Understanding information governance

Access to person-level data in health care

Research summary

Benedict Rumbold, Geraint Lewis and Martin Bardsley

How we govern the use of health care data is of growing concern to members of the public, researchers and policy-makers alike. In recent years, the quality and volume of data recorded about individual patients has increased considerably. Rules and regulations help to protect individual privacy and confidentiality. However, overly restrictive regulations can also stifle valuable research and analysis. The Nuffield Trust recently commissioned a review, *Information Governance in Health*, to explore the current regulations and reflect on the social values that underpin them. This paper summarises the report's findings and discusses them in the context of current government policy.

Key points

- At present, the use of data collected on individual users of NHS-funded care is determined by several laws, including the Data Protection Act 1998 and the National Health Service Act 2006. It is also governed by various interpretations of those laws as set out in guidelines issued by a number of organisations, including the Information Commissioner's Office and the National Information Governance Board.
- A number of 'social values' are relevant to the issue of information governance. These are: consent, privacy, autonomy, property rights, confidentiality, public benefit and fairness or reciprocity.
- Currently, information governance arrangements are usually interpreted as requiring data analysts either to obtain explicit consent from individuals for the use of their data, or else they must render all person-level data completely anonymous.
- The current framework does allow the gatekeepers of NHS data to weigh up the rights of the patient to privacy and consent against the rights of the public to benefits from research, and grant the use of data without the consent of each individual.
- One advantage of this 'public benefit' approach is that it is able to deal with studies that do not fit into the typical 'consent or anonymise' model such as those involving pseudonymous data and data linkage.
- Given that the UK data governance policies are currently in a state of flux, there is an opportunity to advance the 'public benefit' approach to information governance. This may lead to benefits to researchers, patients, and the public alike.

Background

Like all health care systems, the NHS generates a large amount of administrative data recorded at the individual patient level. This includes data about hospital activity (such as outpatient appointments and hospital admissions), diagnostic information and clinical information (such as blood test results and X-ray findings).

These datasets perform various roles in the delivery, financing and planning of health care, from organising payments and booking appointments to generating useful intelligence at a regional or national level. For example, in order to monitor the use of medicines in primary care, the Prescribing Support Unit at the NHS Information Centre analyses hospital prescribing data and samples of identity-protected patient records derived from a number of GP practices. Similarly, the NHS Information Centre is working with the National Cancer Intelligence Network to develop a new Cancer Outcomes and Services Dataset; the aim of this will be to provide the NHS with information about the incidence, mortality and survival rates of different cancers.

Data relating to the population's use of health services can be considered under three broad headings based on how identifiable they are: fully identifiable data, pseudonymous data and anonymous data (see Box 1).

Box 1: Types of individual-level data

Identifiable data: allow for the identification of individual patients, through the inclusion of information such as names, addresses and dates of birth.

Anonymous data: have been stripped of any elements that would make identification possible such as names, addresses and dates of birth.

Pseudonymous data: have had all personal identifiers (e.g. names, addresses, dates of birth, NHS numbers) removed but each individual has been allocated a unique code number (or 'pseudonym'). Although anonymous for all intents and purposes, the pseudonym enables analysts to link together information relating to a particular individual from multiple databases. It also allows the potential, under certain circumstances, for the manager of the database to re-identify each individual at a future time, usually via a 'key' that decodes the pseudonym back into the NHS number. In this sense, pseudonymous data are neither identifiable nor anonymous because all personal identifiers have been removed but identification is still possible through the pseudonym and the key.¹

Recent improvements in the quality, quantity and accessibility of person level data and advances in data linkage techniques have led to a growth in research studies using pseudonymous data from multiple databases. For example, the Department of Health has funded an analysis led by the Nuffield Trust to develop a person-based formula for allocating commissioning budgets for all general practices in England (Nuffield Trust, 2011). In recent years, the Department of Health in England has also funded the development of two predictive models (the 'PARR model' and the 'Combined Model')

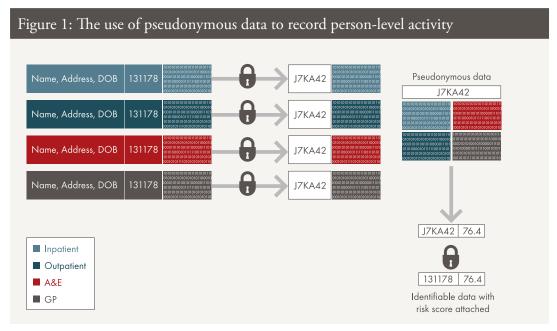
^{1.} In the USA a legal distinction is made between fully and partially pseudonymous data. 'Fully pseudonymous data', are data from which 18 specific identifiers (such as names and dates of birth) have been removed and from which the risk of re-identification is extremely limited. 'Partly pseudonymous data' are data that may contain some identifying information but are made available to users subject to contractual and technical limitations, and with the agreement that no attempt will be made to identify individuals. (House of Commons Health Select Committee, 2007)

which use pseudonymous administrative data to make predictions about future care needs at the individual level. In these models, predictions about each patient's expected future needs are established using pseudonymous information, preventing any



the Department of Health has funded an analysis, led by the Nuffield Trust, to develop a person-based formula for allocating commissioning budgets

possible identification of the individual by the analysts processing the data. Once these predictions have been made, only the patient's own GP is able to identify their patients' risk scores, using the key to convert the pseudonyms back into real patient identifiers. In this way, GPs are able to determine whether an 'upstream' intervention, such as extra support from a community matron, could benefit a particular patient (see Figure 1).



From left to right, identifiable data from multiple sources about the same individual (whose NHS number is 131178) is given a pseudonym (J7KA42) through an encryption key; analysts then calculate a predictive risk score (76.4) for the pseudonym. The pseudonym is then de-encrypted, letting the GP see the predictive risk score of the individual concerned.

An advantage of using pseudonymous data in applications such as these is that it allows data analysts to link records across different data sets, while at the same time protecting patient information by rendering it effectively anonymous. As such, pseudonymous data are particularly useful in data linkage studies, where analysts need to join up multiple sources of information for the same individual, but do not need to know any personal identifiers for each person concerned.

Given the growing importance of these kinds of analyses, the Nuffield Trust asked Sarah Clark and Albert Weale from University College London to examine the various laws and standards that govern the use of these kinds of data usage, and to tease out the different social values that underpin them. The aim was to understand not only the foundations of the current governance arrangements, but also how new standards might be developed to facilitate data linkage studies that could ultimately improve patient care.

Current governance arrangements

At present, laws regulating the use of NHS data include:

- *The common law duty of confidentiality* places data guardians under a duty of non-disclosure when information is disclosed between two parties in a situation of confidentiality (e.g. the doctor–patient relationship).
- Data Protection Act 1998 establishes a series of principles by which personal data must be processed, maintained and transferred if such use of data is to be considered fair, lawful and proportionate.
- *Human Rights Act 1998* enshrines an individual's right to respect for their private and family life.
- Freedom of Information Act 2000 enables people to access information held by public bodies.
- Section 251 of the National Health Service Act 2006 (previously Section 60 of the Health and Social Care Act 2001) – provides a power to allow patient identifiable information needed for 'medical purposes' (defined as including medical research and the management of health and social care services), to be used without the consent of patients.

Together, these laws provide the basic governance framework for the use of data in the UK. The laws set out both the rights of patients, such as the right to privacy, and the responsibilities of NHS organisations and researchers, such as the requirement to obtain consent before certain kinds of data are gathered or used.



these laws provide the basic governance framework for the use of data in the UK

However, taken on their own, they do not tell the whole story about the governance of NHS data. Rather, researchers and analysts also need to navigate various interpretations of these laws, as well the additional guidelines issued by number of different statutory, regulatory and professional bodies. Such bodies include:

- The Information Commissioner's Office (ICO) enforces and oversees the Data Protection Act, the Freedom of Information Act, the Environmental Information Regulations, and the Privacy and Electronic Communications Regulations.
- The National Information Governance Board (NIGB) provides advice on the
 appropriate use, sharing and protection of patient and service user information. For
 example, the NIGB advises on the use of powers under Section 251 of the NHS Act
 2006 to permit the duty of confidentiality to be set aside if other legal routes are not
 available.
- The Caldicott Guardians senior staff working in the NHS and social services who
 are appointed to safeguard the confidentiality of patient/client information and to
 oversee information sharing arrangements.
- *The National Institute for Health Research* develops administrative procedures associated with the regulation and governance of research, including approvals, research administration and reporting arrangements.

In addition, anyone seeking to analyse health service data may also need to consider various professional guidelines. These include guidelines produced by the General Medical Council (GMC; see GMC, 2009), the British Medical Association (BMA; see BMA, 2007), the Medical Research Council (MRC; see MRC, 2000) and the Department of Health (see Department of Health, 2003).

This picture is further complicated by changes over time in the responsibilities of the different bodies and the standards and guidelines that they publish. For example, following an announcement by HM Government in October 2010, the NIGB may be disbanded. Under current proposals, the organisation is to be integrated within the Care Quality Commission (provisionally by 2013), whereas its responsibility for providing advice (e.g. on appeals to Section 251 of the NHS Act 2006) is to be given to the Department of Health and related bodies such as the proposed NHS Commissioning Board. Meanwhile, the Information Commissioner's Office (ICO) has recently launched a new statutory code of practice for data sharing, 'providing a framework for organisations to make good quality decisions about data sharing', with particular reference to the 1998 Data Protection Act (ICO, 2011).

The social values behind governance arrangements

Underlying the legislation and debates about information governance are a set of seven 'social values', namely:

- *Consent* the idea of a waiver, whereby the individual agrees to the use of their information in a way that would otherwise be impermissible.
- *Privacy* the value we place on a sense of separateness from others and 'being apart from'.
- *Autonomy* the claim that it is important to have control over at least a significant portion of one's life.
- *Property rights* a sense of ownership, for example of health data.
- *Confidentiality* the idea that confidents are under obligation not to disclose the information entrusted to them, and in particular not to use it against them to harm or otherwise disadvantage the confider.
- *Public benefit* the claim that there is a strong interest for 'the public' as well as individual patients, in the advances made by medical and health services research.
- Fairness or reciprocity the idea that patients are under an obligation to make data about them available for research as a matter of reciprocity, to help produce goods that are beneficial to all, including themselves.



part of what determines the current governance framework is the importance that we, as a society, place on these values According to this analysis, part of what determines the current governance framework is the importance that we, as a society, place on each of these values, and how we think the balance between them should be struck. For example, the governance framework insists that data analysts seek the consent of individuals: many in society think that it is wrong to use certain kinds of data without first seeking the approval or agreement of the person whose information appears in the data. In a similar way, information governance standards are shaped by the value that our society places on 'privacy': we restrict access to information about a person partly because we perceive certain unauthorized uses of our data as an invasion of our 'informational privacy'. The same line of reasoning could be applied to other social values such as 'autonomy', 'property rights' and 'confidentiality'.

However, if notions such as 'consent' and 'privacy' explain why we limit the availability of data, other social values – in particular 'public benefit' and 'fairness' – show why we think it is important, in certain circumstances, to allow wider access to those data. One example of these social values in use is Schedule 2 of the Data Protection Act. According to this provision, researchers are able to derogate from the standard requirements about processing sensitive data so long as certain safeguards are employed and that the research is in the 'public interest'. This provision reflects society's belief that, in certain circumstances, the public interest in the benefits gained from allowing access to health service data outweighs the private interests of patients in informational privacy.

'Consent or anonymise': balancing private rights with public benefits

The standard interpretation of the current information governance arrangements is that researchers and analysts using health service data must either seek explicit consent or only use anonymous data. In cases where this is impossible, it may be possible to obtain permission to conduct research without explicit consent from the National Information Governance Board using a provision in section 251 of the NHS Act 2006 (previously section 60 of the Health and Social Care Act 2001).

The 'consent or anonymise' approach appeals to many of our social values, yet it also has some serious inadequacies. In many cases, researchers cannot use completely anonymous data and yet cannot feasibly seek consent. This is particularly the case in studies involving very large data sets involving thousands or millions of records.

Box 2: A case study in the use of identifiable patient data without individual informed consent*

In 2006, clinicians in the West Midlands identified that delays in the appropriate referral of patients with anaemia might be resulting in delayed diagnoses of colorectal cancer. Colorectal cancer may cause bleeding into the bowel, resulting in anaemia. So a delayed diagnosis of anaemia may lead to worse cancer survival rates due to delayed diagnosis of the underlying cancer.

Researchers in Birmingham proposed conducting an epidemiological study without explicit consent using routinely collected data. The researchers felt this was the only robust method of describing current patient pathways and of identifying whether suboptimal care exists. They developed a research proposal, which they subjected to peer review. The proposal was considered to be of sufficient importance and of a high enough methodological quality to be supported financially by Cancer Research UK in 2007.

However, before granting permission for the study to commence, the Patient Information Advisory Group (PIAG, the predecessor of NIGB) required the researchers to produce evidence that "...seeking consent would lead to significant data bias". So they conducted a survey of 600 general practice patients with anaemia, obtaining a 70 per cent response rate. 7.8 per cent of responders stated that they would not give consent for their records to be used. Those respondents who said they would refuse to provide consent tended to be younger, female, and from more deprived areas. But they also differed in other, unpredictable ways from those patients who said they would provide consent. This unpredictable variability prohibits statistical adjustment of anonymous datasets. Moreover, other *ad hoc* surveys of cancer patients have demonstrated strong support for the use of their records without seeking consent.

In response the NIGB stated that "...7.8 per cent had actively dissented. This was important as it is a condition of approval for all approved applications that Section 251 support cannot be used to override dissent." Yet the researchers needed to analyse the data of all patients if they were to understand better the subsequent incidence of cancer, describe current care pathways, establish the time to cancer diagnosis and, ultimately, identify the reasons for delays in diagnosis and develop interventions to reduce these delays.

The NIGB decision required the researchers to provide further data relating to the effect of bias on each of their study objectives, and only then might it "...be accepted by the Committee that consent is not practical to achieve some or all of the objectives of the study..." More than four years after identifying a potential means of improving the early diagnosis of colorectal cancer, this research has still not been undertaken because of the information governance hurdles encountered.

^{*} The research team consisted of Professor Sue Wilson, Dr Sarah Damery, Sally Warmington and Professor Heather Draper from the Department of Primary Care Clinical Sciences, University of Birmingham

Equally, the third option given by the 'consent or anonymise' approach, namely to seek a specific exemption from the NIGB, is currently poorly articulated in law, leaving the NIGB the unenviable task of interpreting existing legislation in respect of various cases and with limited resources. In the worst case scenario, important research is delayed while rival interpretations of the law are debated (see Box 2).

However, there is an alternative interpretation of existing standards and guidelines to the traditional 'consent or anonymise' approach. For example, data governance could be viewed as a matter of weighing up certain values to individuals, such as privacy and consent, against the value that the research may generate to the public.

This alternative interpretation is poorly articulated in law at present. As Sarah Clark and Albert Weale explain, if data analysts are to use these arguments of public benefit to justify their use of data, they will need more clarity about what constitutes a public benefit, and, "some form of argument by which the obligations arising from rights and the benefits of public action can be brought into balance". Among other things, a public benefit interpretation of the law would also have to be sensitive to the fact that the 'benefits' of the research may not be evenly distributed. Indeed, the findings may even be felt as harms by certain groups. For example, research might establish that certain population groups are particularly prone to certain diseases (say, the propensity of African—Caribbean populations to hypertension and diabetes). As a result, members of those populations could potentially feel stigmatised by widespread knowledge of information about the entire group's health risks.



data analysts need more clarity about what constitutes a public benefit

Provided the law can be clarified, though, there is the potential for such interpretations to be of great use to data analysts, and ultimately patients, especially in relation to the use of pseudonymous data. Under this 'public benefit' approach, analysts would need to demonstrate:

- that their research is in the public's interest
- that individuals' rights, such as their right to privacy, are protected to the maximum feasible degree
- that the benefits produced by the research are highly likely to outweigh any remaining infringement of the subject's right to privacy
- that the analysts will only use the data subject to a range of safeguards including, for example, restrictions on how many analysts have access to the data.

Provided that they met these criteria, data analysts would be free to use the data without first obtaining explicit consent from data subjects and without stripping it of *every* element that would make identification theoretically possible.

The current policy agenda

Sarah Clark and Albert Weale's analysis provides a powerful argument for expanding and clarifying the 'public benefit' interpretation of information data governance. Their report comes at a crucial moment in the development of information governance policy in England. At present, the government is bringing forth numerous reforms on how health care information is produced, accessed and researched, most notably in *Liberating the NHS: An Information Revolution* (Department of Health, 2010). One ambition is to give patients and the public greater access to information about the health service, and to their own health records. The proposals also include changes that could potentially help facilitate the analysis of health service data (see Box 3). These include providing "a high quality research data service", which could make "aggregated, anonymised data" from electronic care records available to universities and other "authorised researchers". However, following a consultation period which closed in January 2011, these plans are on hold until Autumn 2011.

More recently, the Wellcome Trust has urged the government to go even further on this issue. In its response to the consultation on the White Paper, the Wellcome Trust encouraged the Government to include within *An Information Revolution* a statement that "improved research uses of both anonymous and identifiable patient data [should be] a desired outcome in the final strategy" (Wellcome Trust, 2011). This undertaking would include implementing the recommendations of the 2008 Data Sharing Review undertaken by Richard Thomas (the then Information Commissioner) and Mark Walport (Director of the Wellcome Trust), as well as addressing other barriers to accessing patient data (Wellcome Trust, 2011).



a new service will support researchers by linking together data from various sources

Other changes afoot in data governance include 17 recommendations by the Academy of Medical Sciences for the future regulation and governance of clinical research in the UK (Academy of Medical Sciences, 2011). The Academy's key recommendation is the formation of a new health research agency with two main functions – first to streamline all the current arrangements for ethical approval, and second to provide a national research governance service. This new agency would thereby assume responsibility for the national research ethics service, as well as administering specialist ethical approvals and licences, leaving individual NHS organisations to focus on the assessment of local research (Smyth, 2011).

As part of its Research Capability Programme (RCP), the National Institute for Health Research is currently piloting a new Health Research Support Service (HRSS). This new service will support researchers by linking together information from data sources such as GP patient records and hospital patient records. The NHS Information Centre already offers a 'secondary uses service', which is essentially a repository of hospital data for use in care planning, policy development, performance management, clinical audit and medical research.

Box 3: What the 'information revolution' would mean for data analysts

By making NHS data more accessible to the public, the plan would make NHS data more accessible to researchers. According to *An Information Revolution*, there will be a 'presumption of openness' across health and adult social care. NHS and adult social care organisations will therefore be obligated to deliver "the fullest possible public access to the information they hold as soon as possible, adopting a 'publish and improve' rather than a 'polish and publish' approach". The original report suggested that the following data would be released by April 2011:

- inpatient information at provider level (annually from November 2010)
- NHS Choices Provider Quality Indicators (November 2010)
- outpatient information at provider level (annually from December 2010)
- maternity information at provider level (annually from November 2010)
- Accident and Emergency additional tables (January 2011)
- ambulance status reports (weekly, from February 2011)
- inpatient information at national level for procedures and diagnosis (by April 2011).

In addition, from 2011/12 onwards the following data are set to be released:

- cancer registries and datasets
- information about mixed-sex hospital accommodation
- National Clinical Audit
- NHS Choices directories
- NHS Reference Cost data
- Financial Information Management Systems (FIMS) data
- reference data (e.g. population and demographic data) in more usable forms.

The Information Centre (IC) is set to become a "single, national repository for data collected from NHS and social care organisations". This will include taking on a role in making health and adult social care data more accessible, and will include publishing aggregated data in standard formats. The IC will seek to minimise bureaucracy and eradicate duplication in the data that are collected. It will also routinely publish the aggregate data it collects "in easy-to-use formats so that organisations presenting information to the public, health and social care organisations, researchers, professional bodies, policy makers and others can access this [sic] data for a variety of purposes". The IC will have "the lead responsibility in assuring the accuracy and quality of aggregate data collections against the approved informatics standards set by the NHS".

Box 3: What the 'information revolution' would mean for data analysts

The plan is likely to lead to changes in the kinds of data gathered by the health service and other bodies. According to the plan, the starting point of the 'revolution', will be data that are already collected by the NHS. However, the strategy also sets out proposals to undertake "fundamental reviews of NHS and adult social care data returns". This would be part of a wider examination of "information already circulating within the system to determine whether it is the right information to be capturing". Following these reviews, one change might be an expansion in the collection of Patient Reported Outcome Measures (PROMs) to encompass a much larger set of patients.

The plan also sets out measures to draw together data and intelligence from different sources in order to understand "local needs and preferences". The proposed new public health service Public Health England will play a "critical role" in this respect, bringing together "public health information to support local authorities, commissioners and policy-makers". Equally, information will be shared between local authorities and local HealthWatch programmes, in order "to provide accountability to local people and communities".

While these examples might suggest a broad agreement with the call for a more open and balanced approach to data governance, there are equally many individuals and

organisations that defend the current 'consent or anonymise'
approach. For example, in response to the recent recommendations made by the Academy of Medical Sciences, Brown and colleagues

wrote:

21%

of respondents to an Ipsos MORI poll thought it was never acceptable for researchers to use personal health care information without consent To insist that regulation should not interfere with researchers' access to health records or record linkage capabilities is irresponsible. Though important, anonymity alone cannot be relied on to protect the interests of participants. Providing choices about participation in research through consent remains the most appropriate mechanism to protect people's privacy. (Brown and others, 2009)

Equally, in 2007 an Ipsos MORI poll of the general public for the Medical Research Council found that 21 per cent of respondents thought it was never acceptable for researchers to use personal health care information without consent (MRC, 2007).

Questions for the future

As data governance arrangements continue to evolve in England and other countries, this analysis of the existing arrangements allows us to understand the role played by various social values in debates about how NHS data should be governed. Whilst appealing, the proposed alternative public benefit framework for considering information governance raises a number of important questions, including:

- 1. How can we define what constitutes a 'public benefit'?
- 2. What safeguards are necessary to ensure that patients' rights are protected within a 'public benefit' model?
- 3. How can obligations arising from the rights of patients be balanced against the benefits accrued to the public at large?

Finally, policy-makers will also need to consider how to gain agreement across a range of views. Here, the Organisation for Economic Cooperation and Development (OECD)'s experience in this area may give them some encouragement. In January 2004, the governments of the 30 OECD countries plus those of China, Israel, Russia and South Africa invited the OECD "to develop a set of OECD guidelines based on commonly agreed principles to facilitate cost-effective access to digital research data from public funding to be endorsed by the OECD Council at a later stage". The resulting document was published as *Principles and Guidelines for Access to Research Data from Public Funding* (OECD, 2007). What this example seems to reflect is that a common set of values relating to data governance does exist, even between countries as different as, say, China and Israel. It is therefore to be hoped that agreement among policy-makers in the UK should be achievable.

References

Academy of Medical Sciences (2011) A New Pathway for the Regulation and Governance of Health Research.

British Medical Association (2007) Guidance on Secondary Uses of Patient Information.

Brown I, Brown L, Korff D (2011) 'The limits of anonymisation in NHS data systems', *BMJ* 342: d973.

Clark, S and Weale, A (2011) *Information Governance in Health: An analysis of the social values involved in data linkage studies.* Nuffield Trust.

Department of Health (2003) Confidentiality: NHS Code of Practice.

Department of Health (2010) Liberating the NHS: An Information Revolution. A consultation on proposals.

General Medical Council (2009) Confidentiality.

House of Commons Health Select Committee (2007) *The Electronic Patient Record: Sixth report of session 2006–07.* HC 422–I.

Information Commissioner's Office (2011) *Data Sharing Code of Practice.*

Laurie G (2002) *Genetic Privacy*. Cambridge University Press.

Medical Research Council (2000) Personal Information in Medical Research.

Medical Research Council (2007) The Use of Personal Health Information in Medical Research. General public consultation.

Nuffield Trust (2011) Developing a Person-based Resource Allocation Formula for Allocations to General Practices in England. Project information. www.nuffieldtrust.org.uk/our-work/projects/person-based-resource-allocation-pbra

Organisation for Economic Cooperation and Development (2007) *Principles and Guidelines for Access to Research Data from Public Funding.*

Smyth RL (2011) 'Regulation and governance of clinical research in the UK', *BMJ* 342: d238.

Wellcome Trust (2011) Response to the Department of Health: Liberating the NHS: An Information Revolution.





For more information about the Nuffield Trust, including details of our latest research and analysis, please visit www.nuffieldtrust.org.uk

- Download further copies of this research summary from www.nuffieldtrust.org.uk/publications
- Subscribe to our newsletter: www.nuffieldtrust.org.uk/newsletter
- Follow us on Twitter: Twitter.com/NuffieldTrust

Nuffield Trust is an authoritative and independent source of evidence-based research and policy analysis for improving health care in the UK 59 New Cavendish Street London W1G 7LP Telephone: 020 7631 8450 Facsimile: 020 7631 8451 Email: info@nuffieldtrust.org.uk

www.nuffieldtrust.org.uk

Published by the Nuffield Trust.

© Nuffield Trust 2011. Not to be reproduced without permission.