



Towards the Next Fifty Years of Social Indicators Research: Some Guidance for Authors

David Bartram¹ · Leonardo Salvatore Alaimo² · Eileen Avery³ · Anthony Bardo⁴ · Enrico Di Bella⁵ · Martin Binder⁶ · Ferdi Botha⁷ · Sandra Fachelli⁸ · Andrea Gatto⁹ · Jintao Lu¹⁰ · Adam Okulicz-Kozaryn¹¹ · Angeles Sanchez¹² · Massimiliano Tani¹³

Published online: 9 August 2024

© The Author(s), under exclusive licence to Springer Nature B.V. 2024

Social Indicators Research was founded 50 years ago, in 1974. Alex Michalos was editor-in-chief for 40 years. One of our purposes here is to express gratitude to him and to the other colleagues who have built the journal and contributed to its reputation. Alex's contribution to the social indicators 'movement' is of course much broader – but founding and leading this journal for such an extended period is a core component of that contribution. We all continue to benefit from his enormous intellectual efforts.

Filomena Maggino devoted 10 years as editor-in-chief and has made her own significant mark on the journal. In addition to the core intellectual work (informed by her own substantial contribution to the development of composite indicators methodologies), Professor Maggino recruited a committed and talented team of co-editors. The prospects for taking up that role recently (by Bartram) were attractive in no small measure via familiarity with the

✉ David Bartram
d.bartram@le.ac.uk

¹ University of Leicester, Leicester, United Kingdom

² University of Rome La Sapienza, Rome, Italy

³ University of Missouri, Columbia, USA

⁴ University of Kentucky, Lexington, USA

⁵ University of Genoa, Genoa, Italy

⁶ Bundeswehr University Munich, Munich, Germany

⁷ University of Melbourne, Melbourne, Australia

⁸ Universidad Pablo de Olavide, Sevilla, Spain

⁹ Wenzhou-Kean University (US-China), Wenzhou, China

¹⁰ Taiyuan University of Science and Technology, Taiyuan, China

¹¹ Rutgers University, Camden, USA

¹² University of Granada, Granada, Spain

¹³ University of New South Wales, Sydney, Australia

current state of the journal. The handover has been followed by the addition of some very welcome new colleagues as co-editors.

Our other purpose in this editorial article, the task that informs the remaining text here, is to offer guidance to authors who intend to submit manuscripts to us. In general, our view is that authors should have access to a good understanding of what a journal's editors are looking for – the basis for editors' evaluation of manuscripts. That idea seems especially important in light of some very basic data. We receive an enormous number of submissions – approximately 2,000 in 2023, and the number is sure to increase further in 2024. The vast majority are rejected; we publish barely 15 per cent of the submissions we receive. Some authors will recognise their own experience in those numbers, no doubt with some frustration.

We hope to enhance the efforts made by authors seeking to publish with us (recognising that we as editors will benefit from this as well). We have worked to develop a set of expectations suitable for communication to authors. We already apply many of these expectations when we evaluate manuscripts – so, we now seek to help authors understand them prior to submission (and, ideally, prior to construction of research and composition of manuscripts). Our main hope is to receive better manuscripts. We anticipate that we will also divert some potential submissions away from us – partly because many authors are not in fact doing research relevant to the journal. In the next section, we therefore clarify the concept of 'social indicators research' (a task undertaken in conjunction with revising the 'aims and scope' text on the journal webpage).

In subsequent sections we offer some extended methodological guidance. Almost all manuscripts submitted to the journal report on quantitative research. (We do of course welcome submissions that do qualitative work.) As with quantitative social science research on many topics, there is, in some of the submissions we receive, a lot of underdeveloped work. This is something that can be remedied, at least potentially (in contrast to lack of fit with our aims and scope, where the right solution is to submit to a different journal).

1 Our Focus

The central concept for this journal is people's *quality of life*. This is a broad idea (though it does not encompass *everything* – a point explored here). It incorporates the circumstances of people's lives; in that sense it has an 'objective' manifestation. It also incorporates how people *feel* about their lives, thus a 'subjective' dimension. That latter component has found expression in the development of separate journals, including the *Journal of Happiness Studies*. But the relevance of subjective well-being to 'quality of life' means that it is relevant to *Social Indicators Research* as well.

The fact that 'quality of life' incorporates the circumstances in which people live raises the challenging question of identifying the boundaries of 'social indicators research'. Which circumstances matter? How confident are we that they really contribute to people's quality of life? Considering questions of that sort is one way we try to determine whether an article is 'in scope' for the journal.

We propose the following question as a way of clarifying your own thinking about it: in connection with your particular topic, *how confident can we be that more of something (or indeed less of something) actually makes people's lives better?* For some topics, the answer

is reasonably straightforward. It is likely straightforward for subjective well-being: more happiness/life-satisfaction means better quality of life.

But objective circumstances matter as well (a point well established by Amartya Sen, among others). For some examples in that category, the answer to the question is probably not in dispute. Less poverty makes people's lives better. For economic growth, in contrast, the answer is not as obvious; at a minimum, there are threshold effects and the 'Easterlin paradox' to consider. For some topics (drawn from recent submissions that were desk-rejected), the level of confidence about it would surely be low. There is no *consensus* that 'free markets' make people's lives better. 'Migration intentions' do not obviously make people's lives better. If your goal is to explore the connection between those topics and 'quality of life' (defined in a way that is clearly in scope for us), then fine – but a paper that explores such topics in their own right is not.

To address the question effectively, drawing on existing research about it is a good strategy. If it seems necessary to allay doubts, mere assertion or opinion is not what's needed.

For some topics, we might decide that the manuscript would be better handled in a journal that focuses specifically on your topic. Good health generally makes people's lives better, and we do sometimes publish papers about health. But some papers on health have a more 'medical' focus; we are not well equipped to evaluate such papers and would be likely to suggest submitting them instead to a journal with that particular focus. An analogous view might be taken about papers dealing with education.

Judging from some of the submissions we receive, it's hard to avoid the impression that some authors consider this journal a 'catch-all' venue for topics that *might* be relevant to quality of life. Suggesting that one's topic is relevant to us simply because it could have 'implications' for quality of life amounts to a weak case. Virtually anything could have implications for quality of life. We would prioritize papers that (in contrast) directly explore the way some feature of the social world affects or constitutes people's quality of life. Again, 'social indicators research' is a broad idea – but it does not encompass everything.

A long-standing interest for the journal is the construction of 'indicators' for the purpose of *measuring* people's quality of life. The aims and scope text on the journal webpage has conveyed this interest for many years, and we retain it in the recently revised version. We also welcome manuscripts that do research in what will be (for many) a more familiar mode: seeking to explain quality-of-life outcomes by identifying 'factors' (situations, initiatives, characteristics, processes, etc.) that contribute to higher or lower quality of life.

Lack of fit is a common reason for desk rejections. We encourage authors to contemplate the rigor we seek to exercise in our initial evaluation of manuscripts in these terms. We take this approach in part because we have to, in view of the very large number of submissions we receive.

One way to gain further clarity on our focus is of course to read more widely about the history of the social indicators idea. Among the multitude of available sources, we recommend an overview by Land and Michalos (2018).

2 Quantitative Analysis and Causal Interpretation

Our ‘aims and scope’ reflects the fact that researchers in this area often want to know what contributes to people’s well-being and quality of life. That word – ‘contributes’ – has an inescapably causal meaning. Doing research that gives persuasive evidence for causal effects is challenging – and many of the manuscripts we receive do not excel in rising to that challenge. We encourage researchers in this area to ‘raise their game’ in constructing an analysis intended to offer insight in this mode. Five ideas seem important.

2.1 Clarify and Embrace Your Purpose

If you want to explore causal effects, say so. If (in contrast) you say that you are interested in the ‘relationship’ (or ‘association’) between two variables, you might want to ask: what *kind* of relationship? Perhaps you really do have an ‘effect’ in mind; your manuscript might even use the word ‘effect’ when you interpret your results. ‘Associations’ and ‘correlations’ are likely mere numerical results that don’t tell us much about the social world. If an analysis is framed in these terms, authors should (at a minimum) tell us what can be learned from the associations/correlations.

There is a particular rhetorical strategy that undermines manuscripts. (The word strategy here might overstate the extent to which one’s intentions are explicit/conscious.) Authors sometimes produce an analysis, usually consisting of cross-sectional regression models, yielding results that amount to partial ‘correlations’ (because they do not consider the ideas articulated in the next section). In the author’s conclusion we then see a caveat noting that the results cannot be taken as indicative of a causal relationship. But it is nonetheless evident that a causal relationship is the target of the author’s true interest. The main signal of this interest is the use of language that can only have a causal meaning (e.g., as noted above, some variable ‘contributes’ to some sort of outcome, or ‘shapes’ it, or ‘influences’ it). Another signal is the suggestion of policy implications that apparently follow from one’s ‘correlations’.

If causal language is used, and/or a policy implication is derived, then a cursory caveat disclaiming a causal interpretation is not persuasive. At the same time, a caveat about causal interpretation should be taken seriously if an analysis has not been constructed for that purpose. If a caveat is necessary, then what is really needed is to do a more effective analysis.

2.2 Constructing an Analysis

If you do want to explore causal effects, then some attention is likely needed for a challenging question: what sort of analysis/evidence would support such an interpretation? Many of the manuscripts we receive offer cross-sectional regression models. Whether results from these models could support a causal interpretation is of course debatable. But if we approach the topic with practical considerations in mind (i.e., anticipating that many researchers will continue to use cross-sectional regression models), we would want to see such models constructed in line with a coherent logic.

A key question for construction of these models is: how will you select control variables, and why? Many manuscripts do not address this issue (Wysocki et al., 2022), or they merely

appeal to precedent from previous research (in essence, outsourcing the work). If a criterion *is* articulated, a common idea is to control for ‘other determinants’ of the dependent variable.

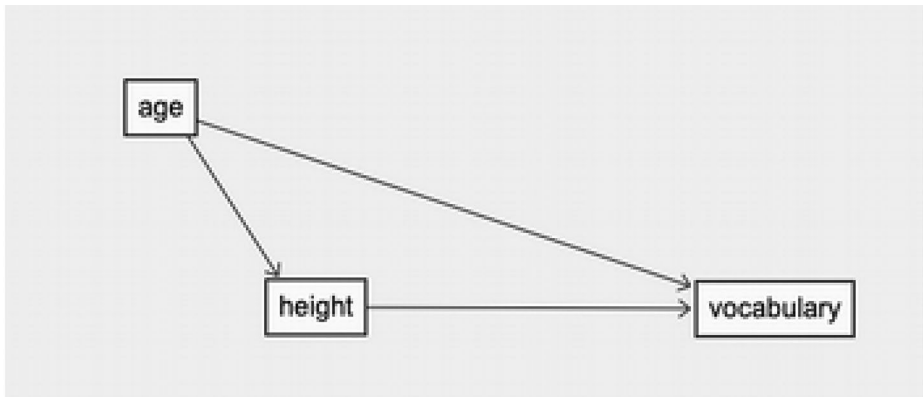
This criterion is insufficient and likely even incoherent. Let’s say we want to know the impact of X on Y (a useful shorthand is $X \rightarrow Y$.) We can use W to denote (potential) controls. The ‘other determinants’ idea says: include controls where $W \rightarrow Y$.

This criterion ignores the relationship between the controls and X . What we really need, to identify/estimate $X \rightarrow Y$, is to control for other determinants of Y that are *also* antecedents of X (so, $W \rightarrow X$). If we include controls where the relationship goes the other way ($X \rightarrow W$), we will exacerbate bias in our estimate of $X \rightarrow Y$. (The relevant term for that situation is ‘overcontrol bias’, see Rohrer, 2018.)

The purpose of controls is to address (i.e., reduce/mitigate) the possibility of bias in our estimates. We want an estimate that does not overstate or understate the true effect. To achieve that purpose, when we construct a model to tell us about a causal effect ($X \rightarrow Y$) we again need controls (W) that are antecedents of X and Y ($W \rightarrow X$ and $W \rightarrow Y$). These are the controls that will take us closer to an unbiased estimate of $X \rightarrow Y$. (See Cinelli et al., 2022 for a good overview.)

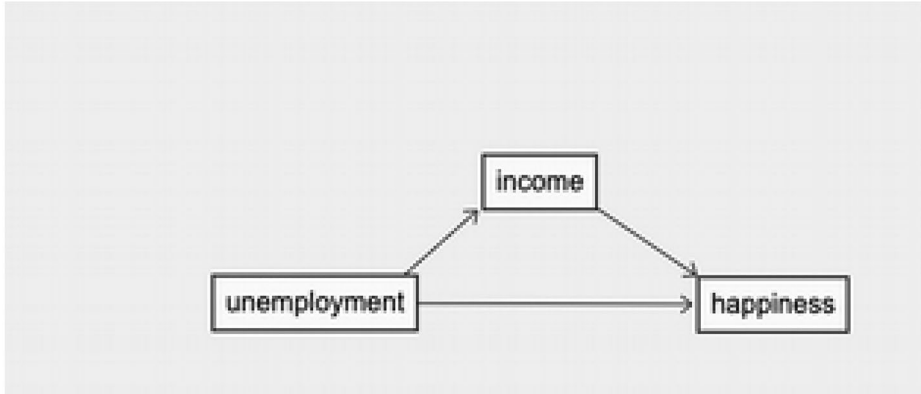
Here’s a ‘toy’ example. There is a strong positive correlation between height and vocabulary size: taller people use a larger number of words. But this does not mean that height *affects* vocabulary. Once we include the right control variable – age – we will see that the impact of height on vocabulary is zero. If we omit the needed controls, we will get a hopelessly biased estimate. Using the right controls, we get the right estimate, one that accords with our correct intuition.

The example works because it’s age (among children) that affects one’s height and also one’s intellectual development. Age is the important antecedent of X and Y . We can be especially clear on the way age is a suitable/necessary control because for that particular W the relationship with X is $W \rightarrow X$. The relevant relationships can be visualised:



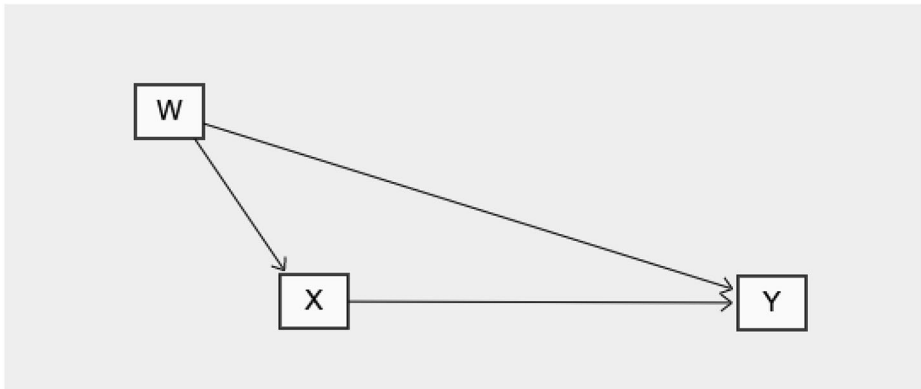
If however we include controls where the relationship goes in the other direction ($X \rightarrow W$), we will *exacerbate* bias rather than mitigating it (‘overcontrol bias’).

Suppose we want to estimate the impact of unemployment on happiness. Income is an ‘other determinant’ of happiness (Y); some researchers are indeed inclined to include it as a control. But what is its relationship with unemployment (X)? Answer: losing your job is very likely to reduce your income ($X \rightarrow W$). Again, a visual presentation is useful:



If we control for income, we are then comparing happiness between unemployed and employed people while holding income constant (i.e., looking at people who earn the same incomes). We will now get a biased estimate of the impact of unemployment on happiness; our coefficient will substantially understate the true impact (Bartram, 2021; this is a specific example of ‘overcontrol bias’ as per Rohrer, 2018). Unemployment reduces income, and reduced income contributes to lower happiness. If we control for income, we will ‘block’ this component of unemployment’s impact on happiness – omitting that component from our estimate.

The letter W is useful to denote controls – because W should ‘come before’ X (as well as Y). If we include controls where X comes before W ($X \rightarrow W$), we’re doing it wrong; the model will yield a biased estimate of $X \rightarrow Y$. Visually, the pattern we need for control variables is:



Using this criterion for control variable selection is however no guarantee of effective causal estimation especially when using cross-sectional data. Among other things, your dataset might not contain variables for all the controls you need. That possibility should be explored. At the same time, researchers can consider whether any of the controls they might initially want to use actually lie on a path between X and Y (i.e., $X \rightarrow W \rightarrow Y$). At a minimum, we would want to see some clarity on *why* a set of control variables has been used – and the ‘other determinants’ idea does not give much clarity.

A brief consideration of propensity-score matching (PSM) is likely useful here. PSM is centred around the idea of propensity for selection into ‘treatment’. Matching on that basis is the means of identifying the effect of the treatment: PSM ‘adjusts’ a bivariate comparison of means, for the purpose of minimizing bias in one’s estimate. The variables used for matching are thus antecedents of the treatment. There is obvious alignment with the idea of selecting control variables (for regression models) that are antecedents of X (so, $W \rightarrow X$). Done correctly, a regression analysis would work with the same logic. (In principle, the two approaches should yield the same results; the reason PSM can’t always be used in place of regression is that PSM is feasible only for categorical independent variables, usually a dichotomy.)

2.3 Towards a Better Set of Terms

To use controls effectively, it likely helps to use a set of terms different from what we commonly see. Many researchers believe that use of controls gives us a ‘net’ effect, and this ‘net effect’ is then understood to be the ‘true’ effect.

What we need to ask is: what is this ‘net effect’ *net of*? If we include controls that lie on the path from X to Y (i.e., $X \rightarrow W \rightarrow Y$), we get a result that is ‘net’ of part of the impact of X itself. A good way to express that idea is that our estimate is *biased*.

A different set of terms is more useful. In the first instance, we almost certainly want to know about the *total* effect of X on Y . To get an unbiased estimate, we again need controls that are antecedents of X and Y .

We might subsequently want to know about *mechanisms* for the effect of X on Y . Here it could in fact make sense to add a variable that lies on a path from X to Y . In the context of path analysis, we could then determine an ‘indirect effect’ of X – the portion of the (total) impact of X that travels through some other variable corresponding to our idea about a mechanism. Loss of income is a mechanism for the impact of unemployment on happiness. We could calculate that component, again using a path analysis framework (or, to use a different term, a structural equation model, SEM).

But if we include income in the model, the coefficient for unemployment itself is no longer *the effect* (the *total* effect) of unemployment. In the language of path analysis, it is the ‘direct effect’. It is far from clear that we can give a sensible interpretation of direct effects, in substantive terms.

The terms used in path analysis are clearer and more easily understood than the idea of a ‘net effect’. That idea is often misunderstood: people often say that it is the effect of X on Y net of all of the *other* variables that affect Y . But if we control for variables (W) that are themselves influenced by X (so, $X \rightarrow W$), then our estimate for X is ‘net’ of part of the effect of X itself. In other words, it is contaminated by ‘overcontrol bias’.

2.4 Don’t Interpret Coefficients for Controls

If we have selected controls properly, in line with the criteria described above, it follows that we cannot interpret the coefficients of the controls as (total) effects of those variables ($W \rightarrow Y$) (see Keele et al., 2020). The estimate for $W \rightarrow Y$ with X in the model reflects the fact that X is being ‘controlled’. Everything written above about overcontrol bias is now relevant to our interpretation of the coefficients of the W s. Having selected controls that

are antecedents of X , we now find X on a path from W to Y : $W \rightarrow X \rightarrow Y$. At best, the coefficients for W (in a model specified to give us an unbiased estimate of $X \rightarrow Y$) will be *direct* effects. These coefficients cannot be interpreted as *total* effects. Trying to interpret them as direct effects will likely involve rhetorical contortions. It would be important (but perhaps difficult) to avoid implying that all the coefficients in a model are equivalent effects. The coefficient for X is the ‘total effect’; the coefficients for the various W s are ‘direct effects’.

The much better choice is to refrain from interpreting the results for the controls. They are very likely not relevant to your research question anyway; discussing them is unnecessary, a distraction.

A closely related idea is that we cannot use a single model to tell us about all of the various ‘determinants’ of a particular dependent variable. We will need different models to explore different research questions. The controls we need will be different, depending on what our focal independent variable is. In the perspective articulated here, we focus on a particular effect, the impact of a focal independent variable on the outcome of interest. The other variables are merely controls. A useful idea in this context is the ‘Table 2 fallacy’ (Westreich & Greenland, 2013). A strong signal that a manuscript has fallen prey to that fallacy is a title that includes the following phrase: ‘The Determinants of ...’.

2.5 Take Seriously the Limitations of Cross-sectional Analyses

Whether statistical models (using ‘observational’ data) can give us results that constitute evidence of causal effects is sometimes disputed. Some people hold that only experiments (‘randomized controlled trials’) can establish causality. That view seems needlessly stringent, in part given the difficulties of applying it in the social world. There is a lot of guidance for quantitative analysis geared towards the development of work intended to identify causal relationships. An important common theme is that causal identification comes not from ‘statistics’ per se, but rather from research design.

In that spirit, the issues we address here (above) are usefully considered a ‘minimum’, pitched at a level relevant to the work being done in many of the manuscripts we receive. So, a caution bears repetition: a cross-sectional model likely will not protect against other threats to causal inference (i.e., even if we have the right control variables). Other methods might be more effective for assessing whether our estimate of $X \rightarrow Y$ really represents the effect of X on Y . If your study involves a cross-sectional analysis, you might want to reflect: how might the results of a longitudinal analysis differ (i.e., if we had the data for it)? What about use of instrumental variables, or propensity-score matching (etc.)? Not every study has to use the most advanced methods – but exploring questions along these lines can lead to informed reflection about your results.

For example: results from longitudinal analyses often (though not always) amount to smaller effect sizes than cross-sectional results, for the same research question. The analytical design, evaluating how change in X is associated with change in Y ‘within’ the individual respondents, is more effective in removing the influence of ‘confounders’ (the time-constant ones). So, a possibility to consider for one’s cross-sectional results is simply that the focal estimate might be biased in the sense of overstating the ‘real’ effect. This is a more useful discussion, in contrast to a vague suggestion that future research should apply a longitudinal analysis.

What's needed above all is some focused thought about what the results of our models might mean. Some researchers move directly from seeing 'significant' results to concluding that there is an 'effect'. This haste carries big risks; it is easy to misinterpret one's results. (Roodman, 2024, describes a telling example where researchers got it wrong even in the context of a longitudinal analysis.) Martin (2018) encourages researchers to consider what sort of underlying social reality might produce the data and then the results we see in our analysis of those data. This advice usefully inverts the more common practice of believing that our data and analyses give us unmediated access to social realities.

3 Proper Use of Statistical Significance

Misuse of statistical significance is very common. Researchers often evaluate their results *solely* by considering whether they are statistically significant (in practice, asking only whether there are asterisks in their output/tables). This is insufficient information. There are also potential fallacies and misinterpretations (Greenland et al., 2016; Carver, 1978).

3.1 Consider Effect Size, not Just 'Significance'

Statistical significance has a precise and limited meaning; at best, it might tell us only whether our results, derived from analysis of sample data, are likely to be found in the corresponding population.¹

In part, the problem is rhetorical. It is all too easy for authors to offer an elision: results that are 'statistically significant' (in an earlier passage) become results that are 'significant' (in a later passage). There's a problem in the implicit notion that statistical significance on its own can underpin some notion of substantive significance. Having 'significant' results requires more than asterisks.

What hypothesis tests and statistical significance do *not* tell us is: *how much* the magnitude of an effect differs from zero. If results are statistically significant, a further question then arises: how big is the effect? One reason that question is potentially important is that, with a sufficiently large sample, an effect can be statistically significant (simply because the standard error is smaller) but nonetheless very small. In that sense a 'significant' result might well be decidedly *non*-significant in substantive terms. The point has been applied to investigation of happiness by Geerling and Diener (2020). More broadly, consider Wasserstein et al. (2019) and Engman (2013).

Please: avoid an elision between 'statistically significant' and 'significant' in a more general sense. Get to know the literature on effect size, and apply it to your results. The point is not that only 'large' effects merit attention and publication. Funder and Ozer (2019), for example, show how ostensibly small effects can accumulate over time. There might also be a role for new studies to show how 'significant' effects in earlier research are actually quite small (i.e., as a counterweight to 'publication bias').

¹ Some statisticians and quantitative researchers hold that statistical significance cannot in fact enable causal inference in that sense. As commonly used, the procedure of 'null hypothesis significance testing' involves a deep logical flaw. For a summary and historical overview of this important perspective, see Gill, 1999.

3.2 Avoid Using Statistical Significance when the ‘Assumptions’ are not Met

As noted, statistical significance might tell us only whether our results, derived from analysis of sample data, are likely to be found in the corresponding population. To be effective in this sense, we must be working with data from a sample that is understood to be representative of the corresponding population (however conceived). For many commonly used datasets (e.g. the European Social Survey and the World Values Survey), researchers are likely on solid ground – though you might want to consider whether non-response bias is potentially affecting your results.

In other situations, the idea of having sample data that are representative of a corresponding population is less robust. Convenience samples are the obvious example. It likely doesn't make sense to use statistical significance when evaluating the results of research using convenience samples (and perhaps other types/methods where representativeness is in question). If you think it *does* make sense, please tell us why; make your case, rather than taking it for granted (or, worse, fostering the impression that you don't understand the issue).

Statistical significance would likely have unclear purpose also when used in analysis of ‘higher-level’ units, e.g. countries or regions (Lucas, 2014). It is very uncommon to see samples of countries or regions that are representative of some larger ‘population’. Typically the reason for inclusion is simply data availability. If one's data includes all (or nearly all) of the countries/regions in a particular category, then extrapolation is irrelevant, and statistical significance is not a meaningful way of evaluating results.

4 Synthetic Indices: Approaches and Analytical Models

The topic of synthetic indices is central to *Social Indicators Research*, as evidenced by the many papers, both methodological and applicative, published over the years (e.g. Alaimo et al., 2021; Casadio-Tarabusi and Guarini 2013; Cherchye et al., 2007; Fattore, 2016; Mazzotta & Pareto, 2016; Ruiz et al., 2022). The analysis and understanding of multidimensional and complex socio-economic phenomena require the definition of systems of indicators. The latter, being complex, require approaches facilitating more concise representations. The guiding concept is synthesis. The right way of understanding socio-economic phenomena is to conceive them as a whole, adopting a synthetic approach. Any synthesis should be a stylization of reality (not an over-simplification). In dealing with systems of indicators, the synthesis must be a meaningful measure, capable of representing the complex system without trivialising or simplifying it (Alaimo, 2022).

From a methodological point of view, synthesis can be achieved via two different approaches: aggregative-compensative and non-aggregative.

As suggested by the term, the aggregative approach consists in the aggregation, by means of a mathematical function, of the basic/elemental indicators. These methodologies are defined as composite indicators (Maggino, 2017; OECD, 2008).

Effective construction of composite indicators is a challenging task. It involves completing a series of stages to ensure a reliable result, as well as the benchmarking of the results. To maximize robustness and validity, the most appropriate methodological choices must be made in each of those steps.

As a guide for researchers working on this topic, we highlight the five stages established in the literature that should be observed in the analytical model for constructing composite indicators (Jiménez-Fernández et al., 2022; Maggino, 2017; Mazziota & Pareto 2021; Nardo et al., 2005; Terzi et al., 2021).

(1) Define the phenomenon to be measured (the latent construct) and the conceptual framework, which in turn requires identifying the nature and direction of the structural relationships between the latent construct and the observed variables. It must be identified whether the measurement model is formative or reflective.

(2) Select a group of variables or individual indicators that represent the phenomenon to be studied according to the conceptual framework. It is not simply about collecting indicators, but about developing a system of indicators as an interconnected set of elements, organized according to the conceptual model. Individual indicators must be clearly defined and their choice justified. What is the statistical nature of the individual indicators? What is the correlation between the individual indicators and the latent construct?

(3) The normalization of the individual indicators offers several methodological paths, each with its statistical advantages and disadvantages. The choice of the type of normalization must be adequately justified, in statistical terms.

(4) Weighting and aggregating the normalized indicators using a mathematical method. In this choice, it must be justified why the chosen method (compensatory, partially compensatory or non-compensatory) is appropriate according to the phenomenon to be studied.

(5) Finally, any composite indicator proposal must be validated by conducting a robustness assessment that evaluates its ability to produce correct and stable measurements.

Composite indicators have been widely used in the literature for assessing social progress and making comparisons between countries in different fields. They have been widely used by various international organisations and actors to measure diverse phenomena. The main purpose of their success is informative. It is easier for the public to understand a synthetic indicator (one single measure) than multiple elementary indicators. For these reasons, the aggregative-compensative approach is the dominant framework in the literature, so much so that the term composite is sometimes mistakenly used as a synonym for synthetic indicator.

There are, as well, methods belonging to the so-called non-aggregative approach, in which the synthesis is achieved without any mathematical aggregation of the elementary indicators. These methods have become very popular in recent years, mainly due to the objective of finding methods suitable for dealing with systems of indicators at different scaling levels. (The composite approach is suitable for use only with cardinal elementary indicators.)

Among the different methodologies belonging to this approach, we can cite Social Choice Theory (Sen, 1977); Multi-Criteria Analysis (Macoun & Prabhu, 1999; Nijkamp & van Delft, 1977; Zopounidis and Pardalos 2010), and Partially Ordered Set (poset) Theory (Negggers & Kim, 1998; Fattore, 2016). The field is therefore moving towards the identification of methods that do not depend on the scale level of the indicators and can therefore be generally used.

Researchers intending to submit manuscripts doing work of this sort are encouraged to consider closely the following points (corresponding to potential pitfalls) in developing their work.

Choice of measurement model: The correct distinction between formative and reflective measurement models is linked to the correct definition of the latent variable and the

conceptual model and allows not only to correctly interpret the relationships between the indicators but also to correctly identify the procedure aimed at their synthesis. The problem is therefore not that it is easier (or less easy) to use one model rather than another, but rather the appropriateness of the model in light of the phenomenon that one intends to study. One common error in the construction of a synthetic indicator is to use a reflexive measurement model to measure formative latent variables or vice versa. The use of a model is not purely a choice of the researcher but depends on the latent variable to be measured. For more details, see: Alaimo & Maggino, 2020; Diamantopoulos & Siguaw, 2006; Diamantopoulos et al., 2008; Maggino, 2017.

Choice of the method of synthesis: sometimes a synthetic index is confused with a composite index, leading to the use of aggregate-compensation methods even for the synthesis of mixed indicator systems, in which there are indicators at different scale levels (not only cardinal, but also ordinal). This is a methodologically (as well as conceptually) incorrect choice that leads to misleading and meaningless results. The choice of synthesis method depends on the nature of the elementary indicators considered and must always be adequately motivated and justified. For more details, see: Alaimo et al., 2021; di Bella et al., 2017; Fattore, 2016; Gatto & Busato, 2020).

Composite index construction: focusing on composite indices (aggregative-compensation approach), many misconceptions, both conceptual and operational, can undermine their effective construction. Among the conceptual pitfalls, one of the most frequent is related to a poor (or sometimes missing) operationalisation of the concept being measured. We have to consider: What is the phenomenon we want to study? Defining a concept is always an abstraction process, a complex phase that requires the identification and definition of theoretical constructs involving the researcher's point of view, the applicability of concepts, the socio-cultural context, and the geographical and historical context. The conceptualisation process allows us to identify and define: the model aimed at constructing the data; the spatial and temporal sphere of observation; the aggregation levels; the models allowing the interpretation and evaluation. This is a challenging exercise, especially when the concept is very complex (as with well-being, sustainability, quality of life). If a phenomenon is poorly defined, then it will certainly be poorly measured. However, the opposite is not true. If the phenomenon is well defined and the matrix is composed of elementary indicators of good quality, it is not always true that the composite index is valid (for example, the methodology used, which must be consistent with the phenomenon studied). Another potential error is related to the 'choice' of measurement model (as noted above).

Another essential consideration is the **selection of the elementary indicators**. At this stage several misunderstandings can arise, related to different issues that need to be addressed.

- *All dimensions of the phenomenon must be represented:* the multidimensional nature of socio-economic phenomena necessitates that, in order to measure them in the best possible way, all identified dimensions should be measurable through the presence of at least one elementary indicator.
- *Generally, the presence of several indicators in a system is useful to increase the reliability of the measurement:* the more numerous the indicators, the smaller the random error of the latent construct measurement. However, we are often faced with systems with too many indicators and synthesis is not possible. It is then necessary to reduce

the number of indicators. There is no universal rule for this. One must always have in mind the theoretical framework and the measurement model. For example, if we were to eliminate one or more indicators from a system, in a reflective measurement model the choice would logically fall on those less correlated to the others (since they do not ‘reflect’ the latent variable); on the contrary, in a training model, the choice to eliminate the indicators should be made among those most correlated with each other, as they are an expression of the same ‘cause’ that forms the latent variable.

- *The polarity of each indicator must be well defined*: the polarity is the sign of the relationship between the elementary indicator and the phenomenon you want to measure. An indicator can have polarity: positive, if it has the same ‘direction’ as the phenomenon you want to measure. For example, in the human development index (HDI) GDP has a positive polarity; negative, if it has the opposite direction compared to the phenomenon to be measured. For example, in the index of poverty (MPI), GDP has a negative polarity. Therefore, the concept of polarity is not absolute, but relative. This is potentially one of the main errors in the polarity issue, i.e. considering the polarity in absolute terms and not relative to the nature of the phenomenon we want to measure. To construct a synthetic index, it is necessary that all indicators have positive polarity. Therefore, in the case of basic indicators originally having negative polarity, these must be inverted.
- *Assumptions about the nature of indicators must be explicit* (substitutability; non-substitutability): This is one of the main assumptions about indicators. The components of a synthetic index are said to be either: (a) substitutable, if a deficit in one component can be compensated by a surplus in another one. The assumption of substitutability of the components involves the adoption of additive aggregation methods (e.g. the arithmetic mean); or (b) non-substitutable, if compensation between them is not permitted. In the case of partial substitutability or non-substitutability of the components, generally, multiplicative methods (e.g. the geometric mean) or non-compensatory methods are adopted. The choice regarding substitutability is linked more to conceptual issues than to methodological ones.
- For more details, see: Alaimo, 2022; Mazziotta & Pareto, 2013; Mazziotta & Pareto, 2016; Maggino, 2017.

5 Towards Research Transparency

In line with developments at other journals (and the ‘open science’ movement more generally, see e.g. Munafò et al., 2017, Loder et al., 2024), we will now ‘strongly encourage’ authors to make their analysis code available to reviewers and readers.

One method is to include (in a manuscript submission) a link created on osf.io, where your code can be downloaded. The link can (and should) be set up as *anonymised*, to preserve blind peer review. Here’s an example of the sort of statement/footnote that could be included in a manuscript: ‘The analysis syntax for this paper is available (anonymously) here: https://osf.io/zpcxj/?view_only=e384bd25ac6f40eaaa1e273cc6417184’.

To ensure that your code is *useful*, it should be annotated (as in the example linked here). You can also include information on software versions as well as any subordinate packages/libraries.

Making your code available is in your interest as an author. Many reviewers are unlikely to inspect the code, when they review a manuscript. But the *availability* of code is likely to enhance reviewers' confidence in the work. That confidence is merited insofar as authors are likely to be more careful when they know that their code will be available. After all, reviewers and editors *might* inspect it. We understand that availability of code might not enable full replication of results (e.g. given restrictions on data availability). Even so, reviewers and readers might learn something useful from it.

We are not insisting on access to data. The publisher has a policy that encourages authors to facilitate access to data where possible and applicable. Please visit the following link (which offers suggestions for data repositories): [https://link.springer.com/journal/11205/submission-guidelines#Instructions for Authors_Research Data Policy](https://link.springer.com/journal/11205/submission-guidelines#Instructions_for_Authors_Research_Data_Policy).

If you can't (or don't want to) make your code available, it will help to tell us why, using the cover letter field in our editorial submission system. We accept that there can be good reasons.

6 General Expectations

There are a number of more basic items that would pertain to evaluation of any scholarly output. We mention items that seem important, but we are not seeking to be comprehensive.

1. Articulate the *contribution* of your work. How does it compel us to change/update our understanding of the topic you are working on? Where and how does previous research lead us astray? How does your own research overcome the deficiencies or limitations of previous research? Why should we adopt *your* approach/perspective?

Passages doing that work would typically appear in the literature review. But it is equally important to develop this discussion in the conclusion, after you have presented your results. How are your findings different/better? Assertions in that mode are effective when they are articulated in dialogue with previous research. It is common – and disappointing – to see concluding sections that have no citations at all. That's a signal that a manuscript is likely not effectively demonstrating the contribution of one's research.

2. Be mindful of 'publication bias'. Having 'significant' results (in the statistical and substantive senses) is not the only basis for a claim to publication. Sometimes 'null' results are important, especially when evaluating previous research where a different conclusion was reached. A related topic is 'p-hacking'. If that term is unfamiliar, start with Simmons et al. (2011).
3. Writing effectively in English: *Social Indicators Research* is an English language journal and as such, we require that submitted manuscripts demonstrate an effective command of the language. Reviewers are less likely to relate well to a manuscript if the quality of writing does not meet that standard. Manuscripts may be desk rejected if they do not meet this expectation. They wouldn't fare well in peer review in any event.
4. The journal website has a link to 'submission guidelines'. One key point: a maximum of 10,000 words (including references) is given, and a manuscript that exceeds this amount more than trivially will likely be rejected on that basis alone.

For a first submission, we are less concerned about format, especially of references; it isn't clear why authors should have to reformat the references every time they submit a manuscript to a different journal. Some journals now allow 'format-free' initial submissions. We can't change (or omit) the publisher's general submission guidelines, but we can de-emphasize them for first submissions.

For any subsequent submission (following an invitation for revisions and resubmission), manuscripts *must* be formatted in alignment with the guidelines.

5. Review timelines: given the volume of submissions and the pressure on the peer review system in general, all journals currently struggle to give timely decisions to authors. The conventional expectation of a 3-month review period is rarely met. If you write to ask when you can expect to receive a decision, this is likely the only reply we will be able to give.

7 Conclusion

The guidance articulated here tells authors about the range of issues we hope to see addressed effectively in manuscripts. To apply it effectively, authors might need to explore further, via the references we have provided (and perhaps others as well); on its own, the text here is likely to be insufficient. That assertion applies in particular to the topic of analysis construction and causal interpretation. Additional citations (among the great many we could choose) include Pearl et al. (2016), Pearl and MacKenzie (2018), Gangl (2010), and Morgan and Winship (2007).

Our final point is a gentler one. We hope that the guidance offered here leads to submission of better manuscripts. Since you've read this far, you are in a position to consider for yourself whether our assertions are persuasive (and then to choose accordingly). At the same time, we genuinely accept that no research is perfect; all research comes with limitations (whether authors recognize them or not). We will strive not to be dogmatic and stubborn in our own application of the guidance as we evaluate your manuscripts – in part because even if you follow that guidance your research will *still* be imperfect (as is our own). In addition, for some manuscripts with good potential in general terms, certain deficiencies could be remedied via the review process.

Author Contributions The authors are editors of this journal; the article is not a peer-reviewed research manuscript.

Funding No funding was received in connection with this manuscript.

Declarations

Competing Interests The authors have no competing interests to declare.

References

- Alaimo, L. S. (2022). *Complexity of social phenomena: measurements, analysis, representations and synthesis*. Università La Sapienza.
- Alaimo, L. S., & Maggino, F. (2020). Sustainable development goals indicators at territorial level: Conceptual and methodological issues—the Italian perspective. *Social Indicators Research*, *147*(2), 383–419. <https://doi.org/10.1007/s11205-019-02162-4>
- Alaimo, L. S., Arcagni, A., Fattore, M., & Maggino, F. (2021). Synthesis of multi-indicator system over time: A poset-based approach. *Social Indicators Research*, *157*(1), 77–99. <https://doi.org/10.1007/s11205-020-02398-5>
- Bartram, D. (2021). Cross-sectional model-building for research on subjective well-being: Gaining clarity on control variables. *Social Indicators Research*, *155*(2), 725–743. <https://doi.org/10.1007/s11205-020-02586-3>
- Carver, R. (1978). The case against statistical significance testing. *Harvard Educational Review*, *48*(3), 378–399. <https://doi.org/10.1177/haer.48.3.t490261645281841>
- Casadio Tarabusi, E., & Guarini, G. (2013). An unbalance adjustment method for development indicators. *Social Indicators Research*, *112*(1), 19–45. <https://doi.org/10.1007/s11205-012-0070-4>
- Cherchye, L., Moesen, W., Rogge, N., & Puyenbroeck, T. V. (2007). An introduction to ‘benefit of the doubt’ composite indicators. *Social Indicators Research*, *82*(1), 111–145. <https://doi.org/10.1007/s11205-006-9029-7>
- Cinelli, C., Forney, A., & Pearl, J. (2022). A crash course in good and bad controls. *Sociological Methods & Research*. <https://doi.org/10.1177/00491241221099552>
- Di Bella, E., Corsi, M., & Leporatti, L. (2017). POSET analysis of panel data with POSAC. In M. Fattore & R. Bruggemann (Eds.), *Partial Order Concepts in Applied Sciences* (pp. 161–176). Springer International Publishing. https://doi.org/10.1007/978-3-319-45421-4_11
- Diamantopoulos, A., & Sigauw, J. A. (2006). Formative versus reflective indicators in organizational measure development: A comparison and empirical illustration. *British Journal of Management*, *17*(4), 263–282. <https://doi.org/10.1111/j.1467-8551.2006.00500.x>
- Diamantopoulos, A., Riefler, P., & Roth, K. P. (2008). Advancing formative measurement models. *Journal of Business Research*, *61*(12), 1203–1218. <https://doi.org/10.1016/j.jbusres.2008.01.009>
- Engman, A. (2013). Is there life after $P < 0.05$? Statistical significance and quantitative sociology. *Quality & Quantity*, *47*(1), 257–270. <https://doi.org/10.1007/s11135-011-9516-z>
- Fattore, M. (2016). Partially ordered sets and the measurement of multidimensional ordinal deprivation. *Social Indicators Research*, *128*(2), 835–858. <https://doi.org/10.1007/s11205-015-1059-6>
- Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: sense and non-sense. *Advances in Methods and Practices in Psychological Science*, *2*(2), 156–168. <https://doi.org/10.1177/2515245919847202>
- Gangl, M. (2010). Causal inference in sociological research. *Annual Review of Sociology*, *36*(1), 21–47. <https://doi.org/10.1146/annurev.soc.012809.102702>
- Gatto, A., & Busato, F. (2020). Energy vulnerability around the world: The global energy vulnerability index (GEVI). *Journal of Cleaner Production*, *253*, 118691. <https://doi.org/10.1016/j.jclepro.2019.118691>
- Geerling, D. M., & Diener, E. (2020). Effect size strengths in subjective well-being research. *Applied Research in Quality of Life*, *15*(1), 167–185. <https://doi.org/10.1007/s11482-018-9670-8>
- Gill, J. (1999). The insignificance of null hypothesis significance testing. *Political Research Quarterly*, *52*(3), 647–674. <https://doi.org/10.1177/106591299905200309>
- Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: A guide to misinterpretations. *European Journal of Epidemiology*, *31*(4), 337–350. <https://doi.org/10.1007/s10654-016-0149-3>
- Jiménez-Fernández, E., Sánchez, A., & Ortega-Pérez, M. (2022). Dealing with weighting scheme in composite indicators: An unsupervised distance-machine learning proposal for quantitative data. *Socio-Economic Planning Sciences*, *83*, 101339. <https://doi.org/10.1016/j.seps.2022.101339>
- Keele, L., Stevenson, R. T., & Elwert, F. (2020). The causal interpretation of estimated associations in regression models. *Political Science Research and Methods*, *8*(1), 1–13. <https://doi.org/10.1017/psrm.2019.31>
- Land, K. C., & Michalos, A. C. (2018). Fifty years after the social indicators movement: Has the promise been fulfilled? An assessment and agenda for the future. *Social Indicators Research*, *135*(3), 835–868. <https://doi.org/10.1007/s11205-017-1571-y>
- Loder, E., Macdonald, H., Bloom, T., & Abbasi, K. (2024). Mandatory data and code sharing for research published by the *BMJ*. *Bmj*, q324. <https://doi.org/10.1136/bmj.q324>
- Lucas, S. R. (2014). An inconvenient dataset: Bias and inappropriate inference with the multilevel model. *Quality & Quantity*, *48*(3), 1619–1649. <https://doi.org/10.1007/s11135-013-9865-x>

- Macoun, P., & Prabhu, R. (1999). *Guidelines for applying multi-criteria analysis to the assessment of criteria and indicators*. CIFOR.
- Maggino, F. (Ed.). (2017). *Complexity in society: From indicators construction to their synthesis* (Vol. 70). Springer International Publishing. <https://doi.org/10.1007/978-3-319-60595-1>
- Martin, J. L. (2018). *Thinking through statistics*. The University of Chicago Press.
- Mazziotta, M., & Pareto, A. (2013). Methods for constructing composite indices: One for all or all for one? *Rivista Italiana Di Economia Demografia E Statistica*, 67, 67–80.
- Mazziotta, M., & Pareto, A. (2016). On a generalized non-compensatory composite index for measuring socio-economic phenomena. *Social Indicators Research*, 127(3), 983–1003. <https://doi.org/10.1007/s11205-015-0998-2>
- Mazziotta, M., & Pareto, A. (2021). Everything you always wanted to know about normalization. *Rivista Italiana Di Economia Demografia E Statistica*, 75, 41–52.
- Morgan, S. L., & Winship, C. (2007). *Counterfactuals and causal inference: Methods and principles for social research*. Cambridge University Press.
- Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Du Sert, P., Simonsohn, N., Wagenmakers, U., Ware, E. J., J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, 1(1), 0021. <https://doi.org/10.1038/s41562-016-0021>
- Nardo, M., Saisana, M., Saltelli, A., & Tarantola, S. (2005). *Tools for composite indicators building*. European Commission Institute for the Protection and Security of the Citizen.
- Neggers, J., & Kim, H. S. (1998). *Basic posets*. World Scientific.
- Nijkamp, P., & van Delft, A. (1977). *Multi-criteria analysis and regional decision-making*. Nijhoff.
- OECD. (2008). *Handbook on constructing composite indicators: Methodology and user guide*. OECD.
- Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. Allen Lane.
- Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal inference in statistics: A primer*. Wiley.
- Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science*, 1(1), 27–42. <https://doi.org/10.1177/2515245917745629>
- Roodman, D. (2024). *Long-term effects of India's childhood immunization program on earnings and consumption expenditure: Comment*. <https://doi.org/10.48550/ARXIV.2401.11100>
- Ruiz, A. B., Luque, M., & Marcenaro-Gutierrez, O. D. (2022). On the use of synthetic indexes based on multi-criteria decision making to study the efficiency of teachers. *Social Indicators Research*, 163(3), 1269–1300. <https://doi.org/10.1007/s11205-022-02944-3>
- Sen, A. (1977). Social choice theory: A re-examination. *Econometrica*, 45(1), 53. <https://doi.org/10.2307/1913287>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359–1366. <https://doi.org/10.1177/0956797611417632>
- Terzi, S., Otoi, A., Pareto, A., Grimaccia, E., & Mazziotta, M. (2021). *Open issues in composite indicators. A starting point and a reference on some state-of-the-art issues*. Roma TrE-. <http://romatpress.uniroma3.it/wp-content/uploads/2021/03/open-togmp-2.pdf>
- Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond ' $p < 0.05$ '. *The American Statistician*, 73(sup1), 1–19. <https://doi.org/10.1080/00031305.2019.1583913>
- Westreich, D., & Greenland, S. (2013). The table 2 fallacy: Presenting and interpreting confounder and modifier coefficients. *American Journal of Epidemiology*, 177(4), 292–298. <https://doi.org/10.1093/aje/kws412>
- Wysocki, A. C., Lawson, K. M., & Rhemtulla, M. (2022). Statistical control requires causal justification. *Advances in Methods and Practices in Psychological Science*, 5(2), 1–19. <https://doi.org/10.1177/25152459221095823>
- Zopounidis, C. (Ed.). (2010). *Handbook of multicriteria analysis*. Springer.