Association chain graphs: modelling etiological pathways

MICHAEL HÖFLER, HANS-ULRICH WITTCHEN, Technical University Dresden, Institute of Clinical Psychology and Psychotherapy, Dresden, Germany and Max-Planck-Institute of Psychiatry, Clinical Psychology and Epidemiology, Muenchen, Germany

ROSELIND LIEB, Max-Planck-Institute of Psychiatry, Clinical Psychology and Epidemiology, Muenchen, Germany.

JÜRGEN HOYER, Technical University Dresden, Institute of Clinical Psychology and Psychotherapy, Dresden, Germany

ROBERT H. FRIIS, California State University Long Beach, Department of Health Science, USA

ABSTRACT Multiple time-dynamic and interrelated risk factors are usually involved in the complex etiology of disorders. This paper presents a strategy to explore and display visually the relative importance of different association pathways for the onset of disorder over time. The approach is based on graphical chain models, a tool that is powerful but still underutilized in most fields. Usually, the results of these models are displayed using directed acyclic graphs (DAGs). These draw an edge between a pair of variables whenever the assumption of conditional independence given variables on an earlier or equal temporal footing is violated to a statistically significant extent. In the present paper, the graphs are modified in that confidence intervals for the strengths of associations (statistical main effects) are visualized. These new graphs are called association chain graphs (ACGs). Statistical interactions cause 'edges' between the respective variables within the DAG framework (because the assumption of conditional independence is violated). In contrast they are represented as separate graphs within the subsample where the different association chains may work within the ACG framework. With this new type of graph, more specific information can be displayed whenever the data are essentially described only with statistical main- and two-way interaction effects.

Key words: disease, etiology, probabilistic models, logistic regressions, graphical model, graphical display, confidence intervals

The challenge of analysing the complex etiology of disease

The derivation and quantification of etiological models in medicine is a complex challenge and, from the statistical point of view, is rarely addressed in a satisfactory manner. Typically, the onset and course of mental disorders result from multiple risk factors that occur in various forms and with varying duration and severity over time and that involve more-or-less pronounced and time-dependent interrelationships. Further, each risk may vary in its effect on the outcome across individuals and the dependencies can be dose dependent. A causal explanation is typically sought by

developing models that attempt to consider all of the assumed relevant factors and by taking into account their temporal structure and interrelationships.

The prerequisites for such modelling exercises are:

- the definition of the relevant target population;
- adequate sampling with a sufficient number of individuals; and
- appropriate measurement of constructs and establishment of their temporal order.

As many factors almost always play an important role in disease etiology, studying only bivariate associations

between a prior factor (explanatory variable – EV) and an outcome is often simplistic or misleading (Blyth, 1972). Bivariate associations are often biased by confounders that affect the exposure (for example, in observational studies – Greenland and Robins, 1985) and are mediated by variables that occur between exposure and disease (for example, heredity and diabetes – Jorgensen et al, 1995). Further, when certain EVs are interrelated (common in observational studies), it is statistically impossible to determine whether they act individually or together in affecting an outcome (Kraemer et al., 2001). Instead, attempts are often made to separate indirect effects from direct effects (for example, whether heredity influences diabetes only via diet and other mediators – Jorgensen et al., 1995). Such a separation, however, is only possible in a sense more closely related to causality when exposure and intermediate variables are independently randomized in experimental studies (Robins and Greenland, 1992). Nevertheless, one can assess the relative importance of each EV in predicting an outcome by adjusting for the remaining EVs that occur on the same or an earlier temporal footing (conditional associations) – for example, within the flexible framework of regression methods (Cox and Wermuth, 1996). One is then able to analyse the relationships among the EVs considering their temporal order (see the application below). Altogether this aims at identifying, quantifying and visualizing the most important direct and indirect pathways into the onset of a disorder. The corresponding statistical approach is called graphical chain modelling (see below).

The challenge in understanding complex associations and evaluating their appropriateness and fit lies in the resulting mass of data and the lack of a coordinated standard procedure to summarize and visualize the findings in an easily conductible, persuasive and statistically sound way. The pressing need for an analytical tool can be underlined by the frequent misuse of so-called structural equation models with latent variables as an exploratory tool (Pearl, 2000). Despite the lack of a straightforward interpretation of the results of these models (Cox and Wermuth, 1996) they are based on a deterministic understanding of causality (Pearl, 2000).

In this paper we present an alternative probabilistic strategy based on graphical chain models (GCMs) that aim to estimate population effects. Graphical chain

models are well established in statistics (Cox and Wermuth, 1996; Lauritzen and Wermuth, 1989; Klein et al, 1995) but are still under-utilized in many fields of empirical research. A new type of graph is introduced that provides more specific information – confidence intervals for strength of association, and separate graphs in the case of heterogeneous association, whenever this essentially describes the data. Graphs previously used to communicate the results of these models (directed acyclic graphs, see below) display only binary information: is the assumption of conditional independence violated or not?

GCMs and directed acyclic graphs

When all variables are binary, a GCM can be constructed by a sequence of univariate multiple logistic regressions assuming that they provide a reasonable fit to the data (for example, no interaction terms are required – see below). Although the dependencies between the regression equations are ignored, this outline strategy provides a full representation of the data (Cox and Wermuth, 1996) and can easily be conducted with standard software packages. The results are summarized in a graph. The use of graphs to communicate statistical results is often recommended. Graphs take advantage of the natural perceptual and cognitive capacities of human beings (Cleveland and McGill, 1985; Kosslyn, 1989; Smith et al., 2002), and are easily and quickly perceived, elaborated and memorized, and complement purely abstract or ‘digital’ information.

The graphs previously used for GCMs are called block-recursive graphs or directed acyclic graphs (DAGs) (Cox and Wermuth, 1996; Greenland et al., 1999). They display variables with nodes and the nodes are arranged in boxes representing their temporal ordering. Conditional associations between pairs of variables are displayed with edges: undirected edges between variables on the same temporal footing and directed edges (arrows) between variables on a different footing. In a binary decision, an edge is added when the assumption of conditional independence between a pair of variables Z and Y – given all variables on the same or on an earlier temporal footing of Z – can be shown to be significantly violated. Conditional independence between Z and Y can mean not only the presence of a statistical main effect of Y on Z but also, for example, a two-way interaction of Y with a third variable X in predicting Y or, that Y

increases the variance of Z but leaves its mean unchanged and so forth (conditional on the other variables). In epidemiology, DAGs were used previously, for instance, to distinguish between direct and indirect effects for causal purposes (Greenland et al., 1999; Robins and Greenland, 1992). Here, our aim is similar, but due to the inherent problems of observational studies, we are concerned with assessing the relative importance of different association pathways for onset of disease. Assessing the complete pathways (including the associations between the EVs) can provide important information such as the presence of only indirect pathways from the first occurring factor in the disorder, and could indicate that intervention does not have to take place at this factor but could also be conducted among intermediate factors.

The general problem is that the appropriate visual depiction of complex multivariate distributions of several variables over time may be too complicated to be fully represented in a graph. Although DAGs are usually supplied with text information and a set of tables, our strategy is to *display* more information. We therefore concentrate on the visual depiction of the parameters that are assumed to be the most important in many applications: main effects and two-way interactions – hoping that they characterize the distribution essentially (as is often the case when most variables are binary as in our application example – see below). Our modifications of DAGs will be twofold:

- displaying all the strength of association values (main effects from the model) that are compatible with the data at the pre-specified confidence level (confidence intervals); and
- separating the visual depiction of two-way interactions and main effects by splitting the graph in the case of a noteworthy interaction.

Why is it insufficient to display binary decisions?

Biomedical researchers often tend to focus on p-values only and the binary decisions that are derived from them (Gardner and Altman, 1986; Rothman and Greenland, 1998; Thompson, 1999; Holman et al., 2001). In our approach, the graphs are modified by displaying all the strength of association values that are compatible with the data (at a pre-specified confidence limit) as indicated by the confidence interval (under the standard interpretation of significance tests) (Rothman and Greenland, 1998). This information is

provided in order to avoid the interpretational pitfalls that are inherent in such binary decisions. Strength of association is especially essential for observational studies. Here, one can apply the heuristic that strong associations are less likely to be due to unconsidered confounders alone than are modest associations.

The limited utility of statistical tests has been debated since they were invented (Goodman, 1983; Cohen, 1994; Thompson, 1999). A statistical test rejects the null hypothesis when the p-value is below a pre-specified significance limit, typically when $p < 0.05$. The two-sided p-value indicates the probability that, under the null hypothesis and in an independent and identically designed study of the same population, a test statistic can be observed that produces a value equal to or higher than the one in the present study (Schervish, 1996). Type-I error (rejecting a true null hypothesis) is thereby controlled for, whereas type-II error (failure to reject a wrong null hypothesis) depends on the specific characteristics of a particular study (such as sample size).

Commonly, small p-values are misinterpreted as indicating the amount of support for the alternative hypothesis (Schervish, 1996), as an indicator of effect intensity (Royall, 1986a), or replicability of a result (Sohn, 1998). Further, p-values are often considered as the most important criterion of causality, even among some epidemiologists (Holman et al., 2001). Binary decisions derived from p-values have various interpretational pitfalls: ‘statistically significant’ does not imply that the association found is strong enough to be of any clinical or public health concern (Kraemer et al, 1997). ‘Statistically not significant’, on the other hand, does not necessarily mean even the lack of strong differences (Freiman et al., 1978) and is therefore not sufficient to establish the null hypothesis (Cohen, 1994). Hence, intensity of association measures (IAMs) are necessary that quantify associations on the individual level in a way that weights differences meaningfully (Kraemer, 1985). When they can be computed from a regression equation, the other variables (variables occurring earlier or on the same temporal footing) can be controlled.

Statistical interactions

If a statistical interaction of Y is found with one such variable, X , in predicting Z (conditional on variables on prior or equal temporal footing as Z), an edge

between Y and Z and also between X and Z would be added in the corresponding DAG. We believe that main effects and interaction effects should not be mixed up in a graph whenever the interaction has a substantial meaning. Our attitude towards statistical interactions, however, is somewhat ambivalent. Basically, the new kind of graphs that we will introduce later are mainly intended to display statistical main effects, hoping that these graphs provide a reasonable fit to the data where, for example, no interaction terms are required. The absence of interactions might be an indication of the stability of the results across different studies (where the distributions of EVs might vary). In the presence of interactions, on the other hand, the interpretation of main effects from a statistical model as the (conditional) population average association on the individual level is not altered. Important heterogeneity in association chains, however, might be overlooked. In general, it is not possible to decide statistically whether different causal dependence chains act, or whether the same dependence chain acts differently, if the EVs are correlated. Correlated EVs are typical in observation studies. A statistical interaction is a property of the model that describes the data, not a property of the data themselves (Kraemer et al., 2001).

We therefore recommend searching first for main effects models that are based on suitable measures for strength of association that provide a reasonable fit to the data. If those do not exist, or if there is external information that different causal chains might indeed work, we allow for separate graphs in the subsamples related to different association chains. Then the difference between the graphs illustrates the heterogeneity in association. In case of doubt, however, one might want to apply the philosophical principle of Occam's razor indicating that simpler models should be maintained until there is evidence for more complexity (Balasubramanian, 1997).

Association chain graphs

The danger of the pitfalls related to binary decisions can be considerably reduced by reporting IAM values along with their confidence intervals (Thompson, 1999; Gardner and Altman, 1986; Cohen, 1994). All values that lie within the limits of the confidence intervals are compatible with the data at the pre-specified confidence limit (typically 95%). Therefore, the strength of association between two variables is

translated here into colours in order to provide a visual assessment of effect size. The IA measure must range from -1 (negative association, white colour), over zero (intermediate level between white and black, grey) to $+1$ (perfect positive association, black colour). The grey level representing independence is chosen as the background colour of the graph. It is therefore the contrast to the background that visualizes the strength in association.

The lower (upper) bound of an edge (left versus right bound for edges in the same box) indicates the lower (upper) boundary of the confidence interval and the variation of the colour within the same edge represents all the values that are compatible with the data – the statistical uncertainty or lack in information gain. Note that the point estimate (usually the maximum likelihood estimate) is not the centre of an edge whenever the distribution is not symmetrical around the point estimate. Our aim is to present the whole range of values that are in line with the data rather than displaying the point estimates. We assume that visualizing the point estimate would move the spectators' attention towards a special value that is subject to imprecision. Displaying all the values within confidence limits, however, is expected to direct spectators' attention towards the often pressing need of larger samples.

This new type of graph is called an association chain graph (ACG). Whenever a statistical interaction is found and considered important (because the interaction is strong or might be of substantive relevance), the graph can be separated – different graphs can be constructed within the sub-samples that are related to different association chains. Figure 1 shows some theoretical chain graphs:

A practical research example: pathways into alcohol dependence via early childhood disorders

As a practical research example, we examined to what extent the association between parental alcoholism and alcohol dependence in young adults is mediated by early childhood disorders that may occur between the condition and the outcome. We simplified this example by restricting the number of EVs, despite important other factors such as broken home, life events, and personality. Data are taken from the EDSP-study (Early Developmental Stages of Psychopathology) (Wittchen et al., 1998; Lieb et al., 2000), an observational longitudinal general

Figure 1. Theoretical association chain graphs.

population study on mental and substance use disorders among adolescents in Munich, Germany, and its suburbs. The variable at the beginning of the dependence chain is a parental alcohol use disorder (PAUD) – alcohol abuse or dependence among mother or father. This factor is purely explanatory because it preceded the disorders in offspring. The second block of variables includes the frequent early childhood disorders, attention deficit or hyperactivity disorder (ADHD), and conduct disorder or oppositional defiant disorder (CONOPP). These two variables are outcomes of PAUD but explanatory for the pure outcome variable alcohol dependence in the offspring (OAD). OAD is defined as fulfilling the lifetime DSM-IV criteria (American Psychiatric Association, 1994) for alcohol dependence among any of the three waves.

All the variables shown in the figure are binary and the associations are modelled with odds ratios (ORs) from logistic regressions as is typical in psychiatric epidemiology. Odds ratios ranging from 0 to infinity can be easily transformed into a measure ranging from -1 to +1 by computing Yule's Q: $YQ = (OR-1)/(OR+1)$. YQ

is calculated from the frequencies of a 2×2 table as $(AD-BC)/(AD+BC)$ where A and D are the absolute frequencies in the diagonal elements (representing concurrence of two events) and C and B are the other elements. Such an IA measure with fixed endpoints often yields an increase in robustness against misclassification errors. Table 1 contains the results for onset of alcohol dependence. Figure 2 shows the corresponding association chain graph.

As can be seen from Figure 2, the direct association of a parental alcohol-use disorder (PAUD) with alcohol dependence in offspring (OAD) ranges from just over zero to moderately positive (corresponding to the 95% confidence interval, CI, limits for YQ from 0.049 to 0.520). Therefore, a negative association as well as a strongly positive association can be ruled out at the 95% level. An interesting indirect pathway from PAUD into OAD could be the one via an intermediate ADHD: the association between PAUD and ADHD might be modest as well as high (0.156 – 0.748) and the association between ADHD and OAD might be small to high (0.111 – 0.819). For the other pathway via CONOPP, only medium to strong

Table 1: EDSP-results on conditional associations for onset of alcohol dependence from logistic regressions (n=915 cases)¹

Dependent variable	Covariate	Odds ratio (OR)	Yule's Q (YQ)	95% confidence interval
OAD ² (Nw = 60 ³)				
	Female gender ⁴	0.381	−0.448	−0.669, −0.160
	Age (in years) ⁴	1.125	0.059	−0.078, 0.194
	ADHD ⁵	3.544	0.56	0.111, 0.819
	CONOPP ⁶	1.503	0.201	−0.303, 0.619
	PAUD ⁷ (Nw = 32 ⁷)	1.695	0.258	−0.049, 0.520
ADHD (Nw = 38)				
	Female gender	0.438	−0.391	−0.691, 0.024
	Age (in years)	1.048	0.023	−0.146, 0.192
	CONOPP	18.705	0.899	0.773, 0.956
	PAUD	3.084	0.51	0.156, 0.748
CONOPP (Nw = 41)				
	Female gender	0.602	−0.249	−0.561, 0.127
	Age (in years)	0.836	0.089	−0.242, 0.068
	ADHD	18.437	0.897	0.765, 0.957
	PAUD	1.211	0.096	−0.279, 0.445

¹ The analysis is based on n = 915 cases who were 17–21 years old at the end of the investigation. Two cases had to be omitted since alcohol dependence began before the age of 14 and the temporal precedence of the early childhood disorders could not be guaranteed.

² OAD: alcohol dependence in offspring.

³ Nw = weighted number. Data were weighted in order to match the distribution of sociodemographical characteristics in the population. This was considered when calculating confidence intervals by applying the robust Huber–White sandwich matrix (Royall, 1986b).

⁴ All the associations were controlled for age and gender of the offspring by using them as additional covariates in logistic regressions. These variables are considered as confounding and their effect not displayed in Figure 2.

All model equations fitted the data reasonably ($p > 0.05$) as indicated by the link test (Pregibon, 1980).

⁵ ADHD: attention deficit or hyperactivity disorder.

⁶ CONOPP: conduct disorder or oppositional defiant disorder.

⁷ PAUD: parental alcohol use disorder.

negative associations and strong positive associations can be ruled out. Strikingly, the early childhood disorders ADHD and CONOPP are very strongly associated and this association is stable ($YQ = 0.765 - 0.957$). These findings still include much statistical uncertainty but there is more evidence for the pathway via ADHD than for the one via CONOPP as important in developing OAD. Therefore, if one were to decide whether to focus either on intervention via ADHD or via OAD, one would recommend ADHD.

Comment

Some limitations should be acknowledged: The usefulness of association chain graphs is limited when a high number of variables is involved, as in observational studies when a variety of potential risk factors for

disease are assessed. In principle, this also holds for DAGs but DAGs generally allow more variables before the graphs become too tangled. In the presence of too many variables when using ACGs one would then have to select variables before fitting a model, again requiring binary decisions that we tried to avoid here.

Association chain graphs might, at least in principle, also be used for quantitative covariates because every association measure can be translated into one ranging from −1 to 1. In this case, however, the approximation of the whole multivariate distribution with statistical main- and two-way interaction effects only could in general be poorer, for example when associations are non-monotonic. Non-linear but monotonic associations could be modelled after transforming the covariates.

Figure 2. Association chain graph for onset of alcohol dependence.

Although DAGs have accumulated a large theoretical background over the last 20 years or more, the richer visualization of ACGs can be profitable in situations where they describe the data sufficiently. An example would be meta-studies where the statistical power varies across the different studies. Despite their limitations in such data associations, ACGs provide a clear and easily applicable rationale to make pathways into disease visible.

References

- American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (4 edn). Washington DC: APA, 1994.
- Balasubramanian P. Statistical inference, Occam's razor, and statistical mechanics on the space of probability distributions. *Neural Comput* 1997; 9: 349–68.
- Blyth C. On Simpson's paradox and the sure-thing principle. *J Am Stat Ass* 1972; 67: 364–81.
- Greenland S, Robins JM. Confounding and misclassification. *Am J Epidemiol* 1985; 122: 495–506.
- Cleveland WS, McGill R. Graphical perception and graphical methods for analyzing scientific data. *Science* 1985; 229: 828–32.
- Cohen J. The earth is round ($p < .05$). *Am Psychologist* 1994; 49: 997–1003.
- Cox DR, Wermuth N. *Multivariate Dependencies*. London: Chapman & Hall, 1996.
- Freiman JA, Chalmers TC, Smith H, Kuebler RR. Importance of beta, type-II error and sample-size in design and interpretation of randomized control trial-survey of 71 negative trials. *New Engl J Med* 1978; 299: 690–4.
- Gardner DG, Altman DG. Confidence intervals rather than p values: estimation rather than hypothesis testing. *Br Med J* 1986; 292: 746–50.
- Goodman SN. P values, hypothesis tests, and likelihood: implications for epidemiology of a neglected historical debate. *Am J Epidemiol* 1993; 137: 485–96.
- Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidem* 1999; 10: 37–48.
- Holman CDJ, Arnold-Reed DE, de Klerk N. A psychometric experiment in causal inference to estimate evidential weights used by epidemiologists. *Epidem* 2001; 12: 246–55.
- Jorgensen LM, Sorensen TIA., Schroll M, Larsen S. *Int Journ Obes* 1995; 19: 909–15.
- Klein JP, Keiding N, Kreiner S. Graphical models for panel studies, illustrated on the data from the Framington heart study. *Stat Med* 1995; 14: 1265–90.
- Kosslyn SM. Understanding charts and graphs. *Appl Cogn Psychol* 1989; 3: 185–225.
- Kraemer HC. The robustness of common measures of 2×2 association to bias due to misclassifications. *Am Stat* 1985; 39: 286–90.
- Kraemer HC, Kazdin AE, Offord DR, Kessler RC, Jensen PS, Kupfer DJ. Coming to terms with terms of the risk. *Arch Gen Psychiat* 1997; 54: 337–43.

- Kraemer HC, Stice E, Kazdin AE, Offord DR, Kupfer DJ. How do risk factors come together? Mediators, moderators and independent, overlapping and proxy risk factors. *Am J Psychiat* 2001; 158: 848–56.
- Lauritzen SL, Wermuth N. Graphical models for associations between variables, some of which are qualitative and some of which are quantitative. *Ann Stat* 1989; 17: 31–54.
- Lieb R, Isensee B, von Sydow K, Wittchen HU. The Early Stages of Psychopathology Study (EDSP): a methodological update. *Eur Addict Res* 2000; 6: 170–82.
- Pearl J, Causality – Models, Reasoning and Inference. Cambridge: Cambridge University Press, 2000.
- Penrose R, Shadows of the Mind: A Search for the Missing Science of Consciousness. Oxford: Oxford University Press, 1994.
- Pregibon D. Goodness of link tests for generalized linear models. *Appl Stat* 1980; 29: 15–24.
- Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiol* 1992; 3: 143–55.
- Rothman KJ, Greenland S. Modern Epidemiology. Philadelphia: Lippincott Raven, 1998.
- Royall RM. The effect of sample size on the meaning of significance. *Am Statistician* 1986a; 40: 313–15.
- Royall RM. Model robust confidence intervals using maximum likelihood estimators. *Int Stat Rev* 1986b; 54: 221–6.
- Schervish MJ. P-values: what they are and what they are not. *Am Statistician* 1996; 50: 203–6.
- Smith LD, Best LA, Stubbs A, Archibald AB, Roberson-Nay R. Constructing knowledge: the roles of graphs and tables in hard and soft psychology. *Amer Psych* 2002; 57: 749–61.
- Sohn D. Statistical significance and replicability. Why the former does not presage the latter. *Theory and Psych* 1998; 8: 291–311.
- Thompson B. If statistical tests are broken/misused, what practices should supplement or replace them? *Theory and Psychol* 1999; 9: 165–81.
- Wittchen HU, Perkonig A, Lachner G, Nelson CB. Early developmental stages of psychopathology study (EDSP). Objectives and design. *Eur Add Res* 1998; 4: 18–27.