Improving measurement models in clinical epidemiology

Journal of Clinical Epidemiology 118 (2020) 119–123

Journal of Clinical Epidemiology

COMMENTARY

Improving measurement models in clinical epidemiology: time to move beyond the inherent assumption of an underlying reflective measurement model

Ludovic G.P.M van Amelsvoort[a,*], Bram P.I. Fleuren[b], IJmert Kant[a]

[a]Department of Epidemiology, Care and Public Health Research Institute (CAPRHI), Maastricht University, Maastricht, The Netherlands
[b]Department of Work- and Social Psychology, Faculty of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands

Accepted 2 November 2019; Published online 7 November 2019

Abstract

Objectives: The nature of a construct's measurement model, most decisively being predominantly reflective or formative, is essential for its development, validation, and use. Differentiating between these types of measurement models cannot be done based on statistics alone, but has to rely on expert judgment, preferably guided by checklists and theoretical assumptions. However, consideration and substantiation of the choices of the measurement models is lacking in most studies describing the validation of measurement instruments in the field of clinical epidemiology.

Study Design and Setting: A convenience sample of 96 clinimetric studies, published from 2017 up until May 17th, 2018 was scored on model use and (mis)specification.

Results: In over 50% of the identified studies in this sample, formative measurement models are considered and/or analyzed as reflective.

Conclusion: Misspecification of formative measurement models as reflective was found to be more rule than exception. It is therefore recommended that model selection and considerations on the theoretical nature of the measurement model should be classified, motivated, and discussed, for example, by using available checklists. Hereby, it can be ensured that the appropriate measurement models and corresponding statistics are used.

Keywords: Clinimetrics; Measurement models; Validation; Reflective constructs; Formative constructs; Patient reported outcomes; Checklist; Psychometrics

1. Introduction

Measurements are an essential element of day-to-day clinical practice and research. They help in quantifying and establishing disease risks and risk factors and diagnosing and follow-up of diseases and health complaints. Furthermore, adequate measurements are essential in the assessment and monitoring of the impact of diseases on health (such as quality of life) and functioning, with often a high relevance for patients. In addition to the broad range of physical- and chemical-based measurements, many other questionnaire- or checklist-based measurement instruments are being developed, tested, and used to capture relevant constructs in health and medicine.

Owing to extensive parallel efforts in several key disciplines such as psychology, economy, epidemiology, and biostatistics, considerable advances have been made in the development, validation, reporting, and critical appraisal of measurement instruments. Within the field of (clinical) epidemiology, advances such as those from the study by Kirshner and Guyatt [1] and the COSMIN group [2,3] have been influential. That is, owing to these efforts, evaluation and development guidelines for measurement instruments (e.g., tools for patient reported outcome assessment) have gained solid ground. However, the explicit specification of measurement models as formative or reflective is too often missing in publications on the development, validation, appraisal, and use of measurement instruments.

In reflective models, the (variations in) observed item scores are considered to be caused by (variations in) the

* Corresponding author: Department of Epidemiology, Care and Public Health Research Institute (CAPRHI), Maastricht University, Debyeplein 1, 6229HA Maastricht, The Netherlands. Tel.: +31433882363; fax: +31 43 388 41 28.

E-mail address: lgpm.vanamelsvoort@maastrichtuniversity.nl (L.G.P.M van Amelsvoort).

https://doi.org/10.1016/j.jclinepi.2019.11.003
0895-4356/

What is new?

Key findings

- Constructs and measures are frequently (>50% of cases) validated using wrongly specified measurement models.

What this adds to what was known?

- A conceptual model to illustrate how to differentiate between reflective and formative models is introduced.
- The distinction between formative and reflective measurement models is insufficiently recognized within clinical epidemiology.

What is the implication and what should change now?

- Whether a construct or measure, as well as its eventual subscales, is of a reflective or formative nature needs careful consideration.
- Careful selection and explication of the appropriate measurement model pertaining to a construct or instrument can be facilitated by using a checklist.
- Motivating and discussing the choice of the measurement model should be an explicit part of studies addressing the construction, validation and/or systematic review of constructs or measures.

underlying latent construct [4]. For example, it is because person A experiences a high amount of fatigue that person A scores high on items 1, 2, and 3, measuring fatigue (see Fig. 1A). As the assumption of unidimensionality prescribes, all indicators (items) measure the same underlying construct, are correlated, and show internal consistency. Consequently, removing individual reflective items does not alter the meaning of the construct. In formative measurement models, constructs are caused by scores on the observed items. For example, person A is considered to have a high socioeconomic status because (s)he has a high income, a high education level, and a prestigious occupation (Fig. 1B). Indicators can, but are not required to, be correlated. Moreover, if one of the formative indicators is removed from the measurement model, the meaning of the construct can change. In addition, where reflective indicators are all of the same type, formative indicators can be differentiated as being of the causal, composite, or covariate indicator type [4]. Moreover, even underlying factors considered to be defining a formative model, can, in turn, be of a formative nature.

Using an inappropriately specified measurement model can lead to poor, suboptimal, or invalid measurements of constructs. Consequently, false conclusions on the validity or generalizability of the construct are likely as well. Although these issues have been highlighted for decades by distinguished scholars such as Wright and Feinstein [7], they are still neglected in many publications. A particularly salient example exists in the field of consumer research, where model misspecification is found to be common [4,8]. Or in the words of Bollen and Diamantopoulos [9], "nearly all of the approaches and tools for measurement in the social and behavioral sciences implicitly (if not explicitly) assume that the indicators depend on the latent variable and do not take account of the possibility of causal/formative indicators."

2. Estimating the extend of model misspecification

To obtain a crude indication of the extent of model specification and potential misspecification within epidemiological or medical research, we assessed a sample of 96 studies from January 1st, 2017, up until May 17th, 2018, from the Web of Science database [10] citing the seminal clinimetric publication by Mokkink et al. [2]. Although this only represents a selective sample of measurement studies, these studies should at least have considered methods for measurement validation as they refer to the key COSMIN publication. The descriptions of the measurement models in these studies and their potential misspecification are scored by the authors. As illustrated in Table 1, 27 of 40 original studies and 26 of 28 systematic reviews failed to correctly specify or substantiate whether the measurement model was reflective or formative. This demonstrates that model

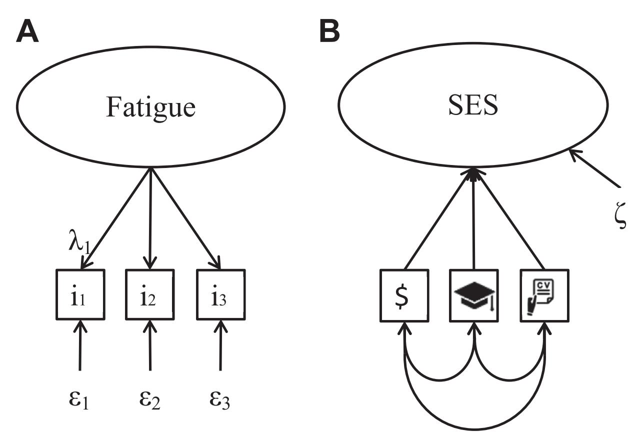

Fig. 1. Examples of a reflective (A) and a formative (with causal indicators) measurement model (B), [5]. In (A), the latent variable fatigue causes scores on the three items, all measuring the same construct. Item-level error is depicted by the ε's and factor loadings by the λ's. In (B), the not-directly observable variable socioeconomic status (SES) is caused by scores on the items, each measuring a different aspect of SES (i.e., income, education level, and occupation). Weights on formative construct are depicted by the γ's, possible correlations by the r's and the error term of the construct by the ζ. In both figures, square boxes represent observed variables, whereas the ovals represent unobserved (latent or indirectly observed) variables as estimated from the items (From Fleuren et al. [6]).

Table 1. Scoring of model specification or misspecification of studies that referred to Mokkink et al. [2] in 2017 and 2018 (up until May 17th)

Original studies $N = 40$	Probably correctly considered model as formative	Probably correctly considered model as reflective and analyzed that way	Likely formative but treated as reflective
Models considering single-factor measurement constructs ($n = 23$).	5 (21%)	3 (13%)	16 (66%)
Multifactorial models with potential of model misspecification of the individual factors ($n = 17$)	1 (6%)	4 (23%)	12 (71%)
Systematic reviews $n = 28$	Discussed formative/reflective nature of construct		Likely covering a (at least partly) formative construct, but no formative model considered/discussed
	2 (7%)		26 (93%)

Number of studies with correct or incorrect formative model use. A detailed overview of codes studies can be found in the Appendix. Twenty-eight studies excluded because the manuscripts were no validation study, translation or covered a one-item instrument.

specification is still often neglected. Moreover, model misspecification was found to be common in both original studies as well as systematic review papers.

So, despite the large consequences of model misspecification, the underlying structural assumptions of the measurement model are rarely a point of consideration, elaboration or discussion. One can often only deduce the assumed underlying measurement model by the measures and methods used (i.e., those methods and statistics which are only appropriate for reflective models not formative models, such as Cronbach's alpha, item response theory, Rasch, and exploratory factor analysis [9]).

A major challenge in differentiating between reflective and formative models is that their distinction cannot be made on the basis of estimated statistics. Instead, model choice should be based on content and theoretical assumptions, thus requiring expertise in both measurement models as well as the subject specific domain. Several checklists have been developed to aid measurement model distinction. For example, the checklist by Fleuren et al. published recently in this journal ([6] provides a practical checklist to differentiate models). Furthermore, this study describes a way to handle formative constructs, when applicable.

3. A conceptual model to help illustrating the difference between formative and reflective models

The example of organic solvent components exposure via air might help illustrating the seemingly small, but

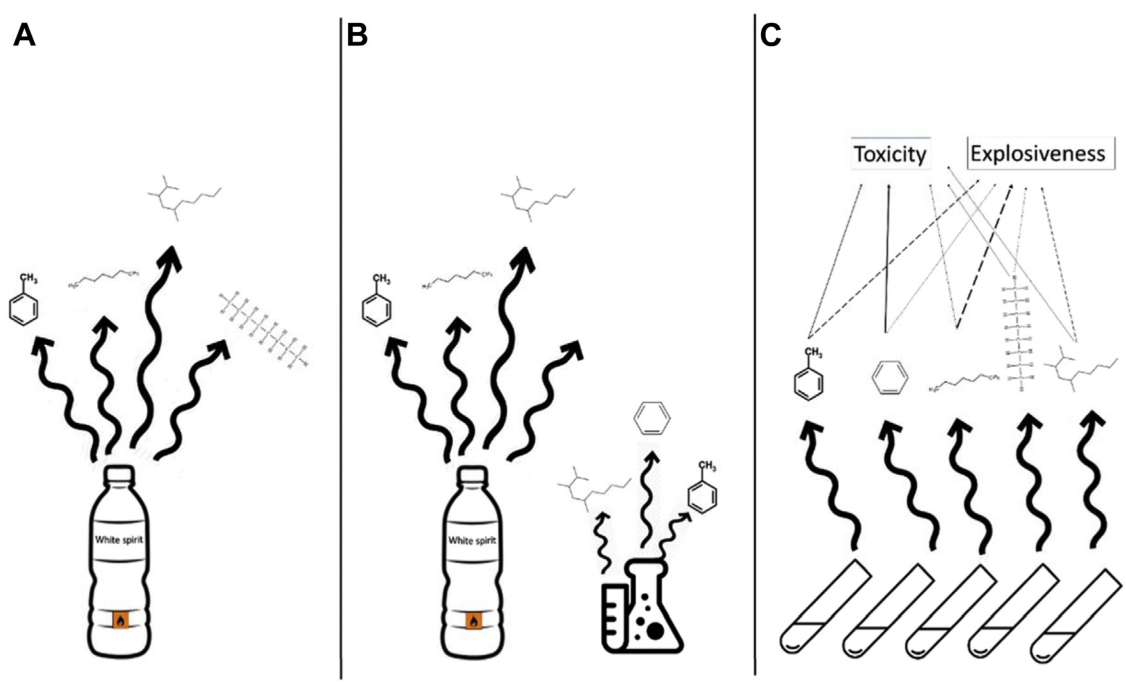

Fig. 2. Graphical representation of white spirit or other organic hydrocarbon exposure, following a reflective, mixed, or formative measurement model. (A). Reflective model. (B). Mixed model (including formative and reflective components). (C). Formative model. Width of arrows indicates different constructs needing different contribution of elements to construct.

Table 2. Application of Fleuren checklist [6] to construct of organic solvent exposure, based on models from Fig. 2

	Application of formative/reflective checklist to Fig. 2 models		
Checklist item	**Model A**	**Model B**	**Model C**
1. Are the indicators (items) (A) defining characteristics or (B) manifestations of the construct? ''A'' indicates a formative and ''B'' a reflective measurement model.	Indicators, that is, concentrations of organic solvents, are manifestations of the concentration of white spirit in the air (exposure).	The construct of organic solvent exposure is defined and characterized by a total of all individual indicators (each of the organic solvents).	The constructs of organic solvent exposure are defined and characterized by a weighted total of all individual indicators (depending on outcome, such as toxicity or explosiveness).
2. Would changes in the indicators/items cause changes in the construct or the other way around? The former indicates formative and the latter reflective.	Changes in the construct cause changes in the indicators.	Changes in the indicators would cause changes in the construct as each indicator might independently contribute to the total construct.	Changes in the indicators would cause changes in the construct as each indicator independently contributes to the total construct.
3. Should each indicator capture exactly the same? ''Yes'' indicates reflective; ''no, but they share conceptual unity in terms of causing a common construct'' indicates causal formative; and ''not at all'' indicates composite formative indicators.	Yes, although fixed different weights might apply for each indicator, depending on their relative presence in the total construct.	No, but components together share unity by being an organic solvent.	No, but components together share unity by being an organic solvent.
4. Would dropping one of the indicators alter the conceptual domain of the construct? ''Yes'' indicates formative; ''no'' indicates reflective.	No. See 6.	Yes, as total exposure to organic solvents is not possible when one of the components is dropped.	Yes, assessment of total organic solvent exposure, organic solvent toxicity or explosiveness is impaired.
5. Should a change in one of the indicators be associated with changes in the other indicators? ''Yes'' indicates reflective; ''no'' indicates formative.	Yes. See 6.	No, not necessarily, only for white spirit component.	No.
6. Are the indicators expected to have the same antecedents and consequences? ''Yes'' indicates reflective; ''no'' indicates formative.	No, as individual indicators might theoretically have different consequences. However, owing to their causal link to the construct, this differences cannot be separated from consequences of other indicators of the construct. Antecedents are the same, that is, their white spirit origin.	No, as different components might have different consequences. Also, antecedents might be different.	No, as different components might have different consequences. Also antecedents might be different.

essential differences between formative and reflective constructs. When we consider white spirit, which is a mixture of hydrocarbons commonly used as solvent for paints, all components individually will reflect the total spectrum of compounds in the mixture (assuming equal evaporation rates for all ingredients). The underlying construct (white spirit) causally links to the individual components, and thus also mutually connects these individual components. This means that the underlying measurement model will be of a reflective nature, as can be deduced by logic reasoning. That is, observing change in one of the components are related to a similar and proportional change in the other components as the levels are linked via the overall construct ''white spirit level'' (see Fig. 2A). In this case, concordance of indicators can be used as an indication for model validity. This means that concordance- or correlation-based methods such as exploratory factor analysis or measures such as the Cronbach's alpha apply [11]. Moreover, whatever the intended use of the construct (such as toxicity assessment or quantification the explosiveness of a mixture of air), the construct can be assumed to remain stable, even when applied to situations, or when the measurement of one item is exchanged for another.

If a mix of organic solvents is not bound together by such a strong underlying ''causal'' factor, as was the case in Fig. 2A, the automatic concordance or correlation between the individual indicators no longer necessarily holds. This is the case for a mixed (Fig. 2B) or pure formative model (Fig. 2C). However, the construct, such as organic solvent exposure, does not necessarily lose its relevance. Consider for example assessing the toxicity of a mixture of organic solvents as a risk factor for chronic solvent-induced encephalopathy (CSE), which is related to organic solvent exposure. Still, as every single component has a separate relevant contribution to the total construct of organic solvent exposure, the underlying correlational structure is of limited value. However, leaving out one of the components might lead to considerable misspecification of the total construct of organic solvent exposure. Moreover, the use and validity of the construct might be limited to a specific application of the construct. This limits the generalizability or applicability of the construct in other situations. A change in setting or application might even necessitate the development and validation of a new model with different relative weights and components for the construct. This is also illustrated with the different widths of arrows in Fig. 2C illustrating different relative contribution to different consequences. For example, consider the potential differences for assessing the explosiveness of organic solvents as compared with CSE-related toxicity. Both uses of the construct of "organic solvents" might end up with different relative contribution of the individual indicators to either the toxicity or explosiveness of the mixture.

Application of the checklist as provided in Table 1 of Fleuren et al. [6] to these three models A, B, and C (Fig. 2) also clearly indicates the distinction between model A as compared with models B and C (as indicated in Table 2). For models B and C, all six criteria would indicate a formative construct, whereas for model A, criteria 1 to 5 would clearly indicate a reflective construct. The application of criterion 6 is less clear cut for model A and its answer might remain indecisive.

4. Conclusion

To conclude, the large advancements in measurement model development and testing are not yet fully reflected in many clinimetric instrument development and validation studies. This means that in the development and validation of measurement instrument, model selection and establishing underlying measurement models deserves more attention. It is advised to consider this as a standard part of clinimetric publications, both in original research papers as well as systematic reviews on clinimetric instruments. The imminent consequences are that for measurement instruments, such as patient-reported outcome measures, model development and validation have to rely on linking to external factors for example by using mimic models [6]. Moreover, substantive subject expertise is needed in assessing model validity and completeness.

CRediT authorship contribution statement

Ludovic G.P.M van Amelsvoort: Conceptualization, Writing - original draft, Investigation. **Bram P.I. Fleuren:** Conceptualization, Writing - original draft, Investigation. **IJmert Kant:** Conceptualization, Writing - original draft.

Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jclinepi.2019.11.003.

References

[1] Kirshner B, Guyatt G. A methodological framework for assessing health indices. J Chronic Dis 1985;38(1):27–36.

[2] Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010;63:737–45.

[3] Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res 2010;19:539–49.

[4] Jarvis CB, MacKenzie SB, Podsakoff PM. The negative consequences of measurement model misspecification: a response to Aguirre-Urreta and Marakas. MIS Q 2012;36(1):139–46.

[5] Diamantopoulos A, Riefler P, Roth KP. Advancing formative measurement models. J Business Res 2008;61(12):1203–18.

[6] Fleuren BPI, van Amelsvoort L, Zijlstra FRH, de Grip A, Kant I. Handling the reflective-formative measurement conundrum: a practical illustration based on sustainable employability. J Clin Epidemiol 2018;103:71–81.

[7] Wright JG, Feinstein AR. A comparative contrast of clinimetric and psychometric methods for constructing indexes and rating scales. J Clin Epidemiol 1992;45:1201–18.

[8] Jarvis CB, MacKenzie SB, Podsakoff PM. A critical review of construct indicators and measurement model misspecification in marketing and consumer research. J Consum Res 2003;30(2):199–218.

[9] Bollen KA, Diamantopoulos A. In defense of causal-formative indicators: a minority report. Psychol Methods 2017;22(3):581–96.

[10] Web of Science. 2018. Available at http://webofknowledge.com/. Accessed May 7, 2018.

[11] Spitzer WO. The teacher's teacher: a personal tribute to Alvan R Feinstein. J Epidemiol Community Health 2002;56:328–9.