22nd meeting of the International Network on Health Expectancy and the Disability Process (REVES)
Havana, Cuba,
19-21 May 2010

Comparison of learning algorithms for Bayesian Networks models: a case study using the World Health Survey data

Andreotti A*, Brogini A°, Minicuci N*

* National Research Council - Institute of Neuroscience - Padova - Italy

° University of Padova, Faculty of Statistics - Italy

BACKGROUND

AIM OF THE STUDY: to compare algorithms for learning Bayesian Networks (BN) using the World Health Survey (WHS) data, a real dataset with numerous interdependences between variables.

WORLD HEALTH SURVEY (WHS):
- World Health Organization (WHO)
- 70 countries
- between 2002 and 2004
- cross-population comparable data on health, health-related outcomes and risk factors

WORLD HEALTH SURVEY

AIM OF THE WHS: to provide valid, reliable and comparable information about the World population health status

SAMPLING DESIGN: probability sampling using multi-stage, stratified, random cluster samples

POPULATION STUDIED: persons aged 18 years and older who lived in households

QUESTIONNAIRE

- Household Questionnaire
- Individual Questionnaire

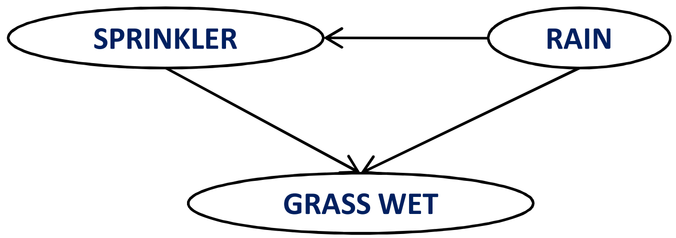
BAYESIAN NETWORK (1)

DEFINITION: BN is a Directed Acyclic Graph (DAG) whose nodes represent variables, and whose arcs describe the conditional in/dependencies between variables.

qualitative part (BN structure)

quantitative part (BN parameters)

Example of the qualitative part

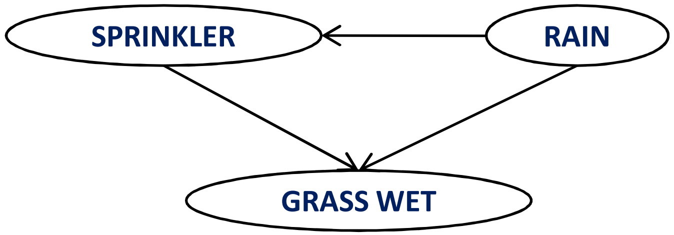


This example shows the <u>structure</u> of a BN. SPRINKLER is called a *child* of RAIN because the RAIN has a direct effect on the use of the SPRINKLER, and RAIN is called a *parent* of SPRINKLER.

BAYESIAN NETWORK (2)

Example of the quantitative part

	SPRINKLER	
RAIN	True	False
True	0.01	0.99
False	0.4	0.6



RAIN	
True	False
0.2	0.8

		GRASS WET	
SPRINKLER	RAIN	True	False
False	False	0.0	1.0
False	True	0.8	0.2
True	False	0.9	0.1
True	True	0.99	0.01

This example shows the <u>parameters</u> of a BN.
The joint probability function of this BN is:
P(RAIN,SPRINKLER,GRASS WET)=P(RAIN)P(SPRINKLER|RAIN)P(GRASS WET|SPRINKLER,RAIN)

BAYESIAN NETWORK (3)

BAYESIAN NETWORK is a graphical representation in the form of a DAG, G, for conditional in/dependencies and for compact specification of full joint distributions

G encodes the **Markov condition**: each node of the BN is probabilistically independent of its non-descendents, given its parents.

The **full joint distribution** is defined as the product of the local conditional distributions of each node of the network:

$$P(x_1,\ldots,X_n) = \prod_{i=1}^{n} P(X_i \mid \mathrm{Parents}(X_i))$$

BAYESIAN NETWORK APPLICATION (1)

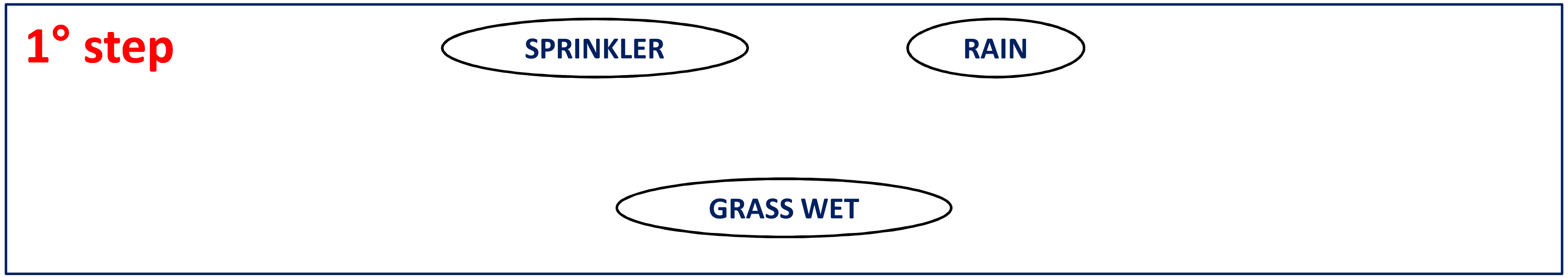
Methods for construct a Bayesian network:
(1) BN specified by an expert
(2) BN learned from data

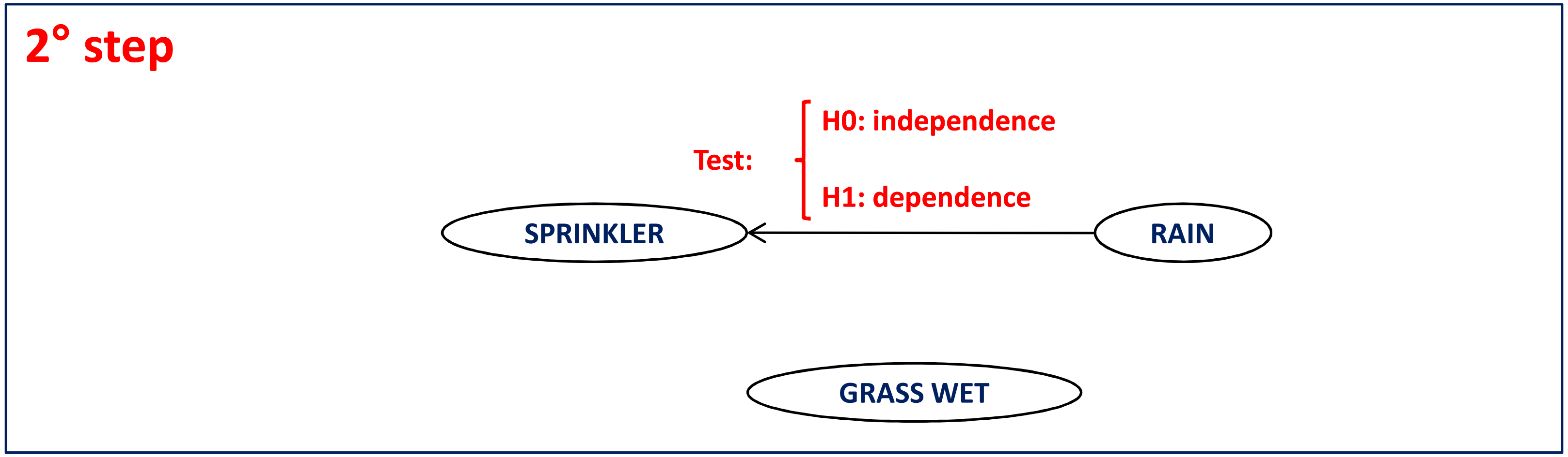
There are two primary approaches for learning a Bayesian Network from data:
(1) the constraint-based algorithms (CI), and
(2) the search and score algorithms (S&S)

The constraint-based evaluates the presence or absence of an arc by testing conditional independencies.

BAYESIAN NETWORK APPLICATION (1a)

Constraint-based method





BAYESIAN NETWORK APPLICATION (2)

- The search and score evaluates the goodness-of-fit of the network to the data maximizing a selected scoring function.

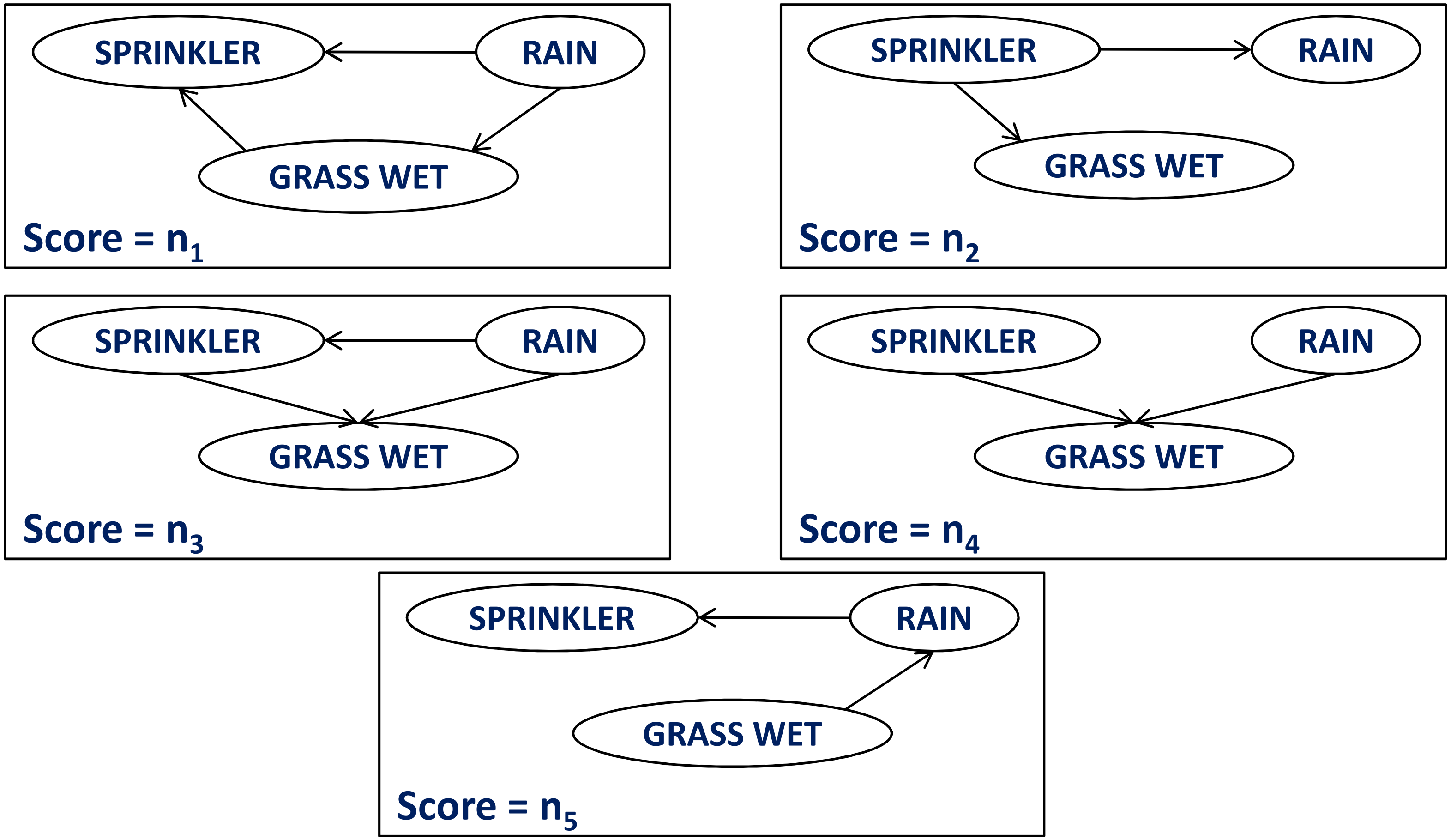
  Typologies of scoring functions :
  (a) Bayesian (based on the Bayes theorem), and
  (b) Information Theory

  The Bayesian scoring functions compute the posterior probability distribution conditioned to the data, starting from different prior probability distributions. The best network is the one that maximizes the posterior probability.

  The scoring functions based on Information Theory select the network structure that best fits the data, penalized by the number of parameters of the network.

BAYESIAN NETWORK APPLICATION (2a)

Search and Score method



BAYESIAN NETWORK APPLICATION (3)

Constraint-based approach

↓

BNPC algorithm
(BNPC software)

Search & Score approach

↓

Tabu search algorithm
(Weka software)

↙ ↘

Scoring function
BDeu
(Bayesian)

↓

With uniform prior distribution

Scoring function
MDL
(Information Theory)

CASE STUDY

- **The dataset used for the analysis contained 26,608 records from 22 countries.**

14 categorical variables:

- **5 socio-demographic characteristics**

(sex, age, marital status, education and employment)

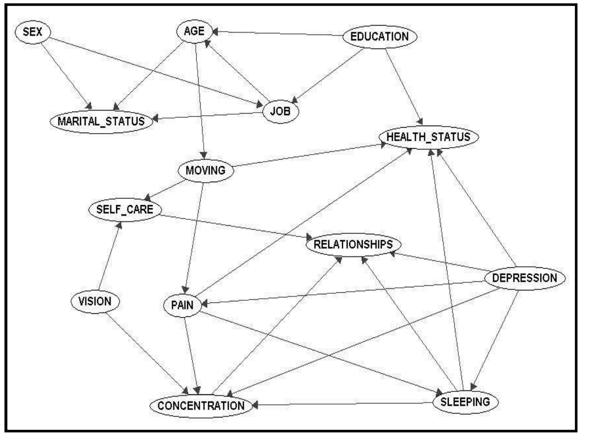
- **1 self-reported overall health status question**

- **difficulties in functioning in 8 health domains**

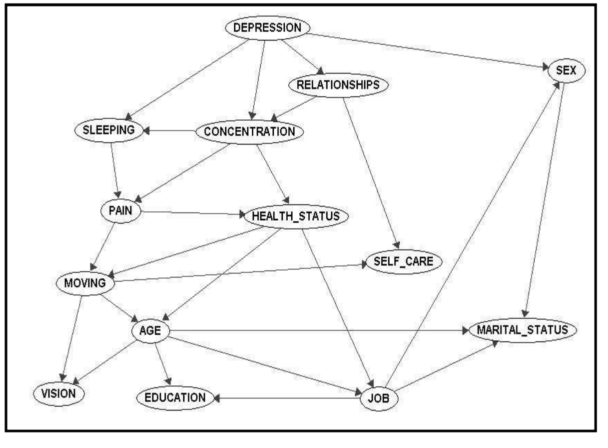
(mobility, self-care, pain and discomfort, concentration, interpersonal relationships, vision, sleeping, and feeling sad or depressed)

RESULTS (1)

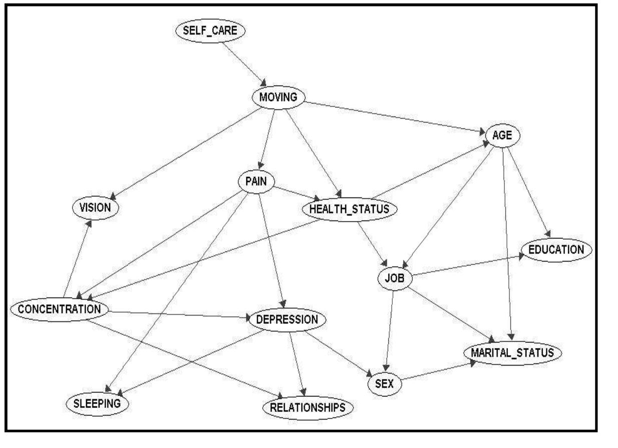
BNPC network (CI)



Tabu search network with scoring function BDeu (S&S)



Tabu search network with scoring function MDL (S&S)



RESULTS (1a)

Health status

	BNPC	Tabu search (BDeu)		Tabu search (MDL)	
	parents	parents	children	parents	children
Socio-demographic domain	**Education**		**Age** **Job**		**Age** **Job**
Psychological domain	**Depression** **Sleeping**	**Concentration**			**Concentration**
Physical domain	**Pain** **Moving**	**Pain**	**Moving**	**Pain** **Moving**	

RESULTS (2)

The principal aim of our study was to compare the different Bayesian networks, in terms of structure.

The comparison process is divided into two typologies:

The comparison of the structure of the network

The comparison of the predictive accuracy on the target variable

RESULTS (3)

Comparison of the structure

	Tabu search (BDeu) vs Tabu search (MDL)	Tabu search (BDeu) vs BNPC	Tabu search (MDL) vs BNPC
Total number of arcs	26 vs 25	26 vs 27	25 vs 27
Coincident arcs*	15	8	11
Inverted arcs**	8	12	9
Added arcs***	2	7	7
Deleted arcs****	3	6	5

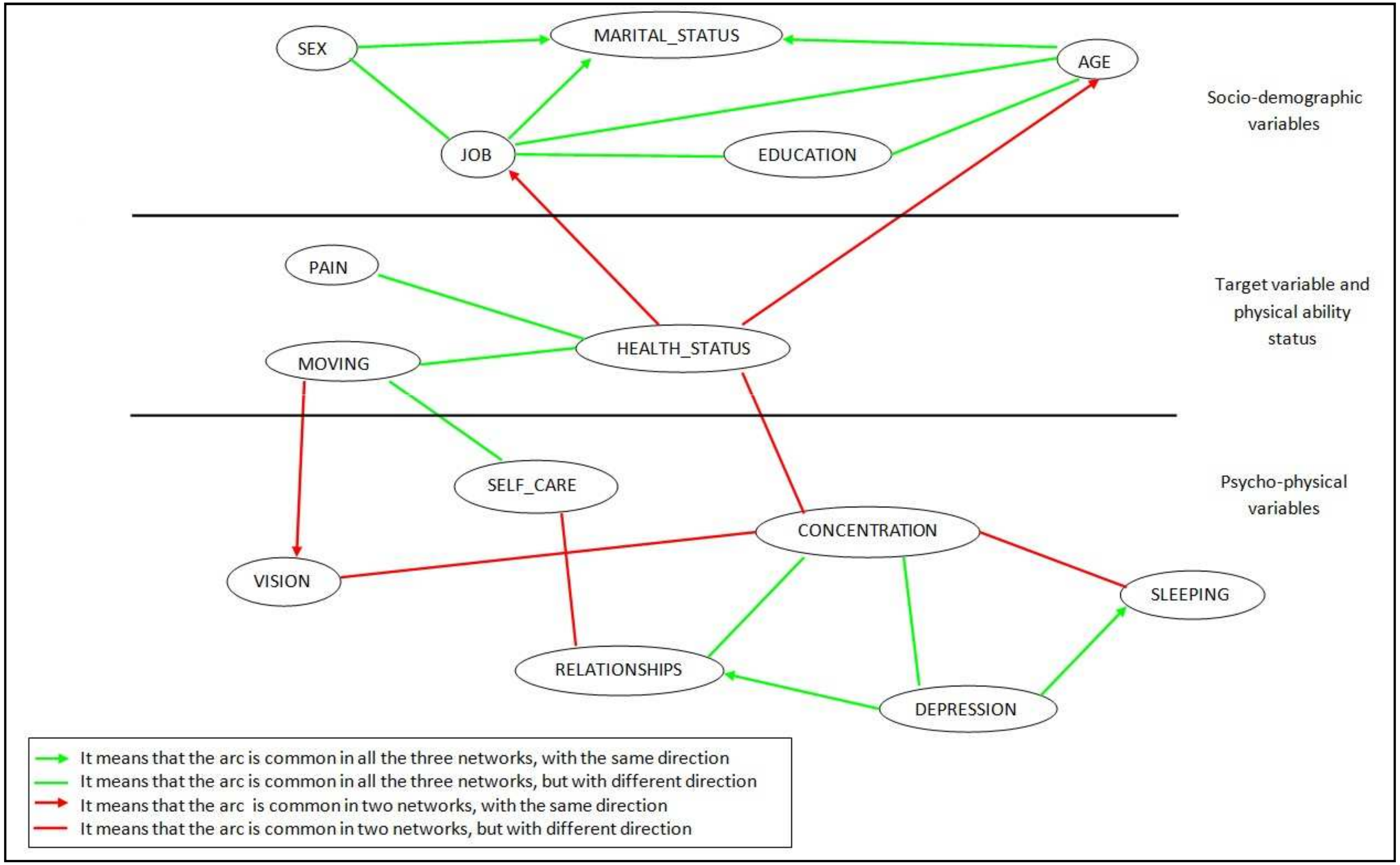
* An arc is coincident if it is present in both networks with the same direction

** An arc is inverted if it is present in both networks but with an inverted direction

*** An arc is added if it is not present in the first network but it is present in the second one

**** An arc is deleted if it is present in the first network but it is not present in the second one

RESULTS (3a)


SEX
MARITAL_STATUS
AGE
JOB
EDUCATION
Socio-demographic variables
PAIN
HEALTH_STATUS
MOVING
Target variable and physical ability status
SELF_CARE
CONCENTRATION
VISION
SLEEPING
RELATIONSHIPS
DEPRESSION
Psycho-physical variables
It means that the arc is common in all the three networks, with the same direction
It means that the arc is common in all the three networks, but with different direction
It means that the arc is common in two networks, with the same direction
It means that the arc is common in two networks, but with different direction

RESULTS (4)

Comparison of the predictive accuracy

Each BN was used to predict the most probable value of the health status variable and the comparison of the predicted with observed values produced the percentage of classification success

	Tabu search (BDeu)	Tabu search (MDL)	BNPC*
Percentage of classification success (SD)	51.77% (0.51)	51.41% (0.52)	49.72% (na**)

* The BNPC software does not support the calculation of the Standard Deviation

** Percentage of classification success was performed with HUGIN software

CONCLUSIONS

- **We have compared two different typologies of algorithms, which are based on different assumptions**

Comparison

Structure **Parameters**

- **Strength: high dimension WHS dataset**

- **Limitation: different characteristics of the software used**

REFERENCES

- Cheng J, Bell DA, Liu W (1997) *An algorithm for Bayesian belief network construction from data*. In: Proceedings of AI and STAT'97, pp. 83-90.

- Chickering DM (1996) *Learning Bayesian Networks in NP-Complete*. In: Learning from Data: Artificial Intelligence and Statistics, Fisher and Lenz Eds., Springer-Verlag, New York.

- Cooper GF, Herskovits E (1992) *A Bayesian method for the induction of probabilistic network from data*. Mach Learn 9: 309-48.

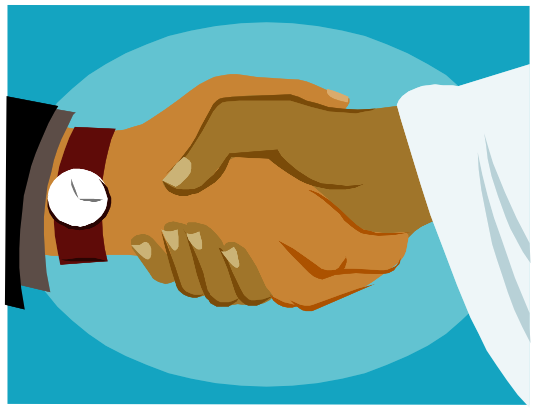
- De Campos L (2006) *A scoring function for learning Bayesian networks based on mutual information and conditional independences tests*. J Mach Learn Res; 7: 2149-87.

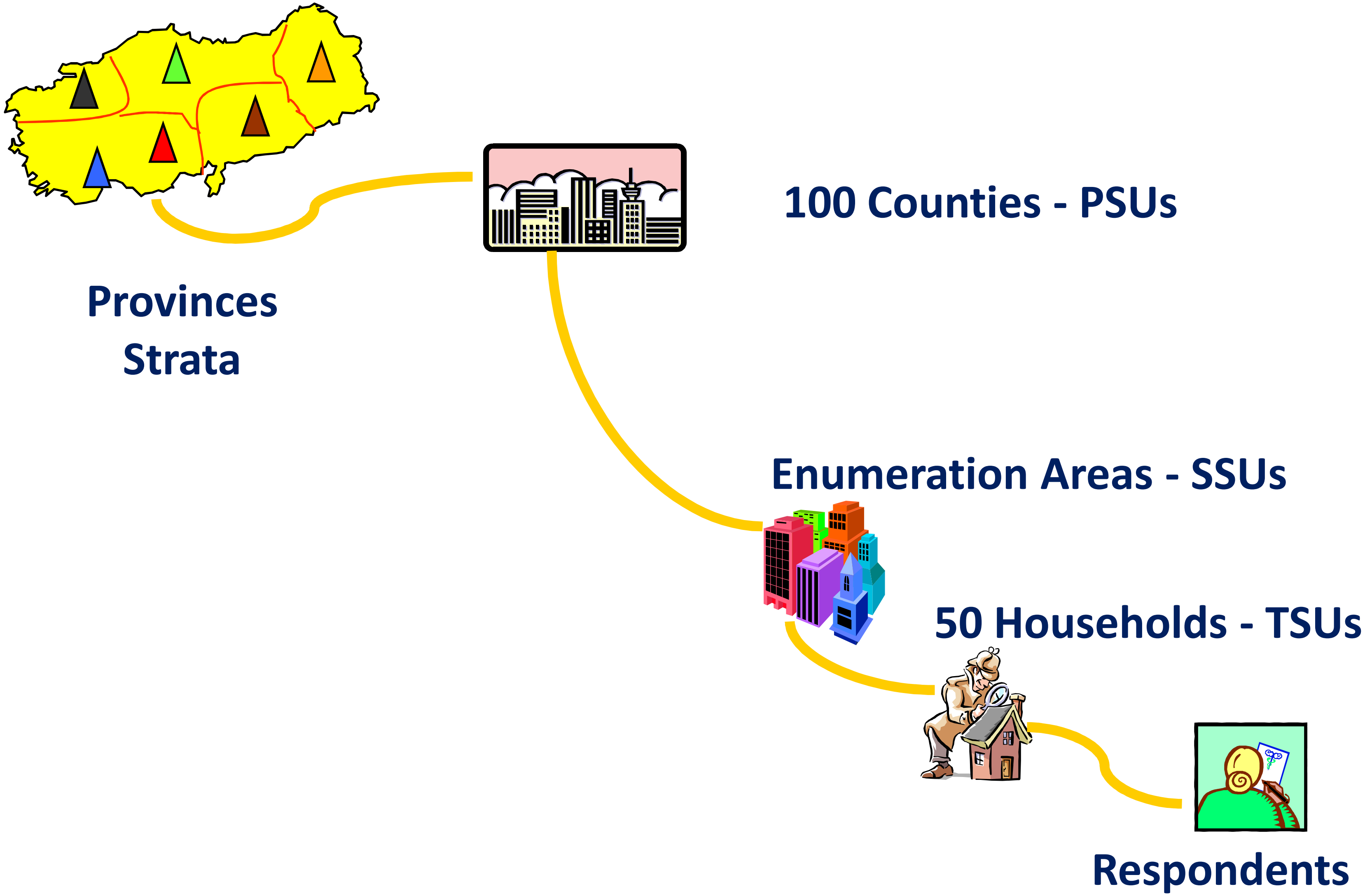
- Edwards D (2000) *Introduction to graphical modelling*. 2nd edition, Springer-Verlag.

- Heckerman D, Geiger D, Chickering DM (1995) *Learning Bayesian networks: the combination of knowledge and statistical data*. Mach Learning; 20: 197-243.

- Lauritzen SL (1996) *Graphical models*. Clarendon Press, Oxford.

Thank you very much for your attention!




100 Counties - PSUs
Provinces
Strata
Enumeration Areas - SSUs
50 Households - TSUs
Respondents

Bayesian scoring functions (based on Bayes theorem)

$$P(G|D) \sim P(G) \quad P(D|G)$$

posterior prior likelihood

Scoring functions based on Information theory

$$MDL(G|D) = LL_D(G) - \frac{1}{2} C(G) \log(N)$$

score log-likelihood network complexity

K-fold cross validation

The overall dataset is randomly partitioned into 10 subsets of approximately equal size. Of the 10 subsamples, a single subsample is retained as the validation data for testing the model (the test set), and the remaining nine subsamples are used as training data (the training set). This process is then repeated 10 times.

Simple cross validation

The overall dataset is split into two subsets: one training set and one test set.