Zurich, September 2005

- Some challenges facing statistical science

- D.R.Cox

- Nuffield College, Oxford, UK

- David.cox@nuf.ox.ac.uk

Scope of statistics

- Variability not all systematic
- Uncertainty

- Measurement
- Data collection (study design)
- Analysis and interpretation
- Decisions

Component parts

- Specific details of application
- Theory: conceptual and mathematical
- Computing

Computing

- History
- Effect on data capture, storage and analysis

- Semi-fallacy of the information explosion
- Have the core principles of statistical analysis changed?

A few specific areas

- Government and Business
- Social (and economic) welfare
- Industry
- Health
- Food
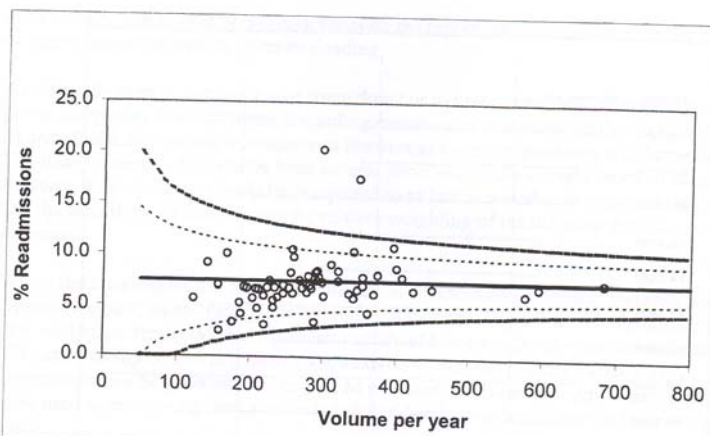- Water
- Environment
- Fundamental science

Government

- Gave the subject its name
- Traditional role of national statistics offices
- Bases for evidence-based policies
  - Performance indicators
  - Role of randomized trials

Performance indicators

- Performance indicators
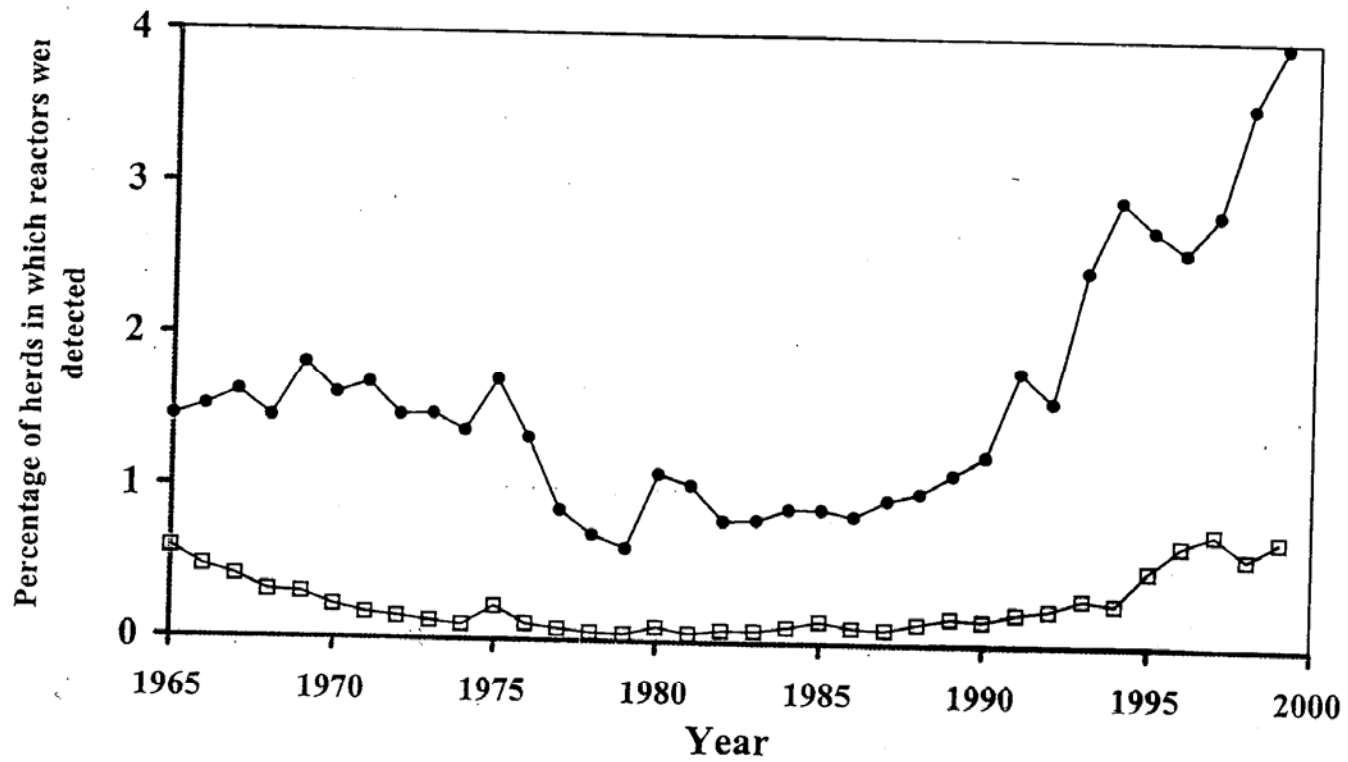- Performance targets
- League tables

- General principles of measurement should apply

Bovine tuberculosis

- Micobacterium bovis
  - Not same as M tuberculosis and M avis
  - Lives in cattle and wildlife

Figure 6.2 Annual Occurrence of Reactor Herds in England and Wales, 1965 – 2000.

Percentage of herds in which reactors were detected

Year

—●— 1) In the South western Region   —□— 2) Elsewhere in England and Wales
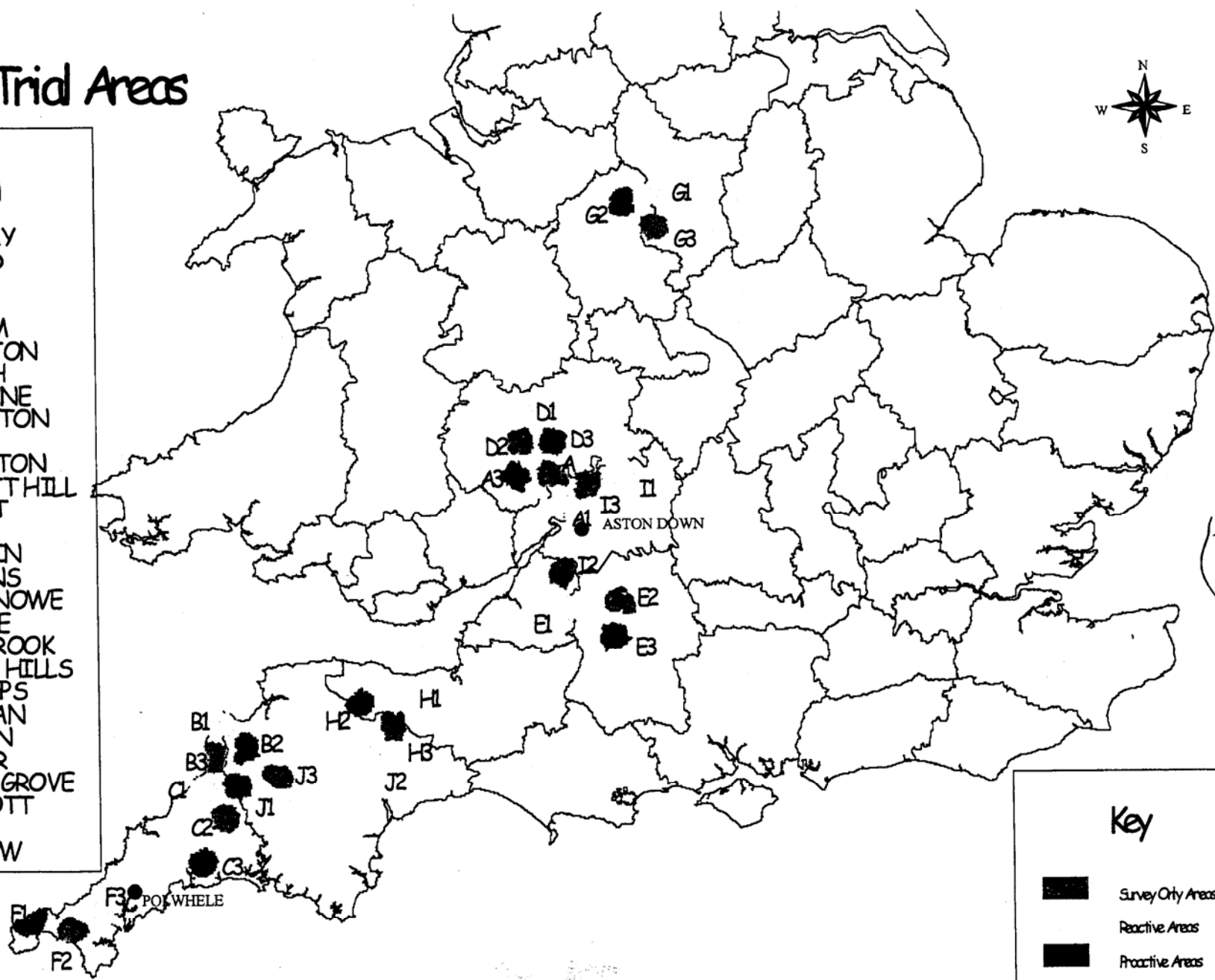
How can it be studied?

- Historical data

- Randomized trial

- Case-control study

- Experimental methods

- Mathematical modelling

Map Of Trial Areas

AL NAMES

- BLAISDON
- DYMOCK
- BROADWAY
- HARTLAND
- PUTFORD
- BUDE
- OTTERHAM
- LAUNCESTON
- LANREATH
- PUDLESTONE
- WITHINGTON
- BOSBURY
- COLD ASHTON
- CHARLCUTT HILL
- POULSHOT
- MADRON
- GODOLPHIN
- STITHIANS
- NETTLY KNOWE
- LADY EDGE
- CUBLEY BROOK
- BRENDON HILLS
- TARR STEPS
- HUNTSMAN
- ALDERTON
- WETMOOR
- APPERLEY GROVE
- LUFFINCOTT
- CADBURY
- NORTHLEW

G1 G2 G3

D1 D2 D3 A3 A1 I3 I1 ASTON DOWN

I2 E2 E1 E3

H1 H2 H3 J2

B1 B2 B3 J3 C1 J1 C2 C3

F1 F3 POLWHELE F2

Key

▨	Survey Only Areas
▤	Reactive Areas
■	Proactive Areas

Some general conclusions

- Large amounts of data from various sources
- Relatively little information on key issues
- Need for considerable care in interpretation

Water

- Many problems
- Extreme values
- Water quality

Reservoir safety

- Estimation of 10,000 yr return period
- That is, probability of 1/10,000 per yr per site
- Roughly 2,000 sites in UK
- How are series of various lengths from nearby and not-so-nearby sites to be combined to produce an estimate for each particular reservoir
- Link with PMP

Challenges to theory

- Strategy of analysis
- Realistic assessment of accuracy and precision
- Transparency of analysis
- More detailed issues
  - Model construction
  - Bridging the gap between analysis and interpretation
  - Nature of causality
  - Graphical Markov models
  - Theories of inference
    - Role of non-frequency concepts of probability
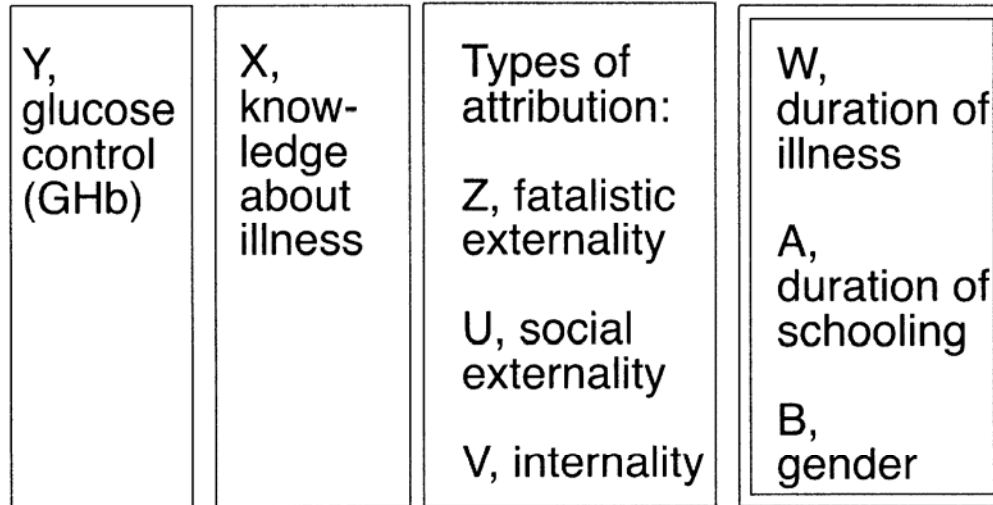
Graphical Markov models

- Development from Sewall Wright's path analysis
- Variables represented by nodes of graph
  - Independency defined by missing edge
  - Need for various kinds of edge and graph

  Joint work with Nanny Wermuth
    (Chalmers/Gothenburg University, Sweden)

A motivating example

What determines the well-being of diabetic patients?

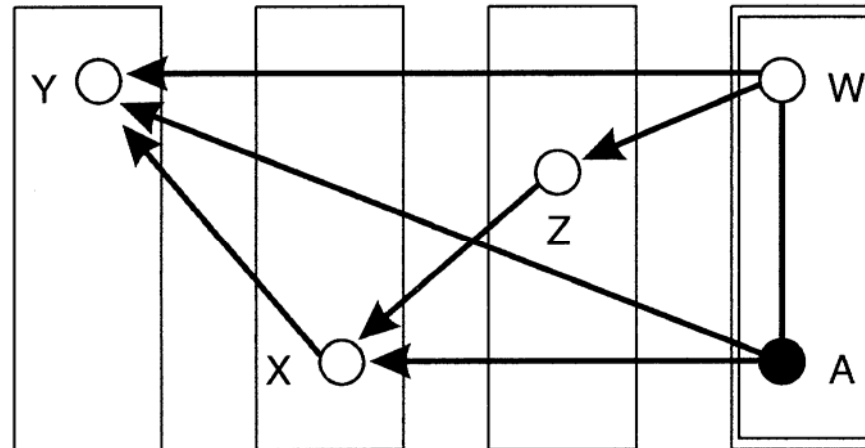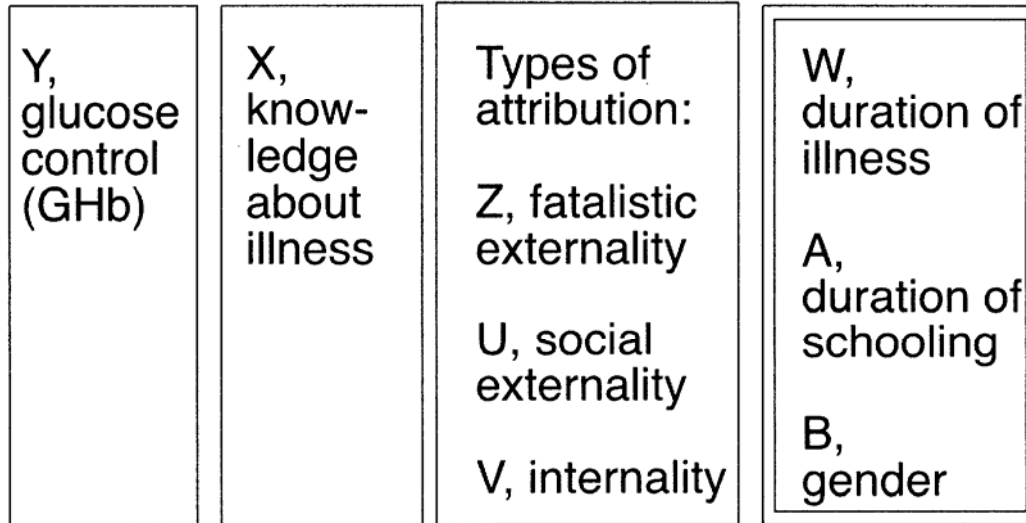| Y, glucose control (GHb)	X, know- ledge about illness	Types of attribution: Z, fatalistic externality U, social externality V, internality	W, duration of illness A, duration of schooling B, gender
primary response		intermediate variables secondary response	purely explanatory variables

6 quantitative variables: X, Y, Z, U, V, W

2 binary variables: A, B
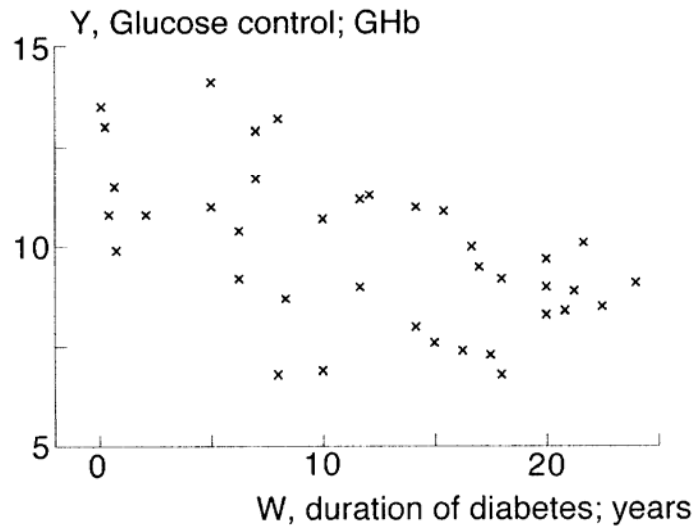
Cross-sectional study of 68 diabetic patients; 1990
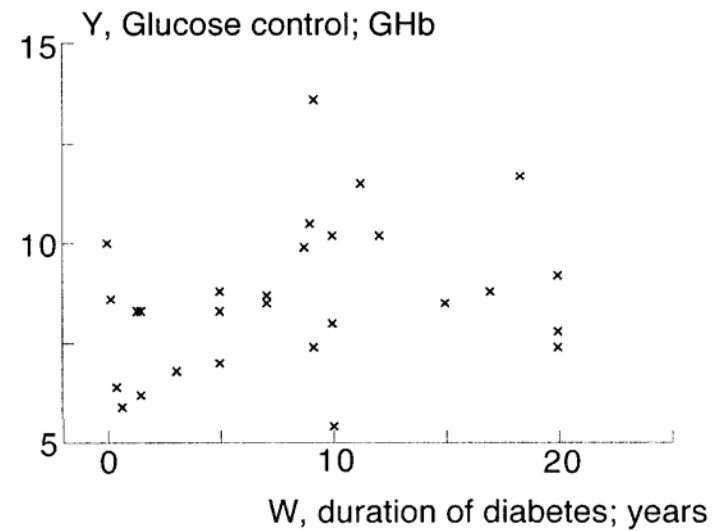
The graphical summary after analysis

| Y, glucose control (GHb) | X, know-ledge about illness | Types of attribution: Z, fatalistic externality U, social externality V, internality | W, duration of illness A, duration of schooling B, gender |

$$Y:\ X + A * W, \quad X:\ Z + A, \quad Z:\ W.$$

The interactive effect $A * W$

A, shorter formal schooling

A, longer formal schooling

Summary

- Statistical problems all pervasive
- Statistical issues cannot be evaded
- Number of statisticians available limited relative to the potential demand
- Challenge of achieving interplay between specifics of applications and new methodological developments
- Challenge to theoreticians and academics of preserving intellectual cohesion in face of ever-increasing specialization

Meta summary

- Presentation of the subject to a general audience?