

ARTICLE

Big Voice or Big Data? The Difficult Birth of care.data

Michael Soljak*

Imperial College London, London

Abstract

In the United Kingdom (UK), a range of historical national healthcare data collections have occurred, in some cases without a very specific legal basis apart from overarching international and European Union data protection commitments expressed in the UK Data Protection Act 1998. In 2012, the English Government announced that the GP Extraction Service (now care.data), a new central flow of patient-identifiable healthcare data from general practice computer systems, would commence to support healthcare commissioning. Data on the whole population would be extracted, and specific patient consent would not be sought. UK primary healthcare data is characterised by its richness, and comprises demographic, diagnostic, clinical, prescribing, test results and a range of other classes of data.

In 2014 the English media and several non-governmental patient organisations began a campaign questioning the care.data initiative, and uncovered quite limited but nonetheless damaging evidence of improper release of patient data from historical data sources. A subsequent national review of information governance and a parliamentary inquiry has delayed care.data implementation, and a patient opt-out is being introduced. Another positive effect is the subsequently much higher public awareness of care.data.

Health information and the law

The United Kingdom (UK) National Health Service (NHS) is fortunate to have comparatively good data relating to the health of its citizens

* DOI 10.7590/221354015X14319325750232

Dr Soljak is funded by Arthritis Research UK, the Department of Health Policy Research Programme and the Imperial NIHR Biomedical Research Centre (BRC). However this paper did not receive external funding.

and their healthcare.^{1, 2} Historically, the NHS has been funded from general taxation and is free at the point of use for all residents, and the vast majority are registered with a single general practice, enabling the production of population-based information. Capitation-based funding for most healthcare, and all hospital care, adjusted for health needs, flows to clinical commissioning groups (CCGs), which are directed by boards of general practitioners (GPs). There is a small private insurance market funded by individual policy-holders, which covers elective surgery, but virtually all emergency care and most elective care is NHS-funded. The previous Labour Government allowed NHS patients to choose between NHS and private hospitals providers, but the latter must return information about the care delivered.

For over 20 years, an extensive Hospital Episode Statistics (HES) electronic inpatient dataset of about 250 data items, including administrative and some clinical data for each hospital admission, has flowed nationally. This includes 20 ICD-10 coded diagnoses and surgical procedures, and more recently extensions covering specialties such as maternity, neonatal and critical care. More recently, central reporting of datasets for NHS outpatient, mental health services and accident and emergency (A&E) care have been mandated, although the content of clinical data is limited, and the A&E dataset has only recently been reported by most hospitals. A dataset for community health services has been established, but national reporting is not yet mandatory.

In addition, a range of other national data collections have been established, ranging from national clinical audits of diabetes and myocardial infarction to ambulance services and immunisation. Recently the Health & Social Care Information Centre (HSCIC), which oversees these systems and processes the data, has begun to undertake the types of data linkage which have long been established in Scandinavian countries, using the unique NHS number identifier.

The specific legal basis for these data collections has been rather sparse. The overarching legislation is the international and European Union data protection commitments expressed in the UK Data Protection Act 1998, which includes (Principle 2) the provision that personal data 'will be obtained only for one or more pre-specified and lawful purposes, and shall not be further processed in any manner incompatible with that purpose or those purposes', al-

¹ A.K. Jha, D. Doolan, D. Grandt, T. Scott & D.W. Bates, 'The use of health information technology in seven nations', *International Journal of Medical Informatics* 77 (2008), 848-854.

² J. Adler-Milstein, E. Ronchi, G.R. Cohen, L.A.P. Winn & A.K. Jha, 'Benchmarking health IT among OECD countries: Better data for better policy', *Journal of the American Medical Informatics Association* 21 (2014), 111-116.

though exemptions are available in relation to personal data that is processed for research, statistical or historical purposes.

In addition, Section 60 of the UK Health and Social Care Act 2001, as re-enacted by Section 251 of the NHS Act 2006, allows the Secretary of State for Health to make regulations to set aside the common law duty of confidentiality for defined medical purposes. This Section has been used to legitimise a number of existing dataflows where obtaining specific patient consent would have disrupted them, for example cancer registration. Progress towards gaining patient consent has been gradual. Currently, all cancer patients should receive an information leaflet, which informs them that they have the right to opt out of their data registration.^{3, 4} Because of their historical roles as administrative and payment mechanisms for NHS providers, no such opt-out exists for HES datasets.

The advent of care.data

What is care.data, and what makes it different from the open data movement in other countries? After all, Scandinavian countries have been linking datasets for years, and primary care electronic health records (EHRs) are widely used in several other countries – for example, Jha et al. found that over 90% of GPs in Denmark, the Netherlands and New Zealand used primary care EHRs.⁵ Major US healthcare organisations such as Kaiser Permanente have achieved similar levels. The biggest difference is in the availability and sheer volume of coded and therefore usable primary care data available in the UK. Two main additional factors were involved in the UK; first, the UK Department of Health has required all GP system suppliers to embed Read codes, which now form the primary care component of the international medical nomenclature SNOMED-CT.⁶ Read codes cover not just diagnoses but also clinical findings, medical history, laboratory tests, prescribing and other key data (see Table 1). This, combined with the use of data entry templates with drop-down menus, has encouraged entry of a wide range of information useful for clinical care. Consistency was also accelerated by using Read codes as payment mechanisms – for example, for the Quality & Outcomes Framework (QOF), the gen-

³ National Cancer Registration Service, 2015, *About Cancer Registration* (online), www.ncr.nhs.uk/patientinfo/ (accessed 2 February 2015).

⁴ B. Roehr, 'US government opens up online datasets', *BMJ* 339 (2009), b3983.

⁵ A.K. Jha, D. Doolan, D. Grandt, T. Scott & D.W. Bates, 'The use of health information technology in seven nations', *International Journal of Medical Informatics* 77 (2008), 848-854.

⁶ International Health Terminology Standards Development Organisation, 2015, *Snomed CT: The Global Language of Healthcare* (online), <http://ihtsdo.org/snomed-ct/> (accessed 2 February 2015).

eral practice pay-for-performance programme (which, it should be noted, only extracts aggregate data).^{7, 8}

Secondly, several GP system suppliers recruited a representative sample of research practices from their system users, and developed the technology to extract anonymised Read-coded data on a monthly basis. This provides a UK population-based sample of 6-8 million (500-600 practices) from each database, which has been used widely by pharmaceutical and healthcare companies worldwide. The Clinical Practice Research Datalink is one of these databases.⁹ It provides online access to primary care data and linked HES and Office for National Statistics mortality data.

The Department of Health and NHS England, the government agency which funds and monitors CCGs and manages the national GP contract, were well aware of the potential of primary care data not only for managing the GP contract in every practice, but also for research. Furthermore, the QOF programme has given HSCIC the technology to extract any data from GP systems. A disadvantage of all the GP research databases is anonymisation of identities of practices as well as patients (and of hospitals). While it had the benefit of further reducing the possibility of patient identification, and helped to reassure and so recruit more practices, this lack of georeferencing below English regional and other UK country levels seriously hampers the utility of the databases in improving health and healthcare. For example, primary care data with provider identification could be used to monitor risk factor prevalence in small populations, or to determine the effects of local policy initiatives such as improved resourcing, or to improve commissioning. But to do this, data from all, or most practices would be required.

In December 2012, as part of 'Securing excellence in GP IT Services: Operating Model', NHS England and HSCIC announced more details about the GP Extraction Service (as care.data was termed at the time), including the intention to extract data from all practice systems.¹⁰ The UK Data Protection Act guidance states that data controllers 'often need to get prior consent to use or disclose

⁷ T. Doran, C. Fullwood, E. Kontopantelis & D. Reeves, 'Effect of financial incentives on inequalities in the delivery of primary clinical care in England: analysis of clinical activity indicators for the quality and outcomes framework', *The Lancet* 372 (2008), 728-736.

⁸ T. Doran & M. Roland, 'Lessons From Major Initiatives To Improve Primary Care In The United Kingdom', *Health Affairs* 29 (2010), 1023-1029.

⁹ Clinical Practice Research Datalink, 2015, *Welcome to The Clinical Practice Research Datalink*, www.cprd.com/intro.asp (accessed 2 February 2015).

¹⁰ NHS Commissioning Board, N. E., 2013, *Securing excellence in GP IT Services: Operating Model*, www.england.nhs.uk/wp-content/uploads/2012/12/pc-it-op-model.pdf (accessed 2 February 2015).

personal data for a purpose that is additional to, or different from, the purpose they originally obtained it for' – in this case, the original purpose was direct clinical care. NHS England announced that, rather than obtaining individual signed consent, the method that practices should use in obtaining such consent was to be patient leaflets and posters in practices. This heightened anxieties amongst some GPs, who were uncertain of the legality of the extracts. At their behest a leaflet drop to every household in England took place, although many public poll respondents claimed not to have seen it.

The backlash

Although care.data counterparts are also planned for Scotland and Wales, the voracious English media began searching for breaches of confidentiality, and as care.data was not yet launched, they used Official Information Act requests to HSCIC to investigate historical releases of HES data. Although HSCIC had proportionate information security for HES data in the form of Data Sharing Agreements and requirements for customers' security policies, in February 2014, *The Telegraph* newspaper published an article entitled 'Hospital records of all NHS patients sold to insurers', which stated:

'The disclosure comes days after controversial plans to extract patient data from GP files were put on hold, amid concerns over the scheme. Those in charge of [care.data] have repeatedly insisted that it will be illegal for information extracted from GP files to be sold to insurers, who might seek to target customers or put up their prices.'

However, a report by a major society of actuaries discloses that it was able to obtain 13 years of hospital data – covering 47 million patients – in order to help insurance companies “refine” their premiums.

The report... details how it was able to use NHS data covering all hospital in-patient stays between 1997 and 2010 to track the medical histories of patients, identified by date of birth and postcode. It boasts that “uniquely” they were able to combine these details with information from credit ratings agencies, such as Experian... to advise companies how to refine their premiums.

The report advises that levels of illness among most customers below the age of 50 were found to be higher than previous calculations had found. As a result of the work, insurers were likely to increase premiums for this group...'

HSCIC immediately launched an inquiry led by its new Chairman, Nicholas Partridge (as part of the new Coalition Government's Health & Social Care Act 2012, HSCIC had replaced its predecessor organisation, the NHS Information

Centre), and published the review in June 2014.¹¹ The review acknowledged that no individual had ever complained that their confidentiality had been breached as a result of data being shared or lost by the Information Centre, and there was no complaint to the Information Commissioner's Office. However it was identified that 'there were four Data Sharing Agreements made by the Information Centre with three re-insurance companies which allow those re-insurers to continue to use the data until the agreements expire in 2015 and 2016'. HSCIC is now putting in place more rigorous (and costly) information governance processes.

There was associated discussion in the media about not releasing data to commercial organisations. However, like universities, pharmaceutical companies also use the data to improve health, and if release was restricted, they could simply set up or fund research charities to carry out the analyses. The furore created by *The Telegraph* also led to a number of hearings by the House of Commons Health Committee, which as of writing has yet to report. A number of confidentiality pressure groups have sprung up and given evidence to the Committee. Amended legislation has also been passed: the Care Act 2014 places HSCIC's Confidentiality Advisory Group on a statutory footing, and secondary legislation will set out the matters to which it must have regard when giving advice to the HSCIC in relation to disclosures. This is likely to include an opt-out option, as for cancer registries. No data collected as part of the care.data programme will be released until these regulations are in place. Meanwhile HSCIC is piloting care.data in around 500 pathfinder practices.

Finally, the Nuffield Trust has recently published a report on the ethical issues involved in the collection, linking and use of data in biomedical research and health care.¹² Recommendations included mandatory reporting of privacy breaches affecting individuals to the individuals concerned (several million people in the case of the example above, including everyone born since HES was established), criminal penalties, including imprisonment, for the deliberate misuse of data, and restriction of access to data to researchers who are subject to institutional oversight (a significant new burden for universities).

¹¹ N. Partridge, *Review of data releases made by the NHS Information Centre*, Health & Social Care Information Centre 2014, www.gov.uk/government/uploads/system/uploads/attachment_data/file/367788/Sir_Nick_Partridge_s_summary_of_the_review.pdf (accessed 2 February 2015).

¹² J. Montgomery, S. Caney, B. Farsides, P. Furness & A. Gallagher, *The collection, linking and use of data in biomedical research and health care: ethical issues* (London: Nuffield Trust 2015), http://nuffieldbioethics.org/wp-content/uploads/Biological_and_health_data_web.pdf (accessed 2 February 2015).

Conclusion

The launch of care.data has provided a number of early insights into the confidentiality issues surrounding the use of very detailed patient data for commissioning and research. It is apparent that the public regard primary care data given to their GP or practice nurse differently to that from other sources. Had the care.data programme been announced in a country with a less active media, it might already be in place: indeed, there has been no such reaction as yet in Wales or Scotland, although health agencies there are viewing implementation of their own programmes with trepidation, and Scotland will offer an opt-out to practices rather than patients. The media revelations relate not to the new data source, but to HES data, a collection which had been operating largely without obvious problems for 20 years. On the positive side, most citizens are now aware of care.data, legal issues have been resolved, and it is likely that patients will now play a more active role in curating their medical data. Hopefully large scale opt-outs will not occur and devalue the information in the process.

Table 1: 'Scope of Read codes used in UK primary healthcare data' on the next page.

Table 1: Scope of Read codes used in UK primary healthcare data

Processes of Care		Diagnoses		Medication	
o	Occupations	A	Infectious /parasitic disease	a	Gastro-intestinal
1	History / symptoms	B	Neoplasms	b	Cardiovascular
		C	Endocrine, nutritional etc.	c	Respiratory
		D	Diseases of the blood	d	Central nervous system
2	Examinations & signs	E	Mental disorders	e	Drugs used in infections
3	Diagnostic procedures	F	Nervous system	f	Endocrine drugs
		G	Circulatory system	g	Obs, gynae, UTI
4	Laboratory procedures	H	Respiratory system	h	Chemotherapy etc.
		J	Digestive system	i	Haematology / dietetic
5	Radiology / physics in medicine	K	Genitourinary system	j	Musculoskeletal
		L	Compl. of child birth etc.	k	Eye
		M	Skin / sub-cutaneous tissue	l	ENT
6	Preventive procedures	N	Musculoskeletal diseases	m	Skin
		P	Congenital conditions	n	Immunology / vaccines
7	Operations, procedures, sites	Q	Perinatal conditions	o	Anaesthetic
		R	Signs, ill-defined conditions	p	Appliances & reagents
8	Other therapeutic procedures	S	Injury & poisoning	q	Incontinence appliances
		T	Causes injury & poisoning	s	Stoma appliances
9	Administration	U	Ext. causes morbidity / mortality	u	Contrast media
		Z	Unspecified conditions	y	Drug release administration