e-Epidemiology:
extraordinary challenges of the e-record epoch

Iain Buchan
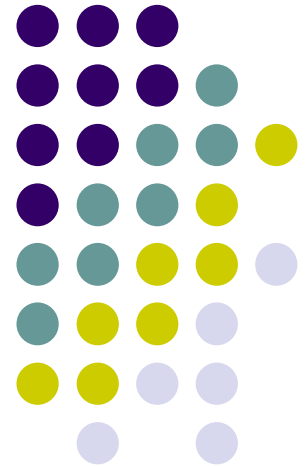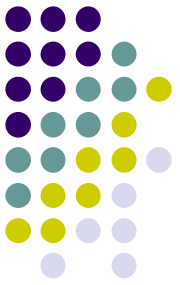Senior Lecturer in Public Health Informatics
buchan@man.ac.uk

School of Epidemiology and Health Sciences Seminar
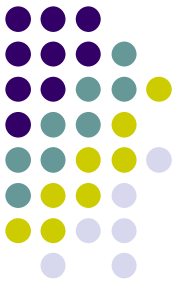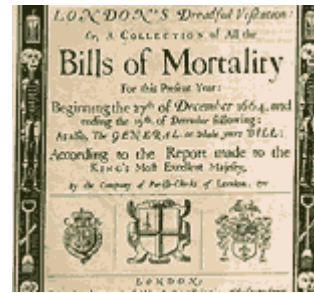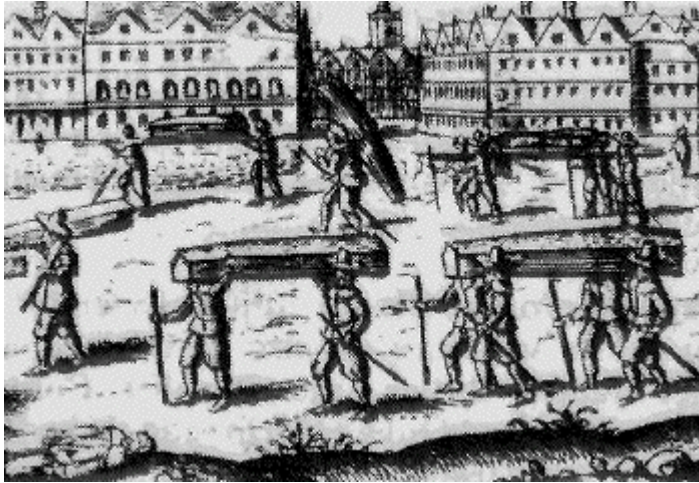University of Manchester
6th June 2003

Aim & objectives of seminar

- To seed debate and collaborations

- What is e-*?
- Can e-* extend the limits of Epidemiology?
- Starting simply; some research ideas
- Potential service impacts

Back to the future

An historical perspective on

Epidemiology in search of **Engineering**…

Health inequalities
in 17th Century London

Parish [% deaths linked to houses in hearth tax 1662-6]	Number of hearths [% dying]							
	1	2	3	4	5	6	7	8+
St James Clerkenwell (44)	31.7	20.8	24.7	9.7	2.6	3.2	1.2	6.2
St Botolph without Aldgate (41)	31.2	25.2	22.3	8.3	4.7	4.7	1.5	2.2
St Dunstan in the West (49)	15.5	11.7	12.6	20.2	8.6	5.0	4.7	21.7
St Michael Queenhithe (20)	35.1	24.3	16.2	5.4	5.4	2.7	2.7	8.1
St Saviour Southwark (42)	14.2	31.8	28.2	12.4	6.9	4.3	0.7	1.5

Adapted from: *Epidemic Disease in London*, ed. J.A.I. Champion (Centre for Metropolitan History Working Papers Series, No.1, 1993): pp. 35-52

Epidemiology in search of engineering

Babbage (1791-1871) the constructionist & Farr (1807-1883) the reductionist
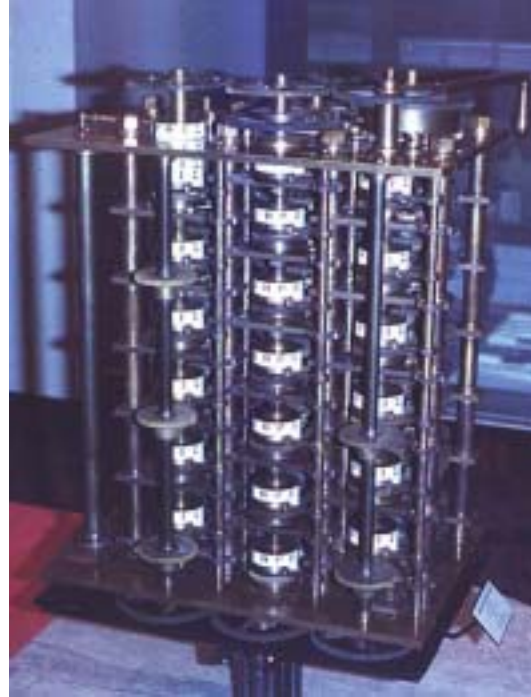
The difference engine could have been built, perhaps also the analytical engine, and translating the design using relays, a stored program computer might have been built one hundred years ahead of time.

A science interface problem? MRC meets EPSRC?

Abb. 7   Titelseite.
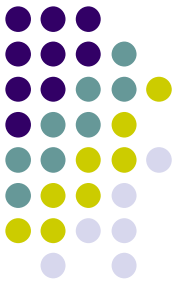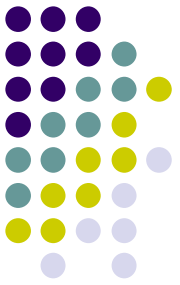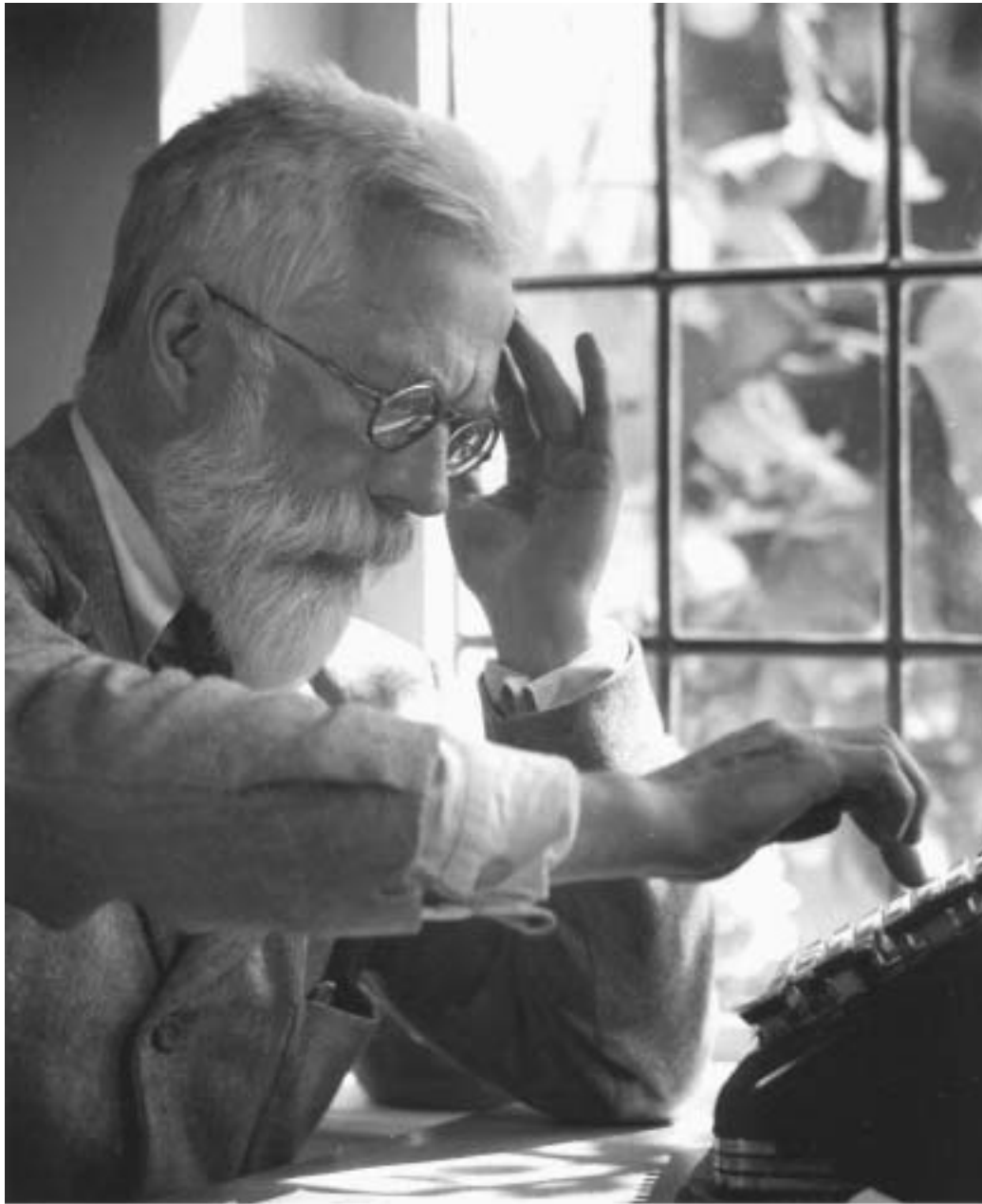Scientific American 63(1890) No. 9.
(August 3o, 189o)

US Census of 1890

Hollerith Tabulator

Company to become IBM

…several steps back from Babbage work

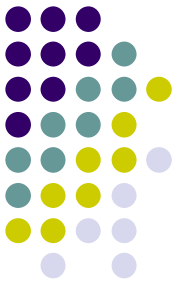Sir Ronald Aylmer Fisher (1890-1962)

using a mechanical calculator

Today…

- Domestic computing power exceeds the requirements of most Epidemiology, with the notable exception of Genomics.

- We live with the paradox of information overload and starvation

- Quantitative methods are becoming less transparent; stochastic, probabilistic networks, information theoretic, graph theoretic etc.

Epidemiology's evolution?

…study of the distribution and determinants of human health and disease

Majority of epidemiological activities are:

continuous health intelligence:
less observation vs. intervention distinction

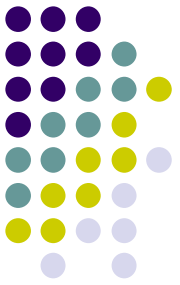discrete studies

public records analysis

time in history and complexity of analysis   2003

Thinking e-*

- more
  - data
  - linkage possibilities
  - diversity of reasoning (knowledge, skills)
  - systematic hypothesis generation
  - collaborative method development
  - output
  - timeliness
- better ? – lets explore scenarios…
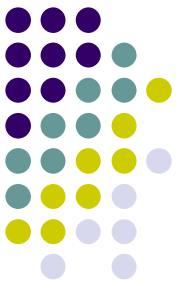
e- is for…?

simplified: e- means collaboration via information and communication technologies

e-* foci of interest:

- **e-Science**
- **e-Health** (including the emergence of electronic records of health and social care)
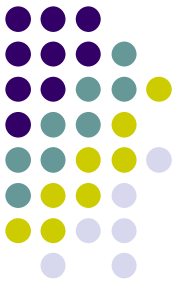
e-Science

- The DTI and the Research Councils are committing £118M to a government-industry programme on **e-Science**.

- **e-Science**
  "means science increasingly done through distributed global collaborations enabled by the Internet, using very large data collections, terascale computing resources and high performance visualisation"

DTI: www.escience-grid.org.uk

GRID

"the word 'Grid' is chosen by analogy with the electric power grid, which provides pervasive access to power and, like the computer and a small number of other advances, has had a dramatic impact on human capabilities and society. We believe that by providing pervasive, dependable, consistent and inexpensive access to advanced computational capabilities, databases, sensors and people, computational grids will have a similar transforming effect, allowing new classes of applications to emerge."

DTI: www.escience-grid.org.uk

GRID in context

- Old web
  - *supermarket*
  - data and pre-packaged information

- New web/GRID
  - *restaurant*
  - processes to inform
  - support reasoning, including semantics: (re-)shape question → find or generate data → process data → variety of visualisation and reporting forms → hooks into other relevant processes
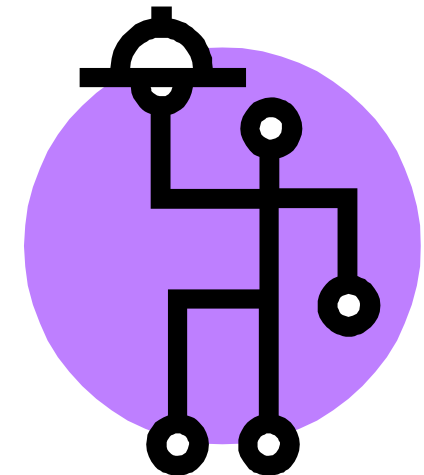
e- in context

- complex collaboration via ICT enabling research that would be
  - infeasible
  - impractical?
  - …without GRID (like) approach

- translation issues
  - inter-discipline
    - GRID has different meaning for particle physics vs. public health
  - intra-discipline
    - multi-scale integration e.g. gene → population
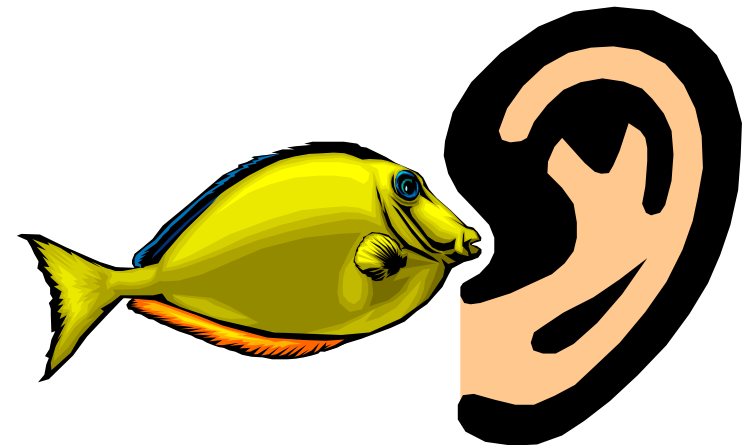
Big picture re. interventions

Resolving population intervention effects from by-products of e-Society:-

Focus: **natural experiments** of health and social policy

Data: **e-records** (↑ quantity, quality, accessibility)

Methods:
 interrogation of data **linked** across systems
 smarter **privacy** protection, replacing blanket bans
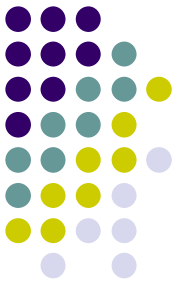 **statistical** developments for large, complex, noisy datasets
 meaningful **visualisation** of complex information
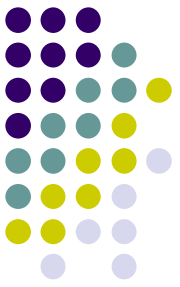 **automation** and **adaptation**

A potential impact of results: decentralisation of policy

Exploring examples…

- NSF equity
- NHS performance vs. needs intelligence systems
- Health visitor data as example of data neglect
- Extending RCTs
- Machine learning for prescribing safety
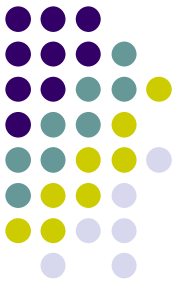- Exception surveillance

NSF effects on service equity

- What effect is the introduction of National Service Frameworks having upon the fairness of distribution of NHS resources?

- macro view using HES, PACT and mortality data
- some micro views where e-records are suitable
- look at concordance with qualitative evaluations
- make use of staggered change over time
- visualise by ranges of 'value issues' e.g. access vs. utilisation
- visualise by ranges of confidence in proxy indicators of need
- consider conclusions before and after extended visualisation

CABG & PTCA

CHD mortality < 75
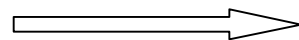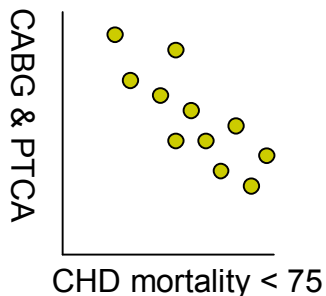
? fluidity of inverse care
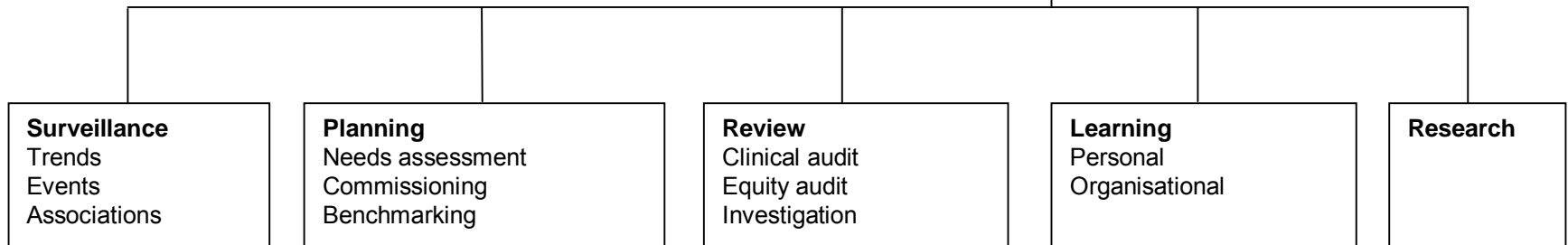
NHS performance vs. needs
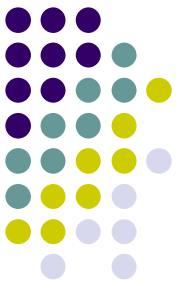
NHS data, including:
>   HES and special returns, e.g. waiting
>   GP prescribing
>   [new GP contract]
>   Performance assessment framework
>   Patient surveys
>   Health promotion returns, e.g. smoking cessation

Other service data, e.g. social services

Census data

Values:
>   Effectiveness
>   Safety
>   Efficiency
>   Equity
>   Transparency

Performance Intelligence Hub

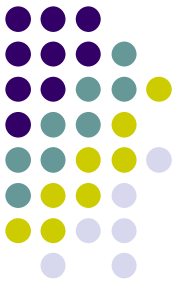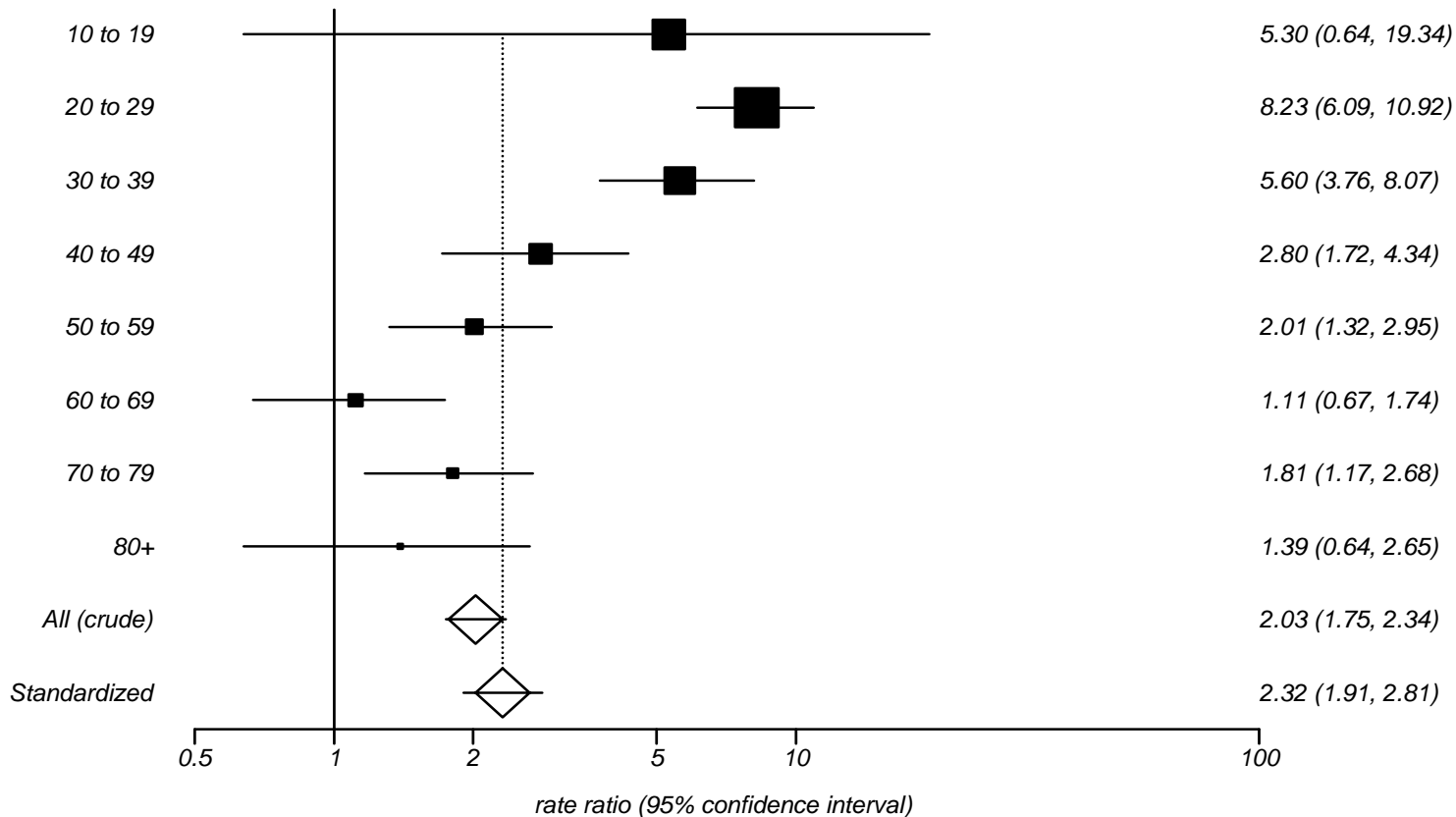| **Surveillance**	**Planning**	**Review**	**Learning**	**Research**
Trends	Needs assessment	Clinical audit	Personal	
Events	Commissioning	Equity audit	Organisational	
Associations	Benchmarking	Investigation		

Practical steps

- NW Public Health Observatory developments at Manchester.

- Starting with equity profiles created on-demand via MS SQL database and ASP.net calling optimised numerical libraries.

- Aiming to improve statistical appreciation as well as providing concise information.

Concise information

Stratified rate ratio plot (direct standardization)

|	rate ratio (95% confidence interval)
10 to 19	5.30 (0.64, 19.34)
20 to 29	8.23 (6.09, 10.92)
30 to 39	5.60 (3.76, 8.07)
40 to 49	2.80 (1.72, 4.34)
50 to 59	2.01 (1.32, 2.95)
60 to 69	1.11 (0.67, 1.74)
70 to 79	1.81 (1.17, 2.68)
80+	1.39 (0.64, 2.65)
All (crude)	2.03 (1.75, 2.34)
Standardized	2.32 (1.91, 2.81)

Retrospective cohort study of mortality in 2122 males treated for schizophrenia in Alberta 1976-1985.
Standard/reference population: Alberta 1981.

Health Visitor data

- Example of little-researched routinely-collected data.

- Would more timely epidemiological analysis of HV data have lead to earlier Public Health action on childhood obesity in England?

- Do we pay too much attention to irrelevant precision of measurement in quests that don't require highly distilled evidence of causality?

Annual mean Standard Deviation Score for height, weight and BMI in Wirral pre-school children from 1989 to 2000.
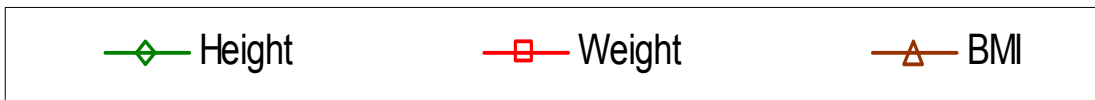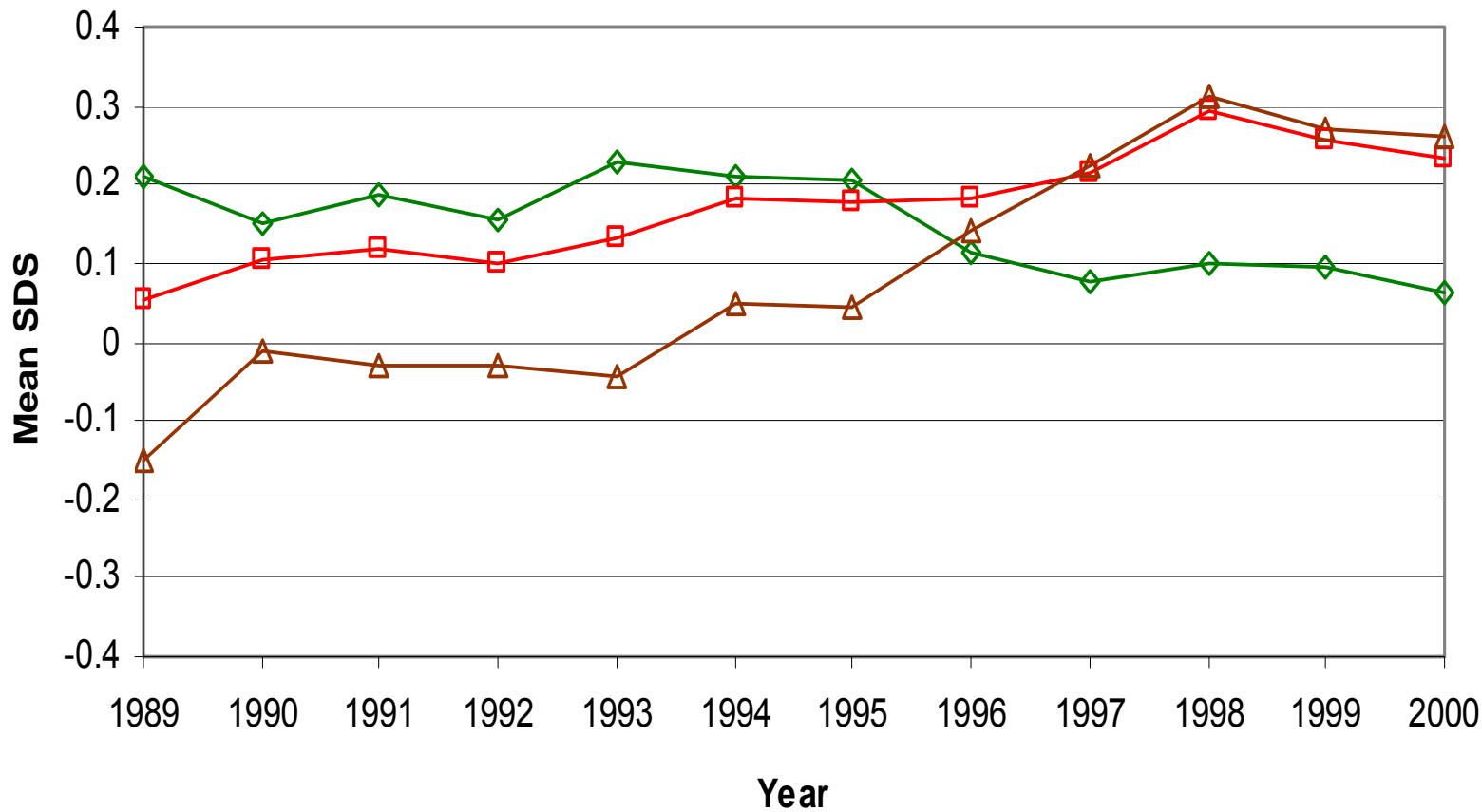
Annual mean Standard Deviation Score for height in Wirral pre-school children from 1989 to 2000 by quartile of Townsend Material Deprivation score.
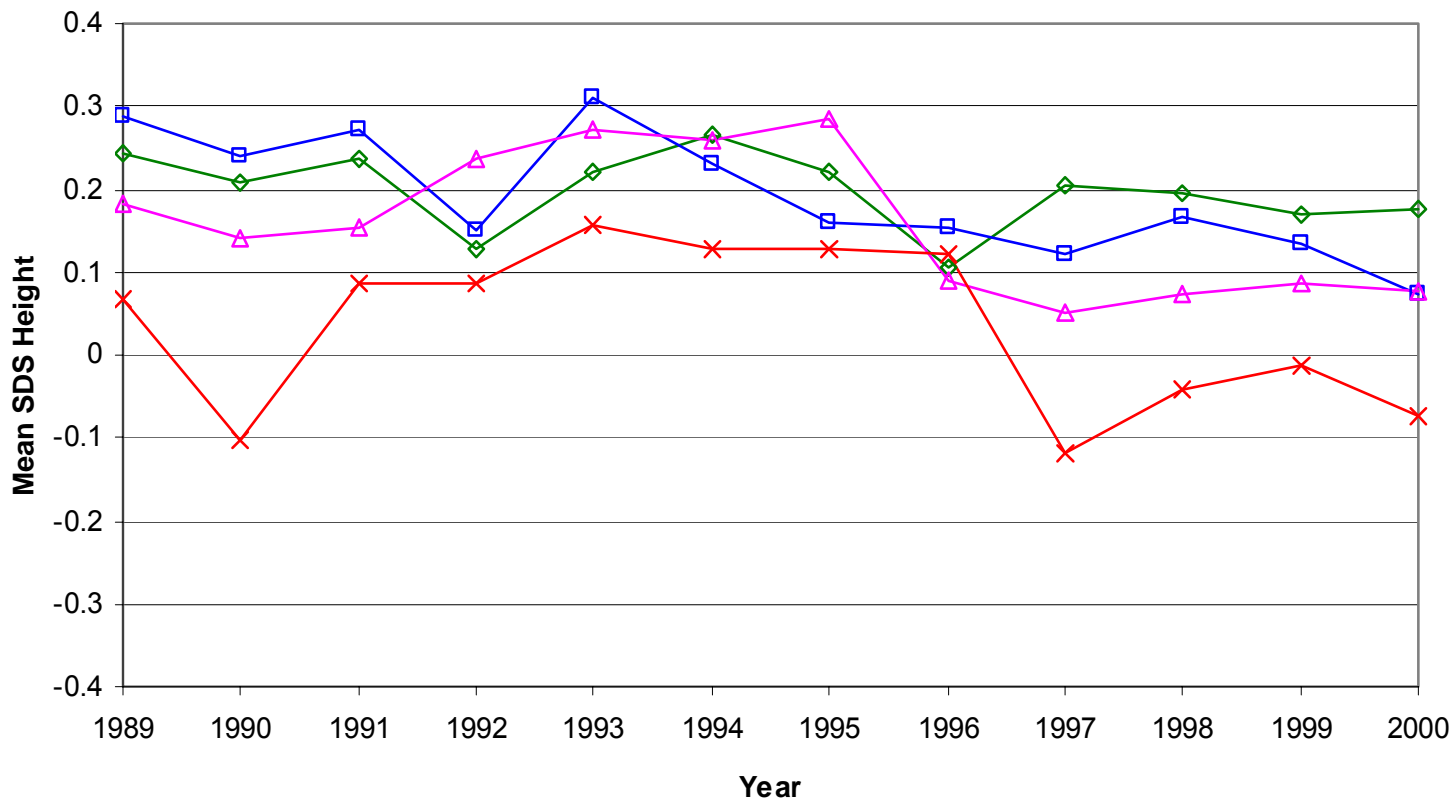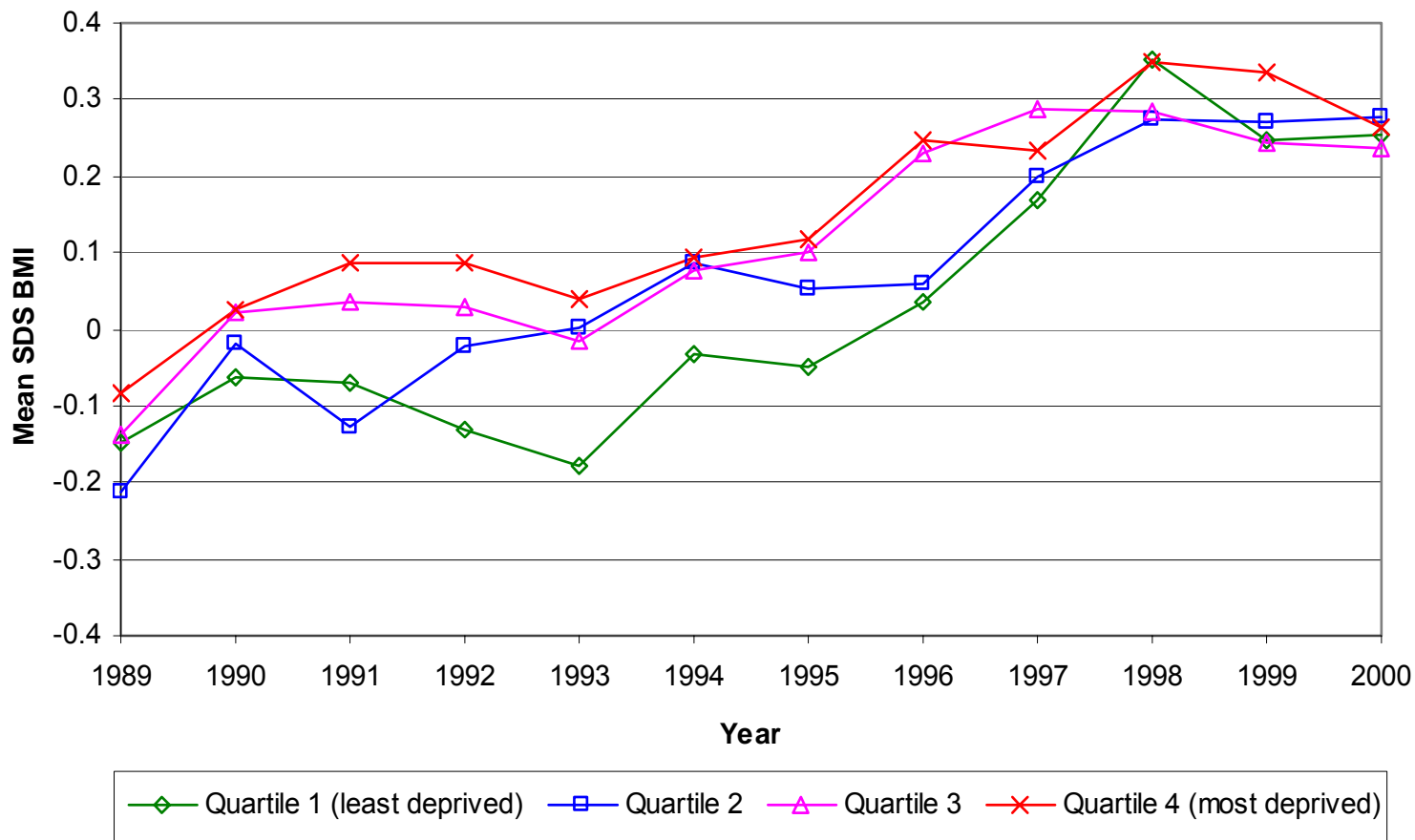
Annual mean Standard Deviation Score for Body Mass Index in Wirral pre-school children from 1989 to 2000 by quartile of Townsend Material Deprivation score.

RCT re-analysis and extension

- What social gradients can be identified in treatment efficacy by systematic re-analysis of RCT data, and how do these compare with the sizes of the relevant associations between morbidity and socio-economic status?

- Could package analytical agents to be given to trial owners for analysis without disclosure of data → MRC e-Science?
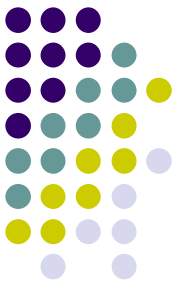
Prescribing intelligence

- Machine learning to classify drug interactions by their clinical relevance.

- Possibly start with one domain, e.g. Rheumatology nested as a special project within the Integrated Care Records Service development.

- Aims
  - To improve understanding of drug interactions.
  - To develop adaptive prompting of the prescriber in order to improve patient safety with respect to interactions.

Exception surveillance
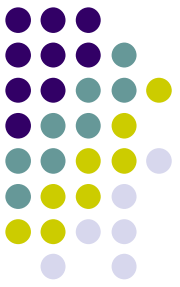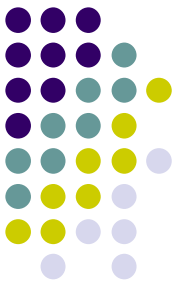
Exception surveillance

- Exceptional events, trends or associations indicating problems with clinical quality.

- Mortality by GP
  - DoH to link GP code (Exeter & NHS Strategic Tracing Service) to cause of death (ONS mortality), and disseminate statistics to SHAs, PCTs, CH(A)I etc.
  - Fixed models of analysis sufficient?
 - Need to investigate adaptive approaches.
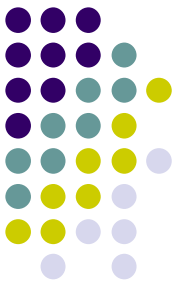
Bais as an information problem

- Can information theory be used to reduce the potential for confounding due to systematic error in the measurement of potential confounders?

- Focus: socioeconomic status confounding aetiological studies, especially where effect sizes are small.
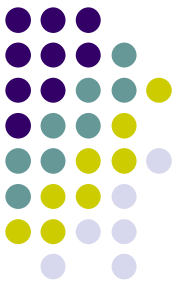
Service impacts of e-Epi

- Time
  - One off looks → automated surveillance triggered
  - Review cycles → response to alerts
  - Shorter times between reviews of policy

- Place
  - More & better quality data → local effects
  - Decentralisation of policy
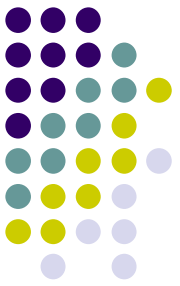
A thought to finish with

Accessible e-epidemiological tools for the exploration of numerous complex data may mobilise an observer-theorist mindset, similar to that which bound Statistics as a discipline around 1900.
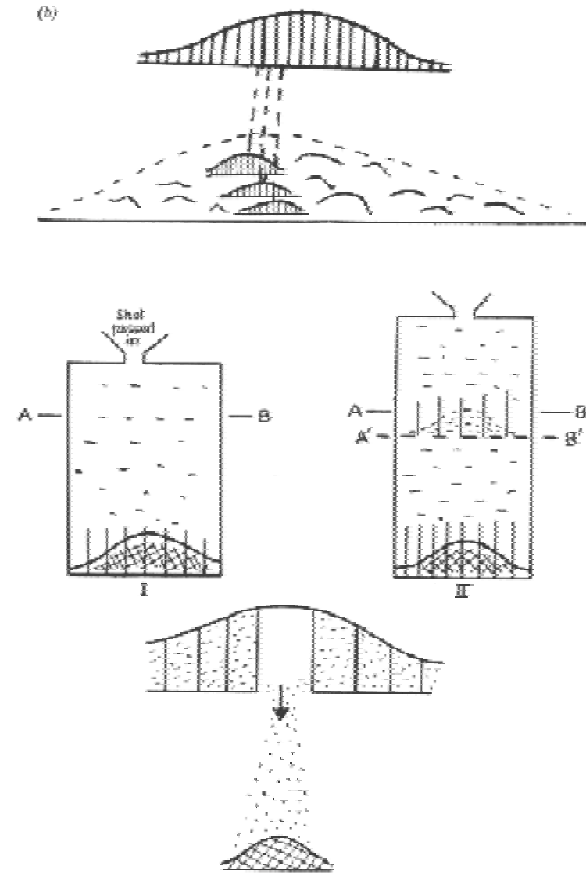
Karl Pearson's drawing of Francis Galton's "quincunx" (around 1873) to show that a normal distribution is normally a mixture of normal distributions.

Thank you for listening