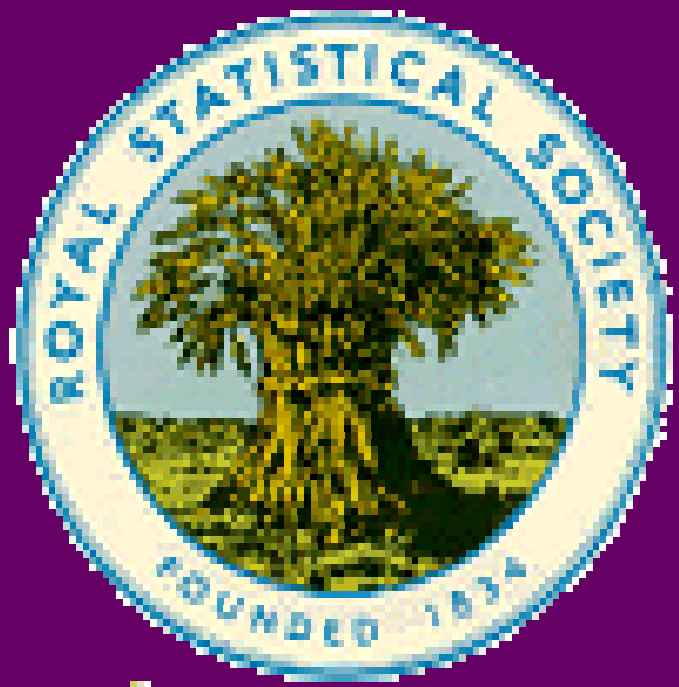
Performance Monitoring in the Public Services



Challenges, opportunities, pitfalls

(failed) Challenge: Performance Monitoring in the Public Services

(missed) Opportunities: Formal Experiments in evaluating new policies

(rescue) Pitfalls: Reporting, Reliance & rMDT

Freedom of Information

Performance Indicators: Good, Bad, and Ugly

Some good examples, but

Scientific standards, in particular statistical standards, had been largely ignored

Royal Statistical Society concern

PM schemes

Well-designed, **avoiding perverse behaviours***,
Sufficiently analysed (context/case-mix)
Fairly reported (measures of uncertainty) &
Shielded from political interference.

***Address seriously criticisms/concerns of those being monitored**

1. Introduction

1990s rise in “government by measurement”

goad to efficiency & effectiveness

better public accountability (financial)

Three uses of PM data

What works? (research role)

Well[/under]-performing institutions or public servants . . . (managerial role)

Hold Ministers to account for stewardship of public services (democratic role)

2. PM Design, Target Setting & Protocol

How to set targets

Step 1 **Reasoned** **assessment of plausible improvement within PM time-scale**

Step 2 **Work out PM scheme's statistical potential**

("power") re this rational target {see p11}

Power matters

Excess power - incurs unnecessary cost

Insufficient power – risks failing to identify effects that matter

Insufficient power – can't trust claims of policy 'equivalence'

How not to set targets **{see p12}**

How not to set targets {p12}

- **Progressive sharpening:** “better of current target & current performance” (ignores uncertainty: prisons)
- **Setting extreme target:** “no-one to wait 4 hours” (abandon)
- **Cascading the same target:** 50% reduction in MRSA within 3 years (most hospitals: < 10 MRSA)

3. Analysis of PM data: same principles

Importance of variability

intrinsic part of real world & interesting per se + contributes to uncertainty in primary conclusions {p15}

Adjusting for context to achieve comparability

{note p17} **incompleteness of any adjustment**

Multiple indicators

resist 1-number summary

(avoid value judgements + reveal intrinsic variation)

4. Presentation of PIs: same principles

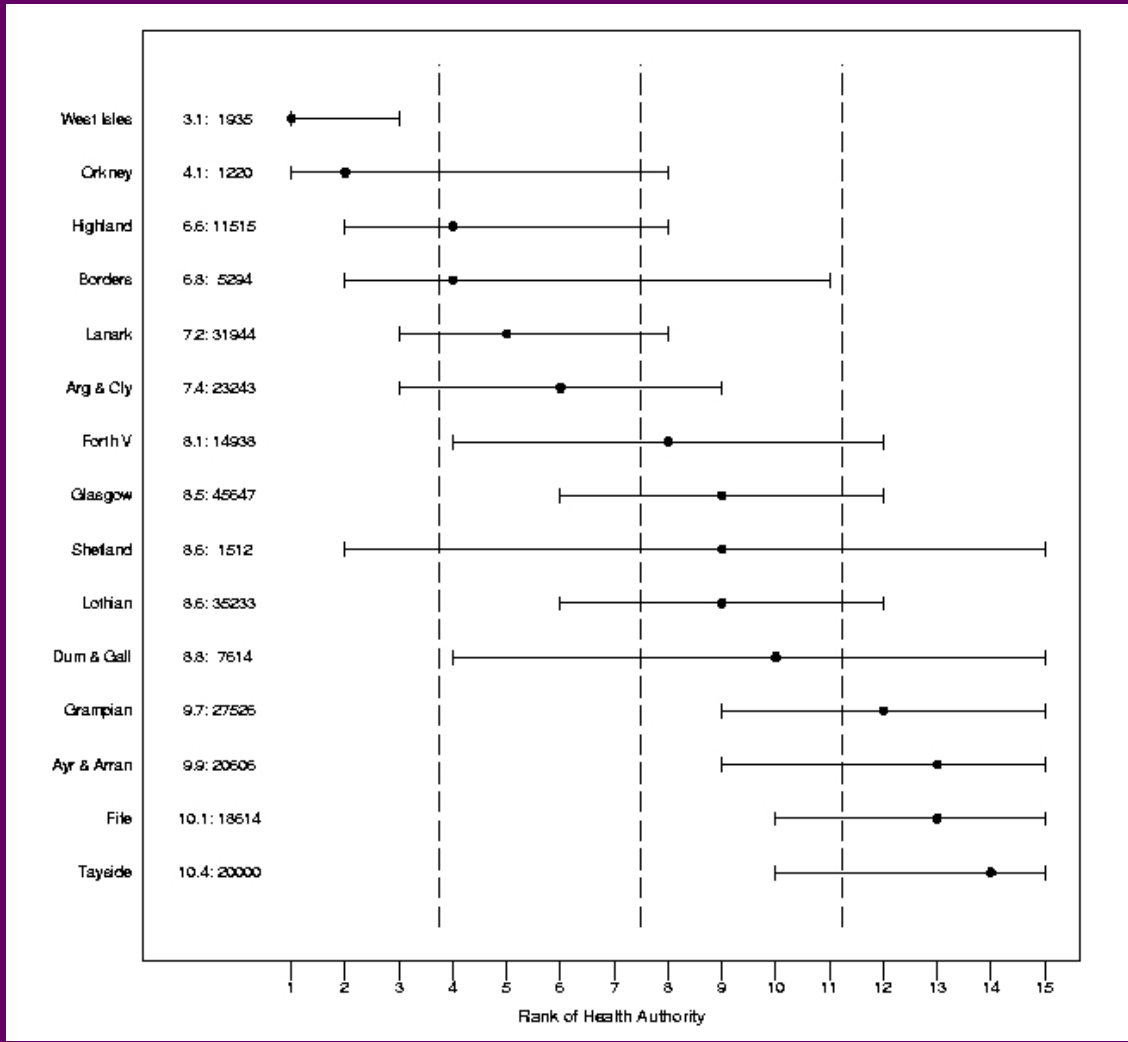
Simplicity =/= discard uncertainty

League tables ⇔ uncertainty of ranking {PLOT 1}

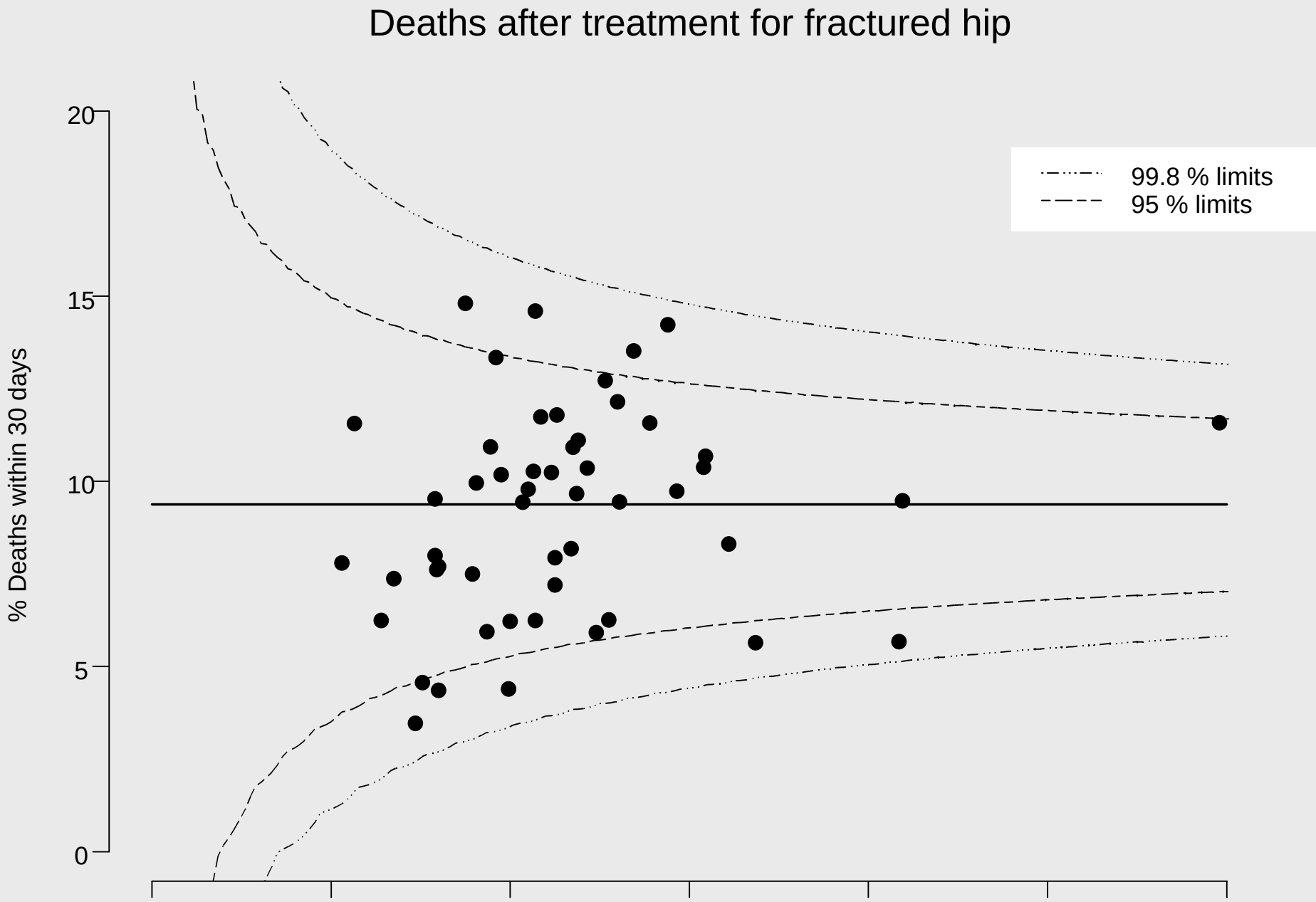
Star 'banding' ⇔ show uncertainty of institution's banding

Funnel plot: variability depends on sample size; divergent hospitals stand out {see PLOT 2}

Plot 1: 95% intervals for ranks

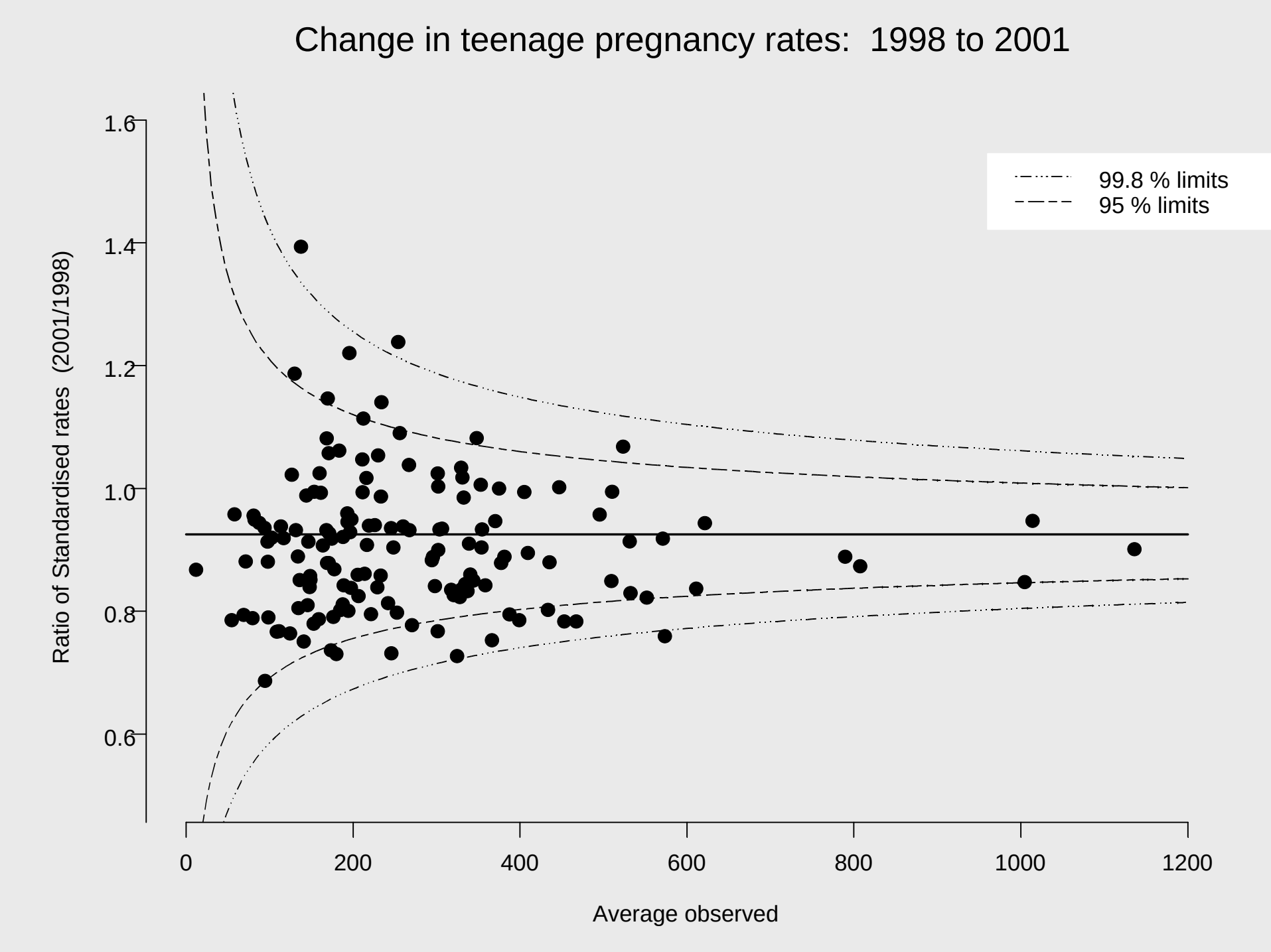


Funnel plot: an alternative to the ‘league table’



Teenage pregnancies

- Government aim to reduce teenage pregnancies
- Target reduction is 15% between 1998 and 2004
- Hope for 7.5% reduction by 2001


Change in teenage pregnancy rates: 1998 to 2001
99.8 % limits
95 % limits
Ratio of Standardised rates (2001/1998)
1.6
1.4
1.2
1.0
0.8
0.6
0
200
400
600
800
1000
1200
Average observed

5. Impact of PM on the public services

Public cost: if PM fails to identify under-performing institutions & so no remedial action is taken

Less well recognised

Institutional cost: falsely labelled as under-performing

Unintended consequences: e.g. risk-averse surgeons

6. Evaluating PM initiatives

Commensurate with risks & costs

How soon to start evaluation

Pre-determined policy roll-out (DTTOs)

Disentangling (several) policy effects

Role of experiments (& randomisation)

“What works” in UK criminal justice?

RCTs essentially untried . . .

Judges prescribe sentence on lesser evidence than doctors prescribe medicines

Is public aware?

7. Integrity, confidentiality & ethics

Integrity (statistical)

For public accountability: PIs need wider-than-government consensus + safeguards, as for National Statistics.

Lacking if: irrational targets, insufficient power, cost-inefficient, analysis lacks objectivity or is superficial.

Royal Statistical Society is calling for

- PM protocols
- Independent scrutiny of disputed PIs
- Reporting of measures of uncertainty
- Research: into strategies other than “name & shame” + better designs for evaluating policy initiatives
- Wider consideration of PM ethics & cost-effectiveness

Application of scientific method

Randomisation: to compare like with like

Adequate study size: for precise estimation

Reporting standards: as in medical journals

Efficacy and costs: rational, prior estimates

Peer **scientific review** of

Study/trial **protocol**

Concept of randomisation

- **Biology,** 1926: Sir Ronald Fisher
- **Medicine,** 1947: Sir Austin Bradford Hill

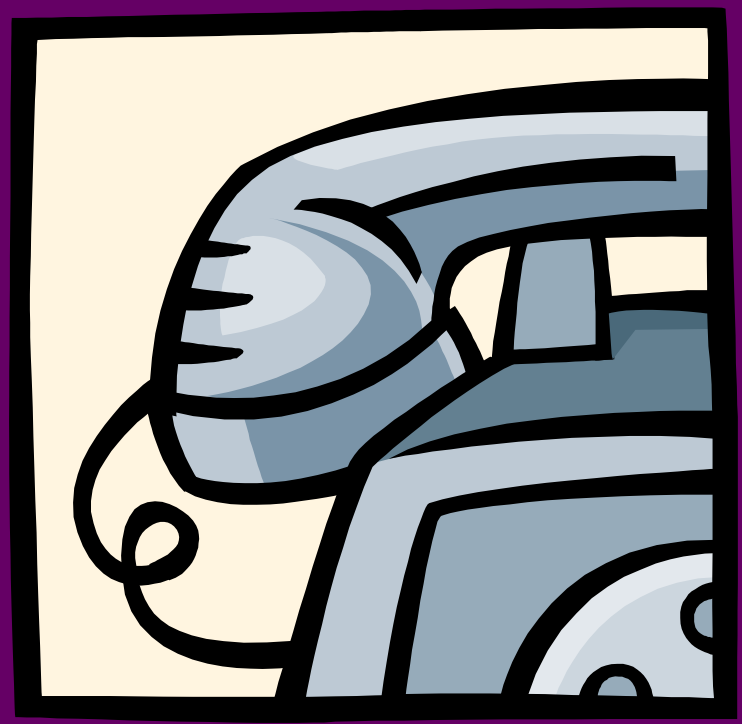
Randomised
Controlled
Trial

- **Criminal justice ?**

Randomisation in medicine

- Toss of coin determines experimental or control treatment ⟹ RCT assignment unpredictable
- Fair *[=> ethical]* allocation of scarce resource
- Balance treatment numbers overall, *in each hospital*, and for major prognostic factors

RCT Telephone randomisation



Experiments & Power matter

Designs for policy evaluations . . .
which respect financial/political constraints

Evaluations-charade

"Public money spent on inferior (usually non-randomised) study designs that result in poor-quality evidence about how well policies actually work"

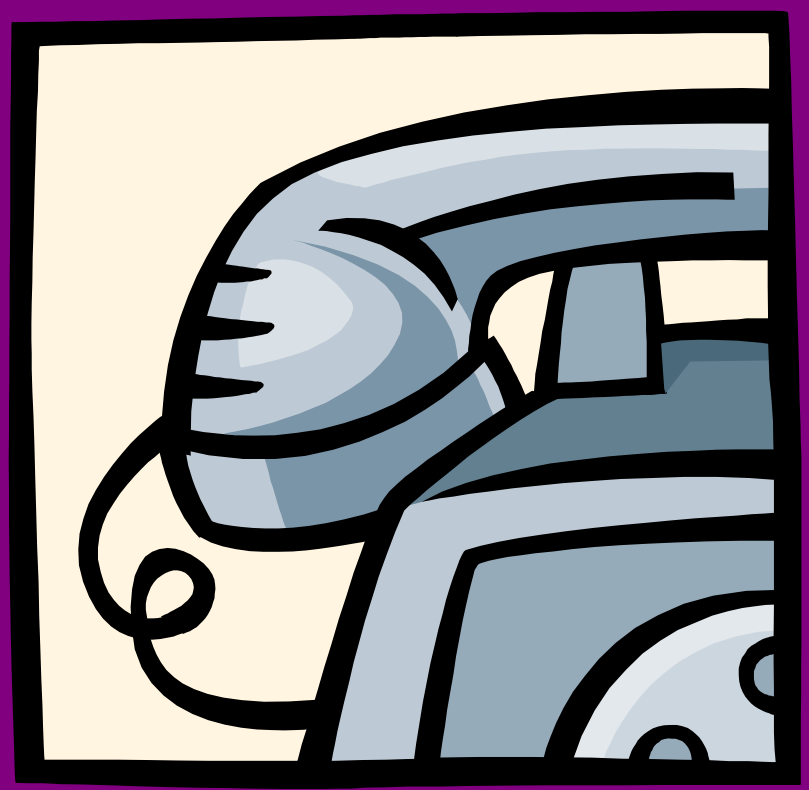
⇔ costly, inefficient by denying scientific method, & a serious loss in public accountability

Missed opportunities for experiments (including randomisation)

Drug Treatment & Testing Orders (DTTOs)

Cost-effectiveness matters!

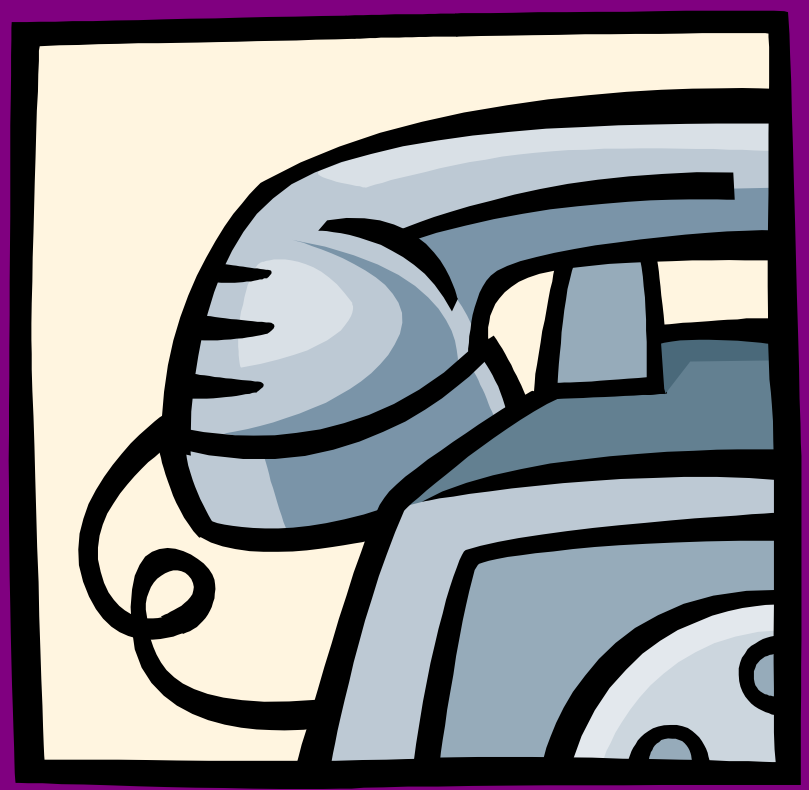
SSRG Court DTTO-eligible offenders: do DTTOs work ?



- Off 1 DTTO
- Off 2 DTTO
- Off 3 alternative =
- Off 4 DTTO
- Off 5 alternative =
- Off 6 alternative =

Database linkage to find out about major harms: offenders' deaths, re-incarcerations & . . .

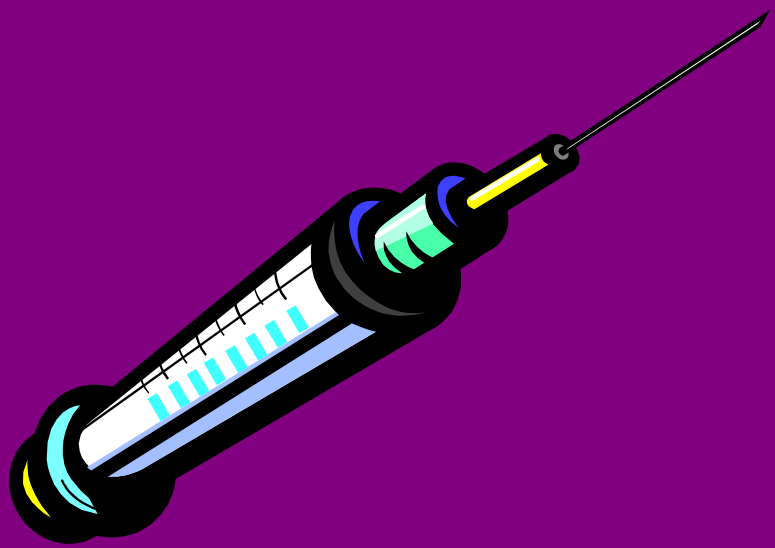
SSRG Court DTTO-eligible offenders: cost-effectiveness ?



- Off 7 DTTO =
- Off 8 alternative =
- Off 9 alternative =
- Off10 DTTO =
- Off11 DTTO =
- Off12 alternative =
- Off13 DTTO =
- Off14 alternative =

Breaches . . . drugs spend?

UK courts' DTTO-eligible offenders: ? guess



- Off 7 DTTO [?]
- Off 8 DTTO [?]
- Off 9 DTTO [?]
- Off10 DTTO [?]
- Off11 DTTO [?]
- Off12 DTTO [?]
- Off13 DTTO [?]
- Off14 DTTO [?]

(before/after) Interviews

versus . . . [?]

Evaluations-charade

- **Failure to randomise**
- **Failure to find out about major harms**
- **Failure *even* to elicit alternative sentence ⇔ funded guesswork on relative cost-effectiveness**
- **Volunteer-bias in follow-up interviews**
- **Inadequate study size re major outcomes . . .**

Power (study size) matters!

Back-of-envelope sum for 80% power

Percentages

Counts

If MPs/journalists don't know,

UK plc keeps hurting

For 80% POWER, 5% significance: comparison of failure (re-conviction) rates

Randomise per treatment group, 8 times

STEP 1 answer =

$$\frac{\text{Success * fail rate for new disposal} + \text{Success * fail rate for control}}{(\text{success rate for new} - \text{success rate for control})^2}$$

DTTO example: TARGET 60% v. control 70% reconviction rate?

Randomise per 'CJ disposal' group, 8 times

STEP 1 answer =

$$\frac{40 * 60 + 30 * 70}{(40 - 30)^2} = \frac{2400 + 2100}{100}$$

DTTOs control

Five PQs for every CJ initiative

- **PQ1: Minister, why no randomised controls?**

- **PQ2: Minister, why** have judges ***not even*** been asked to document offender's **alternative sentence** that this CJ initiative supplants {re CE}?
- **PQ3: What statistical power** does Ministerial pilot have re **well-reasoned targets**?

  {or just kite flying . . .}

- **PQ4: Minister, cost-effectiveness is driven by longer-term health & CJ harms**, how are these ascertained {⇔ database linkage}?
- **PQ5: Minister, any ethical/consent issues?**

BBC NEWS UK EDITION

“If I had 50p for every prisoner that was liberated in error by the Scottish Prison Service and the police when they were doing the job I think I'd be quite a rich man”

Reliance: PIs, thresholds, penalties?

Performance Indicator: severity	Expect monthly	Thresholds >A 0.02	>B 0.001	BLANKED OUT *Appeal*
Late delivery 10	58 SMB	74	82	Scotland's Commission
Key compromise 5	5	9	12	Freedom of Information
Prisoner disorder 3+ 6	5	9	12	
Serious assault 6	2.5	5	8	
Self-harm 3	5	9	12	Confidential Clause . . .
Overdue response 1	50	60	71	

Random Mandatory Drugs Testing of Prisoners: rMDT

Home Affairs Select Committee Inquiry, 2000

ONS contract from Home Office, 2001

Final report, 2003

With Minister . . . raised with **National Statistician, Statistics Commission, 2004**

Publication? . . . Freedom of Information!

Disputed PI: costly, potential danger, impact on parole, underestimates inside-use of heroin, human rights . . .

Restorative Justice: Youth Justice Board

46 Restorative Justice projects with about 7000 clients by October 2001: Evaluation report for YJB, 2004

“to let 1000 flowers bloom . . . “

Satisfaction rates by victim & offender typically high

(both having been willing for RJ?

eligibility for, & response rate to, interviews?)

YJB Targets: RJ used in 60% of disposals by 2003, & in 80% by 2004; 70% victims taking part to be satisfied!

Specific Recommendations

Royal Statistical Society

Working Party on Performance Monitoring in the Public Services

Royal Statistical Society: 11 Recommendations

1. PM procedures need detailed protocol

2. Must have clearly specified objectives, achieve them with rigour; & input to PM from institutions being monitored

3. Designed so that counter-productive behaviour is discouraged

4. Cost-effectiveness given wider consideration in design; & PM's benefits should outweigh burden of collecting quality-assured data

5. Independent scrutiny – as safeguard of public accountability, methodological rigour, and of those being monitored

Royal Statistical Society: 11 Recommendations

6. Major sources of variation - due to case-mix, for example – must be recognised in design, target setting & analysis

7. Report measures of uncertainty: always

8. Research Councils: to investigate range of aspects of PM, including strategies other than “name & shame”

9. Research: into robust methods for evaluating new government policies, including role of randomised trials . . . *In particular, efficient designs for ‘roll-out’ of new initiatives*

Royal Statistical Society: 11 Recommendations

10. Ethical considerations may be involved in all aspects of PM procedures, and must be properly addressed

11. Wide-ranging educational effort is required about the role and interpretation of PM data

Scotland's Airborne score-card: 11/11 . . . wrong!

Statistician's role in PM: both

Strenuously to safeguard from misconceived reactions to uncertainty those who are monitored

Design effective PM protocol so that data are properly collected, *exceptional* performance can be recognised & reasons further investigated

⇔***Efficient, informative random sampling for inspections***

(PM) Protocol

- Assumptions / Rationale in Choice of PI
- Objectives
- Calculations (power) & consultations + piloting
- Anticipated perverse consequences + avoidance
- Context/case-mix + data checks
- Analysis plan & dissemination rules
- Statistical performance of proposed PI monitoring + follow-up inspections
- PM's cost-effectiveness?
- Identify PM designer & analyst to whom queries . . .