

Toward Evidence-Based Urban Planning

Integrating Quality Assessments in Literature Reviews

Léa Ravensbergen & Ahmed El-Geneidy

To cite this article: Léa Ravensbergen & Ahmed El-Geneidy (2023) Toward Evidence-Based Urban Planning, Journal of the American Planning Association, 89:3, 389-398, DOI: [10.1080/01944363.2022.2074872](https://doi.org/10.1080/01944363.2022.2074872)

To link to this article: <https://doi.org/10.1080/01944363.2022.2074872>



© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC.



[View supplementary material](#)



Published online: 11 Jul 2022.



[Submit your article to this journal](#)



Article views: 1562



[View related articles](#)



[View Crossmark data](#)



Citing articles: 1 [View citing articles](#)



Viewpoint

Toward Evidence-Based Urban Planning: Integrating Quality Assessments in Literature Reviews

Léa Ravensbergen Ahmed El-Geneidy

ABSTRACT

Literature reviews can play a pivotal role in designing urban policies. Here we introduce two tools used by public health specialists to assess the quality of studies and quantify the evidence derived from them: the Risk of Bias Assessment (RoB) and Evaluation of Certainty of Evidence (ECE). The RoB scores articles on several domains (e.g., selection bias, study design, etc.) to provide an appraisal of how rigorous the study is, whereas the ECE tool provides a framework to clearly state how much certainty there is in the outcomes under study. Both tools can be used to enhance literature review articles in urban planning to better inform practitioners on how to best develop policies using a rigorous approach.

Keywords: evaluation of certainty of evidence, literature review, practice-oriented, quality assessment, risk of bias assessment

A long and rich history of collaboration exists between the fields of urban planning and public health (Jackson et al., 2013). Both are practice-oriented disciplines, and both have strong advocates for evidence-based approaches wherein research informs practice and policy development (Brownson et al., 2009; Krizek et al., 2009). A key component of evidence-based approaches is to gather all the relevant findings on a topic by, say, completing a literature review (Krizek et al., 2009). Indeed, K. Stevens (2001) called systematic reviews in health research “the heart of evidence-based practice” (p. 529).

Although both urban planning and public health conduct literature reviews, their impact seems to vary across the disciplines. Literature reviews in urban planning have been critiqued for lacking rigor (Xiao & Watson, 2019), but recent studies have found that they are used directly in public health decision making (Dobbins et al., 2004; South & Lorenc, 2020). Further, the approach taken when conducting literature reviews can vary between urban planning and other disciplines (Xiao & Watson, 2019). For instance, the public health field has developed various methods and tools to conduct reviews in a rigorous manner. These tools have, to the best of our knowledge, yet to be incorporated into urban planning. These assessments speak to the quality

of studies and the evidence derived from the reviewed articles in a systematic manner and therefore have much to contribute to urban planners who champion evidence-based policy development.

To help urban planning reviews better inform practice, we focus here on how to adapt and integrate research tools used by public health professionals when conducting literature reviews for quantitative research in the field of urban planning. In this Viewpoint, we begin by introducing the role and impact of literature reviews in the field of urban planning. We then present the Risk of Bias Assessment (RoB) and Evaluation of Certainty of Evidence (ECE) tools for literature reviews. These tools are used to assess the quality of the existing scholarship and provide clear evidence to practitioners in the public health field. We then provide an example of each tool that is applicable to survey research studies in urban planning and discuss the possibilities of applying these tools across different types of urban planning research. We conclude by outlining the foreseen benefits of incorporating RoBs and ECes in our discipline. Given the practice-oriented nature of urban planning, we believe that adapting these tools offers great potential to move toward a more evidence-based planning approach that many planning authorities have begun adopting in recent years.

Table 1. Typical methods in four types of literature reviews common in urban planning.

	Critical reviews	Scoping reviews	Systematic reviews	Meta-analyses
Search strategy	Will often identify most significant articles	Comprehensiveness determined by constraints	Exhaustive and comprehensive search	Exhaustive and comprehensive search
Appraisal	No formal appraisal; may consider impact of articles	No formal appraisal	Quality assessments	Appraisals can be used to determine inclusion
Synthesis	Usually narrative; can be conceptual	Tabular and narrative synthesis	Narrative and tabular synthesis	Numerical data with narrative synthesis
Analysis	Develop conceptual contribution; derive new theory	Characterize literature by key features	Determine what is known/gaps, certainty of evidence, recommendations for practice	Numerical analysis

Source: Adapted from Table 1 in Grant and Booth (2009).

Literature Reviews: Important Tools Yet to Realize Their Full Potential

Literature reviews not only provide a coherent and well-structured overview of the research that has been done in an area, but also add value through, for instance, identifying research gaps, putting forward research agendas or conceptual models, or critically evaluating the methods or frameworks used to study a topic (De Vos & El-Geneidy, 2021; van Wee & Banister, 2016). Literature reviews are a research tool that have had a high impact on many fields. For example, some transport and land-use planning review papers have been cited extensively (for example, Cao et al., 2009, received more than 600 citations on Scopus and 588 on Web of Science).

Many different types of literature reviews exist (De Vos & El-Geneidy, 2021; Grant & Booth, 2009; van Wee & Banister, 2016). Grant and Booth (2009) outline 14 types of reviews in their article, all of which differ when it comes to search strategy, appraisal, synthesis, and analysis. From our experiences reading, publishing, and peer-reviewing literature review articles in the field of urban planning, four common types of reviews in our field are critical reviews, scoping reviews, meta-analyses, and systematic literature reviews (Table 1).

The aim of the tools we introduce in the following section is to assess the quality of the articles included in a literature review. Xiao and Watson (2019) discussed how quality assessments can be used to help authors decide which articles to include in a literature review or to help authors know which articles' results to emphasize. The specific tools we present in this Viewpoint are mainly used in systematic reviews, which are distinguished by their exhaustive search strategy. This comprehensive search is needed to appraise and synthesize *all* the evidence (often including gray literature evidence) to establish what is known and to make

recommendations for practice (Adkins et al., 2017). Systematic reviews are also known for incorporating quality assessments (Grant & Booth, 2009). In our experience, however, urban planning researchers frequently conduct this type of review but omit these quality assessments.

Two tools typically used for quality assessments are the RoB and the ECE. Below we provide an overview of both tools and discuss how they can be adapted for urban planning research to help in deriving more evidence-based urban policy. We also present a version of each tool that can be used to assess quantitative studies using survey research in urban planning. These tools are usually introduced in manuscripts after a detailed description of each study included in the review is introduced to the reader because they help in assessing the quality of the reviewed articles and provide an assessment of the evidence identified across the reviewed manuscripts.

Risk of Bias Assessment

An RoB, also called a *quality assessment* or a *critical appraisal*, aims to establish the quality or rigor of the studies that exist on a topic. Typically, each study included in a literature review will be scored on potential sources of bias (e.g., selection bias, detection bias, or reporting bias). Though researchers completing literature reviews cannot measure the presence of bias in each study, this tool can be used to assess the *risk* that the results are biased based on what is stated in the methods sections of articles. Doing so can help literature review authors avoid generating misleading literature reviews because they can amplify the results of rigorous studies and minimize those from studies with greater risk of bias (Al-Jundi & Sakka, 2017; Ma et al., 2020).

RoBs are a typical step in many review protocols. For instance, "assess studies for risk of bias" is the

seventh step in the Cochrane review (The Cochrane Collaboration, 2011), and considerations of RoB are in five items (nos. 11, 14, 18, 20a, and 21) of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist, a popular tool in literature reviews that has been in use since 2009 (PRISMA, 2021). Considering how often we have come across literature review articles in urban planning that state they follow these guidelines, particularly the PRISMA checklist, it is surprising how infrequently we have seen RoBs used in our discipline (for instance, see Abu Hatab et al., 2019; Basu et al., 2021; and Calderón-Argelich et al., 2021).

Sample RoB Adapted for Survey Research in Urban Planning

Many different RoB tools exist; however, because many of these tools originate from health sciences (Krizek et al., 2009), the bulk are created for study designs rarely (or never) used in urban planning studies, such as clinical trials or cohort studies. This leads to components of the tool that are not applicable to our field. For instance, the Cochrane RoB evaluation designed for clinical trials includes a performance-bias component (whether the trial inadvertently introduced differences other than the intervention being evaluated; The Cochrane Collaboration, 2011). This type of bias is rarely relevant in urban planning studies, a field known for using a wider variety of methods than many other fields.

Given this, Table 2 presents an RoB tool for survey-based research in the field of urban planning that we adapted from the Effective Public Health Practice Project (EPHPP; Effective Public Healthcare Panacea Project, 2021) for urban planning studies. The original EPHPP is also presented in Technical Appendix A. This version of the tool is best suited for survey-based research and incorporates eight potential sources of bias relevant to survey studies in urban planning. Table 2 outlines the types of bias assessed, guiding questions, and grading criteria. This tool also provides a global rating score that incorporates all eight bias scores.

When literature review authors are completing data extraction for the studies included in their literature review, they must simply also respond to the guiding questions in Table 2 for each article. Most answers can be provided by carefully reading the methods sections of the included articles. Therefore, these assessments require nominal additional effort. From our experience, we highly recommend including a supporting statement for each of the types of bias to complement the answers to the guiding questions. Once the guiding questions (and supporting statements) have been answered for all articles included in the review, the author can score each article for each type of bias using the criteria put forth in Table 2.

The results of the RoB can then be reported in the literature review article in tabular or graphical form. For instance, Figure 1 shows the results from an assessment that used the EPHPP. Although the guiding questions and criteria reported in Table 2 will often be relegated to the appendix of the literature review, the results of the assessment (i.e., Figure 1) will generally be reported in the article after the narrative synthesis of the articles. The author(s) can then comment on the overall quality of the articles, for instance by reporting on the global ratings or reporting how well articles scored on each type of bias (in this case, the literature scores highly on Withdrawal and Dropout but very low on Blinding). Other trends can be reported as well; for instance, one could report if gray literature reports or older studies received lower scores on average than peer-reviewed or more recent studies.

Because RoBs are not yet used in our discipline extensively, we see this tool as a starting point: We hope that other researchers contribute to this tool and adapt it to meet their assessment needs. Technical Appendix A includes several RoBs that may be relevant for literature reviews in urban planning. The Critical Appraisal Skills Program (CASP) website also includes many excellent tools to help perform a critical appraisal of articles (CASP, 2021).

It is important to note that most of the tools in Technical Appendix A were designed for quantitative research, with an emphasis on survey-based research. However, different research designs tend to require different types of quality assessments (Kitchenham & Stuart, 2007) and urban planners make use of a wide array of research tools, including interview-based research, ethnography, case studies, archival research, theoretical modeling, and legal analyses. These different approaches mean that many of the types of bias assessed in the RoB we present here are simply inappropriate for research using these methods. Unfortunately, to the best of our knowledge, fewer RoB tools have been designed to assess these methods, which may be due to their lack of prominence in the health sciences. Quality assessment tools for qualitative research may be less prevalent because this research draws from different epistemological, ontological, and methodological foundations. These different framings result in different ways to assess rigor. As Small (2009) argued, qualitative research should not be assessed by the same tools and concepts as quantitative research. For instance, in terms of the data collected, qualitative research tends to collect *rich* data (quality), whereas quantitative research (Fusch & Ness, 2015) tends to emphasize *thick* data. Further, qualitative research tends to seek logical rather than statistical inference, and data saturation rather than representativeness (Small, 2009).

Table 2. Sample risk of bias analysis criteria.

Criteria	Guiding questions	Evaluation
	Components rating	
Selection bias	(Q1) Are the individuals selected to participate in the study likely to be representative of the target population? 1. Very likely 2. Somewhat likely 3. Not likely 4. Can't tell (Q2) What percentage of selected individuals agreed to participate?	Strong = Sample's sociodemographic characteristics compared with the general population (e.g., through the census) and little discrepancy is identified OR 60% or more agreed to participate Moderate = Sample characteristics compared with the general population and some discrepancy is identified OR use multiple data sources, some of which are highly representative (e.g., the census) OR representative sampling strategy Weak = No comparison with general population and very small sample (< 100)
Study design	Indicate the study design Was the study described as randomized? No/Yes If Yes, was the method of randomization described? No/Yes If Yes, was the method appropriate? No/Yes	Strong = Natural experiment design Moderate = Randomized cross-sectional design Weak = Cross-sectional design with no randomization
Confounders	(Q1) Were there important differences between groups prior to the intervention? Yes/No/Can't tell (Q2) If yes, indicate the percentage of relevant confounders that were controlled (either in the design [e.g., stratification, matching] or analysis)? 80%–100% (most)/60%–79% (some)/less than 60% (few or none)/Can't tell	Strong = Compared groups and found minimal differences Moderate = Compared groups and found some differences OR controlled well for confounders Weak = Did not compare groups OR did not control well for confounders
Blinding	(Q1) Was (were) the outcome assessor(s) aware of the intervention or exposure status of participants? Yes/No/Can't tell (Q2) Were the study participants aware of the research question? Yes/No/Can't tell	Strong = Authors state the participants were not aware of research question during data collection Moderate = Unclear whether participants were not aware of research question during data collection Weak = Authors state participants were aware of research question during data collection
Data-collection methods	(Q1) Were data-collection tools shown to be valid? Yes/No/Can't tell (Q2) Were data collection tools shown to be reliable? Yes/No/Can't tell	Strong = data-collection tools result in measured data (accelerometry, GPS) Moderate = self-reported but rigorous survey Weak = Self-reported through poorly described survey OR data originate from different sources
Withdrawals and dropouts	(Q1) Were withdrawals and dropouts reported in terms of numbers and/or reasons per group? Yes/No/Can't tell/Not applicable (i.e., one-time surveys or interviews) (Q2) Indicate the percentage of participants completing the study. (If the percentage differs by groups, record the lowest.) 80%–100%/60%–79%/Less than 60%/Can't tell/Not applicable (i.e., retrospective case-control)	Strong = Withdrawal and dropouts reported (numbers and reasons), minimal differences identified between withdrawals and sample Moderate = Withdrawal and dropouts reported (numbers and reasons) with some differences identified OR 60% or more participants completed the study Weak = Withdrawals and dropouts not reported. Less than 60% completed the study

(Continued)

Table 2 (Continued).

		Not applicable = All cross-sectional research designs
Intervention integrity	(Q1) What percentage of participants received the allocated intervention or exposure of interest? 80%–100%/60%–79%/Less than 60%/Can't tell (Q2) Was the consistency of the intervention measured? Yes/No/Can't tell (Q3) Is it likely that subjects received an unintended intervention (contamination or co-intervention) that may influence the results? Yes/No/Can't tell	Strong = Consistency of intervention measured AND no reason to believe participants received an unintended intervention beyond regular daily life Moderate = No reason to believe participants received an unintended intervention beyond regular daily life Weak = Reason to believe participants received an unintended intervention beyond regular daily life Not applicable = All cross-sectional research designs
Analyses	(Q1) What statistical methods are used? (Q2) Are the statistical methods appropriate for the study design? Yes/No/Can't tell	Strong = Uses statistical modeling Moderate = Uses descriptive statistics Weak = Presents results with no statistical analysis or unclear analysis
Global rating		
List all component ratings		Strong = no weak ratings on any component Moderate = 1 weak rating on a component Weak = 2+ weak ratings on any components

Adapted by the authors from Effective Public Healthcare Panacea Project (2021). Criteria and guiding questions are from original EPHPP tool (Effective Public Healthcare Panacea Project, 2021). Specific evaluation criteria were developed by the authors to appraise survey-based urban planning research.

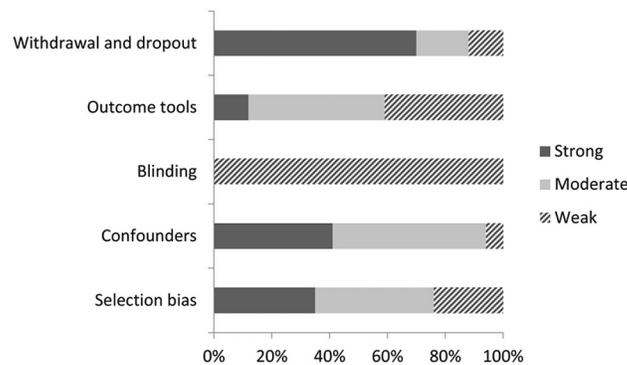


Figure 1. Sample risk of bias assessment from Figure 4 (p. 17) in Desrosiers et al. (2020), which is licensed under Creative Commons Attribution 4.0 International License: <https://creativecommons.org/licenses/by/4.0/>. Half of the original Figure 4 is presented herein.

However, qualitative research makes important contributions in our field, informs policy, and is often the subject of systematic reviews. Therefore, we encourage others to adapt tools to better suit qualitative research methods. *The Cochrane Handbook* recently included a chapter on qualitative quality assessments that emphasized four criteria: credibility, transferability, dependability, and confirmability (Hannes, 2011). The CASP has also developed a tool to appraise qualitative work which (see Table 5 in Technical Appendix A), and Kitchenham and Stuart (2007) provided a checklist for assessing qualitative studies. There is also vast literature on rigor in qualitative studies (Fusch & Ness, 2015; James, 2006; Small, 2009; Stratford & Bradshaw, 2016).

These tools and this literature may be important starting points for those willing to adapt RoBs in qualitative or even mixed-methods reviews.

Evaluation of Certainty of Evidence

The second tool frequently used in public health literature reviews that we have seldom seen in urban planning studies is the ECE (Guyatt et al., 2011). Whereas RoBs provide a quality assessment for each article included in a review, the ECE allows the authors to state how confident they are about the evidence as a whole. This evidence is grouped into all the different outcomes included in the review. ECE tools provide explicit, transparent, comprehensive, and structured processes for

Age group	Review characteristics	Effect	Certainty in the evidence
Pre-school (≤5 years)	–	No reviews identified for pre-school aged children.	N/A
Children (~5–11 years)	1 moderate + 1 critically low quality N = 4,255 (14 CS) 5 countries/regions CAC overlap = 0.0%	Among children, findings generally identified a positive (1 review ++, 1 review ??) association between AT to school and total PA, though there was some mixed evidence. There does not appear to be any clear differences by sex. We were unable to assess whether findings differed by measurement method of AT/PA. Study quality was not assessed for the AT studies.	●○○○ VERY LOW Initial rating: low – all studies cross-sectional RoB: one review of critically low quality and one of moderate quality (–0.5 pts) Inconsistency: the direction of effect was not the same between reviews, but could be partially explained by differences in AT and PA measurement (–0.5 pts) Indirectness: limited global representation and some studies included youth (–0.5 pts) Imprecision: few studies focused specifically on this age group, total sample size <10,000 (–1 pt)
Youth (~12–17 years)	1 moderate + 1 low + 2 critically low quality N = 34,133 (23 CS) 11 countries/regions CAC overlap = 2.9%	Among youth, findings identified a consistent positive (2 reviews ++, 1 review +, 1 review ??) association between AT to school and PA. There does not appear to be any clear differences by sex or method of AT/PA. Study quality was not assessed for the AT studies.	●○○○ VERY LOW Initial rating: low – all studies cross-sectional RoB: review quality ranged from critically low to moderate (–0.5 pts) Inconsistency: the direction of effect was consistent across reviews Indirectness: good global representation, but inclusion of younger ages in some reviews (–0.5 pts) Imprecision: total >10,000 across reviews
Children & youth (≤18 years)	1 high + 1 low + 3 critically low quality N = 389,684 (73 CS, 11 PCS, 1 RCT, 2 QE) 22 countries/regions CAC overlap = 28.2%	Among children and youth combined, findings showed a generally consistent positive (4 reviews ++, 1 review ??) association between AT to school and other destinations and PA. There does not appear to be any clear differences by age, sex or measurement of AT/PA. Study quality was assessed in the two larger reviews and was found to be of moderate-to-high quality.	●●●○ MODERATE Initial rating: high – while most evidence was cross-sectional, evidence from longitudinal and experimental studies found consistent positive associations with AT RoB: quality ranged from critically low to high (–0.5 pts) Inconsistency: the direction of effect was consistent across reviews Indirectness: good global representation and little-to-no inclusion of other ages Imprecision: total >10,000 across review)

Figure 2. Sample portion of evaluation of certainty of evidence. Source: Prince et al. (2021, p. 16).

rating the quality of the evidence. Therefore, ECEs provide clear guidance to practitioners about how confident the literature is about the identified relationship. The greater the number of studies included on a topic, and the higher quality the studies are, the higher the ECE will grade the topic (Guyatt et al., 2011; Mercuri et al., 2018).

GRADE (Grading of Recommendations, Assessment, Development and Evaluations) is a framework typically used to do this assessment (Guyatt et al., 2011). Using the GRADE approach, each outcome is assigned an initial certainty of evidence score based on study design.

Randomized controlled trials, natural experiments, and quasi-experimental studies are assigned an initial level of *high* certainty, whereas cross-sectional studies are assigned an initial level of *low* certainty. Then, different factors can lead to rating up or down the quality of evidence. The criteria that can rate the evidence down are those often included in RoBs: inconsistency, indirectness, imprecision, risk of bias, and publication bias. When the certainty of evidence is not downgraded, it can be upgraded if the following are observed: large effect, dose response, or opposing bias and confounders. Ultimately, the quality of

evidence for each outcome falls into one of four categories from high to very low (high, medium, low, and very low; Balshem et al., 2011; Guyatt et al., 2011).

Sample ECE Adapted for Survey Research in Urban Planning

As an example, Table 3 presents the ECE used in Prince et al.'s (2021) study, which examined the association between active transport and physical activity across the life course. The domain, judgment, scoring, and criteria are outlined in Table 3. To follow this table, one must first group all articles examining an outcome together. Then, for each outcome, natural experiments are separated from observational studies (e.g., cross-sectional studies). Depending on this study design, the evidence is given an original score of *high* or *low*.¹ Then, each domain is assessed using the criteria in Table 3. The resultant scores are added to the original certainty. The sum produces the final level of certainty. This must be calculated for each outcome and presented in the literature review.

Figure 2 illustrates the results of the ECE following the criteria in Table 3 (Prince et al., 2021). As was the case with the RoB, the criteria used to generate the ECE results (Table 3) are usually presented in an appendix or as supplementary material, while the results of the ECE (Figure 2) are embedded in the article in tabular form alongside a narrative synthesis. For instance, in this case the author could state that the evidence is graded as *very low* for children and youth, but *moderate* when studies consider both age categories together.

As was the case for RoBs, many adaptations can be made to the ECE we present here to better suit the field of urban planning. For instance, the total number of participants cutoffs in Figure 2 were generated from a textbook on survey methods that seek rigor through big samples and large-scale studies (Daniel, 2012). This can bias against smaller scale studies, many of which explore new and important topics. When this is the case, the author(s) of the literature review can be careful to highlight the impact of more original work in their narrative synthesis to counter this bias. Further, these criteria can be modified to best suit the types of studies included in a review, for instance, by reducing the sample-size cutoff criteria (in which case we recommend justifying the new criteria).

We envisage other potential modifications as well. For example, in the sample ECE presented here, all criteria are graded equally on a scale from -1 to 0 and all deductions are by 0.5 or 1 intervals. Depending on the study design, perhaps certain criteria can have more weight than others (e.g., indirectness scored on 2 points, whereas imprecision scored on 1 point). Further, to add more nuanced scores, deductions can be made by smaller intervals such as 0.25 or even 0.1 . These

adaptations can be modified to suit the study designs and methods included in a review.

Further, given that that criteria in ECEs are based off RoBs, and the RoB presented here is best suited for survey-based research, we urge researchers using other methods to also develop ECEs. We recommend GRADE as a starting point (Guyatt et al., 2011). However, modifications might be necessary. For instance, in studies using qualitative methods the number of participants may be less appropriate than, say, the richness of the interview or the study's sampling strategy.

Conclusions

In this Viewpoint we present two tools frequently used in public health literature review articles that hold great potential to help move the field of urban planning toward a more evidence-based approach to planning and to evaluating policies and projects. Because these tools have yet to be adapted for our discipline, we also showcase a version of each tool that is best suited for survey research in urban planning. We see three primary benefits to the incorporation of these tools into urban planning literature reviews. First, no research is without bias, regardless of discipline. The quality of studies in urban planning varies, and thus the evidence they produce should not be weighed equally. These tools provide clear guidance on how to rate the quality of articles and evaluate their impact on the emergent results. Given that the results of select past reviews and meta-analyses in urban planning have been quite contentious, spurring multiple responses to the author (e.g., M. Stevens, 2017), perhaps incorporating these tools into future reviews will result in more rigorous presentation of the evidence.

Second, many of the most pressing contemporary challenges, such as climate change, noncommunicable diseases, and pandemics, will require multidisciplinary collaboration. Urban planners will need to collaborate with public health and other city building officials to tackle many of these issues. To fully capitalize on this collaboration, the integration of the two fields remains a challenge. Learning, adapting, and incorporating some of the tools used in public health into urban planning is one of the small steps we can take to strengthen and ease this collaboration.

Finally, a frequent goal of literature reviews is to highlight policy implications, especially in practice-oriented fields such as urban planning (De Vos & El-Geneidy, 2021). We believe the tools discussed in this Viewpoint hold great potential to improve the generation of policy recommendations from urban planning reviews. By assessing each article included in a review (through an RoB) and by giving a clear level of certainty of the results (through an ECE), we can provide policymakers with more accurate syntheses of the literature on a topic and to help them

Table 3. Sample evaluation of certainty of evidence criteria.

Domain	Judgment	Scoring	Criteria
Risk of bias (review quality)	No serious RoB	0	≥ 75% of the reviews were considered high quality as assessed by the AMSTAR2 ^a
	Not serious, borderline RoB	-0.5 points	<ul style="list-style-type: none"> ≥ 75% of the reviews were considered low/moderate quality as assessed by the AMSTAR2^a < 75% of the reviews were considered critically low quality as assessed by the AMSTAR2^a
	Serious RoB	-1 point	≥ 75% of the reviews were considered critically low quality as assessed by the AMSTAR2 ^a
Inconsistency	No serious inconsistency	0	<ul style="list-style-type: none"> The direction and magnitude of the effect was consistent across reviews (> 75% in same direction) Heterogeneity could be explained by measurement of AT and PA (self-report vs. device)
	Not serious, borderline inconsistency	-0.5 points	<ul style="list-style-type: none"> The direction and magnitude of effect was inconsistent across reviews Heterogeneity could be partially (but not completely) explained by measurement of AT and PA (self-report vs. device)
	Serious inconsistency	-1 point	<ul style="list-style-type: none"> The direction and magnitude of the effect was inconsistent across reviews Heterogeneity could not be explained by measurement of AT and PA (self-report vs. device)
Indirectness	No serious indirectness	0	<ul style="list-style-type: none"> There was good global representation of primary studies included in the reviews There was little-to-no inclusion of other age groups captured in the reviews
	Not serious, borderline indirectness	-0.5 points	<ul style="list-style-type: none"> There was good global representation of primary studies included in the reviews There was considerable inclusion of other age groups captured in the reviews
	Serious indirectness	-1 point	<ul style="list-style-type: none"> There was limited global representation of primary studies included in the reviews There was considerable inclusion of other age groups captured in the reviews
Imprecision	No serious imprecision	0	<ul style="list-style-type: none"> The total number of participants across all studies was greater than 10,000^b
	Serious imprecision	-1 point	<ul style="list-style-type: none"> The total number of participants across all studies was less than 10,000^b
Publication Bias	No serious risk of publication bias	0	Due to the systematic and comprehensive search strategy including a scan of the gray literature, an a priori decision was made not to downgrade for risk of publication bias.

Notes: AT: active transportation; GRADE: Grading of Recommendations Assessment, Development and Evaluation; PA: physical activity; RoB: risk of bias. The quality of the evidence was upgraded if there is not cause to downgrade, and there was evidence of a large magnitude of effect from meta-analyses. Half-points were combined across domains to yield a total score; if the final scoring included a half-point, we conservatively rounded down (i.e., -0.5 points = -1 point). The quality of the evidence can be interpreted as follows: *High*—we are confident that the true direction of association between AT and PA lies close to the association we have estimated, and further research is very unlikely to change our confidence in the effect; *Moderate*—we are moderately confident that the true direction of association between AT and PA is likely to be close to the association we have estimated, but there is a possibility that it substantially different; further research is likely to have an important impact on the confidence in the direction of association and may change the direction of association; *Low*—we have limited confidence; the true direction of association between AT and PA may be substantially different from our estimate; *Very low*—we have very little confidence; the true direction of association between AT and PA may be substantially different from our estimate. ^aECE based off the AMSTAR2 RoB from Shea et al. (2017). ^bFrom Daniel (2012). Source: Supplemental Table 2 in Prince et al. (2021).

decide what types of policies to pursue. For instance, if the evidence for one intervention is weak because, say, only two studies found the expected relationship and one found no evidence for the relationship, policymakers will know to wait for further research before investing in that intervention.

Though these tools have not yet been tailored for the field of urban planning, we showcase herein a version of each of these tools that is applicable for survey-based research in our discipline. It is our hope not only that these tools are integrated into urban planning literature reviews, but that they are seen as a starting

point, as tools that can be further expanded on and refined to better reflect the needs of urban planning research. An obvious next step is to develop these tools to include different types of methods, especially those more common in qualitative studies, a type of research that makes important contributions to policy in our discipline. The potential impact of these tools on both research and practice is too important to ignore.

ABOUT THE AUTHORS

LÉA RAVENSBERGEN (lea.ravensbergen@mail.mcgill.ca) is a postdoctoral fellow at the Transport Studies Unit at the University of Oxford. **AHMED EL-GENEIDY** (ahmed.elgeneidy@mcgill.ca) is a professor at the School of Urban Planning at McGill University.

ORCID

Léa Ravensbergen  <http://orcid.org/0000-0003-3259-6673>
Ahmed El-Geneidy  <http://orcid.org/0000-0002-0942-4016>

ACKNOWLEDGMENTS

We thank our past collaborators working in public health disciplines, Rania Wasfi and Stephanie A. Prince, for introducing us to these quality assessment tools and encouraging us to incorporate them in our work.

RESEARCH SUPPORT

This research was funded by The Canadian Institutes of Health Research (CIHR) and The Natural Sciences and Engineering Research Council of Canada (NSERC) Collaborative Health Research Projects (CHRP) Program (project nos. CIHR CPG-170602 and CPG-170602 X- 253156, NSERC CHRPJ 549576-20).

SUPPLEMENTAL MATERIAL

Supplemental data for this article is available online at <https://doi.org/10.1080/01944363.2022.2074872>.

NOTE

1. The example in Figure 2 reviewed literature reviews; therefore, those that included less than 10% experimental and prospective study designs were initially assigned as "high" quality evidence, and those relying on cross-sectional evidence or with 10% or more experimental and prospective study designs were initially assigned as "low" quality evidence.

REFERENCES

Abu Hatab, A., Cavinato, M., Lindemer, A., & Lagerkvist, C. (2019). Urban sprawl, food security and agricultural systems in developing countries: A systematic review of the literature. *Cities*, 94, 129–142. <https://doi.org/10.1016/j.cities.2019.06.001>

Adkins, A., Makarewicz, C., Scanze, M., Ingram, M., & Luhr, G. (2017). Contextualizing walkability: Do relationships between built environments and walking vary by socioeconomic context? *Journal of the American Planning Association*, 83(3), 296–314. <https://doi.org/10.1080/01944363.2017.1322527>

Al-Jundi, A., & Sakka, S. (2017). Critical appraisal of clinical research. *Journal of Clinical and Diagnostic Research*, 11(5), JE01–JE05. <https://doi.org/10.7860/JCDR/2017/26047.9942>

Balshem, H., Helfand, M., Schünemann, H. J., Oxman, A. D., Kunz, R., Brozek, J., Vist, G. E., Falck-Ytter, Y., Meerpohl, J., Norris, S., & Guyatt, G. H. (2011). GRADE guidelines: 3. Rating the quality of evidence. *Journal of Clinical Epidemiology*, 64(4), 401–406. <https://doi.org/10.1016/j.jclinepi.2010.07.015>

Basu, N., Haque, M. I., King, M., Kamruzzaman, M., & Oviedo-Trespalacios, O. (2021). A systematic review of the factors associated with pedestrian route choice. *Transport Reviews*. Advance online publication. <https://doi.org/10.1080/01441647.2021.2000064>

Brownson, R., Fielding, J., & Maylahn, C. (2009). Evidence-based public health: A fundamental concept for public health practice. *Annual Review of Public Health*, 30(1), 175–201. <https://doi.org/10.1146/annurev.publhealth.031308.100134>

Calderón-Argelich, A., Benetti, S., Anguelovski, I., Connolly, J., Langemeyer, J., & Baró, F. (2021). Tracing and building up environmental justice considerations in the urban ecosystem service literature: A systematic review. *Landscape and Urban Planning*, 214, 104130. <https://doi.org/10.1016/j.landurbplan.2021.104130>

Cao, X., Mokhtarian, P., & Handy, S. (2009). Examining the impacts of residential self-selection on travel behaviour: A focus on empirical findings. *Transport Reviews*, 29(3), 359–395. <https://doi.org/10.1080/01441640802539195>

CASP. (2021). *CASP checklists*. <https://casp-uk.net/casp-tools-checklists/>

The Cochrane Collaboration. (2011). *Cochrane handbook for systematic reviews of interventions* (Vol. 5.1.0). www.handbook.cochrane.org

Daniel, J. (2012). Choosing the size of the sample. In J. Daniel (Ed.), *Sampling essentials: Practical guidelines for making sampling choices*. SAGE Publications, Inc.

Desrosiers, A., Betancourt, T., Kergoat, Y., Servilli, C., Say, L., & Kobeissi, L. (2020). A systematic review of sexual and reproductive health interventions for young people in humanitarian and lower-and-middle-income country settings. *BMC Public Health*, 20(1), 666. <https://doi.org/10.1186/s12889-020-08818-y>

De Vos, J., & El-Geneidy, A. (2021). What is a good transport review paper? *Transport Reviews*, 41(1), 1–5. <https://doi.org/10.1080/01441647.2021.2001996>

Dobbins, M., Thomas, H., O'Brien, M., & Duggan, M. (2004). Use of systematic reviews in the development of new provincial public health policies in Ontario. *International Journal of Technology Assessment in Health Care*, 20(4), 399–404. <https://doi.org/10.1017/S0266462304001278>

Effective Public Healthcare Panacea Project. (2021). *Quality assessment tool for quantitative studies*. <https://www.epphp.ca/quality-assessment-tool-for-quantitative-studies/>

Fusch, P., & Ness, L. (2015). Are we there yet? Data saturation in qualitative research. *The Qualitative Report*, 20(9), 1408.

Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information and Libraries Journal*, 26(2), 91–108.

Guyatt, G., Oxman, A. D., Akl, E. A., Kunz, R., Vist, G., Brozek, J., Norris, S., Falck-Ytter, Y., Glasziou, P., DeBeer, H., Jaeschke, R., Rind, D., Meerpohl, J., Dahm, P., & Schünemann, H. J. (2011). GRADE guidelines: Introduction-GRADE evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology*, 64(4), 383–394.

- Hannes, K. (2011). Critical appraisal of qualitative research. In J. Noyes, A. Booth, K. Hannes, A. Harden, J. Harris, S. Lewin, & C. Lockwood (Eds.), *Supplementary guidance for inclusion of qualitative research in Cochrane systematic reviews of interventions. Version 1 (updated August 2011)*. Cochrane Collaboration Qualitative Methods Group.
- Jackson, R. J., Dannenberg, A. L., & Frumkin, H. (2013). Health and the built environment: 10 years after. *American Journal of Public Health, 103*(9), 1542–1544.
- James, A. (2006). Critical moments in the production of “rigorous” and “relevant” cultural economic geographies. *Progress in Human Geography, 30*(3), 289–308. <https://doi.org/10.1191/0309132506ph610oa>
- Kitchenham, B., & Stuart, C. (2007). *Guidelines for performing systematic literature reviews in software engineering*. https://www.elsevier.com/_data/promis_misc/525444systematic-reviewsguide.pdf
- Krizek, K., Forysth, A., & Slotterback, C. (2009). Is there a role for evidence-based practice in urban planning and policy? *Planning Theory & Practice, 10*(4), 459–478. <https://doi.org/10.1080/14649350903417241>
- Ma, L.-L., Wang, Y.-Y., Yang, Z.-H., Huang, D., Weng, H., & Zeng, X.-T. (2020). Methodological quality (risk of bias) assessment tools for primary and secondary medical studies: What are they and which is better? *Military Medical Research, 7*(1), 7. <https://doi.org/10.1186/s40779-020-00238-8>
- Mercuri, M., Baigrie, B., & Upshur, R. E. (2018). Going from evidence to recommendations: Can GRADE get us there? *Journal of Evaluation in Clinical Practice, 24*(5), 1232–1239.
- Prince, S., Lancione, S., Lang, J., Amankwah, N., de Groh, M., Garcia, A., Merucci, K., & Geneau, R. (2021). Are people who use active modes of transportation more physically active? An overview of reviews across the life course. *Transport Reviews*. Advance online publication. <https://doi.org/10.1080/01441647.2021.2004262>
- PRISMA. (2021). *Preferred reporting items for systematic reviews and meta-analyses (PRISMA): Transparent reporting of systematic reviews and meta-analyses*. <http://www.prisma-statement.org/>
- Shea, B., Reeves, B., Wells, G., Thuku, M., Hamel, C., Moran, J., Moher, D., Tugwell, P., Welch, V., Kristjansson, E., & Henry, D. (2017). AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *British Medical Journal, 358*, j4008. <https://doi.org/10.1136/bmj.j4008>
- Small, M. (2009). “How many cases do I need?”: On science and the logic of case selection in field-based research. *Ethnography, 10*(1), 5–38. <https://doi.org/10.1177/1466138108099586>
- South, E., & Lorenc, T. (2020). Use and value of systematic reviews in English local authority public health: A qualitative study. *BMC Public Health, 20*(1), 1100 <https://doi.org/10.1186/s12889-020-09223-1>
- Stevens, K. (2001). Systematic reviews: The heart of evidence-based practice. *AACN Clinical Issues, 12*(4), 529–538. <https://doi.org/10.1097/00044067-200111000-00009>
- Stevens, M. (2017). Does compact development make people drive less? *Journal of the American Planning Association, 83*(1), 7–18. <https://doi.org/10.1080/01944363.2016.1240044>
- Stratford, E., & Bradshaw, B. (2016). Rigorous and trustworthy: Qualitative research design. In I. Hays & M. Cope (Eds.), *Qualitative research methods in human geography* (pp. 92–106). Oxford University Press.
- van Wee, B., & Banister, D. (2016). How to write a literature review paper? *Transport Reviews, 36*(2), 278–288. <https://doi.org/10.1080/01441647.2015.1065456>
- Xiao, Y., & Watson, M. (2019). Guidance on conducting a systematic literature review. *Journal of Planning Education and Research, 39*(1), 93–112. <https://doi.org/10.1177/0739456X17723971>