Neurobiology of Aging 31 (2010) 1656–1660

NEUROBIOLOGY OF AGING

www.elsevier.com/locate/neuaging

Commentary

Some guidelines for structural equation modelling in cognitive neuroscience: The case of Charlton et al.'s study on white matter integrity and cognitive ageing

Lars Penke *, Ian J. Deary [1]

Centre for Cognitive Aging and Cognitive Epidemiology, Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK

Received 11 June 2008; received in revised form 6 August 2009; accepted 14 October 2009
Available online 15 January 2010

Abstract

Charlton et al. (2008) (Charlton, R.A., Landua, S., Schiavone, F., Barrick, T.R., Clark, C.A., Markus, H.S., Morris, R.G.A., 2008. Structural equation modelling investigation of age-related variance in executive function and DTI-measured white matter change. Neurobiol. Aging 29, 1547–1555) presented a model that suggests a specific age-related effect of white matter integrity on working memory. We illustrate potential pitfalls of structural equation modelling by criticizing their model for (a) its neglect of latent variables, (b) its complexity, (c) its questionable causal assumptions, (d) the use of empirical model reduction, (e) the mix-up of theoretical perspectives, and (f) the failure to compare alternative models. We show that a more parsimonious model, based solely on the well-established general factor of cognitive ability, fits their data at least as well. Importantly, when modelled this way there is no support for a role of white matter integrity in cognitive aging in this sample, indicating that their conclusion is strongly dependent on how the data are analysed. We suggest that evidence from more conclusive study designs is needed.

Keywords: Structural equation modelling; Model specification; Model comparison; Cognitive aging; White matter integrity; Diffusion tensor imaging; General intelligence

1. Introduction

In an article recently published in this journal, Charlton et al. (2008) utilized structural equation modelling (SEM) to test various predictions that they derived from two dominant hypotheses of cognitive aging—the 'common cause' and the 'specific gain/loss' hypotheses (Span et al., 2004)—in a dataset of over 100 adults between 50 and 90 years of age. The data included a broad array of cognitive tests and measures of white matter integrity based on diffusion tensor imaging (DTI). SEM is a powerful statistical tool for analysing multivariate data, as often found in studies of neurobiology and ageing. Still, SEM remains underused in cognitive neuroscience, which is why we welcome Charlton et al.'s article. However, using SEM is not without pitfalls, and diagrams of SEM models tend to convey a misleading impression of definiteness. In this commentary, we aim to raise the awareness of researchers in the field of neurobiology of ageing to some caveats that need to be kept in mind when applying or reading about SEM analyses, using Charlton et al.'s study as an example. For a more general treatment of the topic, we refer the reader to textbooks like Kline's (2005), Loehlin's (2004), or Bollen's (1989).

2. Guidelines for structural equation modelling

2.1. Prepare your model

In a nutshell, a SEM analysis tests how well a pre-specified set of assumptions about the linear associations

* Corresponding author. Tel.: +44 131 6508482; fax: +44 131 6511771.
E-mail addresses: lars.penke@ed.ac.uk (L. Penke), ian.deary@ed.ac.uk (I.J. Deary).
[1] Tel.: +44 131 6503452; fax: +44 131 6511771.

0197-4580/$ – see front matter
doi:10.1016/j.neurobiolaging.2009.10.019

of the studied variables (the 'model') is able to reproduce the variance-covariance matrix of the observed variables. Thus, SEM is not very suitable for merely describing or exploring data: it is a strictly model-testing approach. What is tested is whether the model is plausible given the data. Most importantly, SEM can never prove or confirm that a model is correct, it can only falsify implausible models (Freedman, 1987, 2005; Tukey, 1954). Arguably, the most crucial part of a SEM analysis is specifying the model that is tested, which should happen *a priori* based on existing theory and evidence (Freedman, 1987).

2.2. Consider latent traits

One of the biggest advantages of SEM is that, whenever several variables are assessed that are meant to measure the same construct (e.g., mental speed or working memory), these so-called manifest variables can be used to model the construct directly as a 'latent variable'. Latent variables contain only the variance that is shared by all measured variables. Therefore, they are free of unsystematic measurement errors and other sources of variance that are specific to any one manifest measure. As a consequence, estimates of relationships involving latent variables will be more reliable (Loehlin, 2004). While Charlton et al. had assessed each of the four cognitive constructs they hypothesised (speed, flexibility, fluid intelligence, working memory) with two to three indicators, they did not take advantage of latent variables in their SEM analysis. They used averages of indicator scores instead (i.e., only manifest variables).

2.3. Consider competing alternative models

It is always possible to specify a SEM model that trivially 'fits' the data perfectly. Such a model simply needs to contain all possible relationships (called 'paths') between all variables, thus inevitably accounting for all the observed covariance. However, this model will be saturated, prohibiting a statistical test against the data. Any less complex model with fewer assumed relationships can be tested, but the number of possible models is usually large, and there can be different models that fit the data equally well. Thus, it is necessary to have theory-guided hypotheses to begin with. When different theories suggest different models, like the common cause and the specific gain/loss theories considered by Charlton et al., SEM allows for testing these models separately and comparing their fit. Charlton et al., however, did not test the alternative models they had derived from the literature against each other, but instead tried to include almost all paths that have been suggested by any theory at the same time (and actually regretted that they were unable to include all of them, p. 1548).

2.4. Apply Occam's razor as appropriate

More complex models (like Charlton et al.'s all-inclusive initial one) might fit the data better, but they entail a great risk of capitalizing on spurious relationships that are specific to the current dataset. Of course, complex models might accurately reflect the complexity of reality and independent replication can then help to make complex models with good theoretical rationales more plausible, but in most cases researchers are better advised to specify parsimonious models, containing only the most relevant relationships, which are more likely to be generalisable. This is especially true in smaller datasets, since SEM results become unstable when the ratio of free parameters to cases is large (Kline, 2005 suggests at least 10 cases per parameter). Since Charlton et al. started by mixing assumptions from various theoretical perspectives, their initial model was highly complex and had 19 free parameters, leaving them with less than 6 cases per parameter in their sample.

2.5. Assess the strength of prior theory in hypothesising causal associations

For statistical reasons, most paths in SEM models have to be unidirectional, implying causality. However, SEM analyses are based on correlational data, which are inherently devoid of causal information (especially when they stem from cross-sectional studies, as most in cognitive neuroscience do) (Freedman, 2005; Tukey, 1954). Indeed, SEM is unable to distinguish a model that, for example, assumes a causal effect of cognitive speed on fluid intelligence (like Charlton et al.'s does) from another that assumes a causal effect of fluid intelligence on cognitive speed. Both models will yield identical statistical results. Thus, causal assumptions in SEM models also need to be based on theories. Charlton et al. rest the many causal assumptions they make in their model on a long list of reference that provide loose suggestions (see p. 1549 of their article), but none of these assumptions can be considered well-established. Indeed, causal links between individual differences in cognitive abilities, including speed and fluid intelligence, have been seriously questioned (Kovas and Plomin, 2006; Luciano et al., 2005; Plomin and Spinath, 2002).

2.6. If possible, use theory to guide changes to the model

Initial SEM Models can be modified in an iterative process towards a better fit to the data by adding or dropping paths from the model. Such modifications can be useful if they are justified by a theoretical rationale. In these cases, they can be understood as a form of comparing alternative models. Charlton et al., however, reached their final model (that started from a mixture of different theories) in a sequential reduction process that relied solely on what was empirically suggested (as automatically done by many SEM programs). Such an exploratory procedure it is not advisable, as it is prone to capitalizing on non-generalisable chance relationships in the data.

3. An alternative model

Taken together, Charlton et al. presented only one of the many possible SEM models that would be supported by their data, and theirs can be questioned for its assumptions and for being overly fitted to their particular dataset. The critical point here is that the conclusion they draw from their model, in particular the age-related effect of white matter integrity on working memory, might fail to receive support from other models that are also supported by their data. To illustrate this, we tested an alternative model against their data. Because Charlton et al.'s original data were not made available to us, we used the correlation matrix in Table 3 of their article provisionally for this purpose. Testing alternative SEMs based on published correlation matrices is a common and accepted procedure that, if interpreted carefully, yields very similar results to models derived from original data (Cudeck, 1989).

The alternative model we propose is depicted in Fig. 1. It is important to note that it was formulated using the established guidelines for SEM that we described above. With only 13 free parameters it is very parsimonious. The model is based on the well-established finding that different tests of cognitive ability tend to share approximately 40–60% of their variance, giving rise to a common or general factor of cognitive ability (usually abbreviated as "*g*"). Since this *g* factor is based on more than a century of cumulative research (Carroll, 1993; Jensen, 1998), has reached consensus among intelligence researchers (Neisser et al., 1996), and has been called the best-replicated finding in all psychology (Deary, 2000), it provides a very solid base for an alternative model. Crucially for the current context, it has been repeatedly found that cognitive ageing effects are principally on *g* (e.g. Salthouse, 1994, 1996; Salthouse and Czaja, 2000). It is also the backbone of the 'common cause' hypothesis of cognitive aging (Christensen et al., 2001; Lindenberger and Baltes, 1994; Mackinnon et al., 2006; Salthouse et al., 1998), on which Charlton et al. partly base their analysis. In our alternative model (Fig. 1), we used all four (aggregates of) cognitive measures reported by Charlton et al. to define a latent variable that represents *g*, which circumvents questionable assumptions of causal relationships between these variables.

We first test the hypothesis that DTI-measured mean diffusivity (MD, an index of white matter integrity) might mediate age effects on this common cognitive factor (*g*). This model fits the data well ($\chi^2(8) = 11.15$, $p = .19$, CFI = .990, RMSEA = .060, SRMR = .037). It shows that one can model a strong common factor in Charlton et al.'s data (explaining 55.7% of the variance of the cognitive tests, and with all tests loading highly on it), which is substantially predicted by age (standardized path coefficient $\beta = -.65$). It also shows that MD is highly correlated with age ($\beta = .77$). So far, so good. However, crucially, MD does not predict *g* over and above what is already predicted by age ($\beta = .01$) in this model. We consequently tested the hypothesis that white matter integrity has no effect on *g* that is independent of age. To do so, we modified our alternative model (depicted in Fig. 1) by fixing the path from MD to *g* to zero, which yielded a slightly improved model fit ($\chi^2(9) = 11.16$, $p = .26$, CFI = .993, RMSEA = .047, SRMR = .037). We accept this even more simple and economical model. It simply states that age is associated with the general cognitive factor and with MD, but that beyond their association with age, MD and the general cognitive factor are not associated.

The models in the present report and that of Charlton et al. are not nested (i.e., subsets of each other). Therefore, formally comparing them is only possible based on comparative fit indices, which do not allow for a significance test (Kline, 2005). Based on one of them, Akaike's Information Criterion (AIC; low values indicating better fit), our alternative (−4.85) and modified alternative model (−6.84) fitted the data better than the full model Charlton et al. proposed based on the theoretical assumptions they made (−3.11), though they did not fit better than their final model (−8.38), which was improved by the questionable method of empirical modification. (Note that the AICs we report here for Charlton et al.'s models differ

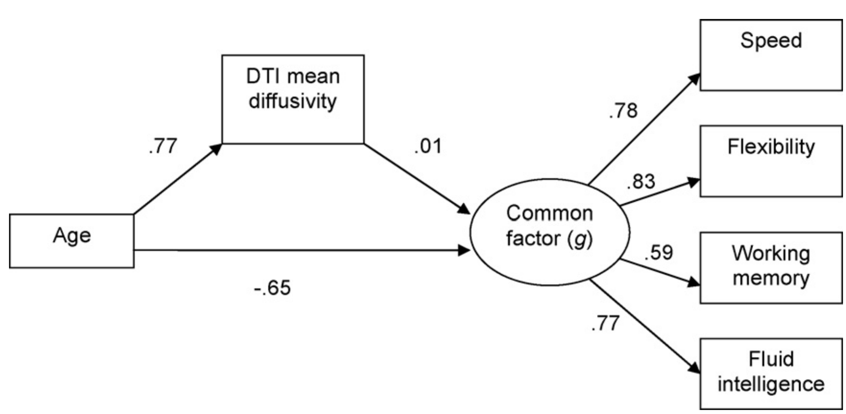

Fig. 1. Alternative common factor model. All path coefficients have been fully standardized and can thus be interpreted as regression coefficients. The four cognitive ability composites were formed by Charlton et al. by averaging *z*-scores of the following tests (for details, see their Table 1): 'Speed': AMIPB information processing speed, WAIS-R digit symbol, and grooved pegboard: 'flexibility': D-KEFS trails, Wisconsin Card Sorting Test, and Stroop; 'working memory': WMS-III digit span backwards and letter-number sequencing; 'fluid intelligence': WASI matrix reasoning and block design.

from those they reported due to the use of different formulas, which do not affect the conclusions.) Also, our models are more parsimonious, which is not rewarded very much in the AIC. Alternatively the Bayes Information Criterion (BIC; again low values indicating better fit) puts a greater penalty on overly complex models and might thus be more appropriate. When this criterion is used, our model (103.87) and the modified version of it (96.15) fit better than either Charlton et al.'s full (139.97) or final (117.52) model. According to Raftery (1995), a BIC difference of 10 or more provides strong evidence that a model fits data better than another. This is given for both of our models compared to both of Charlton et al.'s, indicating that at least judged by the BIC our alternative models fit Charlton et al.'s data better than the models they presented.

While white matter integrity (MD) clearly had no age-independent effect on general cognitive ability (g) in our alternative model, it could be argued that there are such effects on specific cognitive abilities, independent of g. To rule out this alternative hypothesis, we also tested for MD effects on the specific variance of each cognitive ability that is independent of the common factor. To do so, we fixed all parameters to the values depicted in Fig. 1 and introduced paths from MD to each of the four cognitive abilities to the model. All of these path coefficients were smaller than |.09| and within a range of two standard errors, meaning that they failed to reach conventional standards for a significant contribution to the model. Thus, our alternative models indicate that white matter integrity had no effect on cognitive ability independent of age in Charlton et al.'s data.

4. Conclusion

SEM is a powerful statistical tool to analyse complex relationships in multivariate datasets, as are common in the study of neurobiology and aging. However, it can only falsify, but never prove a model. Therefore, it is most powerful when different theoretical models can be clearly specified are compared, so that the one that is most plausible theoretically *and* empirically can be inferred. Whereas Charlton et al. started with two theoretical perspectives, they failed to specify competing models based on differentiating predictions and instead proposed a single complex model that mixed selective assumptions from both perspectives. From this model they concluded that white matter integrity had an age-related effect on a specific cognitive ability. To exemplify how problematic such an approach to SEM is, we presented one alternative model based on a single theoretical perspective which is actually the dominant one in psychometric intelligence and cognitive aging research. Even though it is theoretically clearer, not based on atheoretical modification to increase model fit, and more parsimonious, it fitted their data just as well. However, our alternative model reached this fit even when no relationships between white matter integrity and the common factor or on any specific domain of cognitive ability were assumed. Thus, a model without the specific white matter integrity effect on cognitive ageing that Charlton et al. emphasise cannot be rejected by their data and therefore the effect has not been demonstrated conclusively by Charlton et al.—it depends strongly on how their data is analysed.

Our alternative model does not necessarily imply that there is no effect of white matter integrity on cognitive aging. However, it suggests that it might be difficult to demonstrate such an effect in an age-heterogeneous sample, especially when age and white matter integrity are as highly correlated as in this sample ($r = .77$), making them hard to disentangle. Where establishing causality through experiments, ultimately the only way, is not possible for ethical or practical reasons (as is often the case in human ageing research), it might be necessary to follow Hofer and Sliwinski's (2001) suggestion and study narrow age cohorts (e.g. Deary et al., 2006), preferably longitudinally.

Disclosure statement

We have no actual or potential conflicts of interest to report.

Acknowledgments

LP is funded by the UK Medical Research Council (Grant No. 82800), which is part of the Help The Aged-funded Disconnected Mind research programme. The work was undertaken by The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross council Lifelong Health and Wellbeing Initiative. Funding from the BBSRC, EPSRC, ESRC and MRC is gratefully acknowledged.

References

Bollen, K.A., 1989. Structural Equations with Latent Variables. Wiley, New York.

Carroll, J.B., 1993. Human Cognitive Abilities. Cambridge University Press, Cambridge.

Charlton, R.A., Landau, S., Schiavone, F., Barrick, T.R., Clark, C.A., Markus, H.S., Morris, R.G.A., 2008. Structural equation modelling investigation of age-related variance in executive function and DTI measured white matter change. Neurobiol. Aging 29, 1547–1555.

Christensen, H., Mackinnon, A.J., Korten, A., Jorm, A.F., 2001. The common cause hypothesis of cognitive aging. Psychol. Aging 16, 588–599.

Cudeck, R., 1989. Analysis of correlation matrices using covariance structure models. Psychol. Bull. 105, 317–327.

Deary, I.J., 2000. Looking Down on Human Intelligence. Cambridge University Press, Cambridge.

Deary, I.J., Bastin, M.E., Pattie, A., Clayden, J.D., Whalley, L.J., Starr, J.M., Wardlaw, J.M., 2006. White matter integrity and cognition in childhood and old age. Neurology 66, 505–512.

Freedman, D.A., 1987. As others see us: a case study in path analysis. J. Educ. Stat. 12, 101–128.

Freedman, D.A., 2005. Statistical Models: Theory and Practice. University Press, Cambridge.

Hofer, S.M., Sliwinski, M.J., 2001. Understanding aging: an evaluation of research designs for assessing the inter-dependence of aging-related changes. Gerontology 47, 341–352.

Jensen, A.R., 1998. The g Factor: The Science of Mental Ability. Praeger, Westport.

Kline, R., 2005. Principles and Practice of Structural Equation Modelling, 2nd ed. Guilford Press, New York.

Kovas, Y., Plomin, R., 2006. Generalist genes: implications for the cognitive sciences. Trends Cogn. Sci. 10, 198–203.

Lindenberger, U., Baltes, P.B., 1994. Sensory functioning and intelligence in old age: a strong connection. Psychol. Aging 9, 339–355.

Loehlin, J.C., 2004. Latent Variable Models: An Introduction to Factor, Path, and Structural Equation Analysis, 4th ed. Lawrence Erlbaum, Mahwah.

Luciano, M., Posthuma, D., Wright, M.J., de Geus, E.J., Smith, G.A., Geffen, G.M., Boomsma, D.I., Martin, N.G., 2005. Perceptual speed does not cause intelligence, and intelligence does not cause perceptual speed. Biol. Psychol. 70, 1–8.

Mackinnon, A., Christensen, H., Jorm, A.F., 2006. Search for a common cause factor amongst cognitive, speed and biological variables using narrow age cohorts. Gerontology 52, 243–257.

Neisser, U., Boodoo, G., Bouchard Jr., T.J., Boykin, A.W., Brody, N., Ceci, S.J., Halpern, D.F., Loehlin, J.C., Perloff, R., Sternberg, R.J., Urbina, S., 1996. Intelligence: knowns and unknowns. Am. Psychol. 51, 77–101.

Plomin, R., Spinath, F.M., 2002. Genetics and general cognitive ability (g). Trends Cogn. Sci. 6, 169–176.

Raftery, A.E., 1995. Bayesian model selection in social research. Social. Methodol. 25, 111–163.

Salthouse, T.A., 1994. How many causes are there of aging-related decrements in cognitive functioning? Dev. Rev. 14, 413–437.

Salthouse, T.A., 1996. Constraints on theories of cognitive aging. Psychol. Bull. Rev. 3, 287–299.

Salthouse, T.A., Czaja, S.J., 2000. Structural constraints on process explanations in cognitive aging. Psychol. Aging 15, 44–55.

Salthouse, T.A., Hambrick, D.Z., McGuthry, K.E., 1998. Shared age-related influences on cognitive and non-cognitive variables. Psychol. Aging 13, 445–461.

Span, M.M., Ridderinkhof, K.R., van der Molen, M.W., 2004. Age-related changes in the efficiency of cognitive processing across the life span. Acta Psychol. 117, 155–183.

Tukey, J.W., 1954. Causation, regression, and path analysis. In: Gowen, J.W., Kempthorne, O., Lush, J. (Eds.), Statistics and Mathematics in Biology. Iowa State Press, Ames, pp. 35–66.