# Pseudonymous Data for Research

## and the draft EU Data Protection Regulation

Peter Singleton

BCS, Southampton Street, London

22nd October 2013

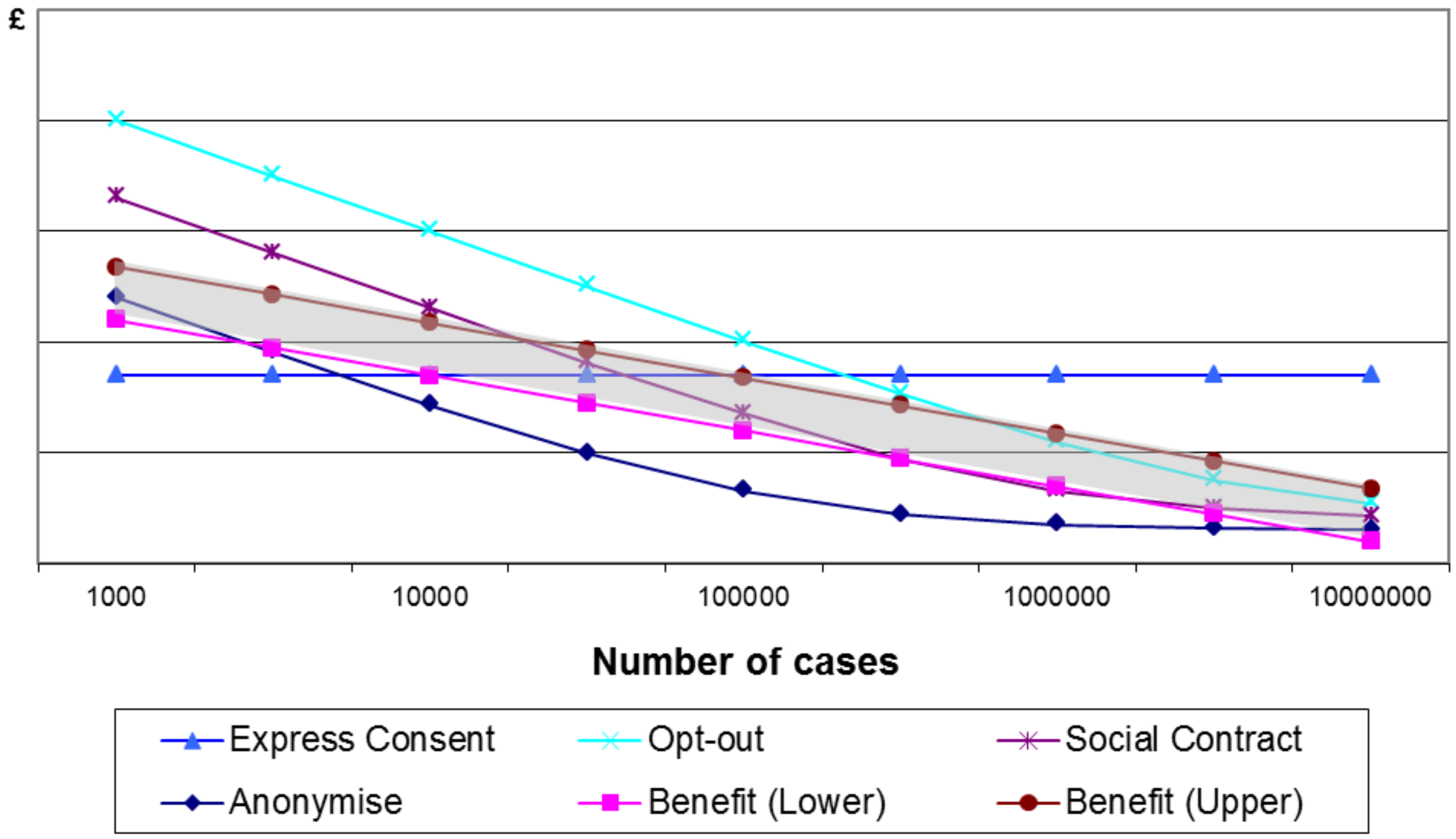
# The Issue

- Current DP Directive (and draft DP Regulation) makes no clear distinction between directly identifiable data and nearly anonymised data
- ‘Secondary uses’ may require further consent before data can be used
- This requirement may prevent much intelligent analysis of the data to improve safety and quality of services and products, particularly in medical research – where need for better use of available data is crucial

# How far should privacy rights run?

- Current EU DP Directive:
  - covers any data that might possibly identified with a person
  - Includes individual rights to see copy and correction
- Proposed EU DP Regulation:
  - Adds individual rights to electronic copy and ‘to be forgotten’
  - May take ‘risk-based approach’
- Should individuals have right to control low-risk uses of data when used for analytic purposes, such as quality & safety, research, audit, etc.?

# Cost/Benefit per case



From: Singleton, P and Wadsworth, M: ‘Practical aspects of obtaining consent for the use of personal medical data in research’, British Medical Journal, Jul 2006; 333: 255 – 258

# Data

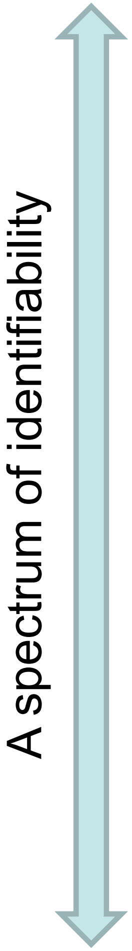
- ‘Non-personal’ data
- ‘Personal data’:

  any information relating to an identified or <u>identifiable</u> natural person – or similar – varies by legislation
- Anonymised data:

  not specifically defined, except as not ‘personal data’
- ‘Pseudonymous data’:

  data where identity has been ‘hidden’ but not necessarily ‘anonymous’

# Identifying Data



A spectrum of identifiability

- **Identified data** – no effort required – someone looking at the record would know (to a reasonable level of certainty) to whom the record related
- **Readily identifiable** – there is an obvious and reliable method by manual or data-set look-up which is easily accessible, e.g. telephone directory, simple Internet search
- **Practically identifiable** – there is a clear and fairly reliable method to re-identify most of the records in a data-set using one or more other sources which may be available (perhaps requiring subscription)
- **Theoretically identifiable** – there is a clear and fairly reliable method to re-identify most of the records in a data-set using one or more other sources which may or may not be available (indeed may be restricted or secret)
- **Not re-identifiable** – there is no clear or reasonably reliable method of re-identifying records or depends on a data-set that no longer exists (e.g. pseudonyms have been destroyed and no further copies exist) – this would not necessarily guarantee that no records might individually be re-identifiable on an ad hoc basis or because of specific peculiarities of the data
- **Anonymised data** – cannot be re-identified, except possibly by unreasonable amount of effort

# Privacy-protected data
## aka ‘pseudonymous data’

- Identity is hidden – data not readily identifiable
- Access is restricted to controlled environments, including confidentiality and non-re-identification clauses
- Adequate security (as may be personal data)
- Re-identification processes may exist but restricted to specific circumstances and controlled & monitored
- Data then free from ‘privacy rights’, including consent for further re-use
- Still need to inform (in broad terms) and uses should be ‘not incompatible’ with original purpose of collection

# Future steps in resolving this challenge

- **Concrete next steps**: Develop definitive paper to prove issue and develop solution/ approach.
- **Who and with whom**: EU DP Supervisor, Article 29 DPWP, UK MoJ, EU Parl Rapporteur.
- **When and where**: Need paper by end of year, or to inform future plans if Regulation fails to meet deadline
- **BCS and/or IFIP involvement**: Help tighten definition; elaborate ethical, legal, and practical issues.