SSM - Population Health 17 (2022) 101028

Contents lists available at ScienceDirect

SSM - Population Health

journal homepage: www.elsevier.com/locate/ssmph

# The challenges of measuring social cohesion in public health research: A systematic review and ecometric meta-analysis

Moritz Oberndorfer [a,b,*], Thomas E. Dorner [a,c], Alastair H. Leyland [b], Igor Grabovac [a], Thomas Schober [d], Lukas Šramek [e], Marcel Bilger [f]

[a] Department of Social and Preventive Medicine, Centre for Public Health, Medical University of Vienna, Vienna, Austria
[b] MRC/CSO Social and Public Health Sciences Unit, University of Glasgow, Glasgow, UK
[c] Karl-Landsteiner Institute for Health Promotion Research, BVAEB-Health Promotion Facility Resilienzpark Sitzenberg, Sitzenberg-Reidling, Austria
[d] School of Government and Public Policy, Department of Politics, University of Strathclyde, Glasgow, UK
[e] Department of Infectious Disease Epidemiology & Surveillance, Austrian Agency for Health and Food Safety, AGES, Vienna, Austria
[f] Health Economics and Policy, Vienna University of Economics and Business, Vienna, Austria

ARTICLE INFO

*Keywords:*
Social cohesion
Social environment
Contextual effects
Contextual characteristics
Ecometrics
Intracluster correlation coefficient
Multilevel modelling
Meta-analysis

ABSTRACT

The relationship between social cohesion and health has been studied for decades. Yet, due to the contextual nature of this concept, measuring social cohesion remains challenging. Using a meta-analytical framework, this review's goal was to study the ecometric measurement properties of social cohesion in order to describe dissimilarities in its measurement as well as bring a new perspective on the empirical usefulness of the concept itself. To this end, we analysed if, and to what extent, contextual-level reliability and intersubjective agreement of 78 social cohesion measurements varied under different measurement conditions like measurement instrument, spatial unit, ecometric model specification, or region. We found consistent evidence for the contextual nature of social cohesion, however, most variation existed between individuals, not contexts. While contextual dependence in response behaviour was fairly insensitive to item choices, population size within chosen spatial units of social cohesion measurements mattered. Somewhat counterintuitively, using spatial units with, on average, fewer residents did not yield systematically superior ecometric properties. Instead, our results underline that precise theory about the relevant contextual units of causal relationships between social cohesion and health is vital and cannot be replaced by empirical analysis. Although adjustment for respondent's characteristics had only small effects on ecometric properties, potential pitfalls of this analytic strategy are discussed in this paper. Finally, acknowledging the sensitivity of measuring social cohesion, we derived recommendations for future studies investigating the effects of contextual-level social characteristics on health.

## 1. Introduction

### 1.1. Social cohesion and public health in brief

Originating from Émile Durkheim's ideas on social dynamics and suicide rates (Durkheim, 2006) in the late 19th century and most prominently revived by a "neo-Durkheimian" research agenda in the 1990s (Muntaner & Lynch, 1999; Wilkinson, 1997), the concept of social cohesion aims to describe aspects of the social environments we live in. Social cohesion is explicitly conceived as an attribute of a contextual unit – e.g. a community, a neighbourhood, a district, a state, or a nation state – and thus, although related, distinguishable from other social determinants of health that operate on the individual level.

Developing coherent and testable multilevel theories of contextual effects on health presents a long-known challenge to public health researchers (Macintyre et al., 2002; Oakes, 2004), with several mechanisms being proposed so far.

Social cohesion is hypothesised to reduce and buffer stress (Chuang et al., 2013; Roux & Mair, 2010) which, when chronic, can induce immune dysregulation through sustained inflammatory responses (Glaser & Kiecolt-Glaser, 2005). Apart from providing access to resources like social support, close social relations may protect from loneliness

* Corresponding author. Department of Social and Preventive Medicine Centre for Public Health Medical University of Vienna Kinderspitalgasse 15/1, 1090, Vienna, Austria.
*E-mail address:* moritz.oberndorfer@muv.ac.at (M. Oberndorfer).

https://doi.org/10.1016/j.ssmph.2022.101028
Received 22 November 2021; Received in revised form 13 January 2022; Accepted 13 January 2022
Available online 21 January 2022

(Holt-Lunstad, 2018), transmit positive health behaviours through established social norms (Smith & Christakis, 2008), and support the diffusion of relevant health information (Chuang et al., 2013). However, living in a tight-knit community can also negatively affect health through the potential spreading of infectious diseases or social transmission of adverse health behaviours (Christakis & Fowler, 2011; Villalonga-Olives & Kawachi, 2017). Further, social entities in which people share close relationships may adversely affect health through excessive group conformity, high demands to provide support to others, feelings of restricted individual freedom, exclusion of outsiders and inter-group conflicts (Portes, 1998, 2014; Villalonga-Olives & Kawachi, 2017). Moreover, social cohesion is related to a sense of belonging (Pérez et al., 2020), which has been connected to reduced suicidality and depression (Durkheim, 2006; Fisher et al., 2015; Hagerty et al., 1992; Hatcher & Stubbersfield, 2013). Another feature of social cohesion, orientation towards the common good, may facilitate collective cooperative behaviour through which communities achieve social order and safety (Sampson et al., 1997), or even influence public health policy (Villalonga-Olives & Kawachi, 2017) as well as modify the built environment in their spatial context (Duncan and Kawachi, 2018).

### *1.2. "Redefining", "revisiting", "reconsidering" social cohesion*

Continuously, studying the effects of social cohesion has been accompanied by a multidisciplinary discussion of the concept itself (Bernard, 2000, p. 26; Chan et al., 2006; Fonseca et al., 2019; Schiefer & van der Noll, 2017). Along with a struggle for unified definitions, different measurements inevitably emerged (Chan et al., 2006; Dickes & Valentova, 2013; Fonseca et al., 2019). Following this debate, a recent literature review on the definition of social cohesion provided an "essentialist" definition. The authors of this review concluded that a cohesive social environment *"is characterized by close social relations, pronounced emotional connectedness to the social entity, and a strong orientation towards the common good."* (Schiefer & van der Noll, 2017).

Although social cohesion is unanimously considered as a characteristic of a collective, the scientific literature about the influence of social environments on health has additionally evolved around different related concepts, such as social capital, or social networks (Berkman et al., 2000; Kawachi et al., 2014; Kawachi & Subramanian, 2018; Lochner et al., 1999; Marmot and Wilkinson, 2006; Moore & Kawachi, 2017). (see Lochner et al., 1999) for a detailed comparison of social capital as an ecological variable and its closely related concepts). Especially the relationship between social cohesion and social capital can be complex because the concept of social capital itself is coined by a persistent multidisciplinary discussion about definition and measurement (Kawachi et al., 2014; Moore & Kawachi, 2017).

Due to the conceptualisation of social capital as a social resource that both individuals and groups have access to, there is a varying degree of overlap between social capital and social cohesion which further depends on the varying definitions of each concept. Following the argumentation of Ichiro Kawachi and other leading figures in social capital research, social cohesion can be seen as an approach to social capital which itself can be further divided into several subcategories (cognitive, structural, bonding, bridging, and linking social capital) (Kawachi et al., 2014; Moore & Kawachi, 2017). At the same time social cohesion is considered as an even broader concept than social capital (Kawachi et al., 2014; Moore & Kawachi, 2017). In this view, social cohesion emphasizes social capital as a collective attribute, whereas the social network approach to social capital focuses on resources that are accessible to individuals through their (egocentric) social networks. Although social capital did not comprise the concept of social cohesion when developed by Pierre Bourdieu in the 1980s and thus still allowed for a conceptual dichotomy of having (social capital) and being (social cohesion), the translation of social capital into public health research inspired by James Coleman's and Robert Putnam's conceptualisations led to a conflation of these concepts (Carrasco & Bilal, 2016). Interestingly, a citation-network path analysis of the public health literature on social capital in 2006 (Moore et al., 2006) showed that, at that time, studies on social capital were mostly using the social cohesion approach while collectively citing Durkheim's work on suicide as a their foundation – much like the social cohesion literature (Muntaner & Lynch, 1999; Wilkinson, 1997). This genealogy of social capital provides another explanation for why social cohesion and this specific approach to social capital are hardly, if at all, distinguishable: both concepts were initially introduced to public health research to meet the same explanatory demand of data on income inequality and its relationship to health in the 1990s (Moore et al., 2006). Additionally, the explanation investigators were looking for, had to be psychosocial in character (because income inequality is proposed to affect health in affluent societies) and operate at the ecological, not individual, level (Moore et al., 2006).

While the multidisciplinary and prominent use of social cohesion in the scientific discourse is welcomed, the obtained breadth of the term "social cohesion" might dilute its explanatory power for empirical analyses (Bernard, 2000, p. 26). Consequently, we argue that after "redefining", "revisiting", and "reconsidering" social cohesion has left a legacy of confusion and an abundance of empirical studies investigating the effects of social cohesion on health (-behaviour) (Glonti et al., 2016; Mair et al., 2008; Samuel et al., 2013; Sawyer et al., 2017), this field of research deserves a methodological review. As a contextual-level characteristic, the measurement and empirical analysis of social cohesion requires a specific set of research methods labelled "ecometrics". Ecometrics is described as a "theory of method" insisting *"(…) that neighbourhood, community and other collective phenomena demand their own measurement logic and are not stand-ins for individual-level traits"* (Sampson, 2012, p. 360). For readers unfamiliar with ecometrics, we provide a brief explanation of the ecometric properties central to our analyses in the following section. This section should not be understood as a comprehensive introduction to ecometrics, but facilitate a common ground for the interpretation of our data.

### *1.3. Ecometrics – "the science of assessing ecological settings"*

Much like psychometrics, ecometrics is interested in traits that are not directly observable, however, its unit of analysis are contexts, not individuals. Ecometrics is a measurement theory and method that aims to assess characteristics of contexts by combining psychometric principles with multi-level modelling. Data sources are often surveys, but the ecometric approach is also well equipped to deal with data generated by systematic social observation. Generally, whenever individual item responses or observations are used to assess characteristics of the contexts to which they belong, ecometrics offer a powerful analytical framework.

Using the ecometric toolbox when studying the effects of social or physical contexts on individual-level variables, like health outcomes, researchers are enabled to deal with two main analytical challenges of contextual effects: perception and same source bias. Both of these biases would distort inferences if contextual characteristics like social cohesion were measured on the individual level via survey responses. As such, given the measurement instrument has acceptable validity, these item responses merely reflect individual perception of social cohesion and do not measure a contextual characteristic. Naturally, it is plausible that these individual perceptions are a function of observable as well as unobservable individual-level traits and do not capture the "true" level of social cohesion within a context (perception bias). Moreover, and especially relevant in public health research, estimating the effect of individual perception of social cohesion on the level of health and well-being reported by the same individual may very well be confounded by individual level traits that drive both perception of the quality of social environments as well as the perception and actual level of health and well-being (same source bias).

#### 1.3.1. The ecometric measurement model

Even though other approaches to mitigate these biases are conceivable, an ecometric measurement model is geared towards such challenges. It conceptualises individuals as informants of the context they belong to. Thus, individuals are nested within contexts while their item responses are nested within individuals. This conceptualisation leads to a three-level measurement model that lies at the core of ecometrics (all equations and their explanation are drawn from Sampson et al. (1997): Wherein, on level 1 (equation (1)), $Y_{ijk}$ is the $i$th response of individual $j$ in context $k$ and $\pi_{jk}$ is the "true" perception of individual $j$ of the focal contextual trait. Similar to item response modelling, $D_{pijk}$ indicates response $i$ to its respective item $p$ of the measurement instrument by individual $j$ in context $k$. Therefore, the coefficient $\alpha_p$ can be understood as the "difficulty" of item $p$ for which "true" individual perception is adjusted. The measurement error $e_{ijk}$ is assumed to be normally distributed with a variance $\sigma^2$.

$$Y_{ijk} = \pi_{jk} + \sum_{p=1}^{p_n} \alpha_p D_{pijk} + e_{ijk}, \quad e_{ijk} \sim N(0, \sigma^2) \tag{1}$$

On level 2, the individual-level, an individual's "true" perception $\pi_{jk}$ is modelled as a function of the "true" level of the contextual trait $\eta_k$ in context $k$ (equation (2)). Additionally, this perception may be adjusted for characteristics $q$ of respondent $j$ that make up the composition of context $k$. Thus, $X_{qjk}$ represents the level of variable $q$ for individual $j$ nested within context $k$ while its coefficient $\delta_q$ is the effect of variable $q$ on individual perception $\pi_{jk}$ of respondent $j$ in context $k$. The random effect $r_{jk}$ is assumed to be normally distributed with the "within-context variance" $\tau_\pi$.

$$\pi_{jk} = \eta_k + \sum_{q=1}^{q_n} \delta_q X_{qjk} + r_{jk}, \quad r_{jk} \sim N(0, \tau_\pi) \tag{2}$$

Finally, on the contextual-level 3, the "true" level of the contextual-level trait $\eta_k$ of context $k$ – our quantity of interest – is a function of the grand mean of the contextual characteristic of interest $\gamma$ and the random effect $u_k$ which is again assumed to be normally distributed with variance $\tau_\eta$. The variance of the random effect, $\tau_\eta$, is the variance in the contextual trait that exists between contexts - the so-called "between-context variance".

$$\eta_k = \gamma + u_k, \quad u_k \sim N(0, \tau_\eta) \tag{3}$$

Note that this is the ecometric measurement model described by Sampson et al. (1997) and that other authors sometimes use different model specifications and notation. In addition, also dependent on the scales of item responses, different estimators have been suggested (Fone et al., 2006; Raudenbush & Sampson, 1999). However, while there might be an argument against the assumptions of the random effects described above (see Hipp & Steenbeek, 2016), the three-level structure does not only attempt to tackle the issues of perception and same source bias outlined above, but also yields estimates of the "within-context variance" and "between-context variance" that allow investigators to study informative qualities of their measurement. The obtained variance components, together with other information about the data structure, can be used to calculate the contextual-level reliability of the measurement and the inter-subjective agreement (ICC) among respondents nested within contexts.

#### 1.3.2. The intracluster correlation coefficient (ICC)

In ecometrics, the ICC uses the variance components obtained by estimating an ecometric measurement model in order to calculate the proportion of the total variance in individual perception that is due to contextual and not individual differences (equation (4)). Thus, it takes values between 0 and 1 and gives us information about the extent of dependence on context in item responses of individuals. The closer to 1, the more variance exists between contexts in relation to variation within contexts.

$$ICC = \frac{\tau_\eta}{\tau_\eta + \tau_\pi} \tag{4}$$

An ICC of 1 would indicate that, within different contexts, every respondent agrees with every other respondent about the level of the contextual characteristic of interest, while, at the same time, different contexts have different levels of this characteristic (Fig. 1). This case would indicate that all of the variation in individual perception is due to contextual level variation. Essentially, for an ICC of 1, the contextual level latent trait that we assume to give rise to item responses has to influence response behaviour strongly enough to eliminate any doubt about its true level among respondents. In this case, it is fair to assume the existence of a latent contextual characteristic that decisively drives individual perception of what we purport to measure. Obviously, such a scenario will most likely never be observed when studying latent traits like social cohesion.

On the contrary, an ICC of 0 indicates that all of the variation in individual perception of the contextual characteristic lies between respondents nested within contexts (Fig. 2). If this is the case, response behaviour of individuals is entirely independent of the chosen contextual level, which violates a central assumption of the ecometric measurement model: observations are not independent from each other, but nested within contexts.

However, an ICC of 0 does not necessarily mean that there is no between-context variation at all. In fact, it solely tells the observer that individual perception is not dependent on the particular contextual level that was chosen for data collection and analysis. Altering the definition of context by changing the spatial unit on which the contextual trait in question is hypothesised to operate could yield entirely different variance components and thus another ICC. Such a scenario is depicted in Fig. 3, wherein we exemplarily rearranged individuals within contexts. It illustrates the extreme case in which the same data can yield an ICC of 0 or an ICC of 1 merely by changing the definition of context itself. This issue in ecometrics is similar to the well-known Modifiable Areal Unit Problem (MAUP) in human geography.

#### 1.3.3. The contextual-level reliability

The contextual-level reliability, often denoted by λ (lambda), does not only depend on the within-context and between-context variance of the contextual characteristic, but also includes information about the underlying data structure of the measurement. Like the ICC, it takes values between 0 and 1. Contrary to the ICC, multiple different approaches to estimate the contextual-level reliability can be found in the literature. Despite differences, to our knowledge, all approaches take the number of respondents per context $n_{jk}$ into account. The more respondents are informing the contextual-level measurement, the more reliable the measurement gets. While Raudenbush and colleagues (Raudenbush & Sampson, 1999; Raudenbush et al., 1991, 2003) as well as Hox (Hox, 2010) and Leyland & Groenewegen (Leyland & Groenewegen, 2020) (who refer to Raudenbush's work) also include the item inconsistency $\sigma^2$ and number of items $n_p$that are included in the instrument - as shown in equation (5) -, some approaches omit these quantities (Mujahid et al., 2007; Stafford et al., 2003). Interestingly, back in 1990, O'Brien discussed the estimation of the reliability of "aggregate-level variables" under different scenarios (O'Brien, 1990). For one scenario, in which multiple respondents rate one context, he proposed a method that omits the number of items and their inconsistency (eq. 7 in O'Brien, 1990). Seemingly coincidental, another method he proposed, for the scenario wherein aggregate scores are based on multiple respondents nested within different interviewers (eq. 11 in O'Brien, 1990), is identical to equation (5) if one exchanges respondents nested within interviewers with item responses nested within respondents.

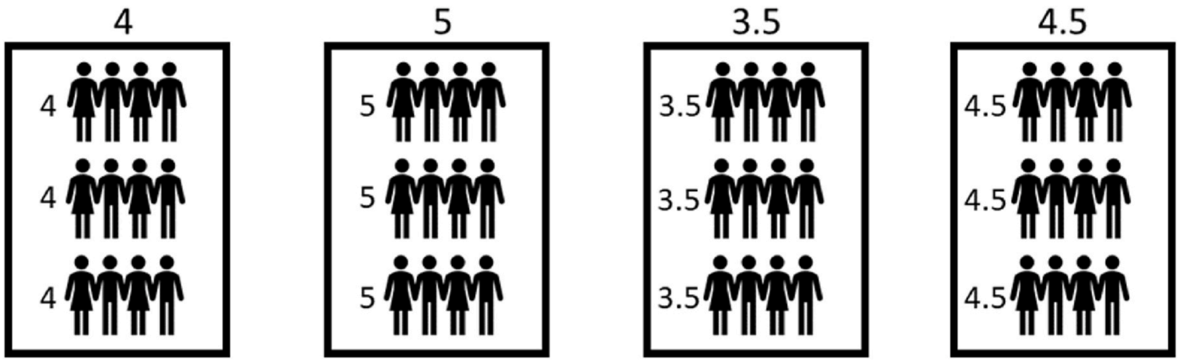


**Fig. 1.** A scenario in which the ICC would be 1. Within each context, all respondents agree on the level of a contextual characteristic (values within each box), thus, the within-variance equals 0. However, contexts vary in their level of this characteristic (values above each box). In the case illustrated above, the between-variance equals 0.3125. Following equation (4) the ICC equals 0.3125/(0.3125 + 0) = 1.

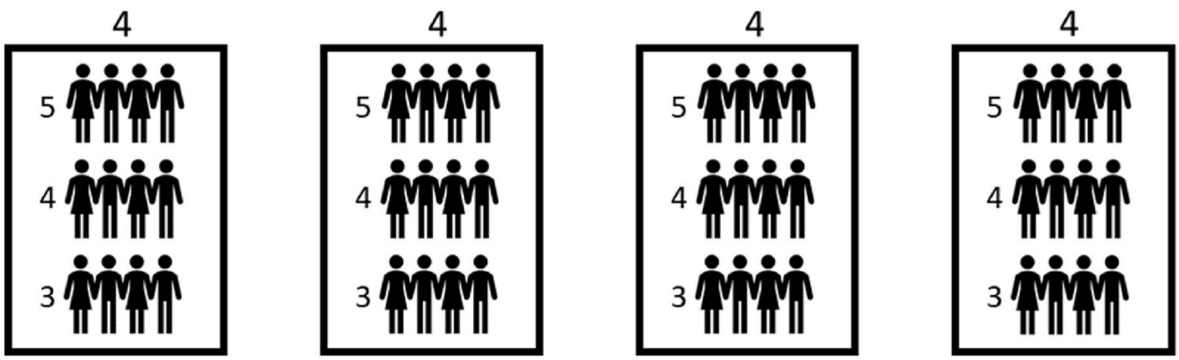


**Fig. 2.** A scenario in which the ICC would be 0. Respondents answer differently within contexts (values next to individuals) but the level of the contextual characteristic (value above each box) does not vary across contexts. Thus, item responses of individuals are independent of the chosen contextual level and there is no variation between contexts, only within contexts.

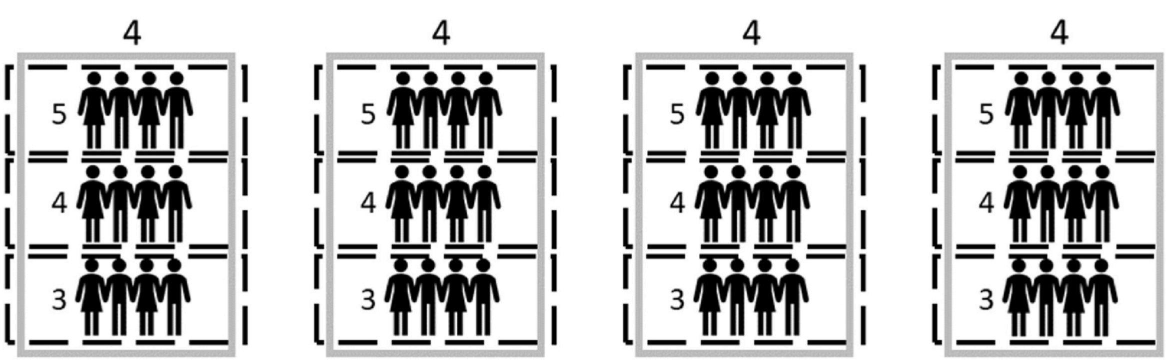


**Fig. 3.** The between-context variation is also dependent on the chosen contextual level. Redefining the contextual level chosen in Fig. 2 (black dashed lines instead of grey lines) yields an entirely different ICC compared to the ICC in the scenario illustrated in Fig. 2.

$$\lambda_k = \frac{\tau_\eta}{\tau_\eta + \frac{\tau_\pi}{n_{jk}} + \frac{\sigma^2}{n_{jk} n_p}} \tag{5}$$

Note that equation (5) estimates the reliability of each context $k$ and thus can also inform the analyst about the relationship between aspects of study design $n_{jk}$, $n_p$, and the reliability $\lambda_k$. Because, often, the average reliability across all contexts will be of higher interest, Raudenbush et al., Hox and Leyland & Groenewegen suggest to exchange number of respondents $n_{jk}$ in context $k$ with the average number of respondents per context in equation (5) (Hox, 2010; Leyland & Groenewegen, 2020; Raudenbush et al., 1991).

Therefore, irrespective of which method is used to calculate the contextual-level reliability, it contains information about the suitability of a study to investigate contextual characteristics. Obviously, a contextual characteristic cannot be reliably measured by only asking a very small number of individuals per context. Depending on the method chosen, $\lambda_k$ also accounts for the fact that more items can increase the reliability of the measurement. However, as illustrated by previous work (Raudenbush et al., 2003; Raudenbush & Sampson, 1999), increasing the number of respondents per context or items included in the measurement yields diminishing returns for the contextual-level reliability.

The contextual-level reliability can also be used to tackle measurement error that arises if the data includes only a small number of respondents in some of the analysed contexts. For a discussion and application of this and other advanced topics of ecometrics, the reader is referred to prior work (Mujahid et al., 2008; Savitz & Raudenbush, 2009). For an accessible and detailed explanation of ecometrics in general, we suggest Raudenbush & Sampson, 1999, Raudenbush, 2003, and Leyland & Groenewegen, 2020; and Fone et al., 2006, Mujahid et al., 2007, and Mujahid et al., 2008 for thoughtful applications of ecometrics in public health research.

### 1.4. From ecometrics to meta-ecometrics

By estimating an ecometric measurement model of the contextual-level trait of interest, researchers can not only overcome some of the analytical challenges of contextual effects, but also obtain valuable empirical information about their measurement, much like in the related field of psychometrics.

Given that researchers have now utilised these methodological

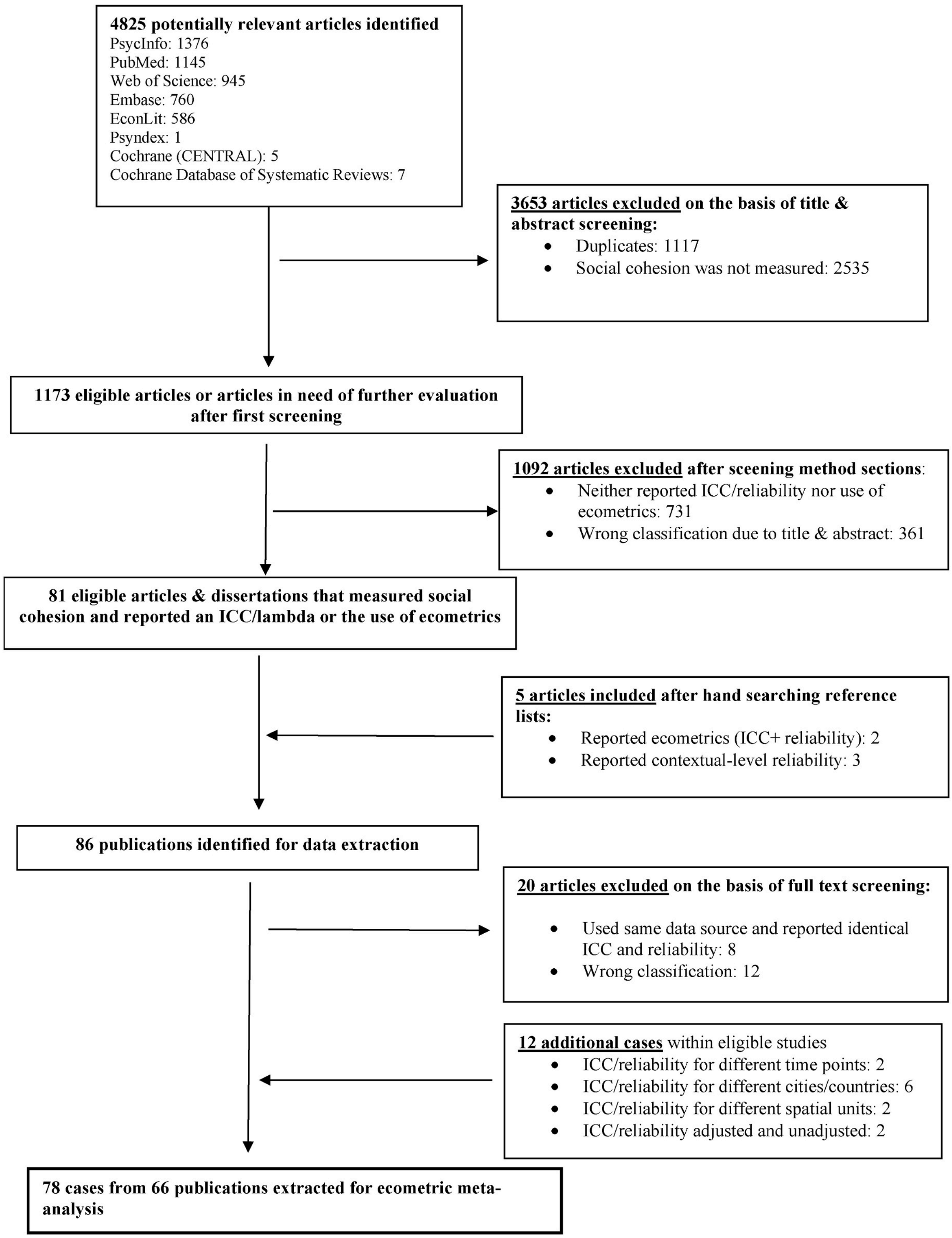


**Fig. 4.** Study selection process.

advances for over two decades, we aimed to systematically review and meta-analyse the ecometric properties of social cohesion measurements estimated by previous studies. We believe that a meta-analytical approach to ecometrics can provide answers to conceptual questions about contextual characteristics like social cohesion that can hardly be dealt with in a single study. By combining the available ecometric properties of social cohesion measurements, especially the intracluster correlation coefficient (ICC), we are able to gain new insights into the usefulness of the concept "social cohesion" in empirical analysis. More precisely, we investigate to what extent the ecometric properties of social cohesion are dependent on study characteristics like measurement instrument, spatial unit, ecometric model specification, and study population. After the presentation of our findings and their careful interpretations, we discuss theoretical as well as methodological considerations that result from our review. Finally, we derive recommendations for future studies investigating the effects of contextual-level social characteristics.

## 2. Methods

### 2.1. Search strategy and selection criteria

Three investigators searched for eligible publications in Web of Science, Embase, Medline, PsychInfo, Psyndex, Cochrane Database of Systematic Reviews, Cochrane Central Register for Controlled Trials, and EconLit. The search strategy was generated in close collaboration with an information retrieval specialist of the Medical University of Vienna. Our search strategy was designed using an iterative process. We started by using the most important search terms ("social cohesion", "measurement") and screened random samples of 25–50 publications in each database for eligibility. When our first searches were too broad and mostly failed to identify eligible studies, we further increased their precision by adding additional terms that hint towards some kind of statistical analysis of social cohesion. Our final search strategy included the search terms "social cohesion", "neighbourhood social cohesion", and "neighbourhood cohesion" (in title, abstract, or key words) in combination with various terms (statistic*, psychometric*, ecometric*, econometric*) which indicate that social cohesion was measured and quantitative analysis was carried out.

The results of these searches were screened by three contributors independently. Title and abstract were first screened simultaneously. In a second step, the method section of those studies with disagreement were screened in order to identify whether these studies included a measure of social cohesion. Any discrepancies were discussed among all three contributors until a decision was reached.

In the course of the screening process, we had to adjust our selection criteria. First, we screened the literature for peer-reviewed studies in which explicitly "social cohesion" or "neighbourhood (social) cohesion" was measured in any population. No limitations for quantitative study designs nor geographical limitations applied, but only publications written in English or German were eligible. Due to an overwhelming number of studies that measured social cohesion (Fig. 4), we focused on those, which measured social cohesion as a contextual level characteristic and thereby engaged in multilevel modelling or applied ecometrics. Narrowing the inclusion criteria, as stated in a contingency statement written in our study protocol, we were able to conduct quantitative analyses on the ecometric measurement properties. In order to increase the sample size of studies that estimated ecometric properties of their measurement, we decided to also include eligible published peer-reviewed dissertations. After arriving at our final set of publications, we hand-searched the reference lists for any further studies that met the inclusion criteria.

Studies published between 01.01.1990 and 15.07.2020 were considered for this review.

### 2.2. Ecometric meta-analysis

The following information was extracted from included studies: how social cohesion was measured (measurement instrument), data source of the measurement (dataset, location, and date), average number of residents in each context, number of contexts included, sample size, ecometric properties of the measurement (intracluster correlation, contextual-level reliability), internal consistency of the measurement (Cronbach's alpha), perception bias correction, study population. In case the authors were not able to extract some of the information listed above, we contacted the corresponding author of the respective study asking for the missing information.

The main goal of our quantitative analysis was to study the variation in ecometric properties of social cohesion measurements and their potential sources. Although we are aware that estimating measures of uncertainty for the ICC is possible, the vast majority of included studies were lacking the necessary information. Therefore, we have decided to weigh every included study equally, but did not provide pooled estimates. Rather, we analysed the distribution of these properties across studies by means of data tabulation and visualisation.

Both ecometric properties – the contextual-level reliability (lambda) and the inter-subjective agreement (ICC) – although related, convey separate information about social cohesion and its respective measurement. The ICC estimated by the ecometric measurement model tells us how much of the variance in individual perception of social cohesion is attributable to the chosen contextual level. In other words, the ICC of a contextual-level trait informs us to what extent individual response behaviour is dependent on the chosen context of respondents. The estimated reliability, on the other hand, additionally includes information about the data structure and measurement instrument. Thus, the contextual-level reliability informs us about the suitability of the data to measure contextual characteristics, whereas the ICC holds subtle, though fundamental, cues for the concept itself. Because meta-ecometrics is meant to uncover new aspects of contextual-level traits themselves instead of just reviewing the data quality of previous studies, our analysis mainly focuses on the ICC of social cohesion measurements.

First and foremost, comparisons of measurement properties become more valid when instruments used to measure social cohesion are identical or sufficiently similar. We assigned ICCs to their respective instrument and thereby created different subgroups of studies. If studies referred to a specific instrument but did not use the exact same items and response scales, they were coded as a 'modified version' of the instrument used in the respective study (measurement instruments and used items are presented in the supplement). Once we established groups of sufficiently similar measurements, we assessed the internal consistency and ecometric reliability of included social cohesion measurements by inspecting their distribution across studies visualised by box plots alongside the observed values. This was done in order to evaluate the quality of the data underlying our meta-ecometric inferences. After this initial analysis, we aimed to study whether there were systematic differences in the ICC due to study characteristics. More specifically, because there is still no conclusion about the spatial units on which social cohesion operates, we investigated to what extent the inter-subjective agreement (ICC) in item responses is dependent on the average number of residents within the spatial units for which social cohesion was measured. As illustrated in Fig. 3, the chosen spatial unit can affect the ecometric properties of a contextual characteristic substantially. We explore this relationship by plotting the observed ICCs against the average number of residents of respective spatial units.

While questioning which contexts matter for contextual social traits is certainly in the focus of the meta-ecometric perspective, other issues cannot be left untouched. One of which is the notion of adjusting for perception bias in the ecometric measurement model. Its goal is to "correct" individual perception of the contextual trait for individual characteristics that might alter perception. In doing so, this approach controls for the different, e.g. demographic, compositional features of contexts. The specification of this part of the ecometric measurement model can detect whether the between-context variation in social cohesion could just be explained by differences in the composition of individuals within each context. Trying to correct for this potential bias is common practice in the application of ecometrics. However, while it is likely that including variables in the individual level equation (eq. (2)) will reduce the between-context variance in social cohesion, and thus affect the ecometric properties of its measurement, less attention has been given to the theoretical legitimacy of the variables usually used for 'perception bias correction'. Therefore, we will review to what extent model specification matters empirically by comparing the ICC from adjusted and unadjusted measurement models and discuss potential pitfalls of this analytic strategy.

Finally, for the same reasons that the concept of social cohesion might be unequally meaningful for different study populations, cultural settings, or points in time, the ecometric properties of its measurement are expected to vary accordingly. Thus, systematic differences in the ecometric properties of social cohesion between study populations, regions, or time points are realistic but only testable in multivariable analysis that can draw from enough cases to control for differences in measurement instrument, spatial unit, and ecometric model specification.

## 3. Results

After title and abstract screening, our search strategy yielded 1173 articles that have either measured social cohesion and conducted quantitative analysis, or needed additional screening of methods sections (Fig. 4). Due to this overwhelming number of studies, the focus of our review shifted towards a meta-ecometric approach using more restrictive inclusion criteria as recorded in our study protocol. To this end, we screened abstracts and, if needed, the methods sections of these 1173 articles to identify those studies, that explicitly reported ecometric properties, the use of ecometrics, or multilevel modelling. Applying these criteria, 81 studies were included. Additionally, we hand-searched the reference lists of included studies, which led to an inclusion of 5 additional studies. However, on the basis of full-text screening we excluded 8 studies that used the same data source and also reported identical ecometric properties and a further 12 studies which reported ecometric properties for measurements other than social cohesion (e.g. collective efficacy). Finally, because some of the 66 included studies reported more than one ICC or contextual-level reliability, we were able to extract a total of 78 cases.

Table 1 describes the frequencies of main study characteristics across our included cases. A complete description of all included studies is provided in Supplementary Table S1. Out of all 78 cases, we extracted from the included 66 publications, the ICC was available for slightly more than half of the cases (55.13%). The contextual-level reliability was reported for 38 (48.72%) cases and 21 (26.92%) of all cases reported both ecometric properties of their social cohesion measurement. On the other hand, in 18 (23.08%) cases neither the ICC nor the contextual-level reliability was available (Table 1).

Inspecting the measurement of social cohesion among this selection of studies, we found that measurements were fairly homogenous (Table 1). Out of all cases, 13 (16.67%) used identical items and respective response scales proposed by Sampson et al. (1997) (items and response scales are shown in Table S2). Another 22 (28.21%) cases measured social cohesion with slightly modified versions of this instrument. We categorised the measurement as "modified version" whenever the instruments only differed by one item (e.g. 4 instead of 5 items) or had the same number of items but with different response scales (e.g. 4-point instead of 5-point response scale). Moreover, 12 (15.38%) cases used instruments that were more heterogeneous but had at least 3 item categories in common with the Sampson et al., 1997 instrument (see Table S3 for a full description of item categories). Seven cases stemmed from measurements with the instrument proposed by Buckner in 1988 (Buckner, 1988) (items and response scales are shown in Table S2) or a modified version thereof. As Buckner's original instrument contained 18 items, we used a more flexible definition of what counts as a "modified version" (Table S1). The remaining 24 (30.77%) cases could not be assigned to either of the two instruments above. They were not sufficiently similar to either of these instruments, nor sufficiently similar to each other to allow further groupings. Thus, these measurements were categorised as "other measurements" and consist of heterogeneous measurements whose ecometric properties are hardly comparable. We, however, included this category in our analyses for completeness but caution against comparisons within this group of observations.

**Table 1**
Descriptive overview of included cases.

| | n (%[a]) |
|---|---|
| **Availability of measurement properties across cases** | |
| ICC | 43 (55%) |
| Contextual-level reliability | 38 (49%) |
| Cronbach's alpha | 46 (59%) |
| Both ICC & contextual-level reliability | 21 (27%) |
| Neither ICC nor contextual-level reliability | 18 (23%) |
| ICC & contextual-level reliability & Cronbach's alpha | 17 (22%) |
| None | 9 (12%) |
| **Measurement instrument** | |
| Sampson et al., 1997 | 13 (17%) |
| Modified version of Sampson et al., 1997 | 22 (28%) |
| At least 3 item categories in common with Sampson et al., 1997 | 12 (15%) |
| Buckner, 1988 | 1 (1%) |
| Modified Buckner, 1988 | 6 (8%) |
| Others | 24 (31%) |
| **Average number of residents per context** | |
| <1000 residents per context | 18 (23%) |
| 1001 to, 5000 residents per context | 36 (46%) |
| 5001 to 10,000 residents per context | 14 (18%) |
| >10,000 residents per context | 5 (6%) |
| not reported | 5 (6%) |
| **Region** | |
| Europe | 33 (42%) |
| North America | 31 (40%) |
| Australia & New Zealand | 8 (10%) |
| Brazil | 3 (4%) |
| South Africa | 3 (4%) |
| **Adjustment for perception bias** | |
| Yes | 37 (47%) |
| No | 41 (53%) |

[a] Percentages are rounded.

Most measurements of social cohesion (46.15%) referred to spatial units with 1001 to 5000 residents while only 5 measurements referred to spatial units that, on average, contained more than 10,000 residents. But social cohesion was also measured for spatial units with a small number of residents. In 18 (23.08%) cases, measurements referred to spatial units with on average less than 1001 residents (Table 1). Reporting of average population size per spatial unit was occasionally vague as authors described their spatial unit with statements like "on average about *X* residents". To prevent misclassification, we created categories instead of using the average number of residents per spatial unit. However, due to the lack of theoretical arguments for this categorisation, we also used this variable in its raw form, which is imprecise and thus prone to misclassification.

The vast majority of included data were collected in Europe (33 cases) and North America (31 cases). It follows, that 82.05% of extracted ecometric social cohesion measurement properties came from a North American or European context, while only 10.26% and 7.7% came from Australia & New Zealand and Brazil & South Africa respectively (Table 1). Moreover, 47.44% of our cases aimed to adjust for perception bias by including individual characteristics into the measurement model, whereas only 2 studies (Araya et al., 2006; Fone et al., 2006) provided adjusted as well as unadjusted variance components.

Generally, internal consistencies of most included social cohesion measurements for which this property was reported ranged between acceptable values of 0.7 and 0.9 (Fig. 5). Along the reported internal consistencies of included social cohesion measurements, Fig. 5 also shows the estimated contextual-level reliabilities (lambda) across measurement categories. According to the contextual-level reliability, most studies were reasonably well suited to assess latent contextual traits like social cohesion. Compared to reported internal consistencies, the contextual-level reliabilities showed higher variation within and between different measurement categories (Fig. 5). However, wide and overlapping confidence intervals between measurement instruments as well as differences in the methods used to calculate the reliability preclude judgments about the superior suitability of a specific instrument.

Complementary to Table S1, Fig. 6 gives an overview of all 43 cases for which the ICC was available and groups them by measurement instrument. It contains the ICC and its graphical illustration alongside vital information about the studies' characteristics that are behind these ICCs. As shown in Fig. 6, the ICCs lay within a range from 0.01 in Fone et al. (2006) to 0.345 in Stafford et al. (2003). Yet, both of these studies were well suited to assess contextual level traits: Fone et al. included 325 contexts with on average 37 respondents, while Stafford et al. (2003) included 259 contexts with an average of 42 respondents per context. First, these two ICCs stemmed from very different measurement instruments – Fone et al. used 8 items from the Buckner, 1988 instrument and Stafford et al. used 19 items that reflected the sub dimensions "trust", "attachment", "practical help", and "tolerance" – which showcases the potential impact of the instrument used. Second, while the data used by Fone et al. measured social cohesion in Caerphilly County Borough's (Wales, UK) enumeration districts with on average 406 adult residents in 2001, Stafford et al. measured social cohesion for electoral wards in England and postcode sectors in Scotland using multiple waves of data between 1994 and 1998. Stafford et al. have not reported the average population size for their chosen spatial units, but the Office for National Statistics reports that, in 1991, postcode sectors had an average population size of 5584.6 (ISD services, 2021) and electoral wards a mean of 5945 residents in 2001 (Ward-level population est, 2021). While the estimated inter-subjective agreement (ICC) on the level of social cohesion within a context may vary due to differences in the underlying usefulness of the contextual-level concept itself within a specific temporal or geographical context (Caerphilly County Borough in Wales of 2001 versus England and Scotland in the 90s), a closer inspection of extreme values in our sample highlights the sensitivity of measuring social cohesion. As ecometric properties should only be compared within groups of sufficiently similar measurement instruments, we present the distribution of ICCs by measurement instruments in Fig. 7. Studies that used the instrument proposed by Sampson et al. (1997) or a modified version thereof found that usually 10%–20% of the variation in perception of social cohesion exists between contexts. Thus, individual perceptions of social cohesion were only modestly dependent on the chosen contexts of respondents. Inter-subjective agreement (ICC) among respondents tended to be lower in studies that used the instrument proposed by Buckner, 1988, however, the sparse data did not allow for conclusions about systematic differences in the ICC between instruments. Confirming this observation, Pauwels & Hardyns (2009) provides ecometric properties for two different measurements of social cohesion from two Belgian cities (Antwerp, Ghent) about 60 km apart. Even though social cohesion was measured by different instruments in each city (Ghent: Sampson et al., 1997; Antwerp: other measurement – see Table S1), the ICC barely varied: 0.123 in Ghent and 0.124 in Antwerp.

We found no evidence that choosing spatial units with a lower

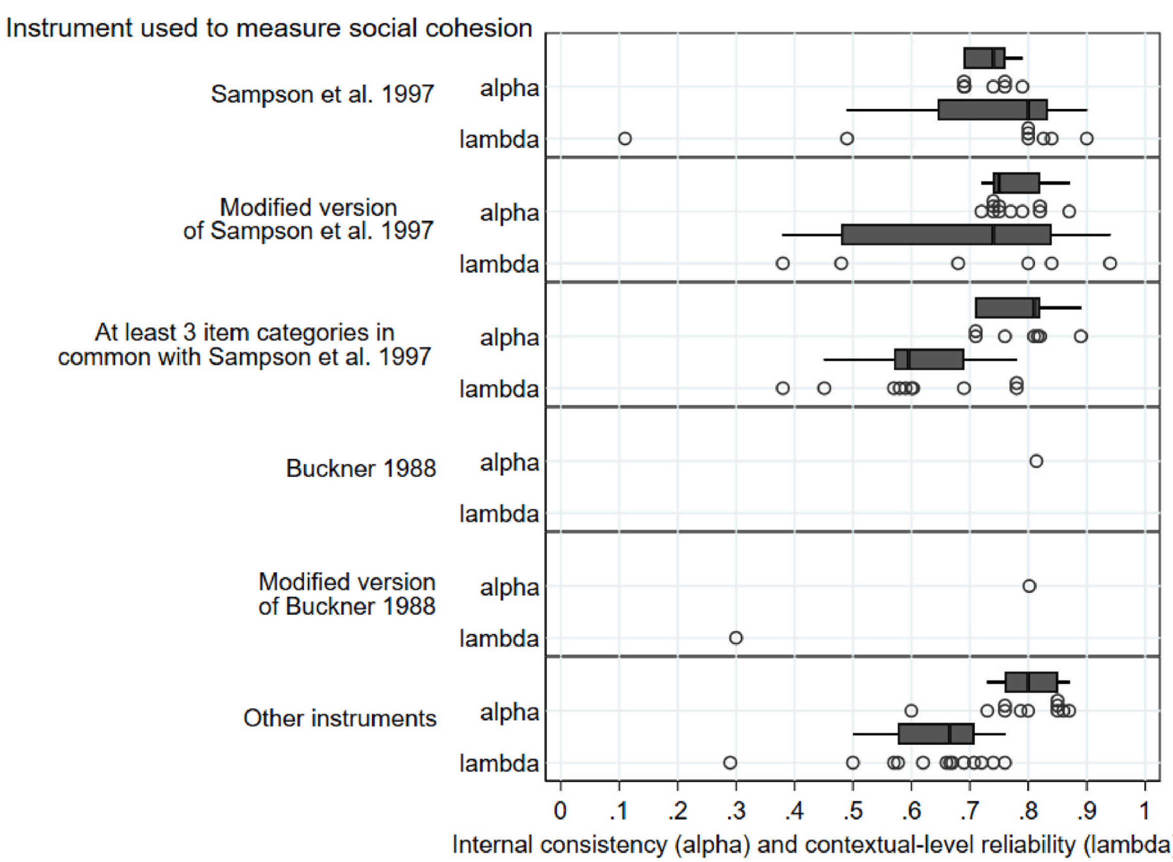


**Fig. 5.** Internal consistencies of social cohesion instruments measured by Cronbach's alpha (alpha) and contextual-level reliabilities (lambda) of social cohesion instruments. Modified versions include measurements that did not use the same items and/or response scales but only differed marginally compared to the original measurement instrument. See supplementary material for explanations of item categories. Contextual-level reliabilities were obtained by slightly different methods – see Fig. S1 or Table S1 for method-specific reliabilities.

| Study | Country | Period of data collection | Study population | # of contexts | respondents per context | avg. # of residents | alpha | Reliability | ICC | adjusted |
|---|---|---|---|---|---|---|---|---|---|---|
| **measured by Sampson 1997** | | | | | | | | | | |
| Wright et al 2015 | USA | | . | 80 | 37.0 | 5,001 to 10,000 residents per context | . | 0.83 | 0.068 | yes |
| Nair et al 2013 | USA | | adults | 154 | 4.9 | 1,001 to 5,000 residents per context | . | . | 0.100 | no |
| Duncan et al 2002 | USA | | adults | 55 | 21.5 | . | . | 0.74 | 0.110 | yes |
| Pauwels et al 2009 | Belgium | 2007 | key informants | 26 | 30.0 | >10,000 residents per context | 0.76 | 0.80 | 0.123 | no |
| du Toit et al 2007 | Australia | 2003-2004 | adults | 154 | 14.3 | ≤ 1,000 residents per context | 0.79 | . | 0.160 | no |
| Brisson et al 2011 | USA | 2002-2003 | adults | 418 | 17.9 | 1,001 to 5,000 residents per context | 0.69 | . | 0.200 | no |
| Rhew 2016 | USA | 2004- 2005 | older adults | 415 | 5.0 | 1,001 to 5,000 residents per context | 0.69 | . | 0.240 | no |
| Raudenbush et al 1999 | USA | 1995 | adults | 343 | 25.6 | 5,001 to 10,000 residents per context | . | 0.80 | 0.240 | no |
| **measured by modified Sampson** | | | | | | | | | | |
| Gerstner et al 2019 | Australia | 2010 | adults | 297 | 30.3 | 5,001 to 10,000 residents per context | . | . | 0.071 | yes |
| Zahnow et al 2013 | Australia | 2008 | adults | 71 | 37.3 | 5,001 to 10,000 residents per context | 0.75 | . | 0.080 | no |
| Barber et al 2016 | USA | 2003 | adults | 111 | 42.3 | 1,001 to 5,000 residents per context | 0.72 | . | 0.090 | yes |
| Zahnow et al 2013 | Australia | 2005 | adults | 71 | 35.3 | 5,001 to 10,000 residents per context | 0.75 | . | 0.106 | no |
| Wickes et al 2014 | Australia | 2007-2008 | adults | 148 | 27.7 | 5,001 to 10,000 residents per context | 0.75 | . | 0.110 | no |
| Gerstner et al 2019 | Germany | | adults | 140 | 27.9 | 1,001 to 5,000 residents per context | . | . | 0.118 | yes |
| Lippman et al 2016 | South Africa | 2012 | adults | 65 | 55.0 | ≤ 1,000 residents per context | . | . | 0.140 | no |
| Oberwittler et al 2001 | Germany | 1999 | adults | 20 | 116.5 | . | 0.82 | 0.94 | 0.155 | no |
| Lippman et al 2018 | South Africa | 2014 | adults | 70 | 43.3 | ≤ 1,000 residents per context | 0.82 | . | 0.191 | no |
| Mujahid et al 2007 | USA | 2004 | adults | 576 | 8.0 | 1,001 to 5,000 residents per context | 0.74 | 0.68 | 0.333 | yes |
| Mujahid et al 2007 | USA | 2004 | adults | 161 | 26.0 | 5,001 to 10,000 residents per context | 0.74 | 0.84 | 0.341 | yes |
| **measured by Buckner 1988** | | | | | | | | | | |
| Wilkinson 2007 | Canada | 2001 | adults | 21 | 82.5 | 1,001 to 5,000 residents per context | 0.81 | . | 0.118 | no |
| **measured by modified Buckner** | | | | | | | | | | |
| Fone et al 2006 | UK | 2001 | adults | 325 | 37.0 | ≤ 1,000 residents per context | 0.80 | 0.30 | 0.010 | yes |
| Fone et al 2006 | UK | 2001 | adults | 325 | 37.0 | ≤ 1,000 residents per context | 0.80 | 0.30 | 0.016 | no |
| Pampalon et al 2007 | Canada | 2004 | adults | 34 | 40.0 | 1,001 to 5,000 residents per context | . | . | 0.050 | no |
| **other** | | | | | | | | | | |
| Friche et al 2013 | Brazil | 2008-2009 | adults | 149 | 27.3 | ≤ 1,000 residents per context | 0.76 | 0.76 | 0.020 | no |
| Lenzi et al 2013 | Italy | 2009-2010 | school-aged | 31 | 12.5 | 1,001 to 5,000 residents per context | 0.86 | . | 0.041 | yes |
| Cutrona et al 2000 | USA | 1996 | adults | 41 | 17.3 | 1,001 to 5,000 residents per context | . | 0.69 | 0.043 | no |
| Cagney et al 2009 | USA | 2001-2002 | older adults | 65 | 17.0 | 1,001 to 5,000 residents per context | 0.76 | 0.60 | 0.050 | no |
| Aminzadeh et al 2013 | New Zealand | 2007 | school-aged | 262 | 18.0 | 1,001 to 5,000 residents per context | . | 0.59 | 0.071 | no |
| Cagney et al 2009 | USA | 2000-2002 | older adults | 82 | 41.0 | ≤ 1,000 residents per context | 0.71 | 0.78 | 0.080 | no |
| Zock et al 2018 | Netherlands | 1999-2000, 2006-2007, 2013-2014 | adults | 181 | 10.3 | ≤ 1,000 residents per context | . | 0.38 | 0.087 | yes |
| Taylor et al 2019 | USA | 2008 | adults | 44 | 97.7 | >10,000 residents per context | . | . | 0.090 | no |
| Mennis et al 2013 | USA | 2006, 2008, 2010 | adults | 520 | 51.0 | 1,001 to 5,000 residents per context | . | . | 0.112 | no |
| Villarreal et al 2006 | Brazil | 2002 | adults | 197 | 19.1 | ≤ 1,000 residents per context | . | 0.67 | 0.120 | yes |
| Araya et al 2006 | UK | 2001 | adults | 51 | 20.7 | ≤ 1,000 residents per context | . | . | 0.122 | yes |
| Pauwels et al 2009 | Belgium | 2004 | key informants | 42 | 7.6 | >10,000 residents per context | 0.73 | 0.50 | 0.124 | no |
| Araya et al 2006 | UK | 2001 | adults | 51 | 20.7 | ≤ 1,000 residents per context | . | . | 0.145 | no |
| Bottoni 2018 | 29 European Countries | 2012 | adults | 29 | 1885.2 | >10,000 residents per context | . | . | 0.146 | no |
| Zock et al 2018 | Netherlands | 1999-2000, 2006-2007, 2013-2014 | adults | 44 | 42.5 | >10,000 residents per context | . | 0.58 | 0.160 | yes |
| Clinton 2013 | USA | 2006, 2008 | older adults | 3316 | 6.0 | 1,001 to 5,000 residents per context | 0.82 | . | 0.212 | no |
| Mackenbach JD et al 2016 | BE, FR, HU, NL, UK | 2014 | adults | 60 | 98.3 | . | 0.79 | 0.48 | 0.214 | no |
| Harding et al 2009 | USA | 1994-1995 | school-aged | 2432 | 8.4 | 1,001 to 5,000 residents per context | . | 0.29 | 0.224 | yes |
| Donelly 2015 | USA | 1994-1995 | school-aged | 2247 | 7.8 | 1,001 to 5,000 residents per context | 0.60 | 0.67 | 0.266 | yes |
| Stafford et al 2003 | UK | HSE 1994-1999, SHS: 1995,1998 | adults | 259 | 42.0 | 5,001 to 10,000 residents per context | 0.81 | 0.60 | 0.345 | no |


**Fig. 6.** An overview of observations for which the ICC was available.

average number of residents yield higher inter-subjective agreement between respondents. In Fig. 8, we used the average number of residents per chosen context extracted from studies as a continuous variable to display its relationship with the ICC without imposing categories. In some cases, we were not able to retrieve precise values for the mean population size or the median instead of the mean, which biases the position of data points on the x-axis. To mitigate this potential bias due to misclassification, we also present the extracted ICCs within categories of average population size that are less prone to misclassification (Fig. S2). However, this did not change our results. If anything, a visual inspection of our data shows that the inter-subjective agreement in response behaviour was highest when spatial units contained, on average, 4000 to 10,000 residents. To maximise the validity of these comparisons, we also looked for single studies who reported ecometric properties for more than one spatial level. Taking this conservative approach, our initial findings were confirmed. Mujahid et al. (2007) provided ecometric properties of their social cohesion measurement ("modified version of Sampson et al., 1997") for 576 census tracts as well as 161 neighbourhood clusters in the US. They reported that census tracts contained an average of 4000 residents and neighbourhood clusters 8000 to 12,000 residents. While the ICC of social cohesion for census tracts was 0.333, inter-subjective agreement among the same respondents nested within neighbourhood clusters was slightly higher: 0.341 (Fig. 6, Table S1). That said, the construction of neighbourhood clusters in this study was not entirely dependent on administrative borders but takes information into account (demographic, socioeconomic and housing characteristics) (Mujahid et al., 2007). Ecometric properties for different spatial units were also estimated by Zock et al. (2018) for Dutch neighbourhoods. They estimated the ecometric properties of their social cohesion measurement for 181 five-digit postal code areas as well as for the 41 municipalities they belong to. Most five-digit postal codes have an area of less than 1 km$^2$ with an average population size of 500 residents whereas municipalities, at the beginning of 2006, contained a mean of 35,664 residents (median: 21,918) (StatLine. https://opendat, 2021). Social cohesion was measured with 10 items (Table S1) out of which 3 items belonged to the same item categories as 3 of the 5 items included in the instrument used by Sampson et al. (1997). In this study, inter-subjective agreement among respondents

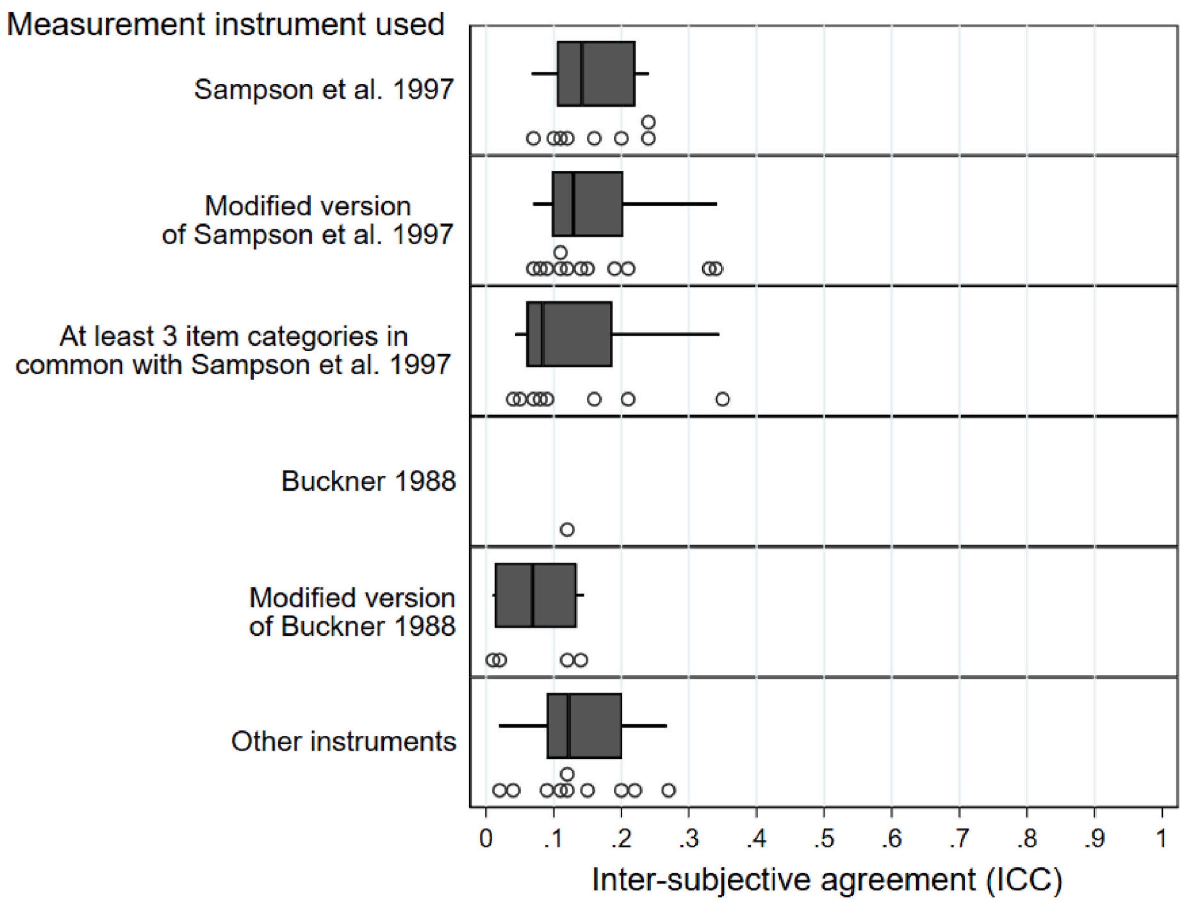


**Fig. 7.** Distribution of estimated inter-subjective agreement (ICC) across different measurement instruments of social cohesion.

within the same municipality was almost twice as high (ICC: 0.160) as the agreement on the level of social cohesion within the much smaller five-digit postal code areas (ICC:0.087, see Fig. 6).

Regarding perception bias correction, two of the included studies reported ICCs that stemmed from adjusted as well as unadjusted variance components. Data from both studies were collected for spatial units with less than 1000 residents in Wales, UK in 2001 and measured social cohesion similarly (modified version of Buckner, 1988; see Table S1). The unadjusted ICC in Araya et al., 2006 and Fone et al., 2006 had a value of 0.145 and 0.016 respectively. Using the same data but a different specification of their individual-level model, these ICCs only marginally changed. By adjusting for age, gender, unaffordable items, financial, and employment status, the variance attributable to postcode areas reduced from 0.145 to 0.122 in Araya et al., 2006. Similarly, Fone et al., 2006 adjusted for age, gender, social class, council tax band, employment status, gross income, and tenancy, which reduced the ICC from 0.016 to 0.01.

Although any further analysis on the effect of study characteristics, like region, on ecometric properties requires multivariable methods that hold important study characteristics constant, one of the included studies is particularly informative about the possible relation between region and the ICC. Ruijsbroek et al. (2017) measured social cohesion using the Sampson et al., 1997 instrument for four different European cities (Stoke-on-Trent, UK; Doetinchem, NL; Barcelona, ES; Kaunas, LT). In three of these cities, social cohesion was measured for similar spatial units that, on average, contained 1,508, 1,538, and 3400 residents. Further strengthening comparability, the data included a similar number of contexts as well as a similar number of respondents per context (Table S1). Despite these similar measurement conditions, the contextual-level reliability of their social cohesion measurements showed high variation: 0.8 in Stoke-on-Trent, 0.84 in Barcelona, 0.49 in Doetinchem, but only 0.11 in Kaunas. Unfortunately, neither the variance components nor the ICCs estimated by their measurement models were available, but, because the similar data structure across cities and Ruijsbroek et al.'s comment on this finding, the considerable difference in contextual-level reliability was due to the low between-context variance in response behaviour in Kaunas.

## 4. Discussion

Improving the measurement of contextual social characteristics remains a prerequisite for the identification of causal effects in research linking social environments to health and other outcomes (Roux, 2008). By now, researchers have collected an enormous amount of data on social cohesion. Despite its explicitly contextual definition, only a small proportion of studies which measured social cohesion eventually used it as the contextual-level characteristic it is and estimated its ecometric properties – the contextual-level reliability and inter-subjective agreement (ICC). Focusing on those which did, we analysed ecometric properties of 78 social cohesion measurements and how they varied under different measurement conditions.

Overall, the internal consistency and contextual-level reliability reported by included studies mostly ranged within acceptable values, which strengthened the basis for meta-ecometric inferences. Large variation in the contextual-level reliability between studies was not surprising, as this ecometric property is sensitive to the respective study design and we would expect that studies are differently suited to measure contextual traits.

Earlier, we stated that the intersubjective-agreement (ICC) on the level of social cohesion among respondents within contexts is not only a property of the measurement, but also reveals information about the empirical usefulness of the contextual-level concept itself. After all, if,

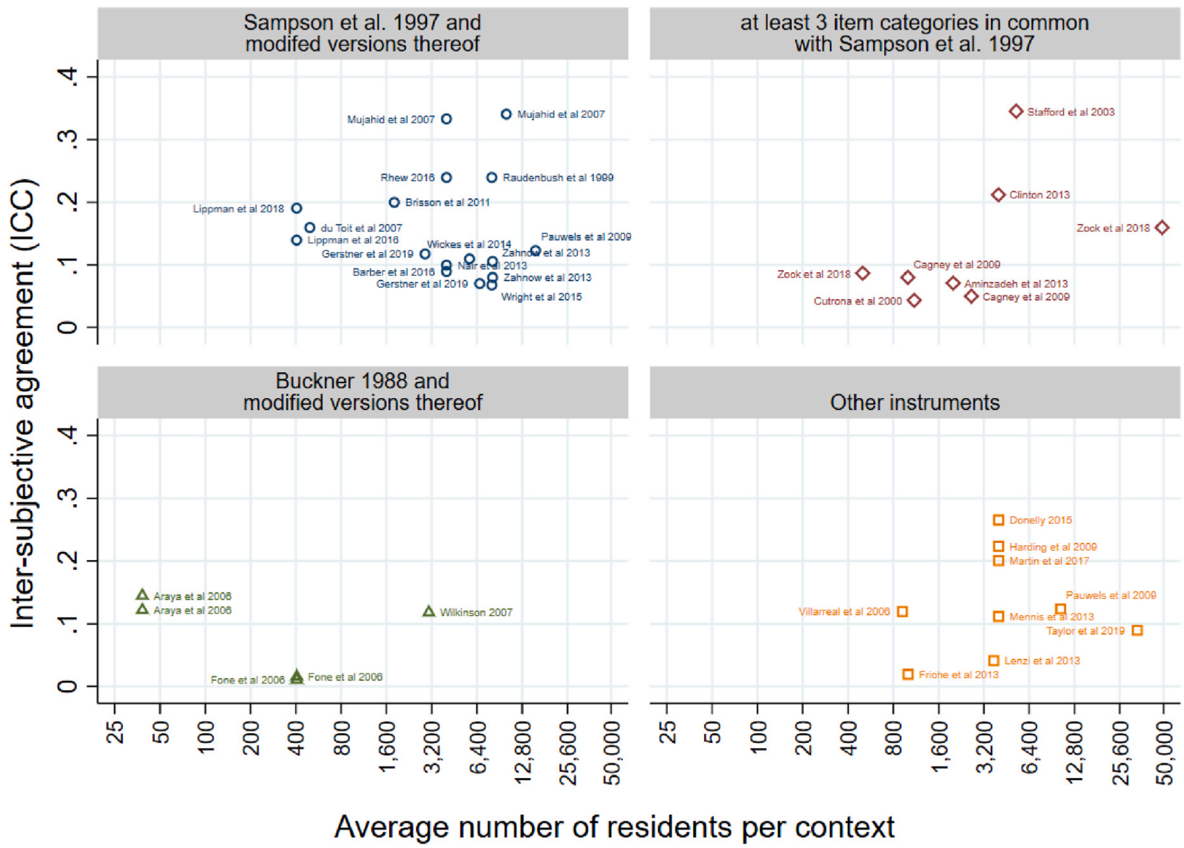


**Fig. 8.** Inter-subjective agreement (ICC) against average number of residents within chosen spatial units by comparable measurements of social cohesion. The scale of the x-axis is logarithmic and the y-axis only ranges from 0 to 0.4 as there were no ICCs greater than 0.4 in our data.

under a variety of different measurement conditions, individual response behaviour is entirely independent of the contexts to which respondents belong, researchers should be very sceptical about the usefulness of the respective concept in empirical research or might even be tempted to question its very existence as a contextual trait. Although we believe in the informative potential of this quantity for the investigation of all contextual social characteristics, there are some assumptions that must not be ignored in order to get the interpretation of the ICC right.

First, a precondition of ecometrics is that individual response behaviour is not independent of the contexts respondents belong to. For social cohesion, this assumption can be confirmed with our data as social cohesion measurements consistently showed that individual response behaviour was indeed dependent on the context of respondents. Even though included studies used a wide range of instruments to measure social cohesion, most approaches were similar to the instrument proposed by Buckner in 1988 or Sampson et al., in 1997. Further strengthening the evidence for the concept's ability to measure aspects of the social environment, no group of measurement instruments was convincingly more successful in detecting contextual dependence in response behaviour. Thus, it appears that the contextual nature of social cohesion is not too sensitive to deviations in item choices. However, as has been noted previously, most of the variation in social cohesion measurements exists between individuals, not between contexts (Fone et al., 2006; Friche et al., 2013; Mujahid et al., 2007; Raudenbush & Sampson, 1999). To render contextual characteristics useful in empirical analyses, it is implicitly assumed that the contextual trait of interest does indeed vary between contexts. But, given that the "true" between-context variance was estimated, a low value can occur for two reasons. One reason could be that the concept under study really does not vary much between contexts in which case the empirical usefulness of this concept as a contextual characteristic is questionable. The other reason for a low between-context variance refers to the second point we want to highlight: the relevance of the chosen spatial unit.

Second, the ICC and its interpretation can be heavily reliant on the spatial unit for which the contextual characteristic was measured. In one study, the ICC almost doubled when researchers merely changed the spatial unit of their measurement (Zock et al., 2018). Counterintuitively, measuring social cohesion in potentially more homogeneous spatial units with a smaller number of residents did not yield systematically higher intersubjective agreement between respondents within one context. This result may reveal a salient quality of the concept "social cohesion" itself: Inferring from our data, there may not be 'a correct' spatial unit on which social cohesion operates that researchers should be looking for. Instead, our data suggest that the relevant spatial unit for social cohesion is neither fixed in time nor space. Admittedly, it should be doubted whether the ICC alone is capable of leading us to the "correct" spatial unit of contextual social characteristics at all. Even if a very low between-context variance severely limits empirical analysis, data-driven approaches are not sufficient to identify the relevant spatial units of causal mechanisms linking social environments to health outcomes. Additionally, although "context" mostly refers to spatial units, the essentialist definition of social cohesion (Schiefer & van der Noll, 2017) also permits its use for contexts that are not necessarily spatial in character. While we doubt that neither meta-ecometrics, nor any other empirical approach can give a conclusive answer on which spatial units or contexts matters for causation, we firmly believe it is essential that researchers who seek to identify the effects of social environments on health outcomes have to specify the spatial unit in the very formulation of their hypotheses. At least, the choice of spatial units should be informed by the mechanism that is assumed to underlie the association between social environments and health outcomes. This is by far not the first time theoretical precision in multilevel research has been called for (Diez Roux, 2001; Diez Roux, Mujahid, Morenoff, & Raghunathan, 2007; Macintyre et al., 2002). Whenever possible, we recommend that researchers should provide ecometric properties and effect estimates for different spatial units and discuss the multiple potential spatial units for their contextual trait under study irrespective of data availability. When the different spatial units under consideration are hierarchical – smaller areas are always strictly nested within larger areas – a single multilevel model can be used to estimate the variance associated with each of the spatial units. Researchers may also choose to fit a series of multilevel models to estimate the variance associated with each spatial unit in isolation (especially, if considered spatial units do not form hierarchies).

Third, an interpretation of both the measurement and its ecometric properties assumes that the ecometric measurement model was correctly specified. The model asserts that individual perception is driven by the true level of social cohesion within a context as well as individual characteristics that bias perception of social cohesion. In almost half of included studies, researchers chose to adjust their social cohesion measurement for perception bias by including sociodemographic or socioeconomic characteristics in their ecometric measurement model, which decreased the between-context variance component. Among studies wherein unadjusted and adjusted variance components were reported (Araya et al., 2006; Fone et al., 2006), the extent of this decrease was only moderate in absolute terms. While the possibility to adjust for individual characteristics that could bias perception of social cohesion is a methodological advantage, previous literature has paid little attention to the subtle challenges of the ecometric model specification. Naturally, contexts may not only differ in their level of social cohesion, but also by compositional characteristics that could bias how it is perceived. It is, for example, quite possible that individuals who have lived in a context for many years have a more accurate but also more positive perception of the level of social cohesion in their context than individuals who have just moved to a place. Perhaps following this argument, several studies included length of residence in their ecometric measurement model. However, length of residence within a context might influence perception of social cohesion for the simple fact that residential turnover may actually affect the true level of social cohesion itself. Assuming unconstrained opportunities to move, the relationship between individually perceived social cohesion and length of residence could also be framed as a selection problem: Individuals might be less likely to move, the more positive their perception of the social environment within their residential context becomes. Therefore, adjusting for length of residence is correct if and only if one can make the case that length of residence affects individual perception directly and not via changing the true level of social cohesion within a context itself. Adjusting for any individual characteristics that do not only affect perception, but also the true level of social cohesion within a context will yield an underestimated between-context variance component and thus a biased ecometric measurement and related ecometric properties. On the other hand, an unconditional measurement model will overestimate the between-context variance and underestimate the within-context variance as long as contexts have a different composition of individual characteristics that influence perception of social cohesion but not social cohesion itself. It follows that accurate theory about which individual characteristics drive social cohesion and which merely bias its perception is critical. Contrary to the notion of perception bias correction, researchers could also make the case that any systematic differences in individual perception of social cohesion are a legitimate reflection of the level of social cohesion that is relevant to the sociodemographic make-up of a context. In this view, controlling for such differences would be conceptually wrong as it would impose the same compositional features on all contexts compared. After all, the distinction between context and composition is not always clear and might even be a "false dichotomy" to begin with (Diez Roux, 2001; Macintyre et al., 2002).

This review has some limitations. Above all, we are aware that some studies which measured related contextual-level concepts like (one approach to) social capital might have been missed by our search strategy but ought to be included due to the similarity of items used for their measurement. Unfortunately, it was beyond the scope of this study to conduct an additional review of explicitly contextual-level social capital measures. However, note that we have also included some measurements of social capital in case these measurements used identical or similar items to social cohesion measurements and applied ecometrics (Mackenbach et al., 2016; Mohnen et al., 2011, 2012; Prins et al., 2012; Rodrigues et al., 2018; Zock et al., 2018). Moreover, we focused on studies which measured social cohesion using survey questionnaires. This is the most common way of assessing social cohesion in the literature, but other methods aiming to measure aspects of the social environment should be explored. Although even more resource-intensive, comparisons with data from large scale field experiments, like lost letter experiments, could provide a new perspective on conventional survey data on social environments. Probably less costly, but equally interesting, computational social science has already shown its usefulness for linking latent traits to health outcomes by using twitter data (Eichstaedt et al., 2015). But, so far, researchers interested in the effects of social environments on health have only started to harness the potential of social media data (Schootman et al., 2016).

It is worth mentioning that we were not able to explore how the average area size of the chosen spatial unit affects the ICC. Given that physical distance between (agglomeration of) residents within a spatial unit could influence the level of intersubjective agreement, reporting area size in addition to population size of a chosen spatial unit is necessary. Also, not only average population and area size of spatial units could impact ecometric properties, but also the variation in these traits among included areal units.

Further, over 80% of eligible studies measured social cohesion in Europe or North America which limits the generalizability of our inferences beyond these geographic contexts. Finally, it is not clear what renders ecometric measurements sufficiently comparable. While we are confident that our sample allowed us to gain new insights regarding measurement instruments and spatial units, more observations are needed to engage in multivariable analysis and employ meta-regressions to further study how ecometric properties vary across time and cultural settings.

Notwithstanding these limitations, two main findings emerge from our review. First, reviewing the inter-subjective agreement (ICC) of 43 social cohesion measurements, we found consistent evidence for the contextual nature of the concept and its measurement. However, we also found that the ecometric properties of social cohesion can be quite sensitive to the conditions under which it has been measured. Acknowledging this sensitivity of measuring social cohesion, we recommend that researchers report both ecometric properties – the ICC and the contextual-level reliability - as well as accurately describe and discuss measurement conditions (measurement instrument, spatial unit, region, data structure, time point of data collection, study population). Unfortunately, contrary to other quantities that describe measurement quality in psychometrics, we are not able to define a meaningful rule of thumb when it comes to the interpretation of the ICC. Such a guidance for the interpretation of the ICC would be inappropriate because it would not only cloud the assumptions behind the ICC, but also conceal the shown sensitivity of measuring social cohesion. Additionally, little variation between contexts does not necessarily preclude the predictive validity of the concept (Raudenbush & Sampson, 1999).

Next, as data alone will not reveal the relevant spatial units of social cohesion, we follow previous calls (Diez Roux, 2001; Macintyre et al., 2002) in reminding that the theoretical specification of the contextual unit is an integral part of formulating hypotheses which link social environments to health outcomes. Finally, we recommend not correcting for perception bias correction, unless a convincing argument can be made that either the correction acts on perception alone without affecting the true level of social cohesion itself or that the potential bias reduction due to correction is higher than the bias this correction could introduce.

Despite its underlying challenges, the inter-subjective agreement (ICC) remains a vital quantity for researchers interested in contextual social characteristics. It may even be tempting to use it as a cue for the very existence of concepts that try to describe what is between, not within individuals. However, attempting to answer whether concepts like social cohesion do or do not exist may not be useful question to start with. While the existence of such concepts cannot be proven, their explanatory power in public health research can, in fact, be evaluated and should ultimately decide whether a concept is useful and not whether it exists. In the end, the question about the usefulness of a theoretical concept like social cohesion in empirical research is one that can only be answered by empirical research itself.

## Funding

The work of AHL and MO was supported by the Medical Research Council (MC_UU_00022/2) and the Scottish Government Chief Scientist Office (SPHSU17).

## Ethical statement

This study is exempt from ethical review as it is a systematic review and meta-analysis.

## Declaration of competing interest

None.

## Acknowledgments

We thank information retrieval specialist Karin Cepicka at the Medical University of Vienna for her support. We also thank contacted researchers who have provided missing information about their measurement of social cohesion.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ssmph.2022.101028.

## References

Araya, R., Dunstan, F., Playle, R., Thomas, H., Palmer, S., & Lewis, G. (2006). Perceptions of social capital and the built environment and mental health. *Social Science & Medicine, 62*(12), 3072–3083. https://doi.org/10.1016/j.socscimed.2005.11.037

Berkman, L. F., Glass, T., Brissette, I., & Seeman, T. E. (2000). From social integration to health: Durkheim in the new millennium. *Social Science & Medicine, 51*(6), 843–857. https://doi.org/10.1016/S0277-9536(00)00065-4

Bernard, P. (2000). *Social cohesion: A dialectical critique of a quasi-concept*. Paper SRA-491.

Buckner, J. C. (1988). The development of an instrument to measure neighbourhood cohesion. *American Journal of Community Psychology, 16*(Dec 88), 771–791.

Carrasco, M. A., & Bilal, U. (2016). A sign of the times: To have or to be? Social capital or social cohesion? *Social Science & Medicine, 159*, 127–131. https://doi.org/10.1016/j.socscimed.2016.05.012

Chan, J., To, H. P., & Chan, E. (2006). Reconsidering social cohesion: Developing a definition and analytical framework for empirical research. *Social Indicators Research, 75*(2), 273–302. https://doi.org/10.1007/s11205-005-2118-1

Christakis, N. A., & Fowler, J. H. (2011). *Connected: The amazing power of social networks and how they shape our lives* (Harper Press paperback edition). London: HarperPress.

Chuang, Y. C., Chuang, K. Y., & Yang, T. H. (2013). Social cohesion matters in health. *International Journal for Equity in Health, 12*(1), 87. https://doi.org/10.1186/1475-9276-12-87

Dickes, P., & Valentova, M. (2013). Construction, validation and application of the measurement of social cohesion in 47 European countries and regions. *Social Indicators Research, 113*(3), 827–846. https://doi.org/10.1007/s11205-012-0116-7

Diez Roux, A. V. (2001). Investigating neighborhood and area effects on health. *American Journal of Public Health, 91*(11), 1783–1789. https://doi.org/10.2105/AJPH.91.11.1783

Diez Roux, A. V., Mujahid, M. S., Morenoff, J. D., & Raghunathan, T. (2007). Mujahid et al.Respond to "beyond the metrics for measuring neighborhood effects. *American Journal of Epidemiology, 165*(8), 872–873. https://doi.org/10.1093/aje/kwm039

Duncan, D. T., & Kawachi, I. (Eds.). (2018). *Neighborhoods and health* (2nd ed.). New York, NY: Oxford University Press.

Durkheim, É. (2006). *On suicide*. London: Penguin.

Eichstaedt, J. C., Schwartz, H. A., Kern, M. L., Park, G., Labarthe, D. R., Merchant, R. M., & Seligman, M. E. P. (2015). Psychological Language on Twitter Predicts County-Level Heart Disease Mortality. *Psychological Science, 26*, 159–169. https://doi.org/10.1177/0956797614557867.

Fisher, L. B., Overholser, J. C., Ridley, J., Braden, A., & Rosoff, C. (2015). From the outside looking in: Sense of belonging, depression, and suicide risk. *Psychiatry, 78*(1), 29–41. https://doi.org/10.1080/00332747.2015.1015867

Fone, D. L., Farewell, D. M., & Dunstan, F. D. (2006). An ecometric analysis of neighbourhood cohesion. *Population Health Metrics, 4*(1), 1–17. https://doi.org/10.1186/1478-7954-4-17

Fonseca, X., Lukosch, S., & Brazier, F. (2019). Social cohesion revisited: A new definition and how to characterize it. *Innovation: The European Journal of Social Science Research, 32*(2), 231–253. https://doi.org/10.1080/13511610.2018.1497480

Friche, A. A., Diez-Roux, A. V., César, C. C., Xavier, C. C., Proietti, F. A., & Caiaffa, W. T. (2013). Assessing the psychometric and ecometric properties of neighborhood scales in developing countries: Saúde em Beagá Study, Belo Horizonte, Brazil, 2008-2009. *Journal of Urban Health, 90*(2), 246–261. https://doi.org/10.1007/s11524-012-9737-z

Glaser, R., & Kiecolt-Glaser, J. K. (2005). Stress-induced immune dysfunction: Implications for health. *Nature Reviews Immunology, 5*(3), 243–251. https://doi.org/10.1038/nri1571

Glonti, K., Mackenbach, J. D., Ng, J., Lakerveld, J., Oppert, J.-M., Bárdos, H., … Rutter, H. (2016). Psychosocial environment: definitions, measures and associations with weight status – a systematic review. *Obesity Reviews, 17*, 81–95. https://doi.org/10.1111/obr.12383.

Hagerty, B. M., Lynch-Sauer, J., Patusky, K. L., Bouwsema, M., & Collier, P. (1992). Sense of belonging: A vital mental health concept. *Archives of Psychiatric Nursing, 6*(3), 172–177. https://doi.org/10.1016/0883-9417(92)90028-h

Hatcher, S., & Stubbersfield, O. (July 2013). Sense of belonging and suicide: A systematic review. *Canadian Journal of Psychiatry*. https://doi.org/10.1177/070674371305800709

Hipp, J. R., & Steenbeek, W. (2016). Types of crime and types of mechanisms: What are the consequences for neighborhoods over time? *Crime & Delinquency, 62*(9), 1203–1234. https://doi.org/10.1177/0011128715604168

Holt-Lunstad, J. (2018). Why social relationships are important for physical health: A systems approach to understanding and modifying risk and protection. *Annual Review of Psychology, 69*(1), 437–458. https://doi.org/10.1146/annurev-psych-122216-011902

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications. 2*. New York: Routledge, Taylor & Francis.

ISD services | geography, population and deprivation analytical support team | population | census | ISD Scotland. https://www.isdscotland.org/Products-and-Services/GPD-Support/Population/Census/. (Accessed 5 July 2021).

Kawachi, I., & Berkman, L. F. (2014). Social cohesion, social capital, and health. In I. Kawachi, & L. F. Berkman (Eds.), *Social epidemiology* (2nd ed., pp. 290–319). New York: Oxford University Press.

Kawachi, I., & Subramanian, S. V. (2018). Social epidemiology for the 21st century. *Social Science & Medicine, 196*, 240–245. https://doi.org/10.1016/j.socscimed.2017.10.034

Leyland, A. H., & Groenewegen, P. P. (2020). *Multilevel modelling for public health and health services research: Health in context*. Springer International Publishing. https://doi.org/10.1007/978-3-030-34801-4

Lochner, K., Kawachi, I., & Kennedy, B. P. (1999). Social capital: A guide to its measurement. *Health & Place, 5*(4), 259–270. https://doi.org/10.1016/S1353-8292(99)00016-7

Macintyre, S., Ellaway, A., & Cummins, S. (2002). Place effects on health: How can we conceptualise, operationalise and measure them? *Social Science & Medicine, 55*(1), 125–139. https://doi.org/10.1016/S0277-9536(01)00214-3

Mackenbach, J. D., Lakerveld, J., van Lenthe, F. J., Kawachi, I., McKee, M., Rutter, H., … Brug, J. (2016). Neighbourhood social capital: measurement issues and associations with health outcomes. *Obesity Reviews, 17*(Suppl 1), 96–107.

Mair, C., Diez Roux, A. V., & Galea, S. (2008). Are neighbourhood characteristics associated with depressive symptoms? A review of evidence. *Journal of Epidemiology & Community Health, 62*(11), 940–946. https://doi.org/10.1136/jech.2007.066605, 8 p following 946.

Marmot, M. G., & Wilkinson, R. G. (Eds.). (2006). *Social determinants of health* (2nd ed.). Oxford ; New York: Oxford University Press.

Mohnen, S. M., Groenewegen, P. P., Völker, B., & Flap, H. (2011). Neighborhood social capital and individual health. *Social Science & Medicine, 72*(5), 660–667. https://doi.org/10.1016/j.socscimed.2010.12.004

Mohnen, S. M., Völker, B., Flap, H., & Groenewegen, P. P. (2012). Health-related behavior as a mechanism behind the relationship between neighborhood social capital and individual health–a multilevel analysis. *BMC Public Health, 12*, 116. https://doi.org/10.1186/1471-2458-12-116

Moore, S., Haines, V., Hawe, P., & Shiell, A. (2006). Lost in translation: A genealogy of the "social capital" concept in public health. *Journal of Epidemiology & Community Health, 60*(8), 729–734. https://doi.org/10.1136/jech.2005.041848

Moore, S., & Kawachi, I. (2017). Twenty years of social capital and health research: A glossary. *Journal of Epidemiology & Community Health, 71*(5), 513–517. https://doi.org/10.1136/jech-2016-208313

Mujahid, M. S., Diez Roux, A. V., Morenoff, J. D., Raghunathan, T. E., Cooper, R. S., Ni, H., & Shea, S. (2008). Neighborhood Characteristics and Hypertension. *Epidemiology, 19*, 590–598. https://doi.org/10.1097/EDE.0b013e3181772cb2.

Mujahid, M. S., Diez Roux, A. V., Morenoff, J. D., & Raghunathan, T. (2007). Assessing the measurement properties of neighborhood scales: From psychometrics to ecometrics. *American Journal of Epidemiology, 165*(8), 858–867. https://doi.org/10.1093/aje/kwm040

Muntaner, C., & Lynch, J. (1999). Income inequality, social cohesion, and class relations: A critique of Wilkinson's neo-Durkheimian research program. *International Journal of Health Services, 29*(1), 59–81. https://doi.org/10.2190/G8QW-TT09-67PL-QTNC

Oakes, J. M. (2004). The (mis)estimation of neighborhood effects: Causal inference for a practicable social epidemiology. *Social Science & Medicine, 58*(10), 1929–1952. https://doi.org/10.1016/j.socscimed.2003.08.004

O'Brien, R. M. (1990). Estimating the reliability of aggregate-level variables based on individual-level characteristics. *Sociological Methods & Research, 18*(4), 473–504. https://doi.org/10.1177/0049124190018004004

Pauwels, L., & Hardyns, W. (2009). Measuring community (Dis)Organizational processes through key informant analysis. *European Journal of Criminology, 6*(5), 401–417, 2009.

Pérez, E., Braën, C., Boyer, G., Mercille, G., Rehany, É., Deslauriers, V., … Potvin, L. (2020). Neighbourhood community life and health: A systematic review of reviews. *Health & Place, 61*, 102238. https://doi.org/10.1016/j.healthplace.2019.102238.

Portes, A. (1998). Social capital: Its origins and applications in modern sociology. *Annual Review of Sociology, 24*(1), 1–24. https://doi.org/10.1146/annurev.soc.24.1.1

Portes, A. (2014). Downsides of social capital. *Proceedings of the National Academy of Sciences of the United States of America, 111*(52), 18407–18408. https://doi.org/10.1073/pnas.1421888112

Prins, R. G., Mohnen, S. M., van Lenthe, F. J., Brug, J., & Oenema, A. (2012). Are neighbourhood social capital and availability of sports facilities related to sports participation among Dutch adolescents? *International Journal of Behavioral Nutrition and Physical Activity, 9*(1), 90. https://doi.org/10.1186/1479-5868-9-90

Raudenbush, S. W. (2003). The quantitative assessment of neighborhood social environments. In I. Kawachi, & L. F. Berkman (Eds.), *Neighborhood and health* (pp. 112–131). Oxford ; New York: Oxford University Press.

Raudenbush, S. W., Rowan, B., & Kang, S. J. (1991). A multilevel, multivariate model for studying school climate with estimation via the EM algorithm and application to U. S. High-school data. *Journal of Educational Statistics, 16*(4), 295. https://doi.org/10.2307/1165105

Raudenbush, S. W., & Sampson, R. J. (1999). Ecometrics: Toward a science of assessing ecological settings, with application to the systematic social observation of neighborhoods. *Sociological Methodology, 29*, 1–41.

Rodrigues, D. E., Cesar, C. C., Kawachi, I., Xavier, C. C., Caiaffa, W. T., & Proietti, F. A. (2018). The influence of neighborhood social capital on leisure-time physical activity: A population-based study in Brazil. *Journal of Urban Health, 95*(5), 727–738. https://doi.org/10.1007/s11524-018-0293-z

Roux, A. V. D. (2008). Next steps in understanding the multilevel determinants of health. *Journal of Epidemiology & Community Health, 62*(11), 957–959. https://doi.org/10.1136/jech.2007.064311

Roux, A. V. D., & Mair, C. (2010). Neighborhoods and health. *Annals of the New York Academy of Sciences, 1186*, 125–145. https://doi.org/10.1111/j.1749-6632.2009.05333.x.

Ruijsbroek, A., Mohnen, S. M., Droomers, M., Kruize, H., Gidlow, C., Grazuleviciene, R., … Groenewegen, P. P. (2017). Neighbourhood green space, social environment and mental health: an examination in four European cities. *International Journal of Public Health, 62*, 657–667.

Sampson, R. J. (2012). *Great American city: Chicago and the enduring neighborhood effect*. Chicago, Ill: Univ. of Chicago Press.

Sampson, R. J., Raudenbush, S. W., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science, 277*(5328), 918–924. https://doi.org/10.1126/science.277.5328.918

Samuel, L. J., Commodore-Mensah, Y., & Himmelfarb, C. R. D. (October 2013). *Developing behavioral theory with the systematic integration of community social capital concepts: Health education & behavior*. https://doi.org/10.1177/1090198113504412

Savitz, N. V., & Raudenbush, S. W. (2009). 5. Exploiting spatial dependence to improve measurement of neighborhood social processes. *Sociological Methodology, 39*(1), 151–183. https://doi.org/10.1111/j.1467-9531.2009.01221.x

Sawyer, A., Ucci, M., Jones, R., Smith, L., & Fisher, A. (2017). Simultaneous evaluation of physical and social environmental correlates of physical activity in adults: A systematic review. *SSM - Population Health., 3*, 506–515. https://doi.org/10.1016/j.ssmph.2017.05.008

Schiefer, D., & van der Noll, J. (2017). The essentials of social cohesion: A literature review. *Social Indicators Research, 132*(2), 579–603. https://doi.org/10.1007/s11205-016-1314-5

Schootman, M., Nelson, E. J., Werner, K., Shacham, E., Elliott, M., Ratnapradipa, K., … McVay, A. (2016). Emerging technologies to measure neighborhood conditions in public health: implications for interventions and next steps. *Int J Health Geogr, 15*, 20. https://doi.org/10.1186/s12942-016-0050-z.

Smith, K. P., & Christakis, N. A. (2008). Social networks and health. *Annual Review of Sociology, 34*(1), 405–429. https://doi.org/10.1146/annurev.soc.34.040507.134601

Stafford, M., Bartley, M., Sacker, A., Marmot, M., Wilkinson, R., Boreham, R., & Thomas, R. (2003). Measuring the Social Environment: Social Cohesion and Material Deprivation in English and Scottish Neighbourhoods. *Environ Plan A, 35*, 1459–1475. https://doi.org/10.1068/a35257.

StatLine. https://opendata.cbs.nl/statline/#/CBS/nl/. (Accessed 6 July 2021).

Villalonga-Olives, E., & Kawachi, I. (2017). The dark side of social capital: A systematic review of the negative health effects of social capital. *Social Science & Medicine, 194*, 105–127. https://doi.org/10.1016/j.socscimed.2017.10.020

Ward-level population estimates (experimental statistics) - Office for national statistics. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/wardlevelmidyearpopulationestimatesexperimental. (Accessed 5 July 2021).

Wilkinson, R. G. (1997). *Unhealthy societies: The afflictions of inequality*. Routledge.

Zock, J. P., Verheij, R., Helbich, M., Volker, B., Spreeuwenberg, P., Strak, M., … Groenewegen, P. (2018). The impact of social capital, land use, air pollution and noise on individual morbidity in Dutch neighbourhoods. *Environment International, 121*, 453–460.