Candidate Genes,

Gene Finding,

SNP Associations

Mixture De-Convolution

S. Stanley Young

GlaxoSmithKline.

Outline

1. Pharmacogenetics
2. Problem statement
3. Statistical methods
4. Software
5. Questions

Why Pharmacogenetics?

Social & Business Drivers

- Healthcare costs
- Adverse drug reactions
- Drug development costs
- Individualized healthcare management

Why Pharmacogenetics?

Knowledge and technology converge

- Human Genome Project(s)
- Sequencing technologies
- Informatics

Why Pharmacogenetics?

- Failed clinical trials
- 100,000 ADR-related deaths per year in U.S. (*JAMA* April 14, 1998)
- Even blockbuster drugs render only 50-70% efficacy (*WSJ* April 16, 1999)
- Genetic insight needed on clinical outcomes
- Allows creation of customized drugs

Three Analysis Problems

1. Candidate gene analysis.

2. SNP associations.

3. SNPs and gene finding.

Pharmacogenetics today

Will genes really predict medicine response?

- Drug metabolizing enzymes
- Drug targets: Enzymes, receptors, transporters

Candidate Genes

Goal: Discover genotype-phenotype relationships

- Single gene effects
- Multiple gene effects
- Multiple, interacting gene effects
- Genotype-environmental interactions.

Method : Direct testing of association!

Why are we failing?

1. Sample size
2. Map not dense enough
3. Gene/gene interactions
4. Gene/environment interactions
5. Multiple mechanisms
6. Statistical methods

What are the problems?

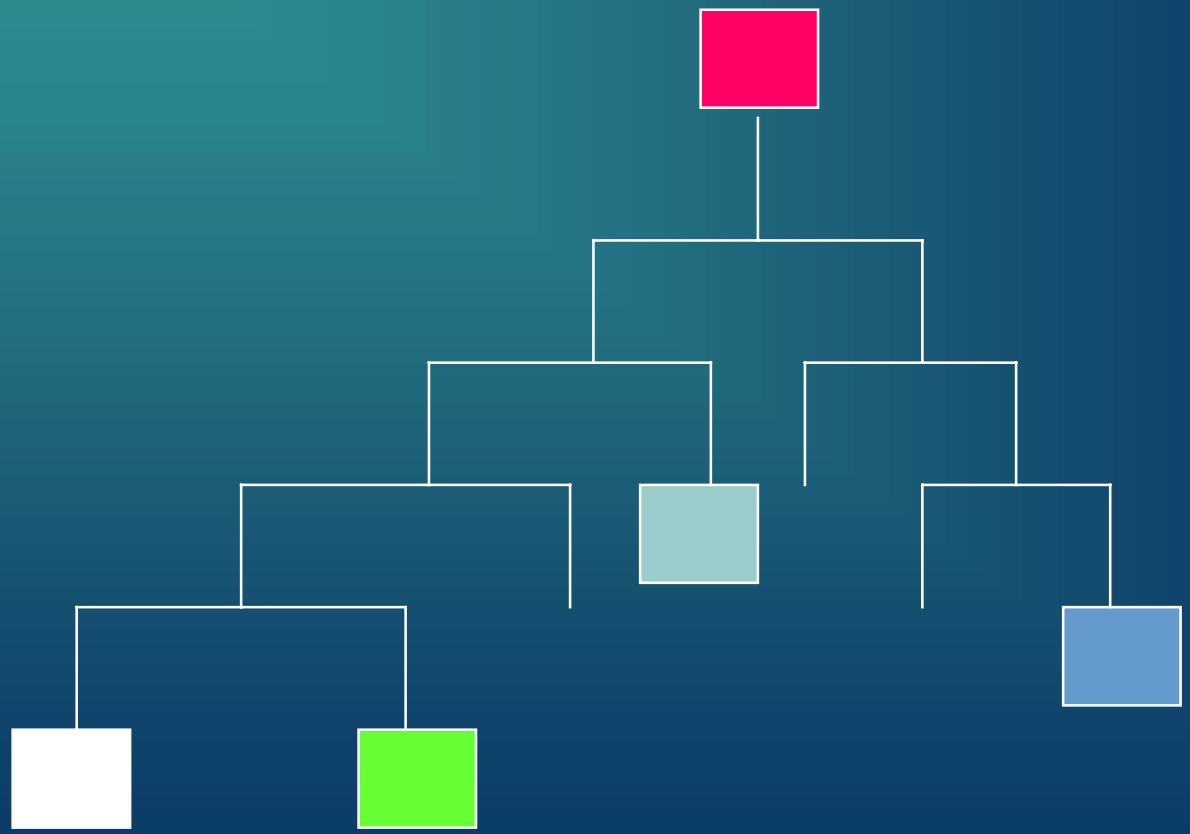
Phenotype is typically a

Mixture of Mechanisms.

Different genes may
give rise to different mechanisms.

Data sets need to be very large.

Divide data into groups following different mechanisms



What are the problems?

False Positives

Association between Attention Deficit Hyperactivity Disorder and DAT1 dopamine transporter is not replicated.

What are the problems?

Coding of gene effects

AA Aa aa : Dominant, additive, recessive and hetrosis

Individual	Dom	Add	Rec	Het
AA	1	2	1	0
Aa	1	1	0	1
aa	0	0	0	0

Candidate gene analysis

Genes and alleles are pre-selected.

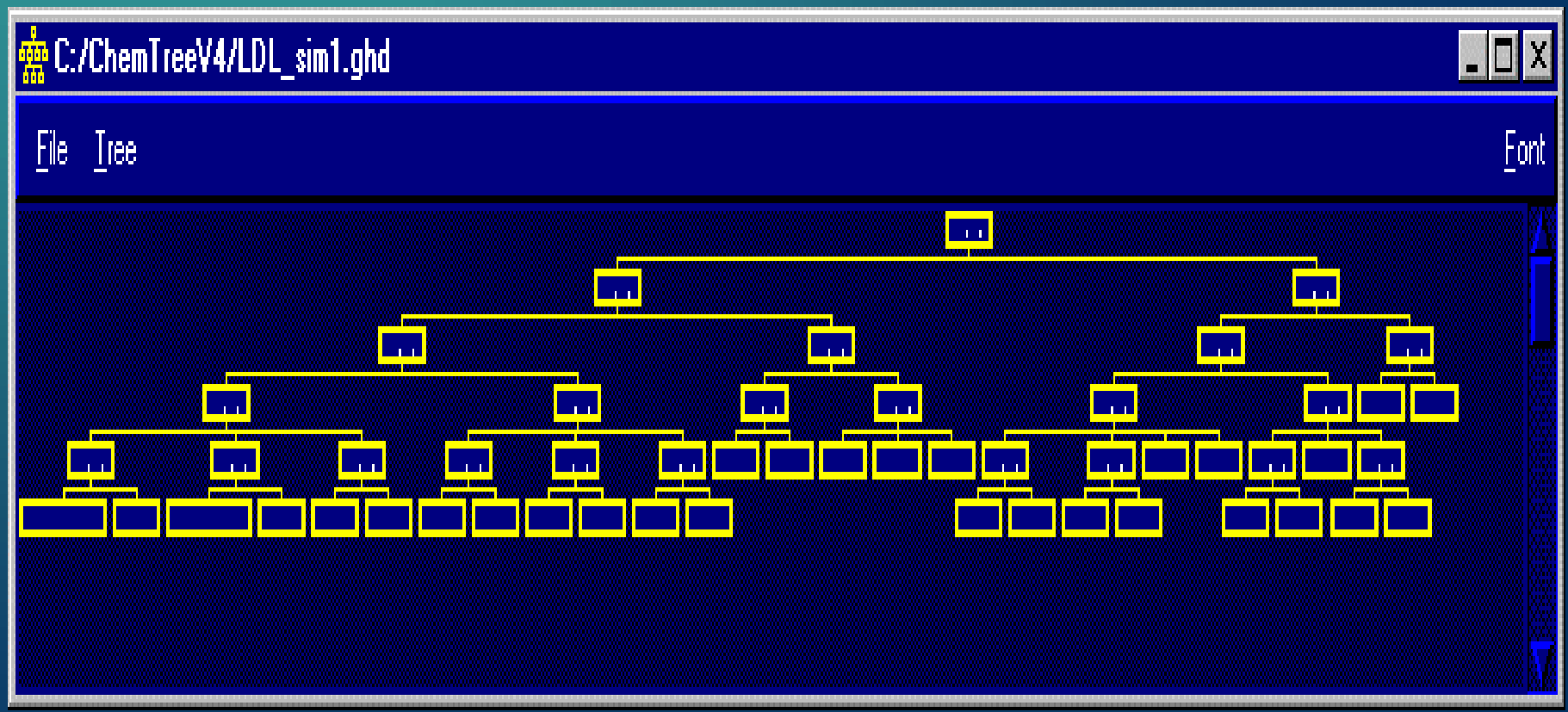
Looks like a logistic regression problem.

Very many independent variables.

Interactions are expected to abound.

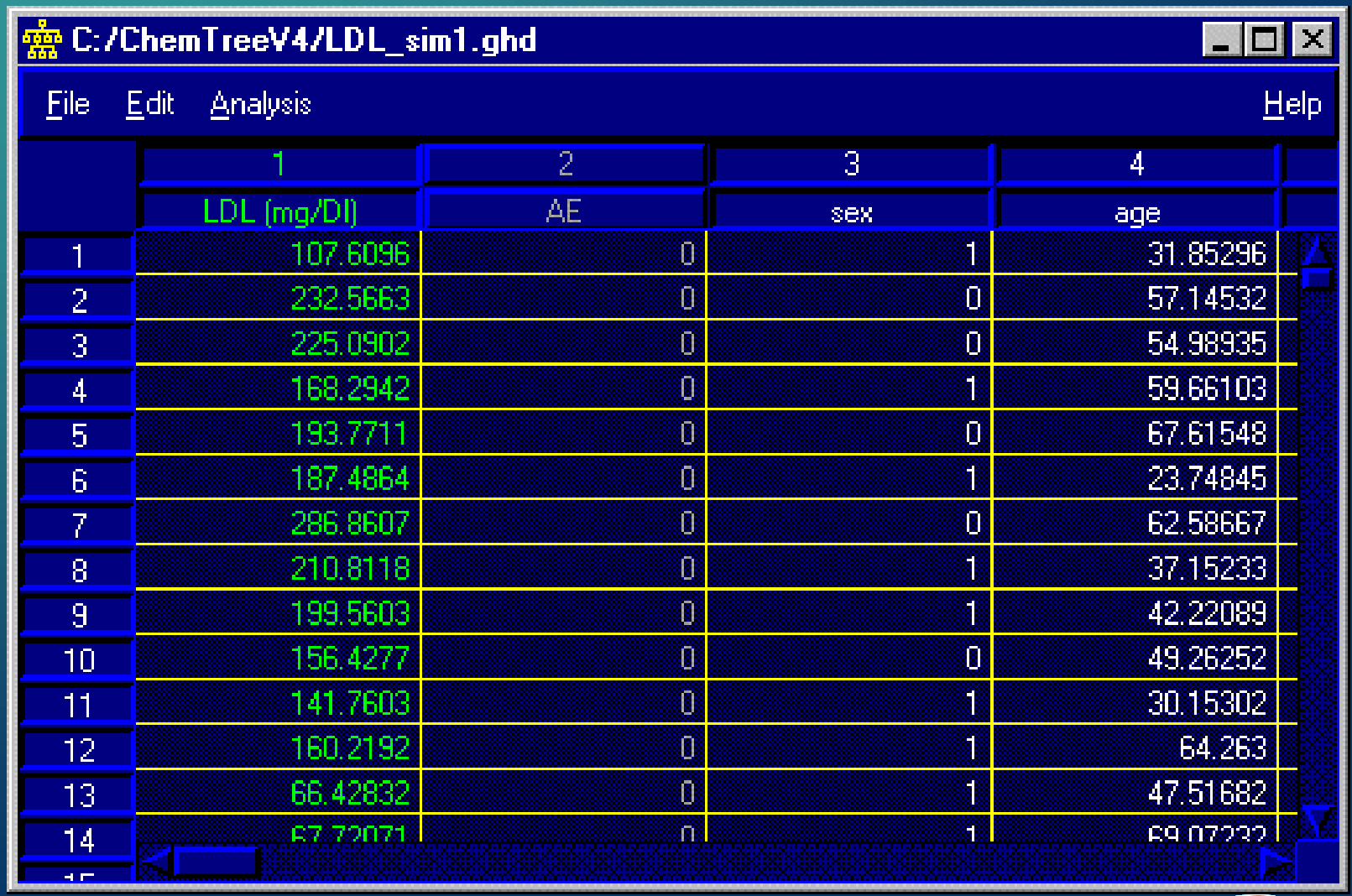
Analysis likely to be “exploratory.”

Complex Relationships

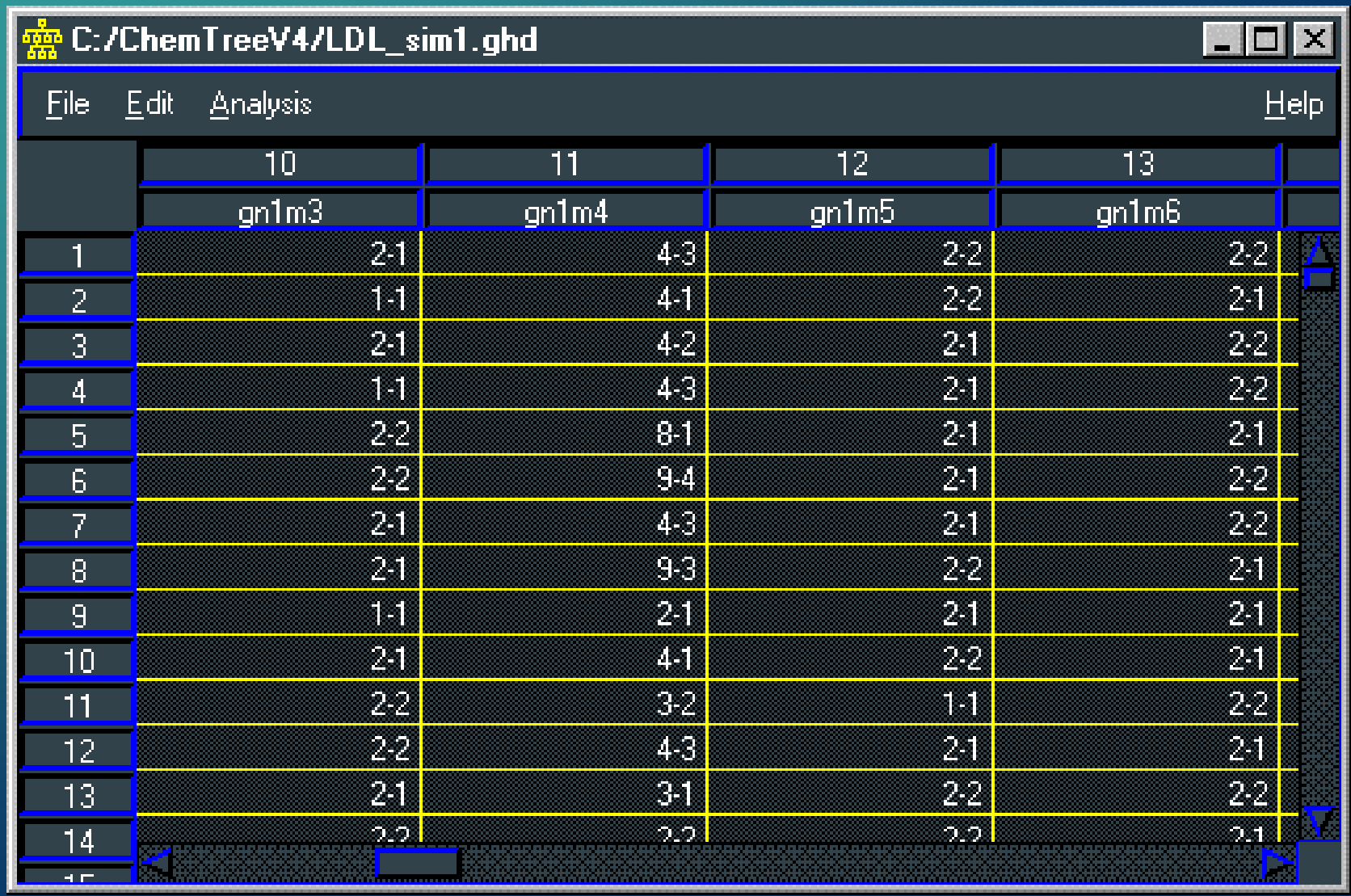


GlaxoSmithKline

HelixTree Software

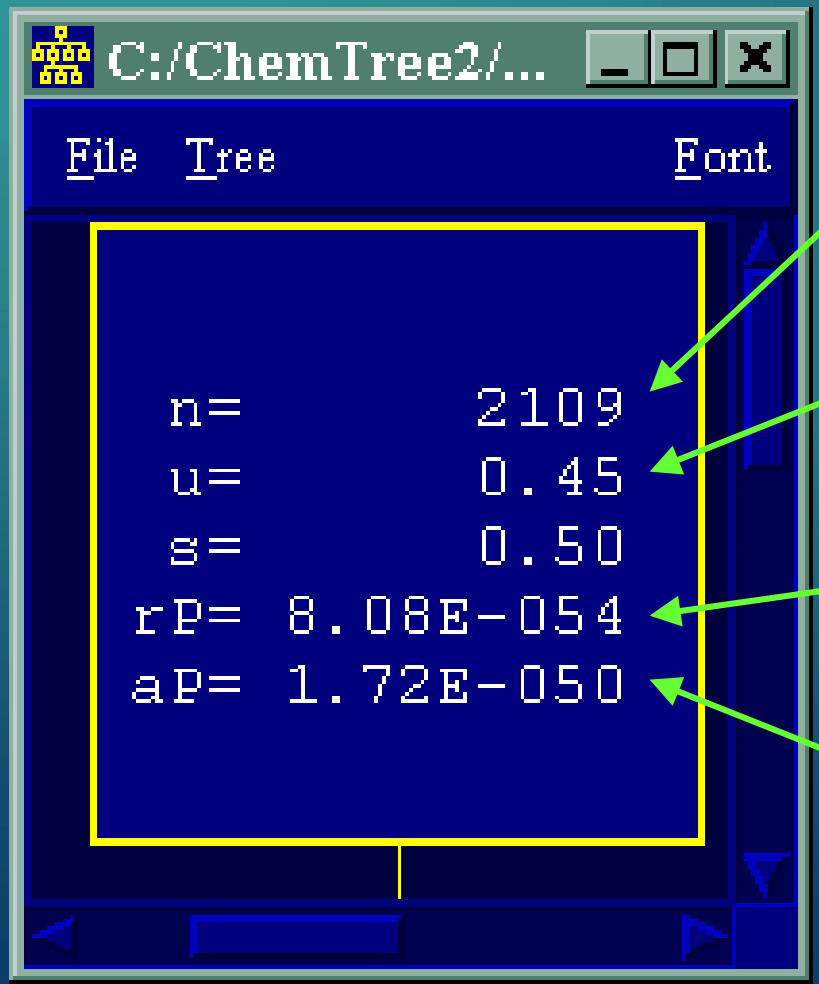


	1	2	3	4
	LDL (mg/Dl)	AE	sex	age
1	107.6096	0	1	31.85296
2	232.5663	0	0	57.14532
3	225.0902	0	0	54.98935
4	168.2942	0	1	59.66103
5	193.7711	0	0	67.61548
6	187.4864	0	1	23.74845
7	286.8607	0	0	62.58667
8	210.8118	0	1	37.15233
9	199.5603	0	1	42.22089
10	156.4277	0	0	49.26252
11	141.7603	0	1	30.15302
12	160.2192	0	1	64.263
13	66.42832	0	1	47.51682
14	67.72071	0	1	69.07233



GlaxoSmithKline

How to Read the Figures

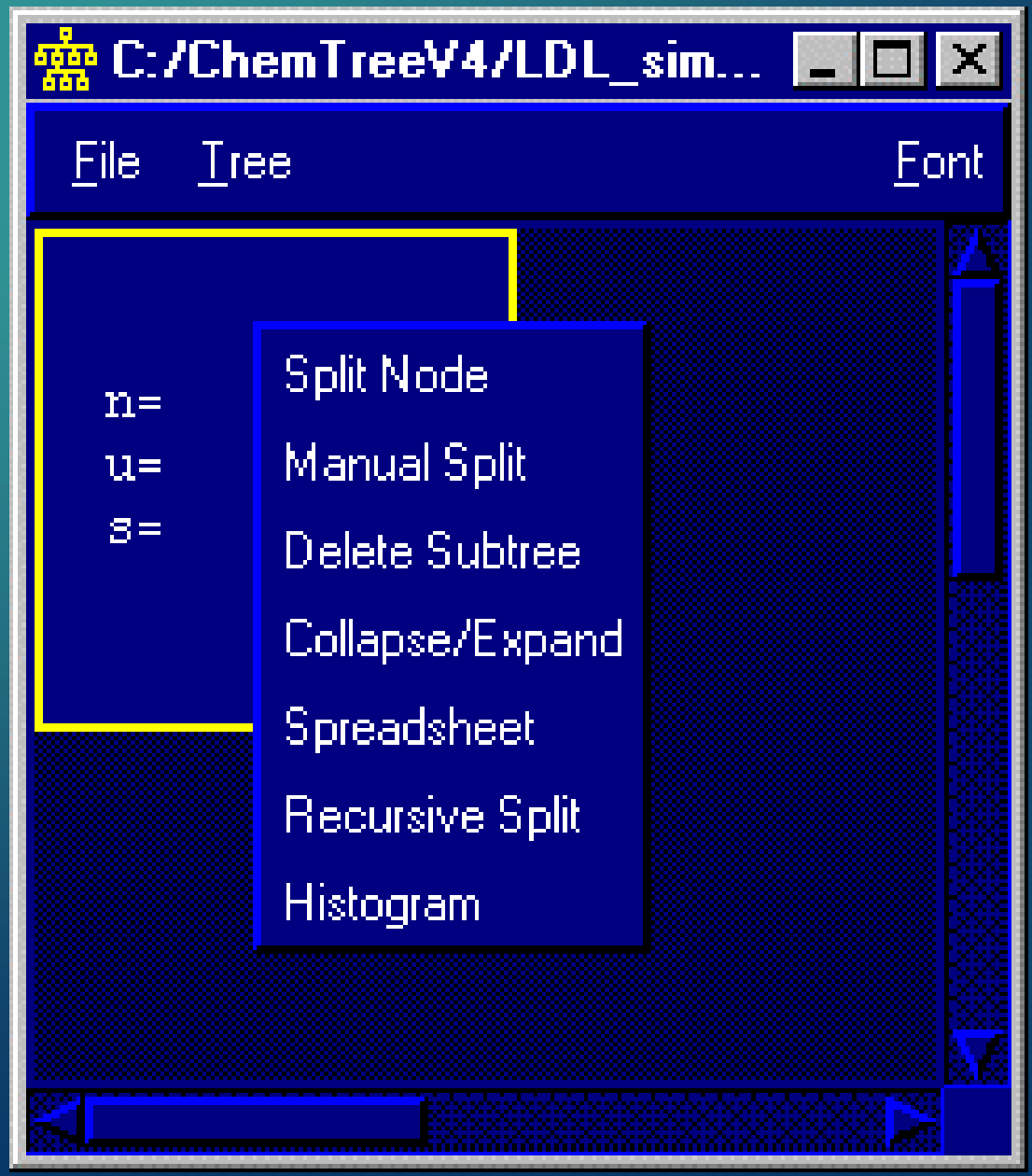


Sample Size

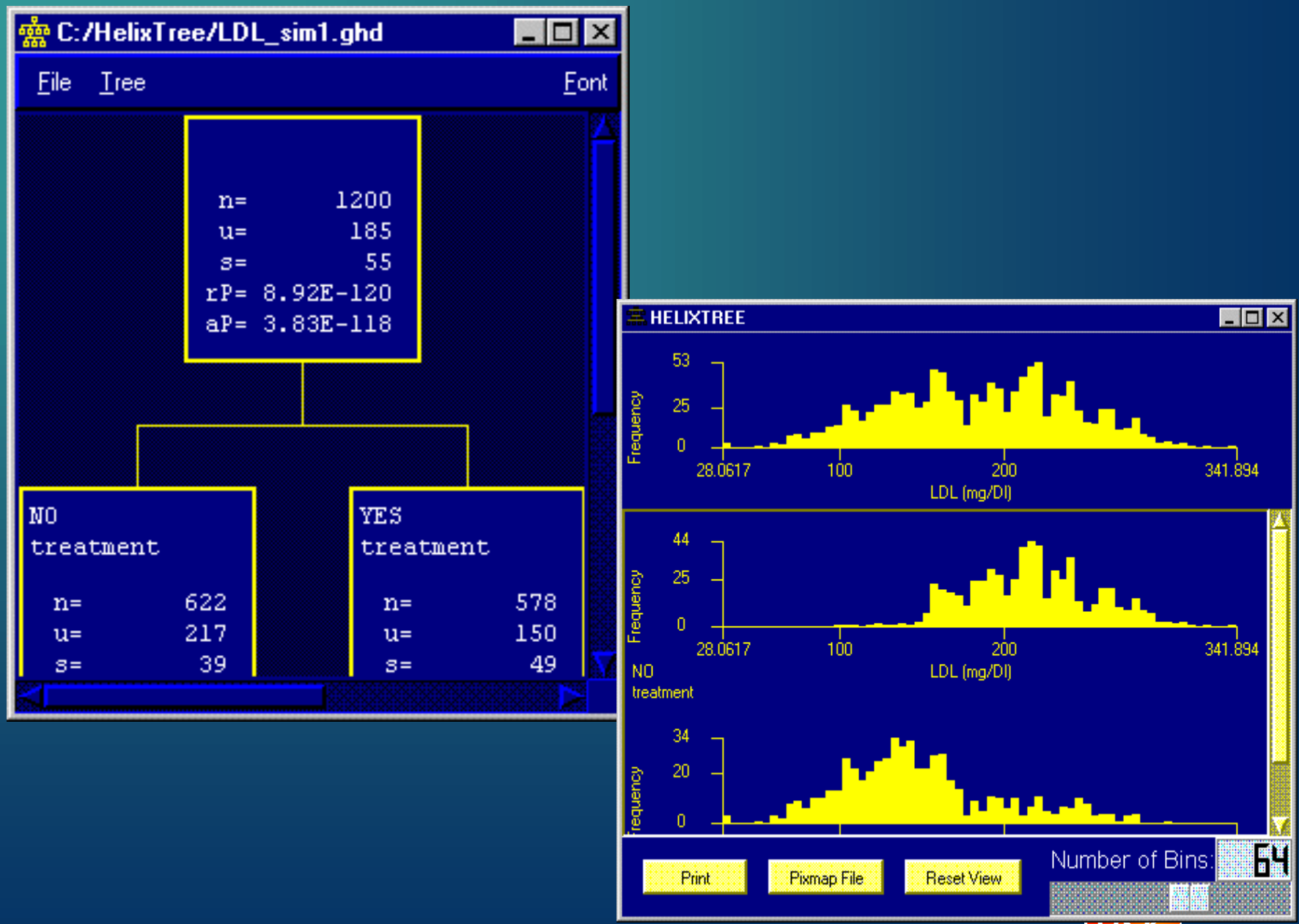
Fraction Positive

Raw P-value

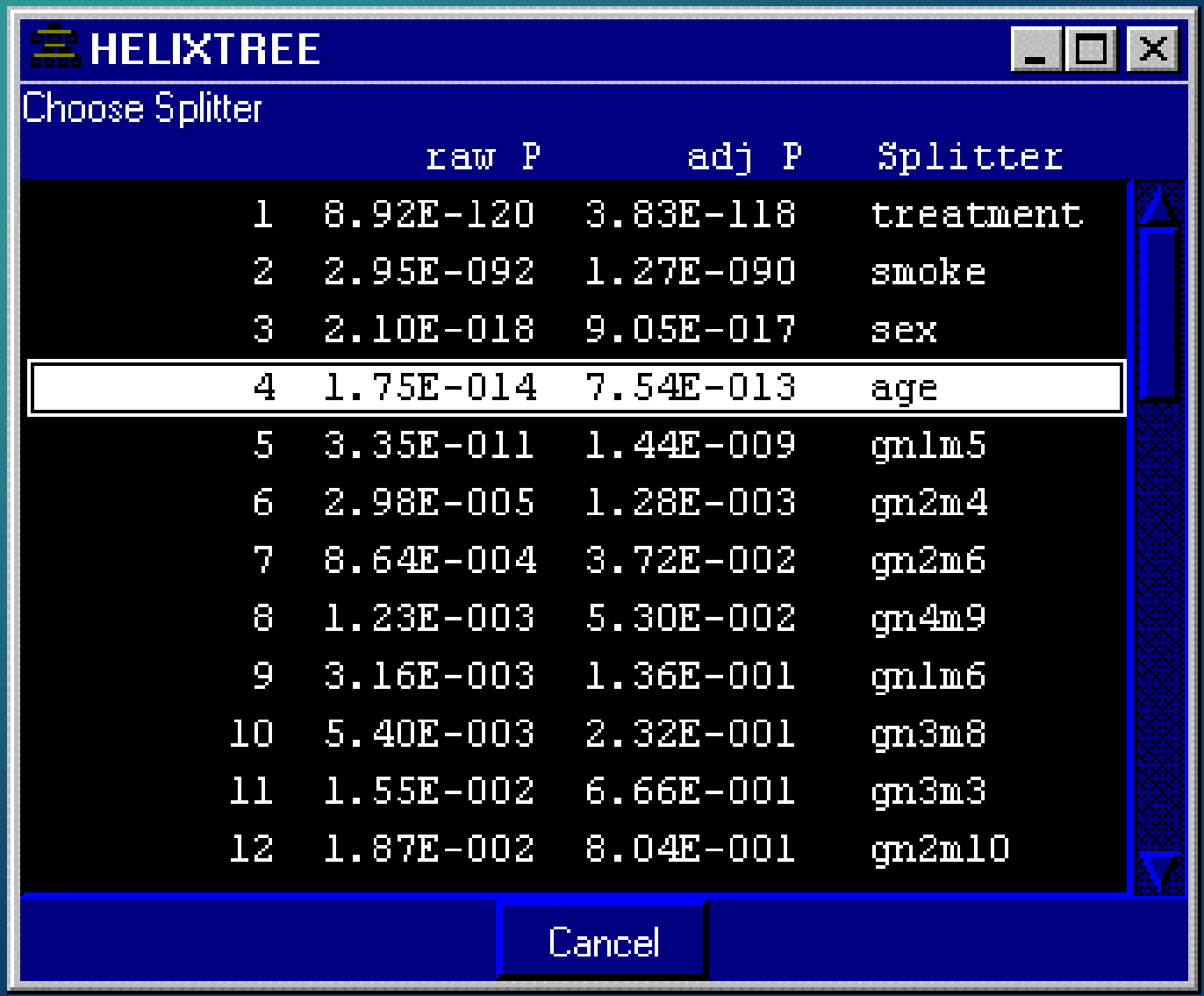
Adjusted P-values



GlaxoSmithKline



GlaxoSmithKline



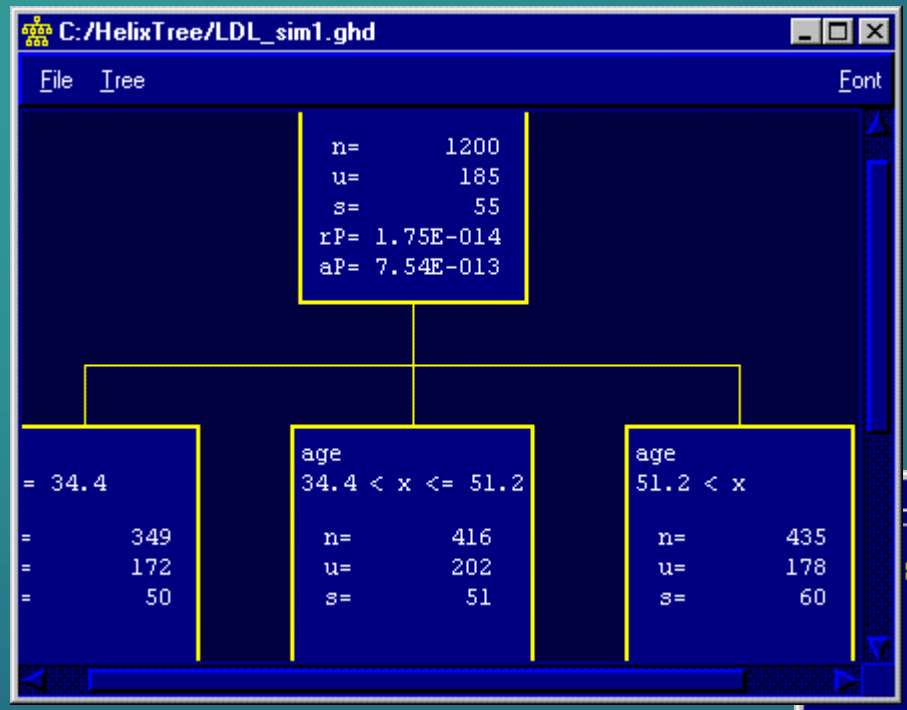
"Cut" Problem for Continuous Variables

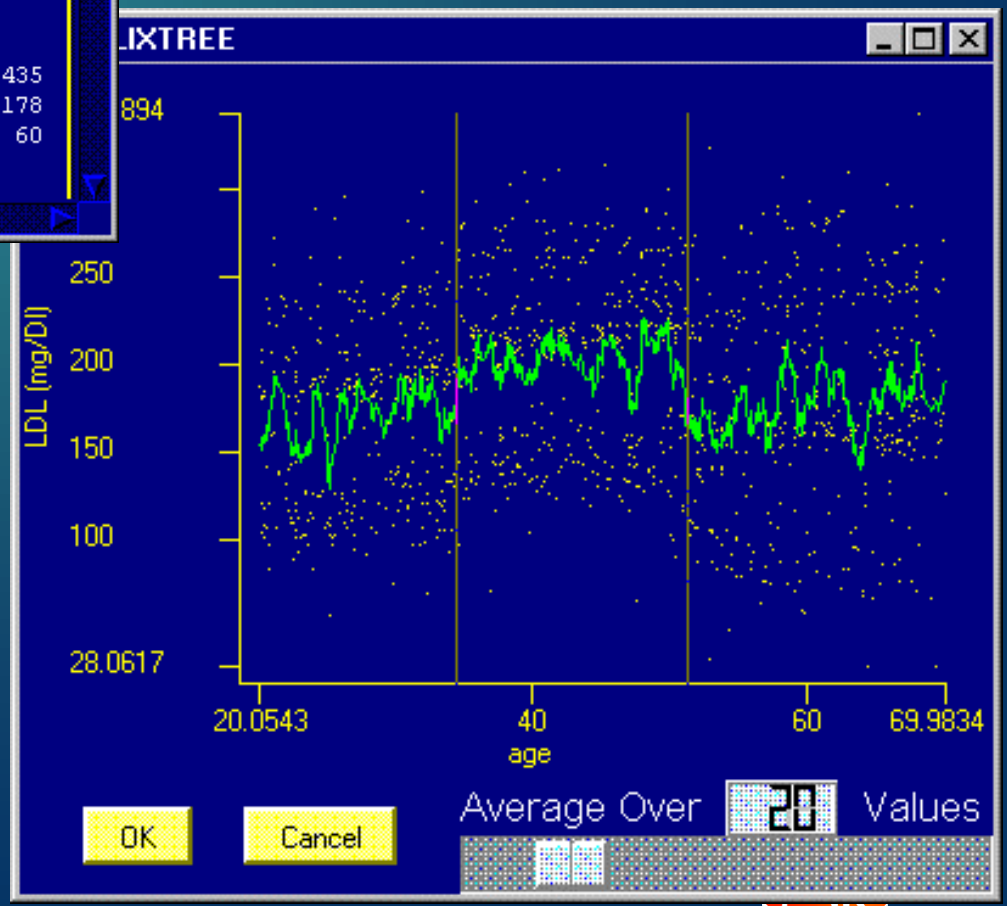
How do you find "cut points"

for an independent continuous variable

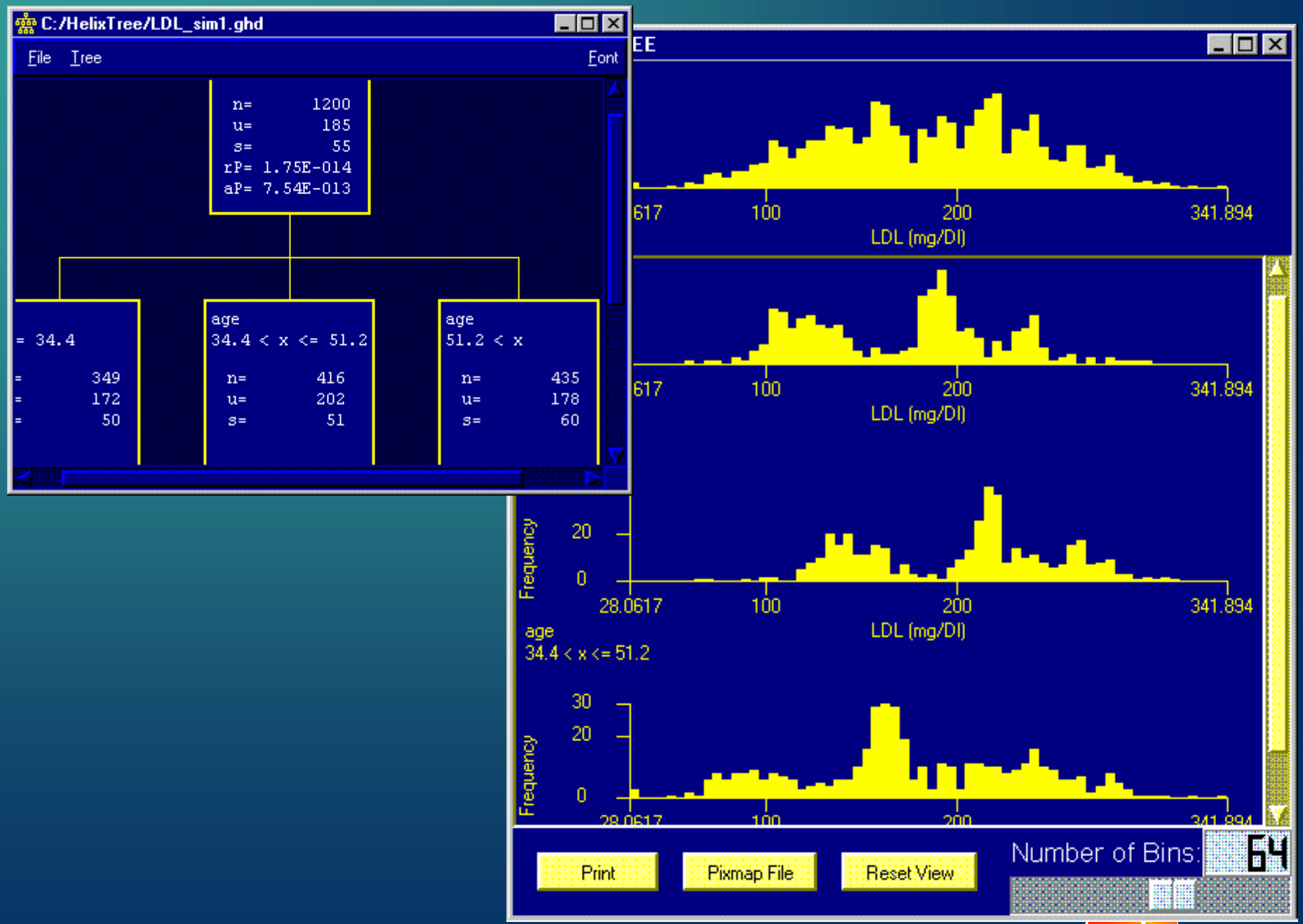
such that

the dependent variable is more homogenous?

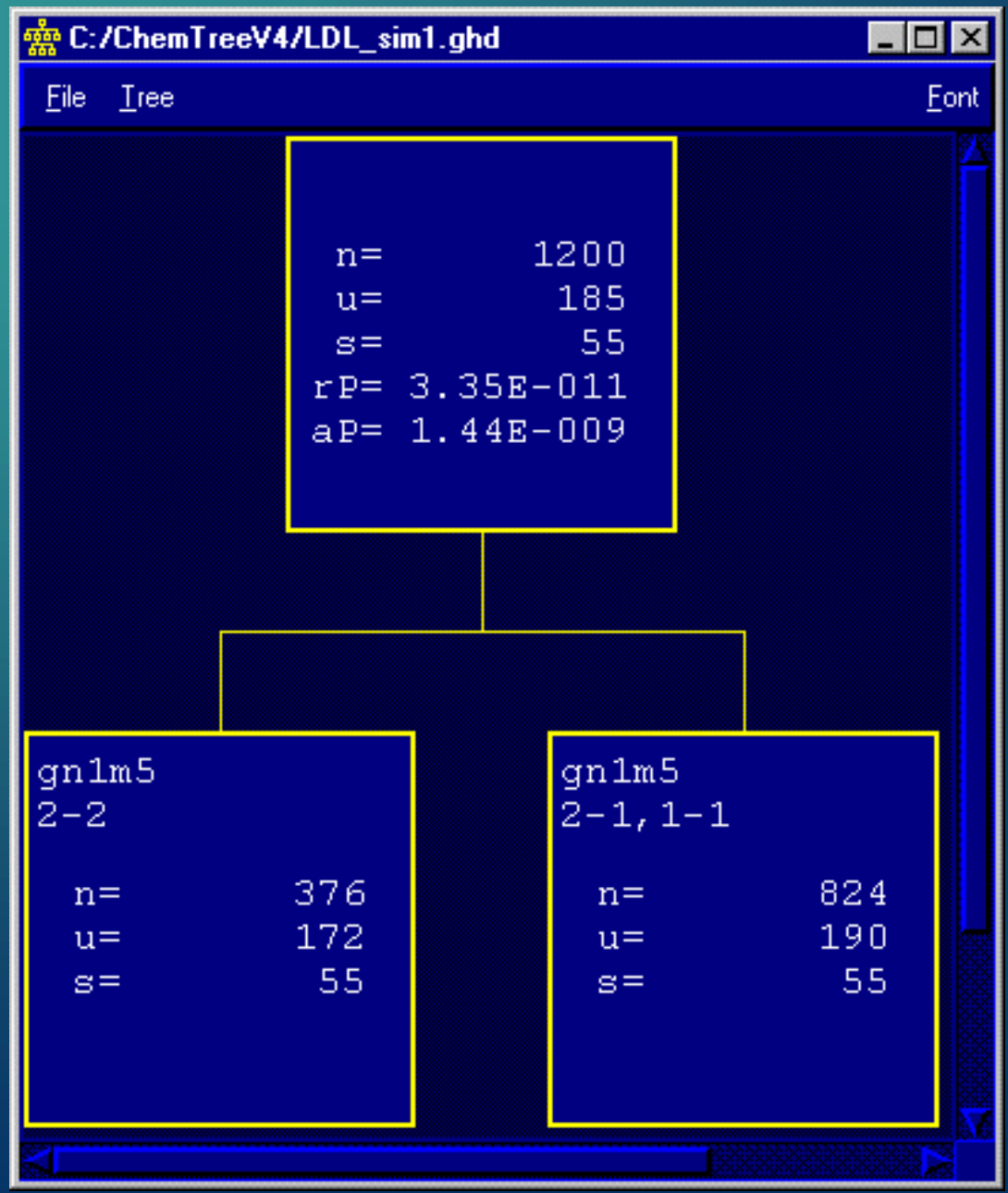




GlaxoSmithKline



GlaxoSmithKline



GlaxoSmithKline

RP Advantages

1. Works for complex situations, mixtures and interactions.

2. Statistical method easy to understand.

3. High statistical power.

4. Produces a valid answer.

Drawbacks

- Data greedy
- Only one view of the data
- Outliers can drive the splitting process
- Highly correlated variables may be obscured
- Higher order interactions may be masked

Literature

D.M. Hawkins and G.V. Kass, “Automatic Interaction Detection”, *Topics in Applied Multivariate Analysis*, ed. Hawkins, (1982).

D.M. Hawkins, S.S. Young and A. Rusinko, “Analysis of a Large Structure-Activity Data Set Using Recursive Partitioning”, *QSAR*, **16**, 296-302 (1997).

Software under development. www.goldenhelix.com

Outline

- Current Efforts in Clinical Trial Simulation
- The Promise of Pharmacogenetics
- Types of Genetic Information
- Genetic Parameters
- Genetic Simulations
- Multiple Genes and Clinical Variables
- Future Work