**Supplement – Statistical causal discovery**

The main article evaluates the factor structure of the WAIS–IV using a causal discovery algorithm to uncover (parts of) the underlying generating system behind the observed test results. This supplement aims to provide an overview of method and assumptions behind statistical causal discovery, as well as a more detailed description of the steps involved in the Bayesian Constraint-based Causal Discovery (BCCD) implementation used to find the models in Figure 2-4. Note that our results could equally be obtained by other principled causal discovery implementations that handle unobserved confounders.

**Introduction to graphical causal models**

Despite the well-known adage ‘correlation does not imply causation’, in the early ‘90s it was shown that, under reasonable assumptions, it *is* possible to learn causal relations from purely observational data.

It starts from the assumption that the underlying system responsible for generating the observed probability distribution corresponds to a *structural causal model* (Pearl, 2009). The interaction structure of this causal model can be represented in the form of a graph with variables depicted as nodes and the causal interactions between them depicted by connecting links, such that e.g. an arc *X* → *Y* represents a causal mechanism in which *X* influences *Y* but not the other way around. In the graph, *X* is then a *parent* of *Y*, and *Y* is a *child* of *X*. Note that not all variables in the causal system need to be unobserved.

Two key additional assumptions are needed in order to validly infer aspects of the causal system from the observed probability distribution:

 the *Causal Markov condition* states that each variable in the system is probabilistically independent of its non-descendants in the model given its parents in the model,

 the *Causal Faithfulness condition* states that all observable (conditional) independencies are implied by the Causal Markov condition.

For example, for the causal system in Figure 5 (left) the Causal Markov condition states that variable *F*, given its parent *B*, is independent of its non-descendants {*A*,*C*,*D*,*E*}, but not of its child *G*.

The Causal Markov assumption makes it possible to reason from model structure to observed conditional independencies between variables, and the combination with the Causal Faithfulness assumption makes it possible to reason backwards form probabilistic independencies to model structure.

The goal of a causal discovery method is now to extract as much *valid* information as possible on the causal interactions between variables from the observed probability distribution.

**Figure 5.**From generating causal system (left), via interaction structure (mid), to output causal model (right).

Figure 5 (right) shows the possible result of such an endeavor: naturally the output causal model cannot fully reconstruct the underlying model including all unobserved variables (left), but it *can* correctly identify the structure of interactions between most observed variables. In particular, it shows that *B* is a direct cause of *E* and *F*, and also an indirect cause of *G* (mediated by *F*). It even identifies that neither *B* nor *D* is a cause of each other, indicated by the bi-directed arc *BD*, meaning that there must be an unobserved confounder that is responsible for the correlation between the two. Also note that it shows when it *cannot* decide on whether a relation is causal or not, indicated by the small circle marks at the links from *A* to *B*, and from *C* to *D*.

In practice, often one or more additional simplifying assumptions are introduced, such as ‘no unobserved confounders’, ‘no feedback mechanisms’, ‘multivariate Gaussian probability distributions’, ‘no selection bias’, ‘missing data MCAR’, etc. These can help to significantly reduce the complexity of the process, and may greatly speed up the performance, but also make the outcomes less reliable. Therefore it is important to stress that the causal model output is only valid with respect to the given assumptions: if these are violated, then the conclusions may not hold.

This holds in particular for the fact that we only have finite data, which means that in practice we do not have access to the true probability distribution. This may obscure weak interactions, resulting in erroneous decisions that in turn affect the final output model. The BCCD algorithm (see below) used in this article handles this by assigning Bayesian reliability estimates to each element in the causal model, making it possible to distinguish between solid and borderline conclusions, which should help tremendously in interpreting the model output.

**Constraint-based causal discovery**

Causal discovery algorithms typically follow one of two main paradigms: the so-called *constraint-based methods* (Spirtes et al, 2000; Pearl, 2009) search for probabilistic (conditional) independences between variables in the data to reconstruct the underlying generating model (bottom-up), whereas the alternative *score-based methods* (Chickering, 2002) incorporate an efficient search strategy over all possible model explanations in order to maximize a score, e.g., AIC or BIC, that weighs model fit versus model complexity (top-down).

In this article we opted for a constraint-based approach, as these are typically better equipped to handle unobserved confounders: essential as we are looking for signs of latent factor structures. We chose for the BCCD implementation, because this method has been successfully applied to similar sized psychometric data sets before (Sokolova et al., 2015), and, as stated above, also quantifies the reliability of the output.

For constraint-based causal discovery algorithms, the first crucial theoretical result is that there should be an edge between variables *X* and *Y* in the output causal model, if and only if there is no subset of observed variables **Z** that can make *X* and *Y* conditionally independent in the probability distribution (Spirtes, 2000). Exhaustive application results in a so-called *skeleton graph*, see Figure 5 (mid), with undirected edges that represent the underlying causal interaction structure. The second crucial result is that based on this skeleton graph and the (conditional) independencies established in constructing it (one per edge removed suffices) all inferable directional (causal) information can be reconstructed using a set of *graphical orientation rules* (Meek,1995; Zhang, 2008).

As a result, a constraint-based algorithm takes the following generic form:

1. Obtain data set D and suitable independence test.

2. Start from fully connected, undirected graph G over (observed) variables in D.

3. For each edge X – Y in G, search for set of variables **Z** that make *X* and *Y* (conditionally) independent in data set D; if found, then:

 remove edge X – Y from G,

 register **Z** as separating set for (X,Y), and

 continue with next edge,

4. Until no more edges can be removed.
*skeleton graph* (Figure 5, mid)

5. Next, exhaustively apply causal orientation rules on graph G, given all the separating sets found in step 3.
*output causal model* (Figure 5, right)

Different constraint-based algorithms typically implement different search strategies and consistency checks that work better depending on the type of problem and/or assumptions employed.

**The BCCD algorithm on the WAIS–IV**

To analyze the WAIS–IV–NL data, we opted for the Bayesian Constraint-based Causal Discovery (BCCD) algorithm (Claassen&Heskes, 2012), as it is robust and powerful enough to handle latent confounders and selection bias, and able to provide meaningful reliability estimates. A suitable, publicly available alternative that also handles unobserved confounders would be the (*Conservative*) *FCI*algorithm (Kalisch et al., 2012; Scheines et al., 1998).

**Figure 6.**Overview statistical causal discovery with the BCCD algorithm.

The BCCD algorithm consists of five distinct phases to construct the final causal model output, as depicted in Figure 6:

(1) The first step involves merging a mixture of discrete and continuous variables into a single uniform representation. For this a Gaussian copula technique is used that rescales each variable by its rank in the corresponding empirical distribution, and then maps this data into Gaussian distributed scores through the inverse Normal cumulative probability density function (Sokolova, 2016b).

(2) From this we obtain a multivariate Gaussian covariance matrix. In case there are missing values, an Expectation Conditional Maximization (ECM) algorithm (Meng&Rubin,1993; Little & Rubin, 2002) is used to estimate an unbiased approximation of this covariance matrix. The covariance matrix is subsequently normalized to unit variance as a standardized Gaussian correlation matrix used in the next steps.

(3) The search stage of the BCCD algorithm employs a PC/FCI based search strategy over increasing sets of variables (Spirtes et al., 2000). For each subset of variables it uses the Bayesian Gaussian equivalence (BGe) score (Heckerman, 1997) to compute the posterior distribution over all possible marginal structures, and turns this into probabilistic estimates on the reliability of structural information (edges in the skeleton), and logical causal statements on the directionality of interaction between variables (Claassen&Heskes, 2012).

(4) The collective information from the BCCD search stage is aggregated into a global list of structural and causal statements, and subsequently processed in decreasing order of reliability, making sure that newly processed information does not contradict already inferred causal conclusions from statements with higher reliability estimates.(Again we refer the reader to Claassen&Heskes, (2012) for details).

(5) The output from step (4) is optionally evaluated against the original statements from the search stage (3), and possibly used to obtain improved estimates to increase consistency. Finally, the resulting output causal model is presented as a combination of two matrices: one with reliability estimates on each edge in the interaction structure (skeleton graph, Figure 5, mid), and one with reliability estimates on each causal relation. These two are then combined in a single graphical model of the form in Figure 5 (right), with line thickness representing edge reliability for easy visualization.

**Note:** in our analysis of the WAIS–IV–NL we know that none of the subtests is a likely driving factor or cause for any of the others, as they are all confounded by the hypothesized latent factors. Therefore all edges between subtests will ultimately appear as bi-directed arrows in the causal model, in combination with directed arcs from FSIQ to all subtests. As we are primarily interested in the residual clustering pattern we decided to unclutter the causal model output by not fully drawing the links originating from FSIQ, and exchange the remaining bi-directed links for straightforward undirected edges.

**Detailed assumptions and parameter configurations**

In running the BCCD algorithm on the WAIS–IV–NL we use the following assumptions and/or parameter settings:

 we assume the standard causal Markov and causal faithfulness conditions apply, in combination with the causal DAG assumption (no feedback cycles),

 we do *not* assume there are no unobserved confounders, and neither that there is no selection bias,

 we use as background information that no other variable in the model can cause either Age or Gender,

 for the missing data step we eliminate rows with >30% missing values,

 we assume a flat prior on all possible marginal model structures in computing the Bayesian posterior estimates in step (3),

 in computing the BGe score we use a maximally uninformative prior, i.e. for the mean we use a precision matrix with unit scale parameter, unit identity matrix for the scale matrix in the normal-Wishart, and degrees of freedom corresponding to the number of variables D + 1,

 in the final output we only use information with reliability >50%, where we ensure all structural/causal statements are evaluated using marginal substructures over at least 5 nodes.