# Multivariate statistical models for categorical epidemiological data

**Svend Kreiner**

**Dept. of Biostatistics, University of Copenhagen**

1) **Recursive models**
2) **Chain graph models**
3) **Loglinear models**
4) **Confounding and effect modification**

# Example

**A panel study of health in Copenhagen County with data collected at ages 40, 45, 51 and 60.**

**Sickness behaviour:**

- **Number of weeks being unable to take care of responsibilities**

**Self-reported health:**

- **Overall evaluation**
- **Tiredness, Headache, Stomach pains, Bronchitis**

**Clinical results:** **BMI & Blood pressure**

**Social habits:** **Smoking, Alcohol consumption and Physical Activity**

**Social problems:** **Work, Economy, House, Family, Personal**

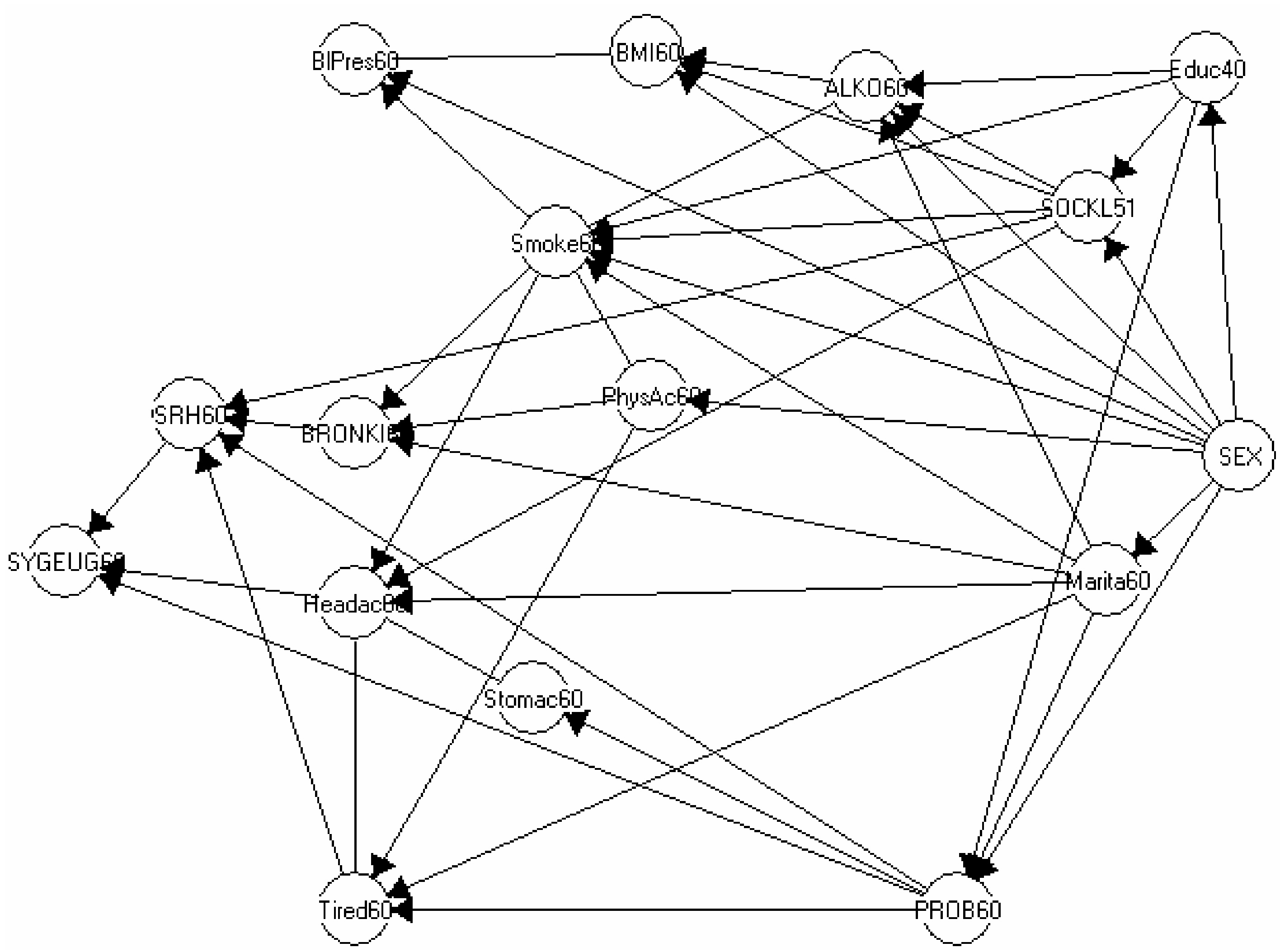
**Socio-demographic variables:** **Marital Status, Social Class, Sex**

## The problem :

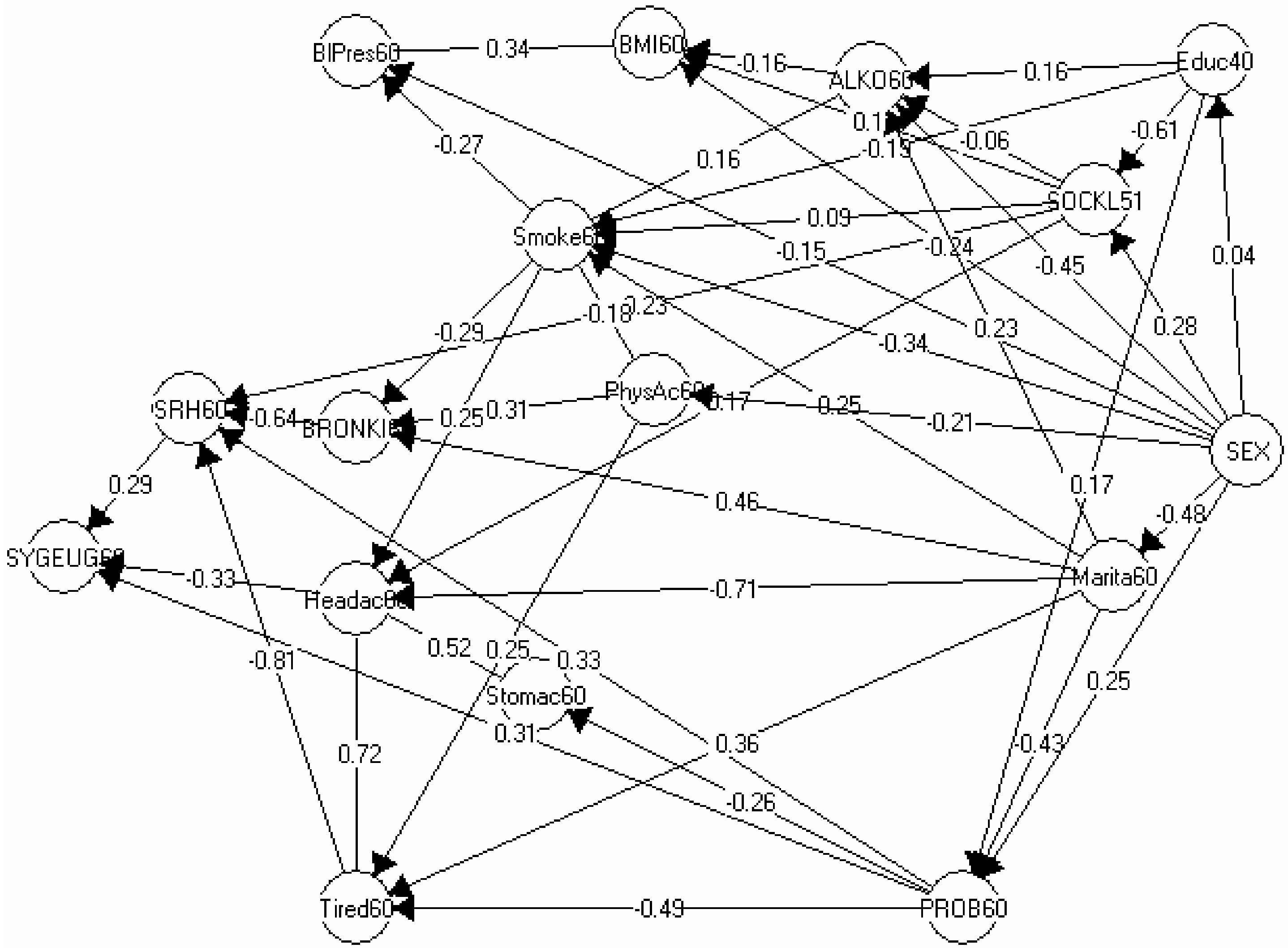
## What are the causal factors shaping SRH?

## The example: Analysis of data from the 60-year Glostrup survey

### The chain graph model



## ***Partial Gamma coefficients measuring the strength of the association***



# Gamma coefficients

**Rank correlations for dichotomous and ordinal categorical variables.**

**Similar to Kendal's rank correlation**

**The gamma coefficient for 2×2 tables is a function of the odds ratio**

$$\gamma = \frac{1 - \mathrm{OR}}{1 + \mathrm{OR}} \qquad \mathrm{OR} = \frac{1+\gamma}{1-\gamma}$$

**Partial γ coefficients are weighted means of γ coefficients in different strata of multivariate tables (similar to the MH estimate)**

# Recursive models

**Random variables : A, B, C**

**Nonrandom design variables : K**

**The joint conditional distribution**

$$P(A,B,C|K)$$

**may be rewritten as a product of conditional distribution in several different ways:**

$$P(A,B,C|K) = P(A|B,C,K)\ P(B|C,K)\ P(C|K)$$

$$P(A,B,C|K) = P(B|A,C,K)\ P(A|C,K)\ P(C|K)$$

$$P(A,B,C|K) = P(B|C,A,K)\ P(C|A,K)\ P(A|K)$$

• • •

$$P(A,B,C|K) = P(C|B,A,K)\ P(B|A,K)\ P(A|K)$$

**A recursive model is a model where the order of the variables has substantial meaning.**

$$C \leftarrow A \leftarrow B \leftarrow K$$

$$\Leftrightarrow$$

$$P(A,B,C,D|K) = P(C|A,B,K)\ P(A|B,K)\ P(D|B,K)\ P(B|K)$$

**Block-recursive models are recursive models where A, B and/or C are multidimensional vectors that cannot be partitioned in any meaningful way.**

**The models are called DAGs if A,B, and C are univariate.**

**DAG = Directed Acyclic Graph**

**Recursive modelling fits regression models to each of the separate factors of the model.**

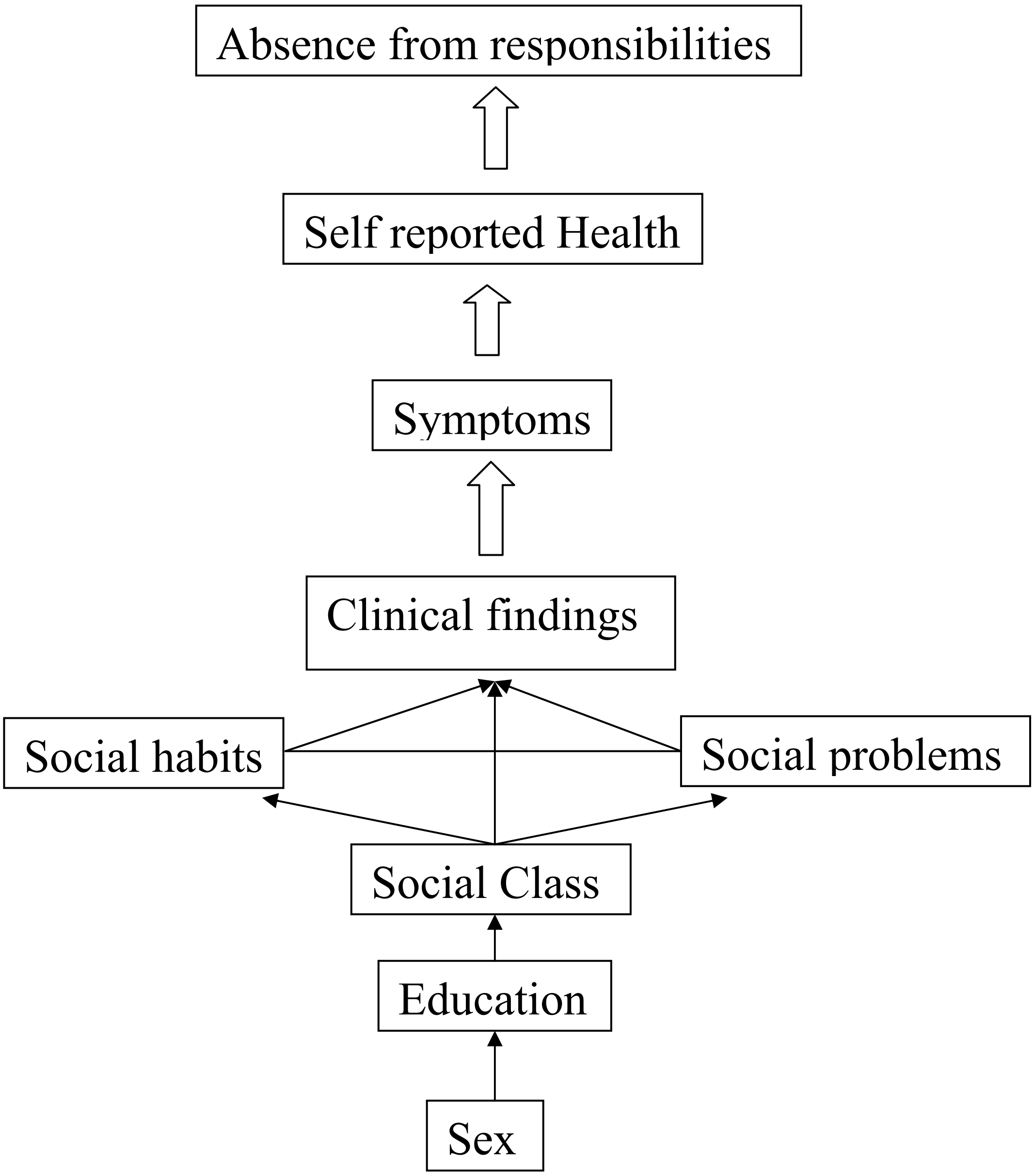
**Conventional epidemiological models are recursive:**

**P(Outcomes,Exposures,Confounders)**

**=**

**P(Outcomes|Exp,Conf)P(Exp|Conf) P(Conf)**

## Initial causal framework for the analysis of the Glostrup data on self-reported health




What are the causal factors influencing SRH?
Can the analysis support causal claims?

**What kind of statistical model should we use?**

# Graphical models

**Three different types of graphical models:**

1) **Conventional graphical models of joint distributions where all variables are assumed to be on the same footing. (Symmetrical relationships).**

2) **Graphical regression models distinguishing between dependent variables and independent explanatory variables.**

3) **Chain graph models are block recursive graphical models with variables in a series of recursive blocks and where each component is a graphical regression model).**

# Graphical models for symmetrical relationships

**The joint distribution of a multivariate set of variables**

***Definition 1.***

**A graphical model is defined by a set of assumptions stating that certain pairs of variables are conditionally independent given the remaining variables of the model.**

**We write $X \perp\!\!\!\perp Y$ when X and Y are assumed to be independent and $X \perp\!\!\!\perp Y \mid Z$ when X and Y are conditional independent given Z in the sense that**

$$P(X,Y|Z) = P(X|Z)P(Y|Z)$$

**or equivalently**

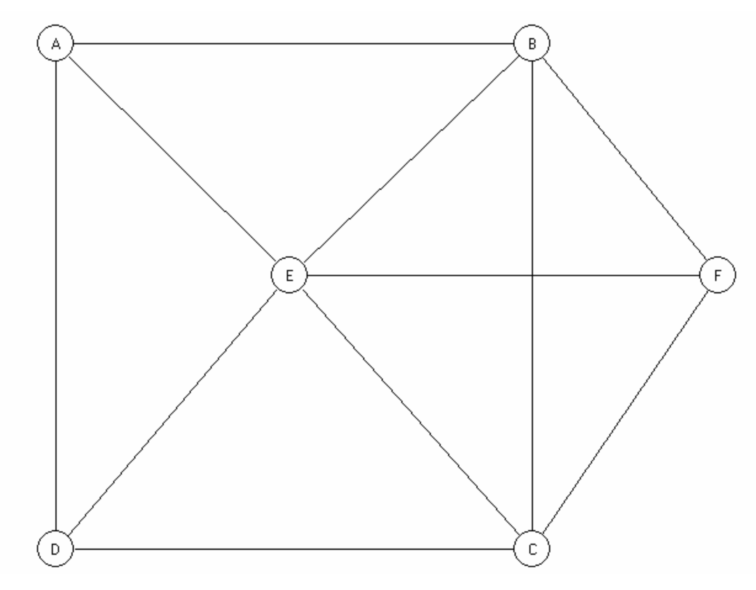
$$P(X|Y,Z) = P(X|Z).$$

## *Example 1.*

**The following four assumptions define a graphical model for variables A, B, C, D, E and F:**

$$A \perp\!\!\!\perp C \mid B,D,E,F \qquad A \perp\!\!\!\perp F \mid B,C,D,E$$
$$B \perp\!\!\!\perp D \mid A,C,E,F \qquad D \perp\!\!\!\perp F \mid A,B,C,E$$

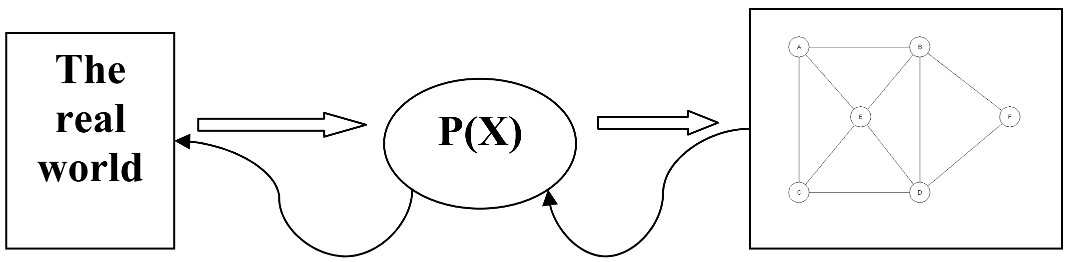
**The model is called a *graphical* model because the assumptions of conditional independence are encoded in a mathematical graph, a set of nodes and edges between nodes, as shown below.**




**A missing edge in the graph means that we assume that the variables are conditionally indepedent**

**Graphs defining graphical models are called independence graphs, interaction graphs or Markov graphs.**

***Markov graphs are second order mathematical models with properties corresponding to properties of the statistical model.***




**⇒ : Modelling  ← : Inference and interpretation**

**The main properties of Markov graphs for discrete data can be summarised in the following way:**

***a)*** ***Graphical models are closed under marginalization and conditioning. Marginal and conditional models may be derived from the graph***

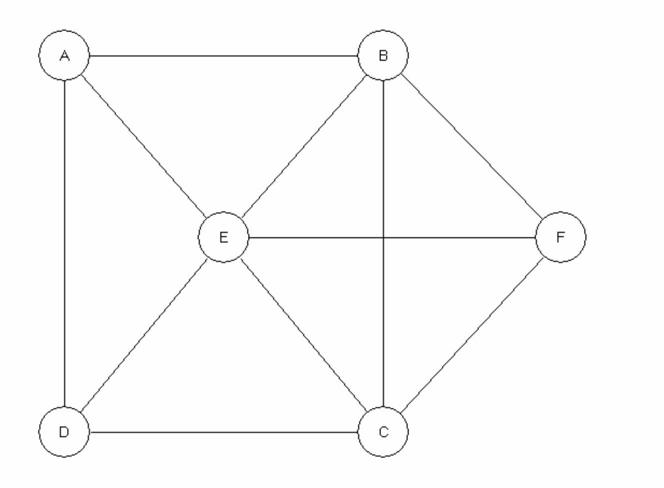
***b) Graphical models for discrete variables are loglinear with generators defined by the Markov graph***

***c) Global Markov properties: Separation implies conditional independence***

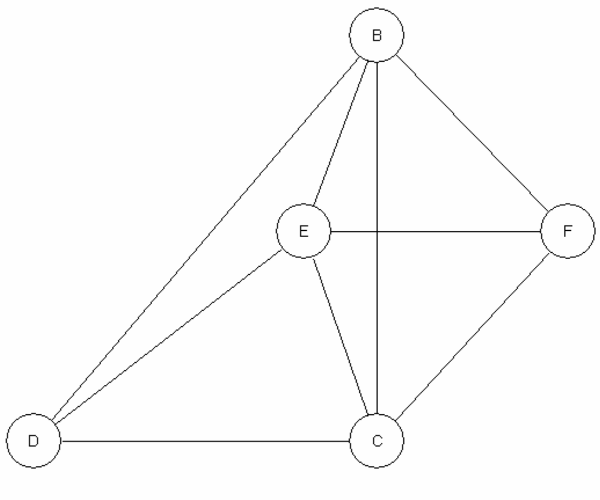
***d) Separation implies parametric collapsibility***

***e) Decomposition implies collapsibility of results from likelihood inference***

**Marginalizing over variables of a graphical model always leads to a new graphical model. Some unconnected variables will be connected in the independence graph of the marginal model.**



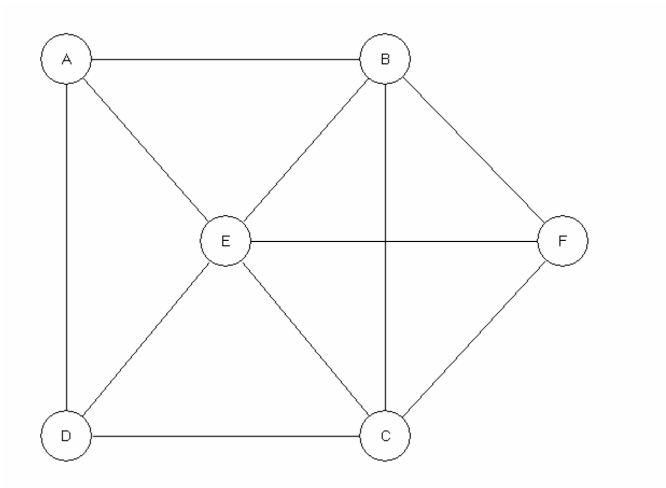

**P(B,C,D,E,F)**



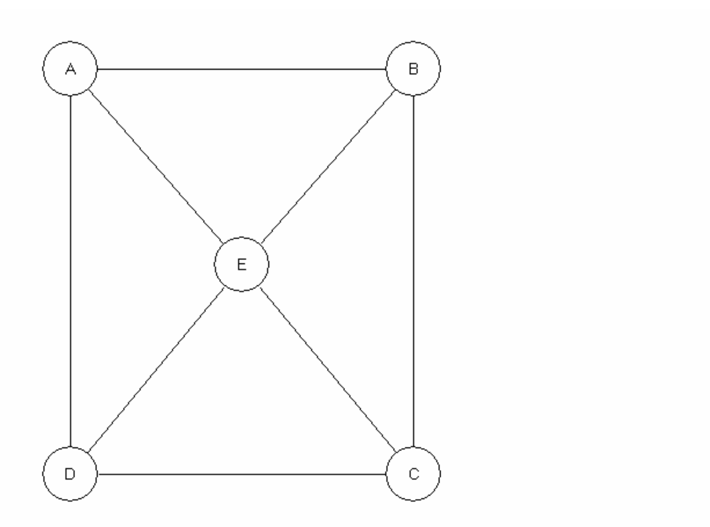

***The conditional distribution of a subset, of the variables given the remaining variables will be a graphical model.***

**Let V be the complete set of variables of the model and $V_1 \subset V$.**

**The independence graph of $P(V_1 \mid V \setminus V_1)$ is equal to the subgraph of $V_1$.**

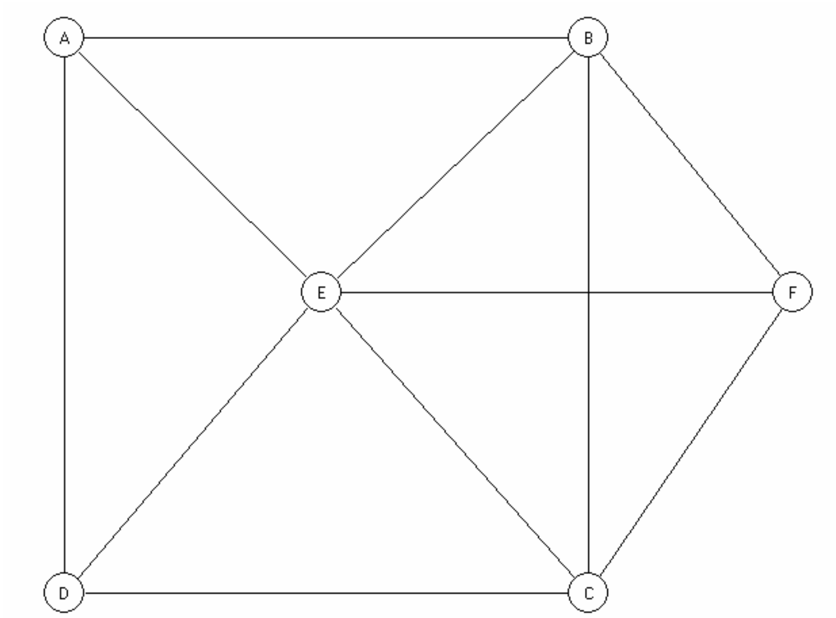



**P(A,B,C,D,E|F)**



# Graphical models for discrete data are loglinear

**The loglinear generators correspond to the cliques of the independence graphs.**




**Cliques/Generators: ABE, ADE, CDE, BCEF.**

$$P(a,b,c,d,e,f) =$$
$$\lambda_0$$
$$+\lambda_{ab}^{AB} + \lambda_{ad}^{AD} + \lambda_{ae}^{AE} + \lambda_{bc}^{BC} + \lambda_{be}^{BE} + \lambda_{bf}^{BF} + \lambda_{cd}^{CD} + \lambda_{ce}^{CE} + \lambda_{cf}^{CF} + \lambda_{de}^{DE} + \lambda_{ef}^{EF}$$
$$+\lambda_{abe}^{ABE} + \lambda_{ade}^{ADE} + \lambda_{cde}^{CDE} + \lambda_{bce}^{BCE} + \lambda_{bcf}^{BCF} + \lambda_{bef}^{BEF} + \lambda_{cef}^{CEF}$$
$$+\lambda_{bcef}^{BCEF}$$

# Properties of loglinear models

**The strength of association is measured by matrices of log-odds ratios which, given the appropriate parameterization[1], correspond to the parameters of the model.**

$$P(a,b,c,d,e,f) =$$
$$\lambda_0$$
$$+\lambda_{ab}^{AB} + \lambda_{ad}^{AD} + \lambda_{ae}^{AE} + \lambda_{bc}^{BC} + \lambda_{be}^{BE} + \lambda_{bf}^{BF} + \lambda_{cd}^{CD} + \lambda_{ce}^{CE} + \lambda_{cf}^{CF} + \lambda_{de}^{DE} + \lambda_{ef}^{EF}$$
$$+\lambda_{cef}^{CEF}$$

## The AB- association

| | A=1 | A=2 | A=3 |
|---|---|---|---|
| B=1 | 0 | 0 | 0 |
| B=2 | 0 | $\lambda_{22}^{AB}$ | $\lambda_{32}^{AB}$ |
| B=3 | 0 | $\lambda_{23}^{AB}$ | $\lambda_{33}^{AB}$ |

**Exp($\lambda_{ij}^{AB}$) = log-odds-ratio in the conditional distribution**

**P(A,B | C,D,E,F,A∈{0,i},B∈{0,j})**

---

[1] $\lambda_{abcd}$ = 0 if a=1, b=1, c=1 and/or d=1

# Three-factor association implies that the strength of association depends on other variables

$$P(a,b,c,d,e,f) =$$

$$\lambda_0$$

$$+\lambda_{ab}^{AB} + \lambda_{ad}^{AD} + \lambda_{ae}^{AE} + \lambda_{bc}^{BC} + \lambda_{be}^{BE} + \lambda_{bf}^{BF} + \lambda_{cd}^{CD} + \lambda_{ce}^{CE} + \lambda_{cf}^{CF} + \lambda_{de}^{DE} + \lambda_{ef}^{EF}$$

$$+\lambda_{cef}^{CEF}$$

## The CE – association given F=f

| | C=1 | C=2 | C=3 |
|---|---|---|---|
| **E=1** | **0** | **0** | **0** |
| **E=2** | **0** | $\lambda_{22}^{CE} + \lambda_{22f}^{CEF}$ | $\lambda_{32}^{CE} + \lambda_{32f}^{CEF}$ |
| **E=3** | **0** | $\lambda_{23}^{CE} + \lambda_{23f}^{CEF}$ | $\lambda_{33}^{CE} + \lambda_{33f}^{CEF}$ |

# Conditioning in loglinear models

**Conditioning leads to a loglinear regression model where the parameters describing the parameters relating to the dependent variables are the same as in the unconditional distribution:**

$$P(a,b,c,d,e,f) =$$
$$\lambda_0$$
$$+\lambda_{ab}^{AB} + \lambda_{ad}^{AD} + \lambda_{ae}^{AE} + \lambda_{bc}^{BC} + \lambda_{be}^{BE} + \lambda_{bf}^{BF} + \lambda_{cd}^{CD} + \lambda_{ce}^{CE} + \lambda_{cf}^{CF} + \lambda_{de}^{DE} + \lambda_{ef}^{EF}$$
$$+\lambda_{abe}^{ABE} + \lambda_{ade}^{ADE} + \lambda_{cde}^{CDE} + \lambda_{bce}^{BCE} + \lambda_{bcf}^{BCF} + \lambda_{bef}^{BEF} + \lambda_{cef}^{CEF}$$
$$+\lambda_{bcef}^{BCEF}$$

$$P(a,b \mid c,d,e,f) =$$
$$\lambda_0$$
$$+\lambda_{ab}^{AB} + \lambda_{ad}^{AD} + \lambda_{ae}^{AE} + \lambda_{bc}^{BC} + \lambda_{be}^{BE} + \lambda_{bf}^{BF}$$
$$+\lambda_{abe}^{ABE} + \lambda_{ade}^{ADE} + \lambda_{bce}^{BCE} + \lambda_{bcf}^{BCF} + \lambda_{bef}^{BEF}$$
$$+\lambda_{bcef}^{BCEF}$$

**Generators(AB|CDEF): ABE,ADE,BCEF**

## Inference in conditional loglinear model

**The conditional loglinear model of P(A,B|CDEF) given by ABE,ADE,BCEF**
**and the unconditional loglinear model of given by ABE,ADE,BCEF,CDEF**
**are inferentially equivalent:**

**Maximum likelihood estimates and likelihood ratio tests of hypotheses concerning parameters relating to A and B are always the same in the conditional and unconditional model.**

**Estimates and tests may differ from thos obtained from the original ABE,ADE,CDE,BCEF model**

**If A is dichotomous then the parameters of the logistic regression model, P(A|BCDEF) may be estimated and tested in the loglinear ABE,ADE,BCDEF model.**

## "Confounding" and "effect modification" in loglinear models

**Marginalization over one or more variables of a loglinear model always leads to a new loglinear model.**

**If the strength of the association between two variables in the marginal model is different from the strength in the complete model, then the measure of association obtained by analysis of the marginal model is confounded by the variables not included in the model.**

**The association between two variables is not confounded by a third variable if the loglinear parameters relating to the two variables are the same in the complete and the marginal table without the third variable.**

**Confounding, as defined here, is a property of the model, not a property of the estimates.**

## “Effect modification” in loglinear models

**The strength of the association between two variables is modified by a third variable if the strength depends on the outcome of the third variable.**

**Effect modification ⇔ Higher order interaction**

**If the association between two variables are modified by a third variable, then marginalization over the third variable leads to confounding since analysis of the table without the confounder cannot provide information on the degree of effect modification.**

**Effect modification ⇒ Confounding**

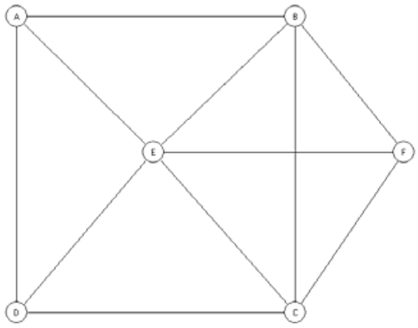
**No confounding ⇔ Parametric collapsibility**

**The inferential problem:**

**Under what conditions can we be sure that there is no confounding?**

**The answer to this question is given by the graphical structure of loglinear models.**

# Graphical models for discrete data are loglinear




**Cliques/Generators: ABE, ADE, CDE, BCEF.**

$$P(a,b,c,d,e,f) =$$

$$\lambda_0$$

$$+\lambda_{ab}^{AB} + \lambda_{ad}^{AD} + \lambda_{ae}^{AE} + \lambda_{bc}^{BC} + \lambda_{be}^{BE} + \lambda_{bf}^{BF} + \lambda_{cd}^{CD} + \lambda_{ce}^{CE} + \lambda_{cf}^{CF} + \lambda_{de}^{DE} + \lambda_{ef}^{EF}$$

$$+\lambda_{abe}^{ABE} + \lambda_{ade}^{ADE} + \lambda_{cde}^{CDE} + \lambda_{bce}^{BCE} + \lambda_{bcf}^{BCF} + \lambda_{bef}^{BEF} + \lambda_{cef}^{CEF}$$

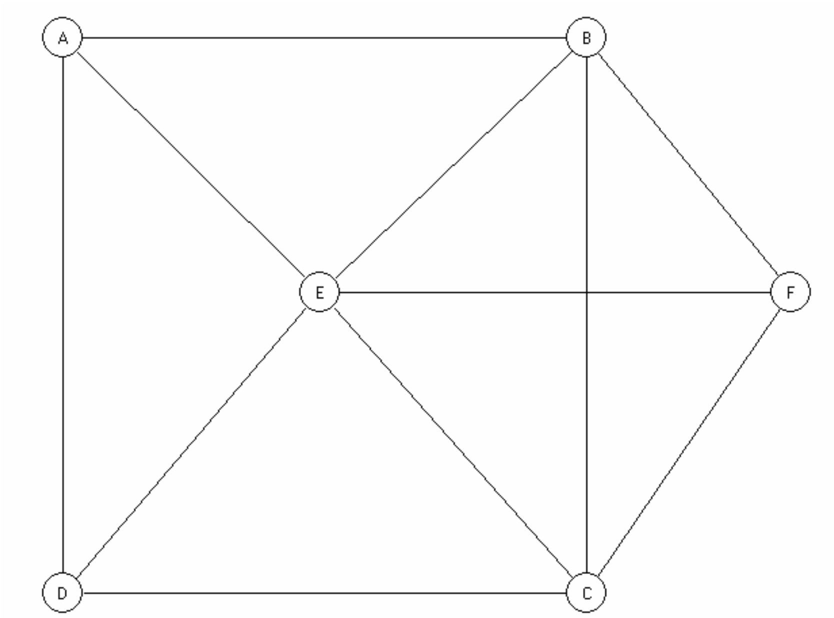
$$+\lambda_{bcef}^{BCEF}$$

***A graphical model always assumes that higher order interactions are present (effect modification) if there are cliques with more than two nodes in the graph!***

# Separation of nodes in graphs

**Two subsets of variables, U and V, are separated by a third subset if all paths from a variable in U to a variable in V goes through one or more variables in W.**

***The separation theorem (global Markov properties):***
***Separation implies conditional independence:*** **If W separates U and V in the graph then U $\perp\!\!\!\perp$ V | W.**




**D and F are separated by both (A,E,C) and (B,C,E).**

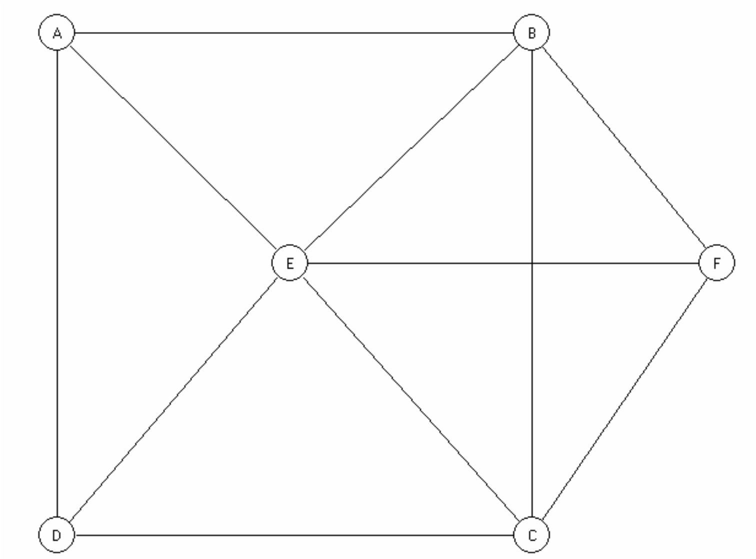
**D $\perp\!\!\!\perp$ F | A,E,C and D $\perp\!\!\!\perp$ F | B,C,E.**

## Separation implies parametric collapsibility in loglinear models.

**If all *indirect* paths between two variables, X and Y, move through at least one variable in a separating subset, S.**

$$\Rightarrow$$

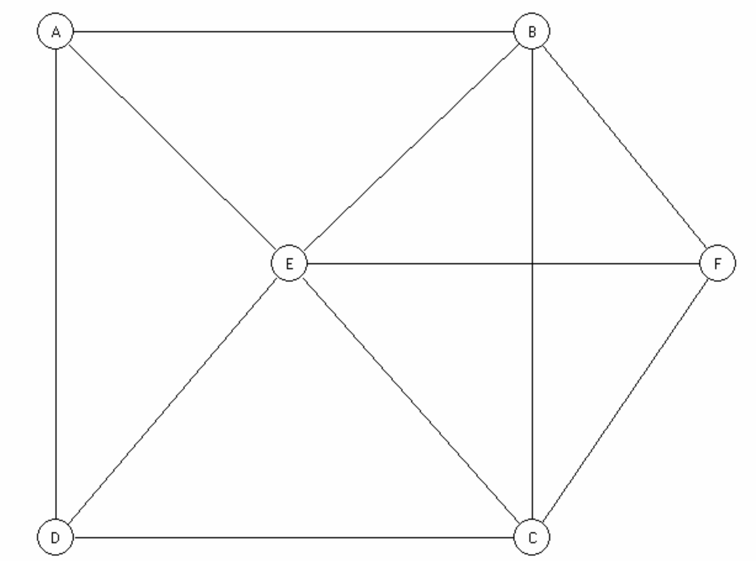
**All parameters pertaining to X and Y are the same in the complete model and in the marginal model, P(X,Y,S).**




**The loglinear association parameters relating to A and D is the same in the complete P(A,B,C,D,E,F) model and in the marginal models of P(A,D,B,E) and P(A,D,C,E) because (B,E) and (C,E) both separate A and D if there is no direct connection between these two variables.**

**Estimation of AD parameters in ADBE and ADCE is therefore not "confounded".**

# Marginal models sometimes have a simpler parametric structure than implied by the marginal graphical model.

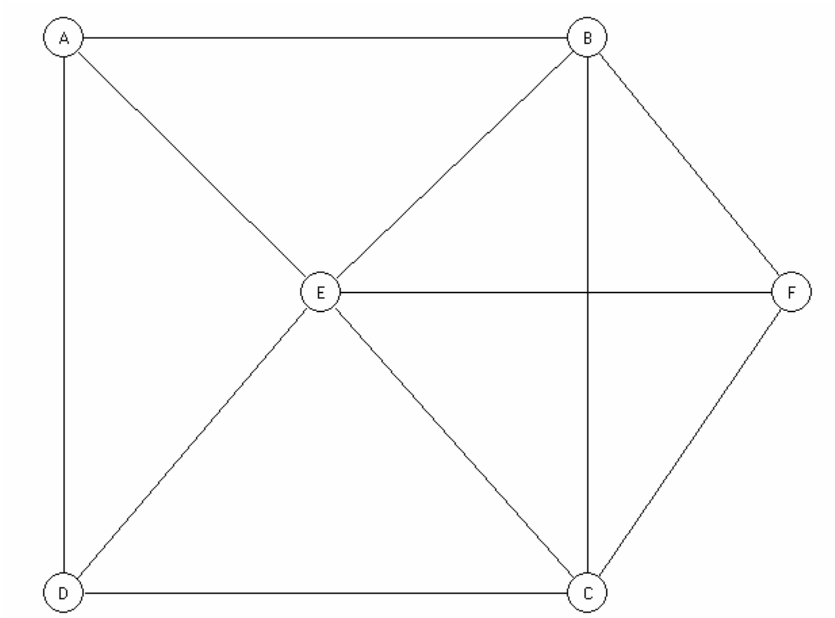



**The graphical model for P(A,D,B,E) is saturated.**

**Parametric collapsibility implies that the ADC-parameters are constant across different levels of B and C not only in the complete model, but also in the marginal model.**

**Therefore, P(A,D,B,E) is loglinear with generators: ADE,ABE,DEB.**

## Decomposition by separation of complete subsets leads to factorizations of statistical models implying collapsibility in terms of likelihood inference for certain types of models.




**The BEC clique separates (A,D) from F implying that (A,D) and F are conditional independent given (B,C,E). The joint distribution can therefore be written as**

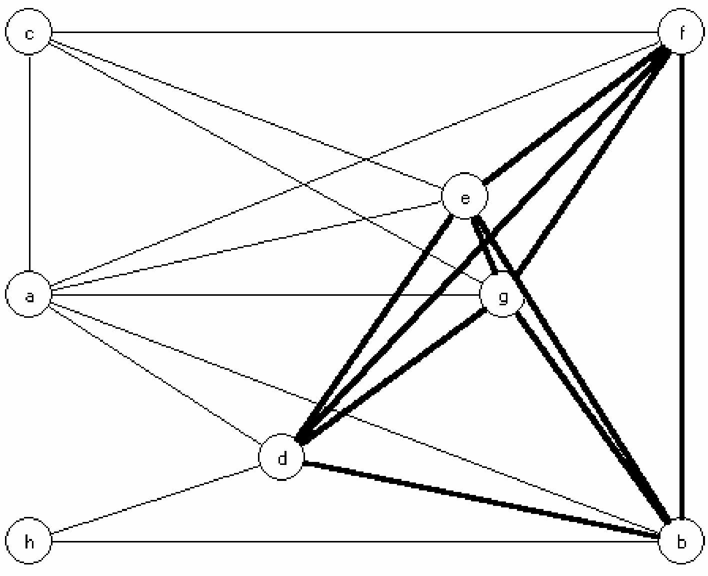
$$P(A,B,C,D,E,F)$$
$$= P(A,D \mid E,B,C)P(F \mid E,B,C)P(B,C,E)$$
$$= \frac{P(A,B,C,D,E)P(B,C,E,F)}{P(B,C,E)}$$

# Graphical regression models

A graphical regression model is a multidimensional multiple regression model, $P(Y \mid X)$ where $Y = (Y_1,..,Y_r)$, and $X = (X_1,..,X_s)$ defined by assumptions concerning conditional independence between two dependent variables or one dependent and one independent variables,

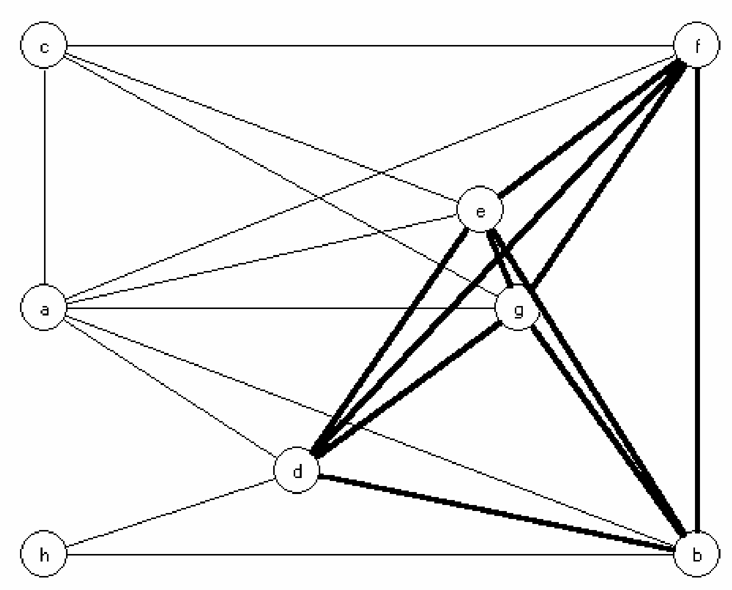
$$Y_i \perp\!\!\!\perp Y_j \mid Y_1,..,Y_{i-1},Y_{i+1},..,Y_{j-1},Y_{j+1}, .,Y_r,X_1,..,X_s$$

$$Y_i \perp\!\!\!\perp X_j \mid Y_1,..,Y_{i-1},Y_{i+1},..,Y_r,X_1,..,X_{j-1},X_{j+1},..,X_s$$




***A Markov graph for a graphical regression model of P(a,c,h|b,d,e,f,g). Edges between explanatory variables have been fixed***

**Graphical regression models are log linear with collapsibility properties defined by the Markov graphs in exactly the same way as for graphical models for symmetrical relationships.**




Loglinear model: **(abdefg),(acefg),(bdh)**

Separation: $A \perp H \mid DB$

Collapsibility onto the ACEFG table with respect to the measure of effect of F on C in terms of both parameters and estimates.

# Chain graph models

## Two sets of assumptions

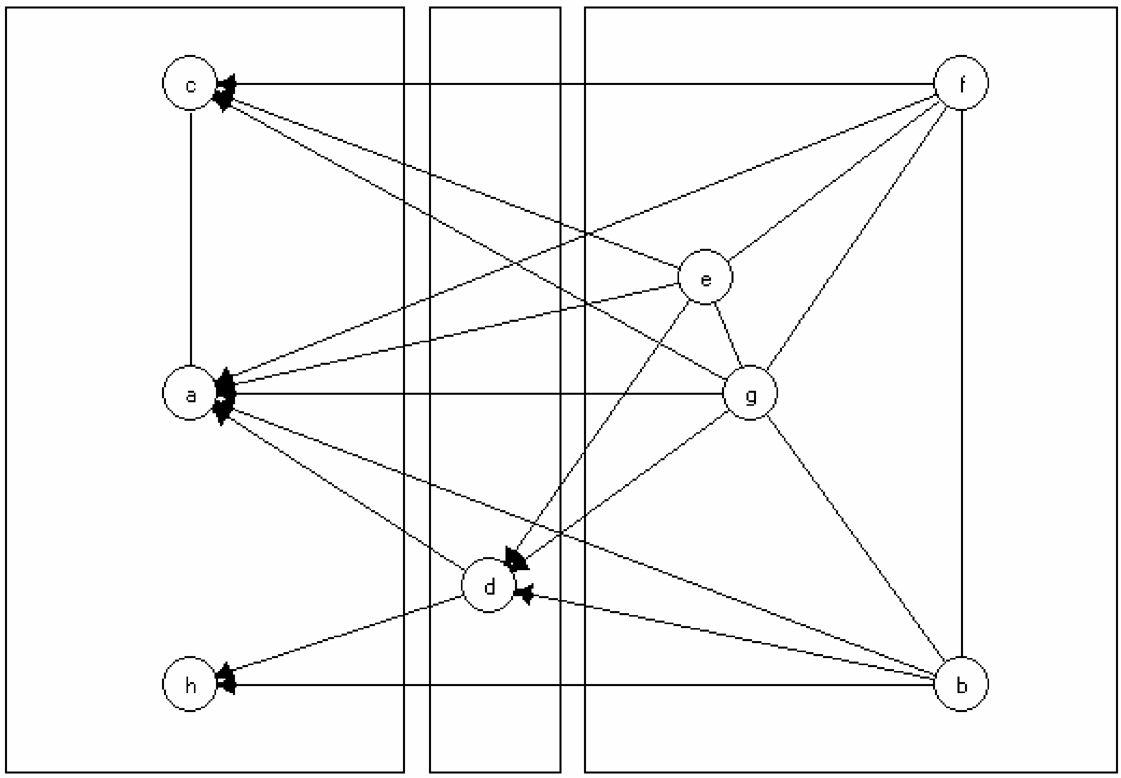
1) **Recursive structure,**

$$P(V) = \prod_{1}^{r-1} P(U_i \mid U_{i+1},..,U_r) \cdot P(U_r)$$

2) **Cond. Ind. assumptions defining graphical regression models, $P(U_i \mid U_{i+1},..,U_r)$**

**Chain graph models are characterized by Markov graphs where variables in at different recursive levels are connected by arrows instead of undirected edges.**

**Properties of chain graph models are most easily found by analysis of the Markov graphs of the separate regression components.**




*Recursive structure*:

**P(a,b,c,d,e,f,g,h) = P(a,g,h | d,b,e,f,g) · P(d | b,e,f,g) · P(b,e,f,g)**

*Conditional independence at Level 1*:

**$a \perp\!\!\!\perp h \mid b,c,d,e,f,g$    $c \perp\!\!\!\perp b \mid a,d,e,f,g,h$    $c \perp\!\!\!\perp d \mid a,b,e,f,g,h$**

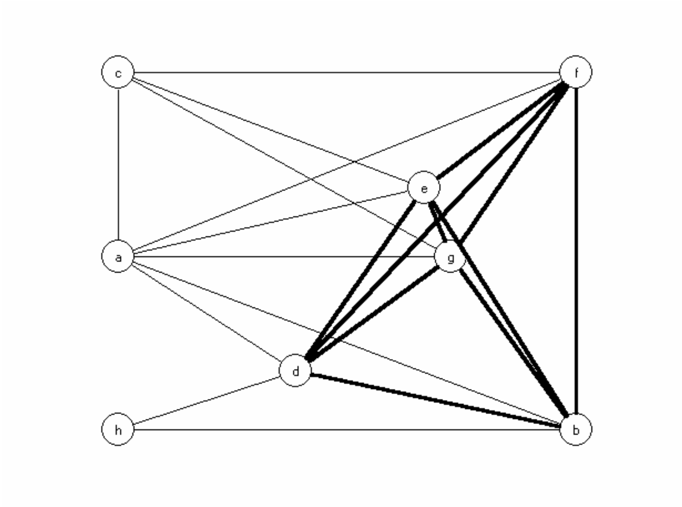
**$c \perp\!\!\!\perp h \mid a,b,d,e,f,g$    $h \perp\!\!\!\perp e \mid a,b,c,d,f,g$    $h \perp\!\!\!\perp f \mid a,b,c,d,e,g$**

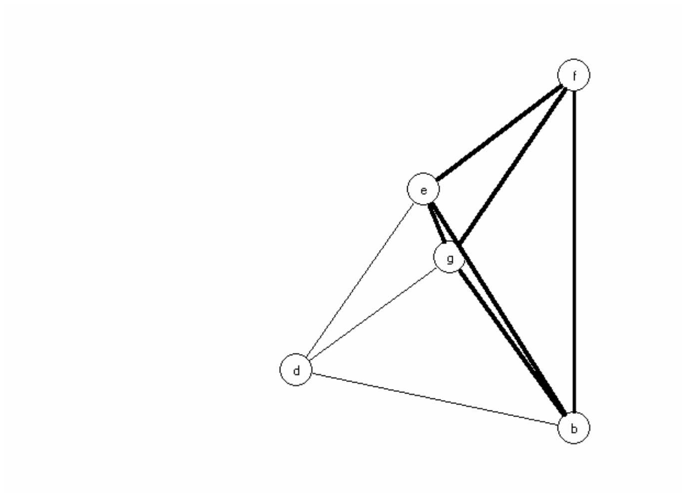
**$h \perp\!\!\!\perp g \mid a,b,c,d,e,f$**

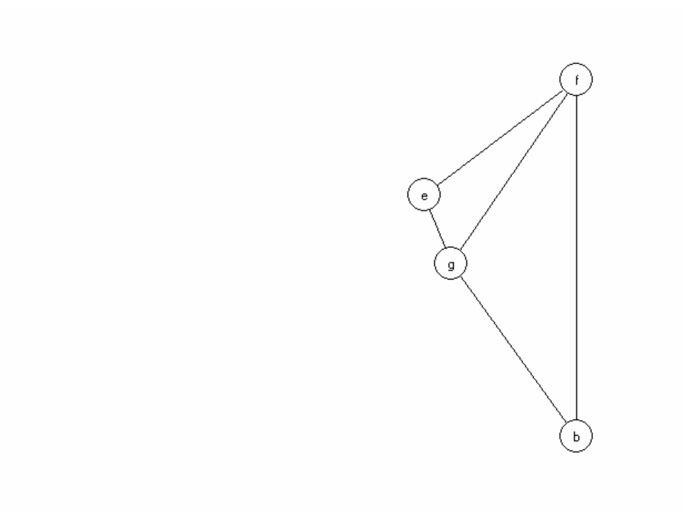
*Conditional independence at Level 2*: **$d \perp\!\!\!\perp f \mid b,e,g$**

*Conditional independence at Level 3*: **$b \perp\!\!\!\perp e \mid f,g$**

# The regression graphs



# Graphical modelling

An idealized overall strategy

**1)** An initial analysis of data (screening) aimed at formulation of a complete base model that you may use as a starting point for your analysis.

**2)** Further specification and simplification of this model using appropriate exploratory model search strategies.

The above two steps should only address secondary problems of model building.

**3)** The definitive analysis aimed at the substantive research problems and formulation of conclusions. This is not - except in special cases - something that you should approach in an exploratory manner.

## Remember that:

***A graphical model always assumes that higher order interactions are present (effect modification) if there are cliques with more than two nodes in the graph.***

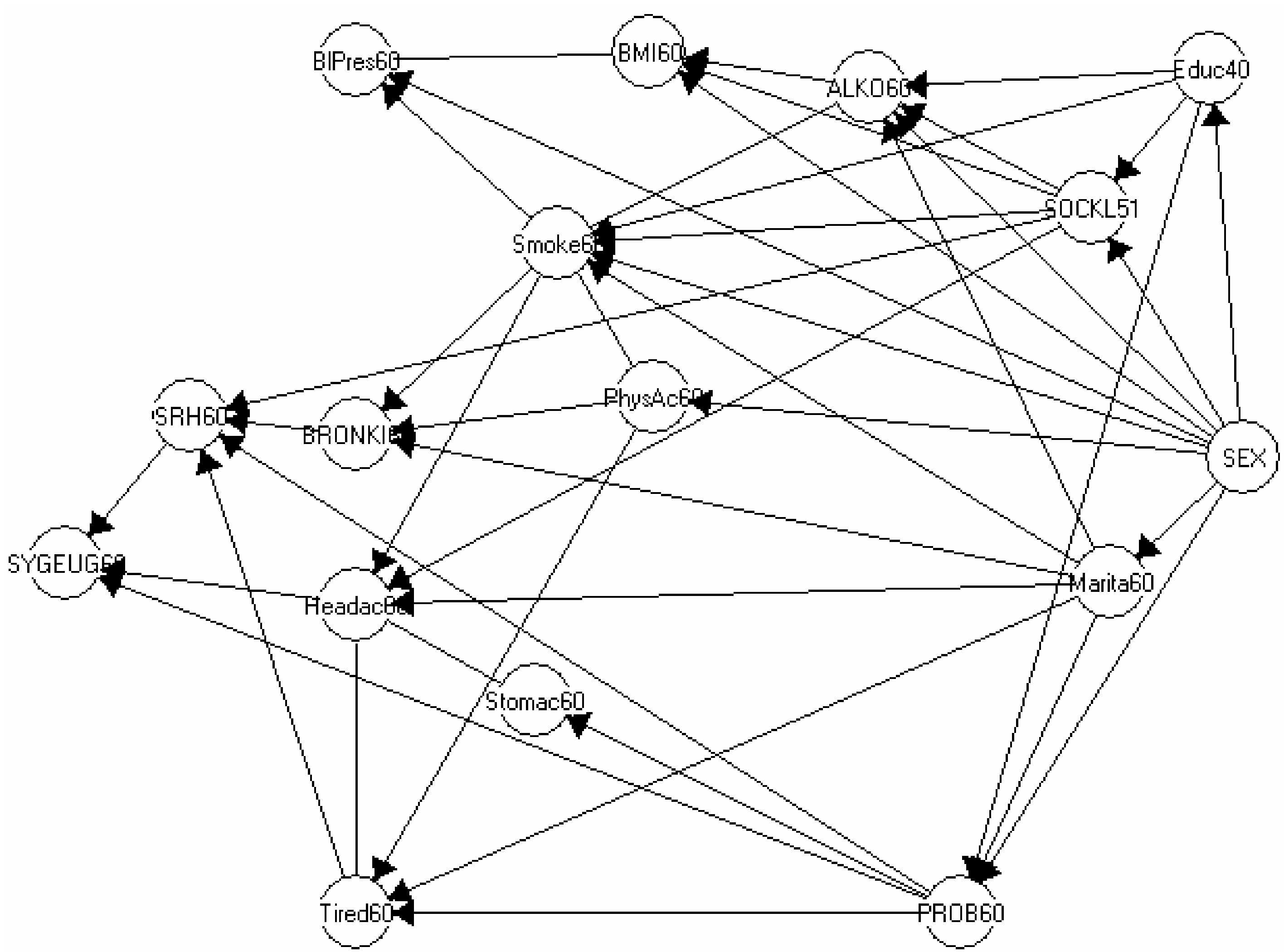
***Tests of vanishing higher order interactions (confounding rather than effect modification) can therefore not be addressed within the framework of conventional graphical models.***

***The final analysis therefore has to be in terms of loglinear rather than graphical models.***

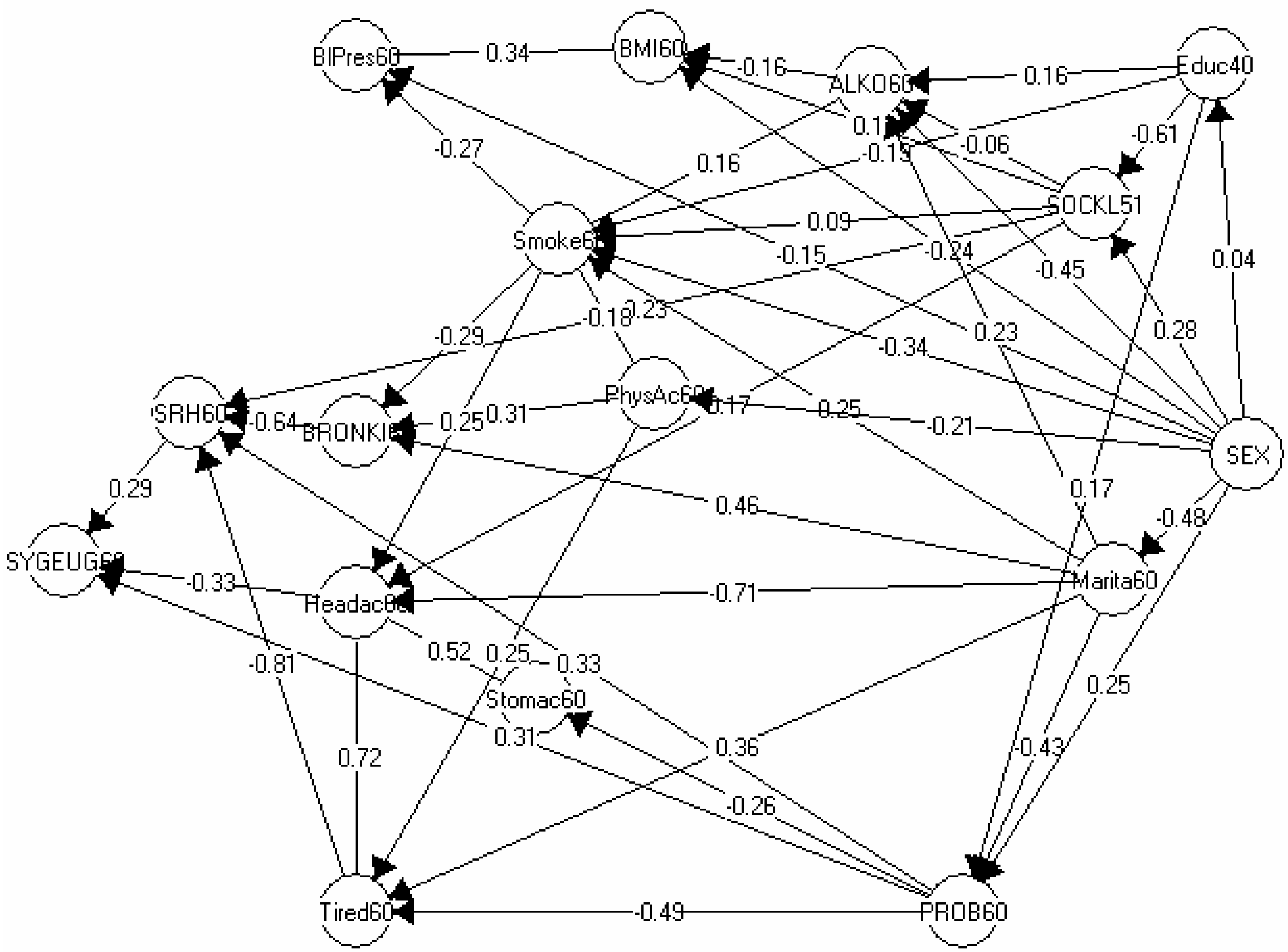
***Graphical modelling is the initial step of high-dimensional loglinear modelling.***

## *The example: Analysis of data from the 60-year Glostrup survey*

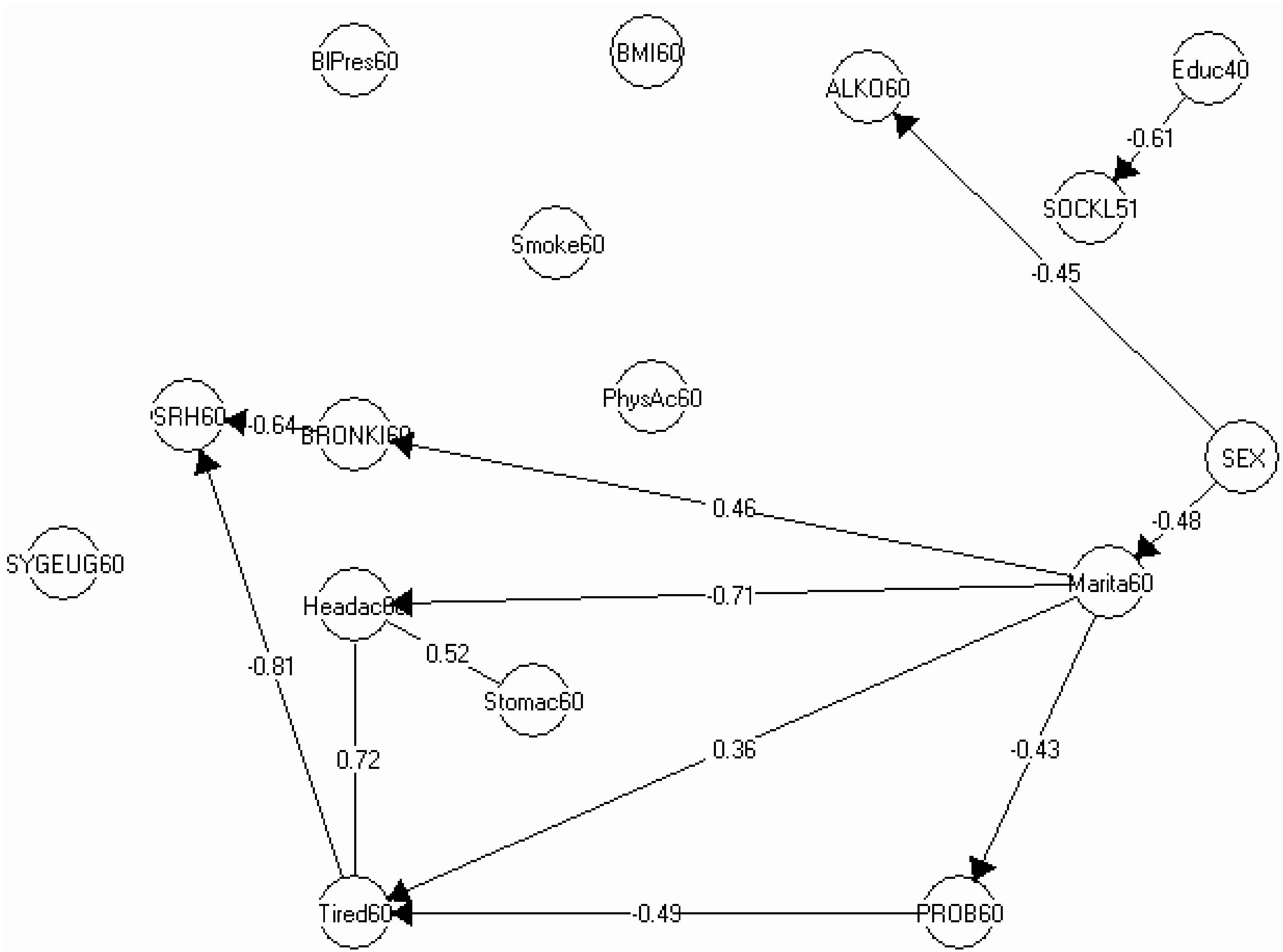
### *The graphical model*



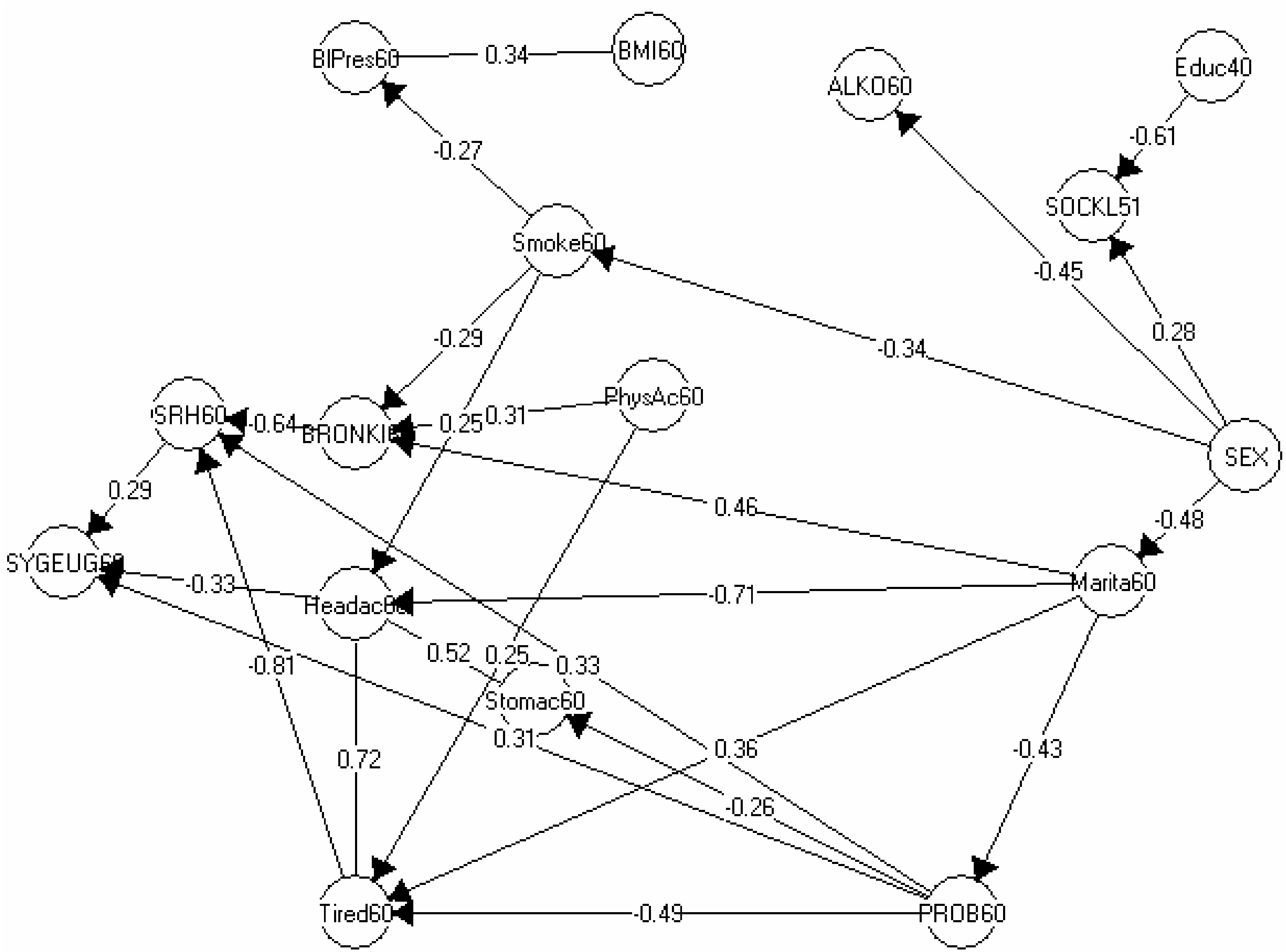
# Partial Gamma coefficient measuring the strength of the association



## ***partial gamma >0.35***



## ***Partial gamma > 0.25***



## The loglinear structure

```
LEVEL 1:  (ABEL),(BCDEFGHIJKLMNOP)
LEVEL 2:  (BCDLM),(CDEFGHIJKLMNOP)
LEVEL 3:  (CKLN),(DIKN),(EIMN),(CEN),(EF),(FL),(GHIJKLMNOP)
LEVEL 4:  (HJMP),(GHP),(GIP),(IJKLMNOP)
LEVEL 5:  (IJMNOP),(LNOP),(IKP)
LEVEL 6:  (MOP),(NP)
LEVEL 7:  (OP)
LEVEL 8:  (P)
```

**Loglinear models may be fitted but degrees of freedom are not to be trusted and the deviance in (very) large and (very) sparse tables is not chi-square distributed, anyway.**

# Analyses of relationships

1) **Collapse on marginal tables with parametric collapsibility on the parameter of interest.**

2) **The graphical model assumes that higher order interactions are present. Eliminate higher order interaction terms if at all possible.**

3) **Calculate partial gamma coefficients in fitted tables to describe the strength of association among ordinal variables.**

**Factors with direct effect on Self Reported Health (B) according to the model:**

**C: Tiredness,**

**D: Bronchitis,**

**L: Problems**

**M: Social class.**

**The model collapses on the 5-dimensional table with these variables.**

**The marginal graphical model is saturated.**

**To which degree is the effect of these variables modified by each other?**

## The Tiredness – SRH association

**The marginal association:**

| Tired60 C | veryg | good | fair | bad | TOTAL |
|---|---|---|---|---|---|
| yes | 4 | 98 | 78 | 8 | 188 |
| row% | 2.1 | 52.1 | 41.5 | 4.3 | 100.0 |
| no | 46 | 387 | 43 | 2 | 478 |
| row% | 9.6 | 81.0 | 9.0 | 0.4 | 100.0 |
| TOTAL | 50 | 485 | 121 | 10 | 666 |
| row% | 7.5 | 72.8 | 18.2 | 1.5 | 100.0 |

(B: SRH60)

$X^2 = 117.1$
df = 3
p = 0.000
Gam = -0.74
p = 0.000

## Test of conditional independence given Bronchitis, Problems and Social Class

| Hypothesis | $X^2$ | df | p-values asymp | p-values exact | 99% conf.int. | Gamma | p-values (1-sided) asymp | p-values (1-sided) exact | 99% conf.int. | nsim |
|---|---|---|---|---|---|---|---|---|---|---|
| 1:B&C\|DLM | 113.9 | 35 | 0.000 | 0.000 | 0.000 - 0.016 | -0.81 | 0.000 | 0.000 | 0.000 - 0.016 | 400 |

## Strong evidence of conditional association.

**The conditional association as measured by the partial gamma coefficient is stronger than the marginal.**

**To which degree is the association modified by DLM?**

## Partial gamma coefficients in strata defined by Problems

** Local testresults for strata defined by PROB60 (L) **

| L: | PROB60 | $X^2$ | df | p-values asympt | p-values exact | Gamma | p-values (1-sided) asympt | p-values (1-sided) exact |
|---|---|---|---|---|---|---|---|---|
| 1: few | 0 | 81.31 | 15 | 0.0000 | 0.0000 | -0.85 | 0.0000 | 0.0000 |
| 2: some | 1 | 25.83 | 10 | 0.0040 | 0.0025 | -0.85 | 0.0000 | 0.0000 |
| 3: many | 2 | 6.81 | 10 | 0.7431 | 1.0000 | 0.27 | 0.2009 | 0.1800 |

## No evidence of association for person with many problems

| L: | PROB60 | Gamma | variance | s.e. | weight | residual |
|---|---|---|---|---|---|---|
| 1: | 0 | -0.85 | 0.0069 | 0.0832 | 0.503 | -0.699 |
| 2: | 1 | -0.85 | 0.0075 | 0.0869 | 0.461 | -0.585 |
| 3: | 2 | 0.27 | 0.0993 | 0.3152 | 0.035 | 3.485 |

**Test for partial association: $X^2$ = 12.1 df = 2 p = 0.002**

**Significant evidence of effect modification**

## “Epidemiological” summary:

```
Summary of analysis of conditional relationship between
SRH60 and Tired60

 D:BRONKI60     Potential modificator - no evidence
 L:   PROB60     Potential modificator - evidence found
 M: SOCKL51     Potential modificator - no evidence


Summary statistics

Marginal Gamma (all cases)          = -0.74   n =    666
Marginal Gamma (missing excluded) = -0.71   n =    538
Partial Gamma                       = -0.81  df =     35
```

## The effect of Bronchitis:

```
Summary of analysis of conditional relationship between
SRH60 and BRONKI60

 C: Tired60     Potential modificator - no evidence
 L:  PROB60     Potential modificator - no evidence
 M: SOCKL51     Potential modificator - no evidence


Summary statistics

Marginal Gamma (all cases)         = -0.58   n =    612
Marginal Gamma (missing excluded)  = -0.62   n =    538
Partial Gamma                      = -0.64  df =     33
```

## The effect of Problems

```
Summary of analysis of conditional relationship between
SRH60 and PROB60

 C: Tired60     Potential modificator - evidence found
 D:BRONKI60     Potential modificator - no evidence
 M: SOCKL51     Potential modificator - no evidence

Global evidence of modification

Summary statistics

Marginal Gamma (all cases)         =  0.39   n =    628
Marginal Gamma (missing excluded)  =  0.40   n =    538
Partial Gamma                      =  0.33  df =     65
```

# The effect of Social Class

```
Summary of analysis of conditional relationship between
SRH60 and SOCKL51

 C: Tired60      Potential modificator - no evidence
 D:BRONKI60      Potential modificator - no evidence
 L:   PROB60     Potential modificator - evidence found


Summary statistics

Marginal Gamma (all cases)          =  0.20   n =    660
Marginal Gamma (missing excluded)   =  0.23   n =    538
Partial Gamma                       =  0.23  df =     62
```

**Test results are inconsistent:**

**The conditional distribution P(B|CDLM)**

**=**

**Either BCL,BD,BM or BCL,BLM,BD**

**The evidence suggesting that L modifies the effect of M is weak.**

# Yet another problem:

**The (partial) gamma coefficient is a non-parametric measure of association between ordinal variables. Gamma coefficients calculated in different strata can therefore not be expected to be the same even though there is no higher order interaction according to the loglinear model[2].**

**Gamma coefficients may therefore be heterogeneous even though the model is a pure 2-factor interaction model.**

**The test of BC,BD,BL,BM against BCL,BD,BM suggests that this is not what have happened here:**

$$LR = 34.1 \; df = 7, \; p = 0.000$$

[2] Except when the gamma coefficient measures association between two dichotomous variables

# Loglinear analysis of non-parametric association.

**Gamma coefficients may be calculated in fitted instead of observed tables.**

**Partial gamma coefficients measuring association between SRH (B) and Tiredness (C) in strata defined by Problems (L)**

| Problems | Observed $\gamma$ | Fitted $\gamma$ under BCL, BD,BM | Fitted $\gamma$ under BC,BL, BD,BM |
|---|---|---|---|
| No | -0.849 | -0.841 (.074) | -0.660 (.117) |
| Few | -0.846 | -0.804 (.104) | -0.697 (.148) |
| many | +0.271 | +0.368 (.298) | -0.700 (.203) |
| $\chi^2$ test of obs=fit | | $\chi^2 = 0.3$ df = 3, p = 0.97 | $\chi^2 = 26.5$ df = 3, p = 0.000 |

**Partial gamma coefficients measuring association between SRH (B) and Social Class (M) in strata defined by Problems (L)**

| Problems | Observed $\gamma$ | Fitted $\gamma$ under BCL, BD,BM |
|---|---|---|
| No | +0.232 | +0.251 (.107) |
| Few | +0.081 | +0.244 (.173) |
| many | +0.552 | +0.302 (.215) |
| Very many | -0.500 | +0.271 (.643) |
| Test of obs=fit | | $\chi^2 = 3.7$<br>df = 4, p = 0.45 |

**The evidence of heterogeneous gamma coefficients is not convincing**