Testing Effective Connectivity Changes with Structural Equation Modeling: What Does a Bad Model Tell Us?

Andrea B. Protzner and Anthony R. McIntosh*

Rotman Research Institute of Baycrest Centre, University of Toronto, Toronto, Ontario, Canada

Abstract: Structural equation modeling (SEM) is a statistical method that can assess changes in effective connectivity across tasks or between groups. In its initial application to neuroimaging data, anatomical connectivity provided the constraints to decompose interregional covariances to estimate effective connections. There have been concerns expressed, however, with the validity of interpreting effective connections for a model that does not adequately fit the data. We sought to address this concern by creating two population networks with different patterns of effective connectivity, extracting three samples sizes ($N = 100, 60, 20$), and then assessing whether the ability to detect effective connectivity differences depended on absolute model fit. Four scenarios were assessed: (1) elimination of a region showing no task differences; (2) elimination of connections with no task differences; (3) elimination of connections that carried task differences, but could be expressed through alternative indirect routes; (4) elimination of connections that carried task differences, and could not be expressed through indirect routes. We were able to detect task differences in all four cases, despite poor absolute model fit. In scenario 3, total effects captured the overall task differences even though the direct effect was no longer present. In scenario 4, task differences that were included in the model remained, but the missing effect was not expressed. In conclusion, it seems that when independent information (e.g., anatomical connectivity) is used to define the causal structure in SEM, inferences about task- or group-dependent changes are valid regardless of absolute model fit. *Hum Brain Mapp* 27:935–947, 2006. © 2006 Wiley-Liss, Inc.

Key words: statistics; functional connectivity; simulation; neuroimaging

INTRODUCTION

As more researchers are adding functional neuroimaging to their research protocol, the interest in moving beyond activation analysis has grown. While activity changes provide clues as to a region's contribution to a given behavior,

many researchers are trying to frame these changes in relation to interactions within large-scale neural systems. One method that has proven useful in this regard is structural equation modeling (SEM). It is similar to a multivariate multiple regression analysis and typically has been used to determine whether a hypothesized causal structure is consistent with observed data [Bollen, 1989; Joreskog and Sorbom, 1979; Loehlin, 1987]. The initial applications of SEM to neuroimaging data (e.g., 2DG, PET, fMRI) made use of the fact that the causal relation between brain regions is essentially a known entity—that being the neuroanatomy. Thus, for neuroimaging data, the focus was on whether there were task-dependent changes in the influences regions have on one another—or the effective connections—within the constraints of the same anatomical network. SEM has been used to explore effective connectivity patterns related to dorsal and ventral stream visual operations [Buchel and Friston, 1997; Kohler et al., 1998; McIntosh et al., 1994], learning-

Contract grant sponsor: Canadian Institutes for Health Research and the Natural Sciences and Engineering Research Council of Canada.

*Correspondence to: A.R. McIntosh, Rotman Research Institute of Baycrest Centre, 3560 Bathurst St., Toronto, Ontario M6A 2E1, Canada. E-mail: mcintosh@psych.utoronto.ca

Received for publication 31 July 2005; Accepted 1 December 2005

DOI: 10.1002/hbm.20233

Published online 23 August 2006 in Wiley InterScience (www.interscience.wiley.com).

TABLE I. Sample correlation matrices

	V1	V2	V3	V4	V5	V6
N = 100						
Network 1						
V1	1					
V2	-0.044	1				
V3	0.093	-0.815	1			
V4	0.966	-0.028	0.064	1		
V5	0.704	0.653	-0.509	0.714	1	
V6	-0.007	0.962	-0.751	0.007	0.635	1
Network 2						
V1	1					
V2	0.039	1				
V3	-0.02	0.759	1			
V4	0.967	0.020	-0.032	1		
V5	0.508	-0.830	-0.644	0.509	1	
V6	-0.048	-0.957	-0.691	-0.048	0.802	1
N = 60						
Network 1						
V1	1					
V2	-0.001	1				
V3	-0.161	-0.685	1			
V4	0.980	0.026	-0.176	1		
V5	0.763	0.622	-0.542	0.777	1	
V6	-0.017	0.947	-0.558	0.003	0.568	1
Network 2						
V1	1					
V2	0.009	1				
V3	-0.124	0.786	1			
V4	0.977	-0.008	-0.13	1		
V5	0.504	-0.849	-0.710	0.521	1	
V6	-0.036	-0.964	-0.695	-0.015	0.817	1
N = 20						
Network 1						
V1	1					
V2	-0.082	1				
V3	-0.023	-0.810	1			
V4	0.976	-0.107	-0.043	1		
V5	0.693	0.646	-0.61	0.677	1	
V6	0.021	0.965	-0.755	-0.015	0.686	1
Network 2						
V1	1					
V2	-0.168	1				
V3	-0.286	0.735	1			
V4	0.987	-0.192	-0.328	1		
V5	0.639	-0.859	-0.721	0.656	1	
V6	0.212	-0.982	-0.687	0.227	0.881	1

related changes in feedback effects on visual cortices [McIntosh et al., 1998], motor behavior [Solodkin et al., 2004], and variations in effective connections related to aging and clinical conditions [Grady et al., 2003; Horwitz et al., 1995; Jennings et al., 1998; Maguire et al., 2000; Northoff et al., 2004; Schlosser et al., 2003]. In all cases, the target of statistical inference was whether the effective connections differed between tasks or groups given the same anatomical network.

Bullmore et al. [2000] explored another application of SEM to neuroimaging data that is more commensurate with its use in other scientific forums: to evaluate competing causal structures. Their application makes use of the long history of neurological models of various behaviors that are testable with SEM and neuroimaging. A model is considered more

plausible if it provides a better fit to empirical data than a competing model having a different causal structure. The inferential focus is primarily on the overall model fit rather than the specific effects within [see also Zhuang et al., 2005]. The approaches of McIntosh and Bullmore are equally valid, but have different goals in mind.

Unfortunately, there seems to be a belief that the ability to make inferences about changes in effective connectivity is compromised if the overall model does not fit the data adequately [e.g., Gonçalves et al., 2001; Gonçalves and Hall, 2003]. The overall absolute model fit may be poor for neuroimaging data, given that most models will only include a few of the total influences on a region [McIntosh and Gonzalez-Lima, 1994]. On the other hand, the inferences about a given effective connection (path coefficient) may not be

meaningful in a poorly fitting model, when the goal is to identify the most plausible causal structure [MacCallum, 1995]. Neither of these points speaks to the question of whether one can detect a change in effective connections despite having an overall model that fits poorly. The present study used simulated data to determine whether such inferences about path coefficients are possible in models that do not fit the empirical data.

METHODS AND RESULTS

Definitions

As there are occasional ambiguities in the terminology for studies of functional networks, we will begin by defining their use in this paper. SEM analysis uses covariances of activity, computed across subjects or across trials for a given subject, among regions of interest to determine path coefficients. These path coefficients indicate the strength of directional influences that would be carried along putative anatomical connections to reproduce the original covariances as closely as possible. The covariances can be considered estimates of *functional connectivity*, which are simply statements that the two regions show some dependency between their activity. The path coefficients derived from SEM are estimates of *effective connectivity*, which indicate the direct influence of one area on another. Unlike functional connectivity, the ability to make inferences about effective connections requires an explicit model within which the effects are estimated. In SEM, the anatomy defines this model. Discussions of the use of these terms in neuroimaging can be found in Friston [1994], Horwitz [2003], and Lee et al. [2003].

The logic behind the simulations was to construct two population models using the same anatomical network but different path coefficients (i.e., different effective connections). The simulations based on samples from the population would then assess whether differences in path coefficients could be detected in instances where anatomical connections were excluded (i.e., the sample model did not fit the data). This simulation approach is identical to that used by McIntosh and Gonzalez-Lima [1994] to determine whether modification indices (Lagrangian multipliers) could be used to identify model exclusions.

We used EQS v. 4.02 [Bentler and Weeks, 1980; Bentler, 1985] to construct the models. Six regions (V1 to V6) with prespecified path coefficients were simulated from a unit normal distribution with three sample sizes: 100, 60, and 20. The three sample sizes allowed us to assess our hypothesis while minimizing concern about sample size issues. The smallest sample size ensured that our results would hold with a more realistic sample size for neuroimaging studies. The correlation matrices used for these analyses are presented in Table I. A graphic depiction of the base network used for the simulations appears in Figure 1 with the population parameters for path coefficients (effective connections) and regionally specific residual influences (PSI). The estimated values for the base networks across the three sample sizes are given in Table II. In the models for each

Population Network 1

$$V1 = 1.0 \cdot \Psi_1$$

$$V2 = 1.0 \cdot \Psi_2$$

$$V3 = -0.3 \cdot V6 + 0.25 \cdot \Psi_3$$

$$V4 = 0.5 \cdot V1 + 0.25 \cdot \Psi_4$$

$$V5 = 0.35 \cdot V1 + 0.7 \cdot V2 + 0.5 \cdot V4 + -0.5 \cdot V6 + 0.25 \cdot \Psi_5$$

$$V6 = 0.5 \cdot V2 + 0.7 \cdot V3 + 0.25 \cdot \Psi_6$$

Population Network 2

$$V1 = 1.0 \cdot \Psi_1$$

$$V2 = 1.0 \cdot \Psi_2$$

$$V3 = -0.3 \cdot V6 + 0.25 \cdot \Psi_3$$

$$V4 = 0.5 \cdot V1 + 0.25 \cdot \Psi_4$$

$$V5 = 0.35 \cdot V1 + -0.7 \cdot V2 + 0.5 \cdot V4 + 0.5 \cdot V6 + 0.25 \cdot \Psi_5$$

$$V6 = -0.5 \cdot V2 + 0.7 \cdot V3 + 0.25 \cdot \Psi_6$$

Figure 1.

Top: "Anatomical" connectivity for base model. Bottom: Population models. Italicized parameters indicate effective connections that differed between the two networks. Ψ indicates residual influence unique to each node.

sample size, PSI effects for those regions that received influences from within the model were fixed between 0.55 and 0.002 based on the estimates from the alternative models, and for those regions that did not receive input from the model, PSIs were set to 1.0. Differences in the populations between the two networks existed in the connections from V2 to V5 (0.7 in Network 1, and -0.7 in Network 2), from V6 to V5 (-0.5 in Network 1 and 0.5 in Network 2), and from V2 to V6 (0.5 in Network 1 and -0.5 in Network 2). The size of these differences varied across the three sample groups, with the difference in the path from V6 to V5 decreasing from that in the population. Nevertheless, in all three samples the differences in effective connections between the two networks were significant. Total effects for the base models are given in Table III. Total effects are the sum of direct effects and the product of indirect effects [Loehlin, 1987], and are important indices of the overall impact that one region has on another [McIntosh and Gonzalez-Lima, 1993].

Our simulations cover most scenarios that one would likely encounter in the application of SEM to imaging data:

TABLE II. Direct effects (path coefficients) for base models

	V1	V2	V3	V4	V5	V6
LISREL estimate of base model for Network 1						
V1						
V2						
V3						-0.86
						-0.68
						-0.85
V4	0.97					
	0.98					
	0.98					
V5	0.39	0.92		0.37		-0.25
	0.47	0.68		0.30		-0.07
	0.39	0.88		0.39		-0.17
V6		1.19	0.28			
		1.13	0.27			
		1.14	0.22			
PSI	1.00	1.00	0.49	0.07	0.02	0.09
	1.00	1.00	0.55	0.05	0.02	0.09
	1.00	1.00	0.49	0.05	0.01	0.07
LISREL estimate of base model for Network 2						
V1						
V2						
V3						-0.78
						-0.83
						-0.76
V4	0.97					
	0.98					
	0.99					
V5	0.45	-0.68		0.10		0.18
	0.23	-0.70		0.30		0.15
	0.15	-0.31		0.34		0.47
V6		-1.13	0.23			
		-1.22	0.32			
		-1.11	0.18			
PSI	1.00	1.00	0.49	0.07	0.02	0.09
	1.00	1.00	0.55	0.05	0.02	0.09
	1.00	1.00	0.49	0.05	0.002	0.07

Each row within a cell corresponds to the estimate from each of the three sample sizes ($N = 100, 60, 20$ from top to bottom in each cell).

elimination of a region that participates to the same degree in the network across groups, and elimination of connections that either do or do not differ between groups. There is another scenario where a region that shows a differential contribution between groups is excluded. Although possible to simulate, we felt this would be an unlikely scenario in practice because the regions one typically selects are those that do show some task-dependent changes in either activity or functional connectivity. It is possible that regions may be excluded because they are not spatially resolvable by the imaging protocol (e.g., thalamic or brainstem nuclei). However, we have shown previously [McIntosh and Gonzalez-Lima, 1994] that the effects of excluding a region that shows strong contributions are usually expressed in high residual influences, either estimated or indicated by modification indices (see below for a discussion of modification indices).

The χ^2 goodness-of-fit statistic was used to assess each model's ability to reproduce the original correlation ma-

trix. The χ^2 value is $N-1$ times the minimum value of the maximum-likelihood fit function, where N equals the sample size [Joreskog and Sorbom, 1984].* A statistically insignificant χ^2 value indicates that the model reproduces the original correlation matrix reliably. As expected, the χ^2 goodness-of-fit statistic indicates that the base model (a model that includes all regions and paths used to simulate the data) reproduces the original correlation matrix reliably (for $N = 100$, $\chi^2(26) = 15.99$ $P = 0.94$, $N = 60$, $\chi^2(26) = 31.57$ $P = 0.21$, and $N = 20$, $\chi^2(26) = 29.91$ $P = 0.27$). The χ^2_{DIFF} test is a hierarchical test that determines if a modification to the model leads to a significant improvement in the fit of the model [McIntosh and Gonzalez-Lima,

*There are other measures of fit that can be used that are better behaved for the purpose of assessing fit per se. However, for present purposes, where relative changes in fit are the primary focus, the χ^2 goodness-of-fit is sufficient.

TABLE III. Total effects estimate for base model

	V1	V2	V3	V4	V5	V6
LISREL estimate of base model for Network 1						
V1						
V2						
V3		-0.83	-0.20			-0.69
		-0.65	-0.15			-0.58
		-0.82	-0.16			-0.71
V4	0.97					
	0.98					
	0.98					
V5	0.74	0.68	-0.06	0.37		-0.20
	0.76	0.61	-0.02	0.30		-0.06
	0.77	0.72	-0.03	0.39		-0.14
V6		0.96	0.23			-0.20
		0.96	0.23			-0.15
		0.96	0.19			-0.16
LISREL estimate of base model for Network 2						
V1						
V2						
V3		0.75	-0.15			-0.67
		0.80	-0.21			-0.66
		0.75	-0.12			-0.67
V4	0.97					
	0.98					
	0.99					
V5	0.54	-0.85	0.03	0.10		0.15
	0.77	-0.85	0.04	0.30		0.12
	0.49	-0.77	0.07	0.34		0.41
V6		-0.96	0.19			-0.15
		-0.96	0.35			-0.21
		-0.98	0.16			-0.12

Each row within a cell corresponds to the estimate from each of the three sample sizes ($N = 100, 60, 20$ from top to bottom in each cell).

1992; McIntosh et al., 1994]. Unlike the goodness-of-fit, the χ^2_{DIFF} test should be statistically significant to affirm the model modification. Here the difference between the χ^2 value for the null model (χ^2_{H0}) can be compared to that of the alternative model (χ^2_{HA}). If the χ^2 value for the null model is significantly larger than that for the alternative model, then the coefficients that varied between conditions are statistically different. For example, for $N = 60$, $\chi^2_{\text{H0}}(34) = 1187.18$ $P < 0.0001$ and $\chi^2_{\text{HA}}(26) = 31.57$ $P = 0.21$. The resultant, and rather large, $\chi^2_{\text{DIFF}}(8) = 1155.61$ $P < 0.0001$ indicates that the modification (i.e., allowing for differences between the two networks) leads to an improvement in the fit of the model.

Within a model, modification indices can be used to determine which paths are correctly defined. Modification indices are computed from the first- and second-order partial derivatives for each path coefficient and residual in the model. They indicate which parameters should be freed to improve the fit of the model. A high modification index suggests that the fit of the model could be significantly improved by allowing that parameter to change. Modification indices can be evaluated as a chi-squared value with 1 degree of freedom [Joreskog and Sorbom, 1984]. The dominant ones are highlighted in Tables IV, V, VI, and VIII. As an

additional evaluation the impact of modification, we examined the total effects involving the regions that were most directly connected to the modified path or area. In addition to the base model (Table III), tables of total effects are presented for cases where there were substantive changes in total effects (Tables VII, IX).

Simulation I

Remove region V4.

The influence of region V4 does not differ between the two networks. Region V4 has positive input from region V1, and positive output to region V5. Direct effects are given in Table IV. Save for the loss of effects related to V4, there were no changes in total effects.

$N = 100$.

When Region 4 was removed from the analysis, the χ^2 goodness-of-fit statistic indicated that the alternative model reproduced the original correlation matrix reliably ($\chi^2_{\text{HA}}(18) = 18.53$, $P = 0.42$). The null model, where path coefficients were constrained to be the same between networks, had a $\chi^2(24) = 1710.43$, $P < 0.001$. The difference between null and

TABLE IV. Direct effects from simulation 1: omission of V4

	V1	V2	V3	V4	V5	V6
LISREL estimate for Network 1						
V1						
V2						
V3						−0.86
						−0.68
						−0.85
V4	X					
V5	0.74	0.93		X	15.36	−0.25
	0.76	0.71			5.07	−0.09
	0.77	0.93			7.22	−0.23
V6	1.19	0.28				
	1.13	0.27				
	1.14	0.22				
LISREL estimate for Network 2						
V1						
V2						
V3						−0.78
						−0.83
						−0.76
V4	X					
V5	0.52	−0.70		X	1.05	0.16
	0.52	−0.69			2.29	0.17
	0.49	−0.39			23.76	0.39
V6		−1.13	0.23			
		−1.22	0.32			
		−1.11	0.18			

Cells with X indicate source of eliminated effective connections. Boldface values are the highest modification indices, in this case for the PSI effects on V5. Each row within a cell corresponds to the estimate from each of the three sample sizes ($N = 100, 60, 20$ from top to bottom in each cell).

alternative models was significant ($\chi^2_{\text{DIFF}}(6) = 1691.09, P < 0.0001$).

In the alternative model, the relative size of most path coefficients was maintained. However, the strength of the positive path from V1 to V5 was higher to account for the direct effect between V1 and V5 as well as the indirect effect from V1 through V4 to V5. The highest modification index for the alternative model was the residual influences on V5, albeit much smaller for Network 2 than Network 1. No other modification indices were significant.

N = 60.

The overall results did not change with the smaller sample of 60. When region V4 was removed from the analysis, the χ^2 goodness-of-fit statistic indicated that the alternative model reproduced the original correlation matrix reliably ($\chi^2_{\text{HA}}(18) = 21.55, P = 0.25$). The null model had a $\chi^2(24) = 1190.43, P < 0.001$, resulting in a significant difference between null and alternative models ($\chi^2_{\text{DIFF}}(6) = 1168.88, P < 0.0001$).

In the alternative model, the relative size of most path coefficients was maintained. Again, the strength of the positive path from V1 to V5 increased because this path accounted for the direct effect of V1 on V5 and the indirect effect from V1 to V4 to V5. The highest modification index for the alternative model was the residual influences on V5,

although this time the indices were more similar between networks. No other modification indices were significant.

N = 20.

When region 4 was removed from the analysis, it led to a poor fit for the alternative model. The χ^2 goodness-of-fit statistic showed that the models for null and alternative did not reproduce the original correlation matrix ($\chi^2_{\text{H0}}(24) = 618.81, P < 0.0001, \chi^2_{\text{HA}}(18) = 35.73, P = 0.008$). Despite the bad absolute fit, the χ^2_{DIFF} test correctly identified that Network 1 and Network 2 were different ($\chi^2_{\text{DIFF}}(6) = 583.08, P < 0.0001$).

In the alternative model, the relative size of most path coefficients were maintained, with the exception of the path from V1 to V5, which was higher, as in the previous cases. It is noteworthy that, as in the previous cases, the modification index for the residual effect on V5 was high. No other modification indices were significant.

Simulation 2

Remove path from V1 to V5.

We removed the path from V1 to V5 to examine whether the omission of a path that does not change between the two networks influenced the ability to detect a change in effective connection. There is an indirect path from V1 to V5: V1 to V4 to V5.

TABLE V. Direct effects from simulation 2: omission of path from V1 to V5

	V1	V2	V3	V4	V5	V6
LISREL estimate for Network 1						
V1						
V2						
V3						−0.86
						−0.68
						−0.85
V4	0.97					
	0.98					
	0.98					
V5	48.03	0.90		0.74	18.56	−0.24
	18.57	0.65		0.76	12.68	−0.05
	10.65	0.80		0.76	6.99	−0.08
V6	1.19	0.28				
	1.13	0.27				
	1.14	0.22				
LISREL estimate for Network 2						
V1						
V2						
V3						−0.78
						−0.83
						−0.76
V4	0.97					
	0.98					
	0.99					
V5	54.52	−0.58		0.53	9.96	0.27
	6.06	−0.71		0.52	3.62	0.14
	2.55	−0.27		0.49	2.23	0.50
V6		−1.13	0.23			
		−1.22	0.32			
		−1.11	0.18			

Boldface values are the highest modification indices in the path from V1 to V5 and the PSI effect on V5. Each row within a cell corresponds to the estimate from each of the three sample sizes ($N = 100, 60, 20$ from top to bottom in each cell).

$N = 100$.

When the positive path from V1 to V5 was omitted from the model, the χ^2 goodness-of-fit statistic indicated that the models for null and alternative were unable to reproduce the original correlation matrix ($\chi^2_{H_0}(35) = 1757.38$ $P < 0.0001$, $\chi^2_{H_A}(28) = 127.7$ $P < 0.0001$). The χ^2_{DIFF} test correctly identified that Network 1 and Network 2 were different from each other ($\chi^2_{DIFF}(7) = 1629.68$ $P < 0.0001$).

The alternative model maintained the relative size of most path coefficients. However, the estimated input from region V4 to V5 increased. This makes sense because there is an alternative effect from region V1 to V4 to V5, so the direct effect from V1 to V5 was accounted for in the indirect effect from V1 to V4 to V5. This was emphasized in the examination of the total effect of V1 on V5 (not shown). Despite the missing path, the total effect of V1 on V5 was the same as in the base model. This arose because of the increased direct effect of V4 on V5, which mediated the indirect effect of V1 on V5. The modification index for the eliminated path was also very large, and, although smaller than the index for the eliminated path, the index for the PSI effect on V5 was also high.

$N = 60$.

When the path from V1 to V5 was omitted from the model, the χ^2 goodness-of-fit statistic indicated that the models for H_0 and H_A were unable to reproduce the original correlation matrix reliably ($\chi^2_{H_0}(35) = 1207.16$ $P < 0.0001$, $\chi^2_{H_A}(28) = 62.66$ $P = 0.00018$), and the χ^2_{DIFF} test correctly identified that Network 1 and Network 2 were different from each other ($\chi^2_{DIFF}(7) = 1144.5$ $P < 0.0001$).

Again, the alternative model maintained the relative size of most path coefficients. However, the estimated input from region V4 to V5 increased. The direct effect from V1 to V5 was accounted for in the indirect effect from V1 to V4 to V5, and was expressed in total effects (not shown). Modification indices were highest for the missing connection from V1 to V5, and, although smaller than the index for the eliminated path, the index for the PSI effect on V5 also was high.

$N = 20$.

When the path from V1 to V5 was omitted from the model, the χ^2 goodness-of-fit statistic indicated that the models for H_0 and H_A were unable to reproduce the original

TABLE VI. Direct effects from simulation 3: omission of path from V2 to V5

	V1	V2	V3	V4	V5	V6
LISREL estimate for Network 1						
V1						
V2						
V3						−0.86
						−0.68
						−0.85
V4	0.97					
	0.98					
	0.98					
V5	0.34	273.71		0.38	519.79	0.64
	0.32	150.16		0.46	196.47	0.57
	0.15	70.64		0.54	210.09	0.69
V6	1.19	0.28				
	1.13	0.27				
	1.14	0.22				
LISREL estimate for Network 2						
V1						
V2						
V3						−0.78
						−0.83
						−0.76
V4	0.97					
	0.98					
	0.99					
V5	0.26	171.53		0.29	133.64	0.83
	0.28	76.59		0.26	43.73	0.83
	0.05	14.31		0.44	28.82	0.77
V6		−1.13	0.23			
		−1.22	0.32			
		−1.11	0.18			

Boldface values are the highest modification indices in the path from V2 to V5 and the PSI effect on V5. Each row within a cell corresponds to the estimate from each of the three sample sizes ($N = 100, 60, 20$ from top to bottom in each cell).

correlation matrix reliably ($\chi^2_{H_0}(35) = 482.31$ $P < 0.0001$, $\chi^2_{H_A}(28) = 47.58$ $P = 0.012$). The χ^2_{DIFF} test correctly identified that Network 1 and Network 2 were different from each other ($\chi^2_{DIFF}(7) = 434.73$ $P < 0.0001$). The alternative model in this instance showed similar patterns as for $N = 100$ and $N = 60$, although the magnitude of the differences from the base model were not as large.

Simulation 3

Remove path from V2 to V5.

We removed the path from V2 to V5 to examine whether the omission of a path that changes between the two groups influences SEM's ability to detect a change in effective connection. The path from V2 to V5 was positive in Network 1, and negative in Network 2. Although the direct path from V2 to V5 was removed, there is an indirect path from V2 through V6 to V5. This simulation was the only case where a task difference was obscured.

$N = 100$.

When the path from V2 to V5 was omitted from the model the χ^2 goodness-of-fit statistic indicated that the

models for H_0 and H_A were unable to reproduce the original correlation matrix reliably ($\chi^2_{H_0}(35) = 1770.1$ $P < 0.0001$, $\chi^2_{H_A}(28) = 506.16$ $P < 0.0001$). Despite the bad fit, the χ^2_{DIFF} test correctly identified that Network 1 and Network 2 were different from each other ($\chi^2_{DIFF}(7) = 1263.94$ $P < 0.0001$).

The alternative model maintained the relative size of most path coefficients. However, for both networks the direct effect of V6 on V5 became more positive, changing the original nature of the effect. The relative differences in the path from V6 to V5 were maintained (i.e., Network 2 had a more positive path), but this was overwhelmed by the rerouted indirect effect from V2. The change in the direct effect from V6 to V5 enabled the total effect of V2 on V5 to remain the same, despite the zero direct effect (Tables VI, VII). Interestingly, despite the increase in the direct effect of V6 on V5, the differences between networks were still carried by V2. As may be anticipated, the path from V2 to V5 had a very large modification index. The PSI influence on V5 was also quite large.

$N = 60$.

When the path from V2 to V5 was omitted from the model the χ^2 goodness-of-fit statistic showed that the

TABLE VII. Total effects estimate for simulation 3: omission of path from V2 to V5

	V1	V2	V3	V4	V5	V6
Total effects for Network 1						
V1						
V2						
V3		-0.65	-0.20			-0.69
		-0.83	-0.15			-0.58
		-0.82	-0.16			-0.71
V4	0.97					
	0.98					
	0.98					
V5	0.71	0.61	0.15	0.38		0.51
	0.77	0.55	0.13	0.46		0.48
	0.68	0.67	0.13	0.54		0.58
V6		0.96	0.23			-0.20
		0.96	0.23			-0.15
		0.96	0.19			-0.16
Total effects for Network 2						
V1						
V2						
V3		0.75	-0.21			-0.67
		0.80	-0.12			-0.66
		0.75	-0.16			-0.67
V4	0.97					
	0.98					
	0.99					
V5	0.55	-0.79	0.15	0.29		0.70
	0.53	-0.80	0.21	0.26		0.66
	0.48	-0.76	0.12	0.44		0.68
V6		-0.96	0.19			-0.15
		-0.96	0.25			-0.21
		-0.98	0.16			-0.12

Italic values indicate changes from estimates obtained in direct effects from simulation 3 (Table VI). Each row within a cell corresponds to the estimate from each of the three sample sizes ($N = 100, 60, 20$ from top to bottom in each cell).

models for H_0 and H_A were unable to reproduce the original correlation matrix reliably ($\chi^2_{H_0}(35) = 1261.51$ $P < 0.0001$, $\chi^2_{H_A}(28) = 269.3$ $P < 0.0001$). The χ^2_{DIFF} test correctly identified that Network 1 and Network 2 were different from each other ($\chi^2_{DIFF}(7) = 992.21$ $P < 0.0001$). The pattern of changes for direct and total effects and modification indices were like those for the $N = 100$ model (Tables VI, VII).

$N = 20$.

When the positive path from 2 to 5 was omitted from the model, the χ^2 goodness-of-fit statistic indicated a poor fit for H_0 and H_A ($\chi^2_{H_0}(35) = 491.51$ $P < 0.0001$, $\chi^2_{H_A}(28) = 142.58$ $P < 0.0001$), but the χ^2_{DIFF} test correctly identified that Network 1 and Network 2 were different from each other ($\chi^2_{DIFF}(7) = 348.93$ $P < 0.0001$). The pattern of changes for direct and total effects and modification indices are like those for the $N = 100$ and $N = 60$ model (Tables VI, VII).

Simulation 4

Remove path from 2 to 6.

We removed the path from V2 to V6 to examine the impact on effective connection between two networks when the omitted path is different between networks, and there is no indirect path that can mediate the missing effect. The path from V2 to V6 is positive in Network 1, and negative in Network 2. Direct effects are presented in Table VIII and total effects in Table IX.

$N = 100$.

When the path from V2 to V6 was removed, the χ^2 goodness-of-fit statistic showed that the models for H_0 and H_A were unable to reproduce the original correlation matrix reliably ($\chi^2_{H_0}(35) = 1693.82$ $P < 0.0001$, $\chi^2_{H_A}(28) = 1116.94$ $P < 0.0001$). The χ^2_{DIFF} test indicated that Network 1 and Network 2 were different from each other ($\chi^2_{DIFF}(7) = 576.88$ $P < 0.0001$).

The paths and total effects that changed involve V3 and V6. In both networks, the positive path from V3 to V6 became

TABLE VIII. Direct effects from simulation 4: omission of path from V2 to V6

	V1	V2	V3	V4	V5	V6
LISREL estimate for Network 1						
V1						
V2						
V3						−0.30
						−0.22
						−0.30
V4	0.97		764.02			
	0.98		1334.15			
	0.98		258.12			
V5	0.39	0.92		0.37		−0.25
	0.47	0.68		0.30		−0.07
	0.39	0.88		0.39		−0.17
V6		156.81	−0.72			764.02
		221.19	−0.54			1334.15
		38.16	−0.73			258.12
LISREL estimate for Network 2						
V1						
V2						
V3						−0.29
						−0.25
						−0.29
V4	0.97		1194.12			
	0.98		683.47			
	0.99		418.14			
V5	0.45	−0.68		0.10		0.18
	0.23	−0.70		0.30		0.15
	0.15	−0.31		0.34		0.47
V6		230.00	−0.66			1194.12
		126.97	−0.67			683.47
		66.49	−0.66			418.14

Boldface values are the highest modification indices in the path from V2 to V6 and the PSI effects on V3 and V6. Each row within a cell corresponds to the estimate from each of the three sample sizes ($N = 100, 60, 20$ from top to bottom in each cell).

negative, and the negative path from V6 to V3 weakened. The total effects between these regions of the networks were similarly affected, with total effects from V2 on V3 and V6 disappearing because of the loss of the direct effect from V2 to V6 (Table IX). Despite the large changes in direct and indirect effects, the network differences were restricted to the effect of V2 on V5. Although the alternative model did not account for the removed path, the modification indices indicated how to alter the model, with a large value for the path from V2 to V6, but also large indices for PSI effects in the V3 and V6 region. The large PSI modification index for V3 is a result of the loss of the indirect effect from V2 that was mediated through V6.

$N = 60$.

When the path from V2 to V6 was removed the χ^2 goodness-of-fit statistic indicated a poor model fit for H_0 and H_A ($\chi^2_{H_0}(35) = 1187.81$ $P < 0.0001$, $\chi^2_{H_A}(28) = 831.25$ $P < 0.0001$). The χ^2_{DIFF} test indicated that Network 1 and Network 2 were different from each other ($\chi^2_{DIFF}(7) = 356.56$ $P < 0.001$). The pattern of changes for direct and total effects and modification indices are like those for the $N = 100$ model.

$N = 20$.

When the path from V2 to V6 was removed the χ^2 goodness-of-fit statistic indicated a poor fit for H_0 and H_A ($\chi^2_{H_0}(35) = 481.32$ $P < 0.0001$, $\chi^2_{H_A}(28) = 302.21$ $P < 0.0001$). The χ^2_{DIFF} test indicated that Network 1 and Network 2 were different from each other ($\chi^2_{DIFF}(7) = 179.11$ $P < 0.0001$). The pattern of changes for direct and total effects and modification indices are like those for the $N = 100$ and $N = 60$ models.

DISCUSSION

Our simulations suggest that it is possible to make inferences in SEM about changes in effective connectivity when the overall model does not fit the data. In all cases, SEM correctly identified that Network 1 and Network 2 were different from each other. The size and strength of the differences in effective connections were maintained when they were part of the model and manifested as both direct and total effects.

The main findings of our simulations are as follows:

- (1) If paths or regions whose effects do not differ across tasks are left out of a model, they have no impact on

TABLE IX. Total effects estimate for simulation 4: omission of path from V2 to V6

	V1	V2	V3	V4	V5	V6
Total effects for Network 1						
V1						
V2						
V3			0.28			−0.39
			0.14			−0.25
			0.28			−0.38
V4	0.97					
	0.98					
	0.98					
V5	0.74	0.92	0.23	0.37		−0.32
	0.76	0.68	0.04	0.30		−0.08
	0.77	0.88	0.16	0.39		−0.21
V6			−0.92			0.28
			−0.61			0.14
			−0.93			0.28
Total effects for Network 2						
V1						
V2						
V3			0.24			−0.36
			0.20			−0.30
			0.23			−0.35
V4	0.97					
	0.98					
	0.99					
V5	0.45	−0.68	−0.15	0.10		0.22
	0.52	−0.70	−0.12	0.30		0.19
	0.49	−0.31	−0.38	0.34		0.57
V6			−0.81			0.24
			−0.80			0.20
			−0.82			0.23

Italic values indicate changes from estimates obtained in direct effects from simulation 4 (Table VIII). Each row within a cell corresponds to the estimate from each of the three sample sizes ($N = 100, 60, 20$ from top to bottom in each cell).

the ability to detect task differences in the rest of the model, regardless of whether the overall model fits the data.

- (2) If a path that carries a task difference is not part of the model, the difference may manifest in the total effects so long as indirect routes are possible. Importantly, while the magnitude of the direct effects along these indirect routes may change, they themselves do not appear to embody the task-dependent change. In simulation 3, where the path from V2 to V5 was omitted, the direct effect from V6 to V5 increased, but did so similarly for both networks. It is thus possible for larger effects to obscure small effects when the direct effect carrying the large effects is not part of the model.
- (3) If a path carries a task difference that cannot be captured in the total effects, the effect is not represented in the model.
- (4) Modification indices can be useful to decide whether regions and/or paths are missing from a model.

The conclusions we draw are possible because the focus of SEM in the present article is to test for differences in path

coefficients within a predefined causal structure. The important point is that the causal structure is independent of the data. In the case where SEM is used to determine causal structure, the interpretation of individual path coefficients is indeed ambiguous in a poorly fitting model. This distinction emphasizes what Bullmore et al. [2000] defined as the relativistic vs. confirmatory approach. If the purpose of SEM is to evaluate whether the hypothesized causal structure is consistent with the data, one should not interpret path coefficients if the model does not fit [MacCallum, 1995]. This is because the coefficients may change when the causal structure is modified. In the case where the causal structure is relatively fixed, the path coefficients are themselves quite interpretable and can be compared between tasks or groups.

Consistent with previous work [McIntosh and Gonzalez-Lima, 1994], we noted that modification indices are able to indicate possible omissions of connections and regions. However, modification indices alone may not be completely accurate because they only provide clues as to possible omissions. Usually, the regions and direct paths that are included in the model are selected beforehand using either statistical criteria or a theoretical guide. The situation for

residual influences (PSI) is less clear. As noted elsewhere [Hayduk, 1987; McIntosh and Gonzalez-Lima, 1994], model instabilities can be expressed in high PSI estimates and is the most direct way minimize discrepancy in model fitting. We have previously suggested a rough rule of thumb for constraining PSI values, but a more empirically based means to derive the estimates [Bullmore et al., 2000] would be preferable. When model modification can be made either on the paths between regions or residual influences, it is preferable to make the modification on the paths (effective connections) first. For example, in simulations 3 and 4 the modification indices for PSI were higher than for the path coefficients. Had the PSI influences been allowed to change, this would have obscured the actual task differences in the effective connections.

Our simulations also emphasize the need for careful model specification both in terms of regional selection and determination of anatomical connectivity. Univariate tests that are used in neuroimaging are quite insensitive to changes in the relation between areas, and thus models based solely on such analyses are likely to miss important task differences. This is particularly true in cases where the level of activity in a region is identical between tasks, yet the interregional interactions change. Such a case was noted in a model of dorsal vs. ventral cortical visual system effective connections where prefrontal cortex showed strong differences in connectivity despite similar activity levels across tasks [McIntosh et al., 1994]. The present simulations, and those published earlier [McIntosh and Gonzalez-Lima, 1994], suggest that these missing influences may manifest as large PSI estimates, but it is advisable to perform additional analyses that are sensitive to regional interactions prior to SEM to ensure that such important effects are not missed [McIntosh, 1999].

The particulars of our conclusions are, of course, dependent on our simulation, and may not hold should models that are more complicated be considered. We do note that more complex models that have larger numbers of areas and more prevalent reciprocal connections have been used with replicable [Horwitz et al., 1995; Kohler et al., 1998; McIntosh et al., 1994] and stable results [McIntosh et al., 1996, 1998]. However, we are confident that the main conclusion of this report—that one can detect differences in effective connectivity with SEM even when the overall model does not fit the data—is accurate.

It is important to remember that SEM is primarily a modeling technique. Thus, the models are, by definition, false. A model embodies the theoretical biases of its user and the decisions that are made in building the model are open for validation. SEM can be considered part of a larger family of theoretical approaches classified under linear Gaussian models, which run the range from simple regression, to SEM, factor analysis, independent component analysis, and Bayesian networks, such as is implemented in Dynamic Causal Modeling (DCM) [Friston et al., 2003]. Many of these methods can be used to estimate effective connectivity, but to do so requires a certain amount of user intervention. In the case of SEM, the user defines a causal structure based on known neuroanatomy. Although there is a concern that the source of anatomical connectivity comes from studies in monkeys rather than hu-

mans, there is little reason to believe from the comparative anatomy literature that the connections per se differ across primate species. Using a relativistic approach, DCM is able to determine the causal structure that is most consistent with the data from a range of prespecified possibilities (although this can be computationally expensive). At a more general level, by specifying a particular model to estimate effective connections, the researcher is making an explicit statement on their belief about the data. Such explicit statements allow for unambiguous evaluation of a researcher's theory, which are then open to the community for further scientific scrutiny.

REFERENCES

- Bentler PM (1985): Theory and implementation of EQS, a structural equations program. Los Angeles: BMDP Statistical Software.
- Bentler PM, Weeks DG (1980): Linear structural equations with latent variables. *Psychometrics* 45:289–308.
- Bollen KA (1989): Structural equations with latent variables. New York: John Wiley & Sons.
- Buchel C, Friston K (1997): Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modeling and fMRI. *Cereb Cortex* 7:768–778.
- Bullmore E, Horwitz B, Honey G, Brammer M, Williams S, Sharma T (2000): How good is good enough in path analysis of fMRI data? *Neuroimage* 11:289–301.
- Friston K (1994): Functional and effective connectivity: a synthesis. *Hum Brain Mapp* 2:56–78.
- Friston KJ, Harrison L, Penny W (2003): Dynamic causal modelling. *Neuroimage* 19:1273–1302.
- Goncalves MS, Hall DA (2003): Connectivity analysis with structural equation modelling: an example of the effects of voxel selection. *Neuroimage* 20:1455–1467.
- Goncalves MS, Hall DA, Johnsrude IS, Haggard MP (2001): Can meaningful effective connectivities be obtained between auditory cortical regions? *Neuroimage* 14:1353–1360.
- Grady CL, McIntosh AR, Craik FI (2003): Age-related differences in the functional connectivity of the hippocampus during memory encoding. *Hippocampus* 13:572–586.
- Hayduk L (1987): Structural equation modeling with LISREL. Baltimore: Johns Hopkins University Press.
- Horwitz B (2003): The elusive concept of brain connectivity. *Neuroimage* 19:466–470.
- Horwitz B, McIntosh AR, Haxby JV, Furey M, Salerno JA, Schapiro MB, Rapoport SI, Grady CL (1995): Network analysis of PET-mapped visual pathways in Alzheimer type dementia. *Neuroreport* 6:2287–2292.
- Jennings JM, McIntosh AR, Kapur S, Zipursky RB, Houle S (1998): Functional network differences in schizophrenia: a rCBF study of semantic processing. *Neuroreport* 9:1697–1700.
- Joreskog K, Sorbom D (1979): Advances in factor analysis and structural equation models. Cambridge, MA: Abt Publishers.
- Joreskog KG, Sorbom D (1984): LISREL VI: user's guide, 3rd ed. Mooresville, IN: Scientific Software.
- Kohler S, McIntosh AR, Moscovitch M, Winocur G (1998): Functional interactions between the medial temporal lobes and posterior neocortex related to episodic memory retrieval. *Cereb Cortex* 8:451–461.
- Lee L, Harrison LM, Mechelli A (2003): A report of the functional connectivity workshop, Dusseldorf 2002. *Neuroimage* 19:457–465.
- Loehlin JC (1987): Latent variable models. An introduction to factor, path, and structural analysis. Hillsdale, NJ: Lawrence Erlbaum.

- MacCallum R (1995): Model specification: procedures, strategies, and related issues. In: Hoyle RH, editor. *Structural equation modelling: concepts, issues, and applications*. Thousand Oak, CA: Sage. p 16–36.
- Maguire EA, Mummery CJ, Buchel C (2000): Patterns of hippocampal-cortical interaction dissociate temporal lobe memory sub-systems. *Hippocampus* 10:475–482.
- McIntosh AR (1999): Mapping cognition to the brain through neural interactions. *Memory* 7:523–548.
- McIntosh AR, Gonzalez-Lima F (1992): Structural modeling of functional visual pathways mapped with 2-deoxyglucose: effects of patterned light and footshock. *Brain Res* 578:75–86.
- McIntosh AR, Gonzalez-Lima F (1993): Network analysis of functional auditory pathways mapped with fluorodeoxyglucose: associative effects of a tone conditioned as a Pavlovian excitator or inhibitor. *Brain Res* 627:129–140.
- McIntosh AR, Gonzalez-Lima F (1994): Structural equation modeling and its application to network analysis in functional brain imaging. *Hum Brain Mapp* 2:2–22.
- McIntosh AR, Grady CL, Ungerleider LG, Haxby JV, Rapoport SI, Horwitz B (1994): Network analysis of cortical visual pathways mapped with PET. *J Neurosci* 14:655–666.
- McIntosh AR, Grady CL, Haxby JV, Ungerleider LG, Horwitz B (1996): Changes in limbic and prefrontal functional interactions in a working memory task for faces. *Cereb Cortex* 6:571–584.
- McIntosh AR, Cabeza RE, Lobaugh NJ (1998): Analysis of neural interactions explains the activation of occipital cortex by an auditory stimulus. *J Neurophysiol* 80:2790–2796.
- Northoff G, Kotter R, Baumgart F, Danos P, Boeker H, Kaulisch T, Schlagenhauf F, Walter H, Heinzel A, Witzel T, Bogerts B (2004): Orbitofrontal cortical dysfunction in akinetic catatonia: a functional magnetic resonance imaging study during negative emotional stimulation. *Schizophr Bull* 30:405–427.
- Schlösser R, Gesierich T, Kaufmann B, Vucurevic G, Hunsche S, Gawehn J, Stoeter P (2003): Altered effective connectivity during working memory performance in schizophrenia: a study with fMRI and structural equation modeling. *Neuroimage* 19:751–763.
- Solodkin A, Hlustik P, Chen EE, Small SL (2004): Fine modulation in network activation during motor execution and motor imagery. *Cereb Cortex* 14:1246–1255.
- Zhuang J, LaConte S, Peltier S, Zhang K, Hu X (2005): Connectivity exploration with structural equation modeling: an fMRI study of bimanual motor coordination. *Neuroimage* 25:462–470.