2018.05.007>

- Liu, C., Susilo, Y.O., Karlström, A.: The influence of weather characteristics variability on individual's travel mode choice in different seasons and regions in Sweden. *Transp. Policy* **41**, 147–158 (2015). <https://doi.org/10.1016/j.tranpol.2015.01.001>
- Liu, Y., Ji, Y., Shi, Z., He, B., Liu, Q.: Investigating the effect of the spatial relationship between home, workplace and school on parental chauffeurs' daily travel mode choice. *Transp. Policy* **69**, 78–87 (2018). <https://doi.org/10.1016/j.tranpol.2018.06.004>
- Mahpour, A., Mamdoohi, A., HosseinRashidi, T., Schmid, B., Axhausen, K.W.: Shopping destination choice in Tehran: an integrated choice and latent variable approach. *Transport. Res. F: Traffic Psychol. Behav.* **58**, 566–580 (2018). <https://doi.org/10.1016/j.trf.2018.06.045>
- Manoj, M., Verma, A.: Activity-travel behaviour of non-workers belonging to different income group households in Bangalore, India. *J. Transp. Geogr.* **49**, 99–109 (2015). <https://doi.org/10.1016/j.jtrangeo.2015.10.017>
- Mattson, J., Hough, J., Varma, A.: Estimating demand for rural intercity bus services. *Res. Transp. Econ.* **71**, 68–75 (2018). <https://doi.org/10.1016/j.retrec.2018.11.001>
- Maxwell, S.E.: The persistence of underpowered studies in psychological research: causes, consequences, and remedies. *Psychol. Methods* **9**(2), 147–163 (2004). <https://doi.org/10.1037/1082-989X.9.2.147>
- Mccloskey, D.N., Ziliak, S.T.: The standard error of regressions. *J. Econ. Literat.* **34**(1), 97–114 (1996)
- Mehdizadeh, M., Nordfaern, T., Mamdoohi, A.R.: The role of socio-economic, built environment and psychological factors in parental mode choice for their children in an Iranian setting. *Transportation* **45**(2), 523–543 (2018). <https://doi.org/10.1007/s11116-016-9737-z>
- Mitra, R., Buliung, R.N.: The influence of neighborhood environment and household travel interactions on school travel behavior: an exploration using geographically-weighted models. *J. Transp. Geogr.* **36**, 69–78 (2014). <https://doi.org/10.1016/j.jtrangeo.2014.03.002>
- Mitra, R., Buliung, R.N.: Exploring differences in school travel mode choice behaviour between children and youth. *Transp. Policy* **42**, 4–11 (2015). <https://doi.org/10.1016/j.tranpol.2015.04.005>
- Moniruzzaman, M., Farber, S.: What drives sustainable student travel? Mode choice determinants in the Greater Toronto Area. *Int. J. Sustain. Transp.* **12**(5), 367–379 (2018). <https://doi.org/10.1080/15568318.2017.1377326>
- Motoaki, Y., Daziano, R.A.: A hybrid-choice latent-class model for the analysis of the effects of weather on cycling demand. *Transp. Res. Part A Policy Pract* **75**, 217–230 (2015). <https://doi.org/10.1016/j.tra.2015.03.017>
- Orozco-Fontalvo, M., Arévalo-Támara, A., Guerrero-Barbosa, T., Gutiérrez-Torres, M.: Bicycle choice modeling: a study of university trips in a small Colombian city. *J. Transp. Health* **9**, 264–274 (2018). <https://doi.org/10.1016/j.jth.2018.01.014>
- Paleti, R., Faghih Imani, A., Eluru, N., Hu, H.H., Huang, G.: An integrated model of intensity of activity opportunities on supply side and tour destination & departure time choices on demand side. *J. Choice Model.* **24**, 63–74 (2017). <https://doi.org/10.1016/j.jocm.2017.03.003>
- Parady, G., Ory, D., Walker, J.: The overreliance on statistical goodness-of-fit and under-reliance on model validation in discrete choice models: a review of validation practices in the transportation academic literature. *J. Choice Modell.* **38**, 100257 (2021). <https://doi.org/10.1016/j.jocm.2020.100257>
- Parady, G., Suzuki, K., Oyama, Y., Chikaraishi, M.: Activity detection with Google Maps Location History data: factors affecting joint activity detection probability and its potential application on real social networks. *Travel Behav. Soc.* **30**, 347–357 (2023)
- Paulssen, M., Temme, D., Vij, A., Walker, J.L.: Values, attitudes and travel behavior: a hierarchical latent variable mixed logit model of travel mode choice. *Transportation* **41**(4), 873–888 (2014). <https://doi.org/10.1007/s11116-013-9504-3>
- Pnevmatikou, A.M., Karlaftis, M.G., Kepaptsoglou, K.: Metro service disruptions: how do people choose to travel? *Transportation* **42**(6), 933–949 (2015). <https://doi.org/10.1007/s11116-015-9656-4>
- Qin, H., Gao, J., Guan, H., Chi, H.: Estimating heterogeneity of car travelers on mode shifting behavior based on discrete choice models. *Transp. Plan. Technol.* **40**(8), 914–927 (2017). <https://doi.org/10.1080/03081060.2017.1355886>
- Qin, H., Gao, J., Kluger, R., Wu, Y.J.: Effects of perception on public bike-and-ride: a survey under complex, multifactor mode-choice scenarios. *Trans. Res. Part F: Traffic Psychol. Behav.* **54**, 264–275 (2018). <https://doi.org/10.1016/j.trf.2018.01.021>
- Rahman, M.L., Baker, D.: Modelling induced mode switch behaviour in Bangladesh: a multinomial logistic regression approach. *Transp. Policy* **71**, 81–91 (2018). <https://doi.org/10.1016/j.tranpol.2018.09.006>
- Rose, J.M., Bliemer, M.C.J.: Sample size requirements for stated choice experiments. *Transportation* **40**(5), 1021–1041 (2013). <https://doi.org/10.1007/s11116-013-9451-z>
- Rose, J.M., Hensher, D.A.: Demand for taxi services: new elasticity evidence. *Transportation* **41**(4), 717–743 (2014). <https://doi.org/10.1007/s11116-013-9482-5>

- Rose, J.M., Bliemer, M.C.J., Hensher, D.A., Collins, A.T.: Designing efficient stated choice experiments in the presence of reference alternatives. *Trans. Res. Part B: Methodol.* **42**(4), 395–406 (2008). <https://doi.org/10.1016/j.trb.2007.09.002>
- Rossi, J.S.: Statistical power of psychological research: What have we gained in 20 years? *J. Consult. Clin. Psychol.* **58**(5), 646–656 (1990). <https://doi.org/10.1037/0022-006x.58.5.646>
- Rotaris, L., Danielis, R.: Commuting to college: the effectiveness and social efficiency of transportation demand management policies. *Transp. Policy* **44**, 158–168 (2015). <https://doi.org/10.1016/j.tranp.2015.08.001>
- Sarkar, P.P., Chunchu, M.: Quantification and analysis of land-use effects on travel behavior in smaller Indian cities: case Study of Agartala. *J. Urban Plann. Dev.* **142**(4), 04016009 (2016). [https://doi.org/10.1061/\(asce\)up.1943-5444.0000322](https://doi.org/10.1061/(asce)up.1943-5444.0000322)
- Sarkar, P.P., Mallikarjuna, C.: Effect of perception and attitudinal variables on mode choice behavior: a case study of Indian city, Agartala. *Travel Behav. Soc.* **12**, 108–114 (2018). <https://doi.org/10.1016/j.tbs.2017.04.003>
- Satiennam, T., Jaensirisak, S., Satiennam, W., Detsamrong, S.: Potential for modal shift by passenger car and motorcycle users towards Bus Rapid Transit (BRT) in an Asian developing city. *IATSS Res.* **39**(2), 121–129 (2016). <https://doi.org/10.1016/j.iatssr.2015.03.002>
- Schoner, J.E., Cao, J., Levinson, D.M.: Catalysts and magnets: built environment and bicycle commuting. *J. Transp. Geogr.* **47**, 100–108 (2015). <https://doi.org/10.1016/j.jtrangeo.2015.07.007>
- Standen, C., Crane, M., Collins, A., Greaves, S., Rissel, C.: Determinants of mode and route change following the opening of a new cycleway in Sydney, Australia. *J. Transp. Health* **4**, 255–266 (2017). <https://doi.org/10.1016/j.jth.2016.10.004>
- Stone, M., Larsen, K., Faulkner, G.E.J., Buliung, R.N., Arbour-Nicitopoulos, K.P., Lay, J.: Predictors of driving among families living within 2km from school: exploring the role of the built environment. *Transp. Policy* **33**, 8–16 (2014). <https://doi.org/10.1016/j.tranpol.2014.02.001>
- Sun, G., Han, X., Sun, S., Oreskovic, N.: Living in school catchment neighborhoods: perceived built environments and active commuting behaviors of children in China. *J. Transp. Health* **8**, 251–261 (2018). <https://doi.org/10.1016/j.jth.2017.12.009>
- Thigpen, C.G., Driller, B.K., Handy, S.L.: Using a stages of change approach to explore opportunities for increasing bicycle commuting. *Transp. Res. Part D: Transp. Environ.* **39**, 44–55 (2015). <https://doi.org/10.1016/j.trd.2015.05.005>
- Tilahun, N., Thakuriah, P.V., Li, M., Keita, Y.: Transit use and the work commute: analyzing the role of last mile issues. *J. Transp. Geogr.* **54**, 359–368 (2016). <https://doi.org/10.1016/j.jtrangeo.2016.06.021>
- Vij, A., Walker, J.L.: Preference endogeneity in discrete choice models. *Transp. Res. Part B: Methodol.* **64**, 90–105 (2014). <https://doi.org/10.1016/j.trb.2014.02.008>
- Vij, A., Gorrupaty, S., Walker, J.L.: From trend spotting to trend 'splaining: understanding modal preference shifts in the San Francisco Bay Area. *Transp. Res. Part a: Policy Pract* **95**, 238–258 (2017). <https://doi.org/10.1016/j.tra.2016.11.014>
- Wang, Y., Correia, G.H.A., de Romph, E., Timmermans, H.J.P.: Using metro smart card data to model location choice of after-work activities: An application to Shanghai. *J. Transp. Geogr.* **63**, 40–47 (2017). <https://doi.org/10.1016/j.jtrangeo.2017.06.010>
- Yamamoto, T., Takamura, S., Morikawa, T.: Structured random walk parameter for heterogeneity in trip distance on modeling pedestrian route choice behavior at downtown area. *Travel Behav. Soc.* **11**, 93–100 (2018). <https://doi.org/10.1016/j.tbs.2018.02.006>
- Yang, C.-W., Tsai, M.-C., Chang, C.-C.: Investigating the joint choice behavior of intercity transport mode and high-speed rail cabin with a strategy map. *J. Adv. Transp.* **49**(3), 297–308 (2015). <https://doi.org/10.1002/atr>
- Yang, L., Shen, Q., Li, Z.: Comparing travel mode and trip chain choices between holidays and weekdays. *Transp. Res. Part a: Policy Pract* **91**, 273–285 (2016a). <https://doi.org/10.1016/j.tra.2016.07.001>
- Yang, Y., Yao, E., Yang, Z., Zhang, R.: Modeling the charging and route choice behavior of BEV drivers. *Transp. Res. Part C: Emerg. Technol.* **65**, 190–204 (2016b). <https://doi.org/10.1016/j.trc.2015.09.008>
- Zaidan, E., Abulibdeh, A.: Modeling ground access mode choice behavior for Hamad International Airport in the 2022 FIFA World Cup city, Doha Qatar. *J. Air Trans. Manage.* **73**, 32–45 (2018). <https://doi.org/10.1016/j.jairtraman.2018.08.007>
- Zhang, G., Wang, Z., Persad, K.R., Walton, C.M.: Enhanced traffic information dissemination to facilitate toll road utilization: a nested logit model of a stated preference survey in Texas. *Transportation* **41**(2), 231–249 (2014). <https://doi.org/10.1007/s11116-013-9449-6>
- Zhang, N., Zhang, Y., Zhang, X.: Pedestrian choices of vertical walking facilities inside urban rail transit stations. *KSCE J. Civ. Eng.* **19**(3), 742–748 (2015). <https://doi.org/10.1007/s12205-012-0331-4>

- Zhang, L., Chen, C., Zhang, J., Fang, S., You, J., Guo, J.: Modeling lane-changing behavior in freeway off-ramp areas from the shanghai naturalistic driving study. *J. Adv. Transp.* **2018**, 1–10 (2018). <https://doi.org/10.1155/2018/8645709>
- Ziliak, S., McCloskey, D.: *The cult of statistical significance. How the standard error cost us jobs, justice and lives.* The University of Michigan Press, Ann Arbor (2007)

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

Giancarlo Parady is a lecturer at the department of Urban Engineering, the University of Tokyo. He received his Ph.D. in Urban Engineering from the University of Tokyo in 2014. His main research interest is the connection between social networks the built environment and travel behavior.

Kay W. Axhausen has been a Professor of transport planning at the Institute for Transport Planning and Systems at the ETH Zurich since 1999. He had previously held academic positions at the University of Oxford, Imperial College London and the University of Innsbruck. He has been involved in the measurement and modeling of travel behavior for the past 30 years, contributing in particular to the literature on stated choice methods, micro-simulation of travel behavior, valuation of travel time and its components, activity scheduling, and travel survey data collection.