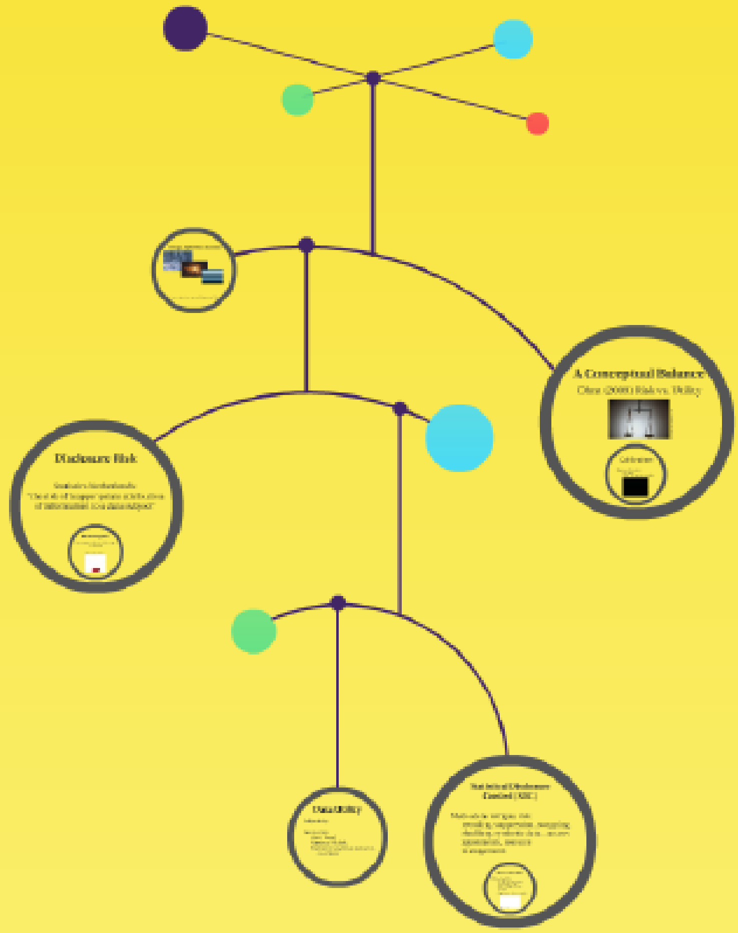
Examining Disclosure Risk and Data Utility:

An Administrative Data Case Study

Michael Comerford

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

University of Glasgow



Deluge, Explosion, Tsunami

photo credits: thewritingroom, The Guardian, IUGG

A Conceptual Balance

Ohm (2009) Risk vs. Utility

pic:flickr.com/sepehrehsani

Calibration

More sticks than carrots
- Penalties?
- Benefits?... are they tangible?

Calibration

More sticks than carrots

Penalties?

Benefits?... are they tangible?

pic:flickr/kaptainkobold

Disclosure Risk

Statistics Netherlands:
"the risk of inappropriate attribution of information to a data subject"

Measuring Risk

k-anonymity + t-closeness + l-diversity
= **n-confusion**

eScience Perspective

Measuring Risk

k-anonymity + t-closeness + l-diversity
= n-confusion

eScience Perspective

Statistical Disclosure Control (SDC)

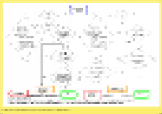
Methods to mitigate risk:

recoding, suppression, swapping, shuffling, synthetic data… access agreements, resource management.

SDC in Practice

NHS Scotland ISD

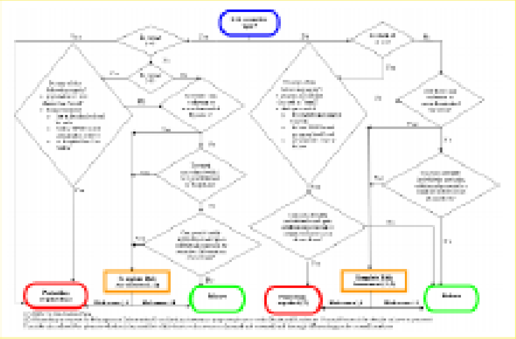
- Recoding/Suppression almost exclusively
- Ad hoc application by analysts
- Few systematic approaches



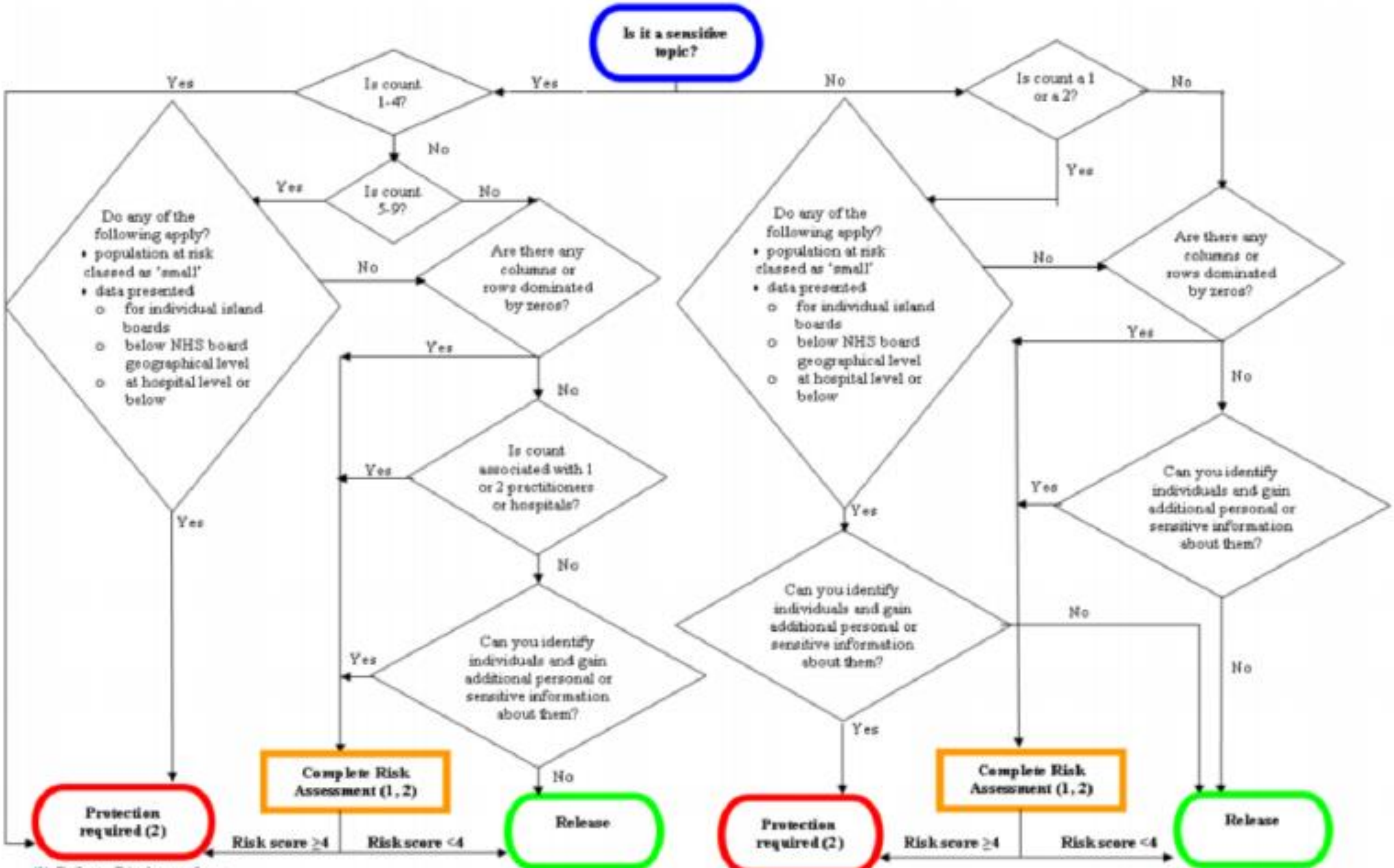
SDC in Practice

NHS Scotland ISD

- Recoding/Suppression almost exclusively
- Ad hoc application by analysts
- Few systematic approaches



NHS Scotland Information Services Division



(1) Refer to Disclosure form
(2) If handling a request for Management Information/Data Quality Assurance purposes please see the Statistical Disclosure Control Protocol for details on how to proceed
Consideration should be given to whether it is possible to link data to other sources (internal and external) and through differencing produce small numbers

Data Utility

Subjectivity

Approaches

- Query based
- Statistical Models
- Real-world analytical scenarios
  - Case Study