This response was submitted to the consultation held by the Nuffield Council on Bioethics on *The linking and use of biological and health data* between 17 October 2013 and 10 January 2014. The views expressed are solely those of the respondent(s) and not those of the Council.

The linking and use of biological and health data

Evidence written in a personal capacity by biostatistician Sheila M. Bird OBE MA PhD CStat FFPH FRSE

Reasons for giving evidence: As a biostatistician, I have designed high response-rate surveys (so-called WASH/C surveillance) that gave *demonstrable anonymity* to prisoners in their self-reporting of HIV/Hepatitis C virus risk behaviours (including inside prison) and in the linking of their self-completion questionnaire and saliva sample (to be tested for HIV/HCV antibodies). For the past 20 years, I have designed record-linkage studies which respected prisoner, drug treatment client and medical confidentiality and which made important discoveries, such as prisoners' high risk of drugs-related death in the fortnight after prison-release. Well-designed studies *demonstrate respect for confidentiality by demonstrating how anonymity is safeguarded*.

1.1 **Medical data are different**: they are obtained by doctors under a strong duty of confidentiality. In responding to a survey, I can choose to be truthful of not. But, when I give permission for tests to be done on a blood or other sample, for imaging to be performed, or for surgery, I do not know what will thereby be revealed about me, and I have no opportunity to withhold the information because I receive it second-hand – from my doctor. Moreover, subsequent falsification of my medical record would not be in my interest. *We must not create perverse consequences that risk the distortion of medical records.*

1.2 **Well-designed data-collection, for peer-reviewed and ethically-approved research purposes, is different from the data that, daily, doctors choose to record or not record**. For example, GPs routine recording of body mass index (BMI) is grossly incomplete – not least because the patient sits before the GP who – for the most part - can 'size up' the over/under-weight problem by looking at the patient. Recording of BMI for specific patients, in many but not all practices, is likely to start only when a weight-problem has been diagnosed. However, when a patient has been recruited into a research-study, the study-record for the patient will generally contain complete data on all study-questions posed whether the question was specifically germaine to this patient or not – not least because study-procedures will typically have been designed to identify missing data, and to instigate follow-up to obtain them.

1.3 **Data in administrative records are not subject to research-level checking:** and back-checking may occur in highly selected circumstances. For example, when we began analyses of the UK outcome of kidney grafting, we observed a hugely significant association between the availability of the patient's HLA-DR type and graft failure – for the simple reason that the then-new technique of HLA-DR-typing was deployed first to help improve the match of next kidney for those patients who had already suffered graft loss. The HLA-DR types had been back-filled, preferentially for patients who had already experienced graft failure, but these HLA-DR types had not been available at the time of the first transplant . . .

1.4 **Substantial improvements in information technology are necessary to deal with free text in medical records**: and to do so in a manner that does not lead to deductive disclosure – for example, about patient networks. In specific contexts, such as HIV disease transmission, a

patient's network of injecting or sexual contacts can be critical for the identification of a highly viraemic "transmitter", especially when analysed alongside molecular epidemiological findings from blood samples. There is a clear public health need for such data to be assembled but the very gathering of such confidential data is put at-risk if the infectees cannot be guaranteed that their membership of a specific transmission network shall remain confidential – except by court order. We ceased the longer-term without-consent follow-up by record-linkage of those who had been in HMP Glenochil in 1993 when HIV seroconversions had occurred when it became clear that no such guarantee could be given by the research-team. The cessation of, or inability to publish, other carefully-designed studies which are in the interest of public health may be a perverse consequence of over-emphasis on open access to data and the availability of even no-names data-sets.

1.5 **Disclosure happens:** during a review of 97 fatal accident inquiries into deaths in Scottish prison custody, I unmasked the 14th HIV-infectee from HMP Glenochil who, in life, had chosen not to seek confirmation of his HIV diagnosis. Why was he so fearful? His mistrust may not be unique. Sadly, his unwillingness to have his HIV diagnosis confirmed and to seek medical care cost him many life-years.

1.6 **Database content determines usefulness more than technological advances:** technological advances allow data (including free text) to be abstracted but if the content of the database was ill-designed at the outset then the "right" data will not be there. Administrative databases may have been under-analysed for years, and their faults will have gone undetected – as was the case for NHS Organ Donor Register. Competent analysis leads directly or indirectly to improvement in administrative databases, and this is welcome – such as the effort by Scottish Drug Misuse Database to record in-treatment-follow-up and end-of-episode data.

1.7 **Statistical imputation is not a testably robust solution when over half the data are missing** . . .

1.8 **Predictive analytics need to demonstrate**: that they have validated their prediction-credentials (to warrant the name) . . .

1.9 **Global interoperability standards:** to increase the statistical power of research studies generally require that **a protocol should be agreed between research studies** on how the pooling of data shall proceed. The transparency that such a protocol implies is to be welcomed.

1.10 **Good practice:** Just as most randomized controlled trials have a Data Safety and Monitoring Board, many research studies have appointed a Protocol and Data Access Board whose members can review research protocols proposed by those who wish to access data and/or samples from participants in the research study for which the board has oversight.

1.11 I am a practitioner of, and strong advocate for, research-led record-linkage studies for the public good. As a biostatistician, I follow professional codes of conduct (such as the Royal Statistical Society and the Medical Research Council) but none of us is sole judge of "the public good" and it generally behoves research-scientists to justify their study-plans through both peer-scientific-review and to ethics committees. For example, Dr Ray Brettle and I both appeared before a Lothian research ethics committee in the mid-1990s which approved the first confidential record-linkage study that I designed to link Edinburgh's HIV-clinical-cohort and prison-records. We were asked: "what if news got out about the study?" We explained that news would, and was intended to, get out as we would publish the study method in detail alongside the findings – and that Edinburgh was uniquely placed, globally, to enable the proposed study of whether incarceration affected the morbidity and mortality of HIV infected

injector-prisoners. Approval was given and led to the first quantification of the very high risk of drugs-related death soon after prison-release, which is now corroborated internationally.

1.12 **It is most unfortunate:** that the current publicity about NHS data sharing coincides with the issuing of survey-forms for the first 2014 wave of the GP Patient Survey. The covering letter for the GP Patient Survey that I received included my name & address and my survey reference-number; and was image-signed by Tim Kelsey who *does not have my permission* to know my name and address. The GP survey form included my reference-number. Bizarrely, the pair demonstrates to the public that linkage can be, and has been made, between my personal data and my survey reference-number; and thus could be made between my responses, name and address. What a dire demonstration to the public of how NHS England understands **respect for the recipient's confidentiality** and how inadequate was the **deployment of statistical science for the maintenance of anonymity of my answers**. Smart survey-design methods exist, and should be deployed, whereby reminders can issued to those who have not responded and yet the responses received-in are not identifiable to the sender (only to the sender's GP practice, say).

1.13 **Online completion, another option with the GP Patient Survey, presents different challenges in terms of anonymity.** The current set-up ensures that ONLY some-one with access to my online password can respond to my GP survey-form; it does not demonstrably reassure me about the use that could be made of my email-address, NOW in potential linkage with my name and postal address and survey-reference. Demonstrable anonymity has been denied to the public. Instead, there has been demonstration that the public has to 'take on trust' those who have shown by their very study materials that they should not be 'taken on trust'. I hope that the GP Patient Survey team is trustworthy. However, they have failed to demonstrate that they have respected my right to confidentiality.

1.14 **Transparency about study methods is needed:** if there is to be public trust in the linking of biological and health data. Outside of the biomedical research community, what guarantees are there that purchased data are used for per-protocol purposes only?

1.15 Selling of NHS data, and payment to safe-havens for linkages between held-databases, is a new revenue stream.

1.16 **No-names is not sufficient to guarantee non-identifiability**: if the data from databases A (GP), **B** (hospitalizations) & C (criminal justice) are linked and the linked-data are returned to the holder of database **B**, who rightfully knows client-names and has thereby learned about the criminal-justice record of some (though not necessarily all) patients. For such reasons, typically only 3rd-party researchers {that is: not A, B or C} have access to the no-names linked data OR **B** would be expected to work on the linked data-set only in a safe haven where **B**'s access to client-names is blocked, unlike if B could work on the linked dataset alongside **B**'s named-version of database-B.

1.17 **Making services available only on condition that patients' personal data can be re-used for other undisclosed purposes is foul**: this bullying tactic threatens the basis of informed consent for medical research wherein there is a guarantee that those who do not wish to take part will not be disadvantaged and shall continue to receive the best available treatment.

1.18 In summary, biomedical data are different and have special significance. New privacy issues do arise because of the potential for deductive disclosure and a potential to threaten dissenters that their NHS services will be restricted. There should be transparency about the ways that linked data are being used, including and especially by government, and there should be

transparency about linkage-protocols and associated analysis plans, much as for randomized controlled trials.

1.19 **Greater use of sampling methods** should be made when there is risk of deductive disclosure so that the risk pertains for only as many persons as need be studied to meet the scientific objectives.

1.20 **Hence, record-linkage studies are not necessarily data-driven research**: on the contrary, prior hypotheses may be the strong justification for how extensive a record-linkage study needs to be, see Bird & Hutchinson Addiction 2003.

1.21 We are told that extracting value from rich data resources has become a priority for the knowledge economy. Extracting value has always been a priority but public good has been balanced by respect for individuals' confidentiality and rights. Independent assessment of **my judgement of that balance** not only serves the public good, but serves me well too.

1.22 Holders of a linked data-set do not, in general, have a right to pass-it-on unless the holding-licence permits them to do so, which would be unusual.

1.23 A key hope from biomedical record-linkage is that pharmaco-epidemiological studies will generate information about the frequency, apparent (in-) effectiveness and serious adverse-events associated with poly-pharmacy; and do so not only by gender and age-group but also by constellation of diagnoses and genotype.

1.24 Approaches to risk-profiled patients should have the patients' - not commercial - interest as their primary purpose. Biomedical researchers' approaches to patients also have constraints.

1.25 **Biomedical research proceeds by public consent.** Consent does not have to be individual for study-methods to have public approval. However, study methods that have not sufficiently justified themselves to the public are almost surely methods that need design-improvement.

References

Prisoners' risks from blood-borne viruses & demonstrable anonymity

BIRD AG, GORE SM, JOLLIFFE DW and BURNS SM. Anonymous HIV surveillance in Saughton Prison, Edinburgh. *AIDS* 1992; 6: 725-733.

GORE SM and BIRD AG. No escape: HIV transmission in jail. Prisons need protocols for HIV outbreaks. *British Medical Journal* 1993; 307: 147-148.

BIRD AG AND GORE SM. Inside methodology: HIV surveillance in prisons (editorial comment). *AIDS* 1994; 8: 1345-1346.

BIRD AG, GORE SM, CAMERON S, ROSS AJ and GOLDBERG DJ. Anonymous HIV surveillance with risk factor elicitation at Scotland's largest prison, Barlinnie. *AIDS* 1995; 9: 801-808.

GORE SM, BIRD AG, BURNS SM, GOLDBERG DJ, ROSS AJ and MACGREGOR J. Drug injection and HIV prevalence in inmates of Glenochil prison. *British Medical Journal* 1995; 310: 293-296.

HUTCHINSON SJ, GORE SM, GOLDBERG DJ, YIRRELL DL, MCGREGOR J, BIRD AG, LEIGH-BROWN AJ. Method used to follow-up previously undiagnosed infections in the HIV outbreak at Glenochil prison. *Epidemiology and Infection 1999;* 123: 271 - 275.

BIRD SM, Leigh Brown AL. Criminalisation of HIV transmission: implications for public health in Scotland. *BMJ 2001;* 323: 1174 - 1177.

BIRD SM. Fatal accident inquiries into 97 deaths in prison custody in Scotland (1999-2003, or during first five years of operation of Scotland's only private prison): elapsed time to end of inquiry or written determination, issues and recommendations.
Howard Journal for Criminal Justice 2008 (published online 22 May ahead of print version); 47: 343 – 370.

BIRD AG, GORE SM, HUTCHINSON SJ, LEWIS SC, CAMERON S and BURNS S on behalf of the European Commission Network on HIV infection and hepatitis in prison. Harm reduction measures and injecting inside prison versus mandatory drugs tests: results of a cross sectional anonymous questionnaire survey.
British Medical Journal 1997; 315: 21 - 24.

GORE SM, BIRD AG, CAMERON SO, HUTCHINSON SJ, BURNS SM, GOLDBERG DJ. Prevalence of Hepatitis C carriage in Scottish prisons: Willing Anonymous Salivary Hepatitis C surveillance linked to self-reported risks. *Quarterly Journal of Medicine 1999;* 92**:** 25 - 32.

Seaman SR, BIRD SM, Brettle, RP. Historical HIV prevalence in Edinburgh Prison: a database-linkage study. *Journal of Epidemiology and Biostatistics 2000; 5: 240 - 245.*

High risk of drugs-related death in the fortnight after prison-release

SEAMAN SR, BRETTLE RP, GORE SM. Mortality from overdose among injecting drug users recently released from prison: database linkage study. *British Medical Journal 1998*; 316: 426 - 428.

BIRD SM, Hutchinson SJ. Male drugs-related deaths in the fortnight after release from prison: Scotland, 1996 - 1999. *Addiction* 2003; 98: 185 - 190.

Merrall ELC, Kariminia A, Binswanger IA, Hobbs M, Farrell M, Marsden J, Hutchinson SJ, BIRD SM. Meta-analysis of drug-related deaths soon after release from prison. *Addiction* 2010; 105: 1545 – 1554.

Strang J, BIRD SM, Parmar MKB. Take-home emergency naloxone to prevent heroin overdose deaths after prison release: rationale and practicalities for the N-ALIVE randomized trial.
Journal of Urban Health 2013; 90: 983 – 996.

Confidential audit, surveillance (by international protocol for pooled analysis) and performance monitoring

GORE SM, CABLE DJ and HOLLAND AJ. Organ donation from intensive care units in England and Wales: two year confidential audit of deaths in intensive care. *British Medical Journal* 1992; 304: 349-355.

MAP Workshop (GORE SM and FAREWELL VT co-convenors). Extending public health surveillance of HIV infection: information from a five cohort workshop. *Statistics in Medicine* 1993; 12: 2065-2085.

Royal Statistical Society Working Party on Performance Monitoring in the Public Services (SMB: chair). *Performance Indicators: Good, Bad, and Ugly.* London: Royal Statistical Society, 23 October 2003 (see Reports at www.rss.org.uk).

BIRD SM. Recipients of blood or blood products "at vCJD risk". We need to define their rights and responsibilities and those of others (Editorial). *British Medical Journal* 2004; 328: 118 – 119.

BIRD SM. Please join the organ donor register: you could save a politician's life. *Straight Statistics* 2010; 25 October. {http://straightstatistics.org/article/please-join-organ-donor-register-you-could-save-politician's-life}

Morbidity and mortality of injectors, hepatitis C diagnosed persons, or drug treatment clients

King R, BIRD SM, Brooks SP, Hutchinson SJ, Hay G. Prior information in behavioural capture-recapture methods: demographic influences on drug injectors' propensity to be listed in data sources and their drugs-related mortality. *American Journal of Epidemiology* 2005; 162: 1 – 10.

McDonald SA, Hutchinson SJ, BIRD SM, Mills PR, Dillon J, Bloor M, Robertson C, Donaghy M, Hayes P, Graham L, Goldberg DJ. A population-based record linkage study of mortality in hepatitis C diagnosed persons with and without HIV co-infection in Scotland. *Statistical Methods in Medical Research* 2009; 18: 271 – 283.

McDonald SA, Hutchinson SJ, BIRD SM, Mills PR, Robertson C, Dillon J, Hayes P, Williams T, Goldberg DJ. Hospitalisation of hepatitis C diagnosed persons in Scotland for decompensated cirrhosis: a population-based record linkage study. *European Journal of Gastroenterology and Hepatology* 2010; 22: 49 - 57.

Merrall ELC, BIRD SM, Hutchinson SJ. Mortality of those who attended drug services in Scotland 1996-2006: record linkage study. *International Journal of Drug Policy* 2012; 23: 24 - 32.

Merrall ELC, BIRD SM, Hutchinson SJ. A record linkage study of drug-related death and suicide after hospital discharge among drug-treatment clients in Scotland, 1996-2006. *Addiction* 2013; 108: 377 - 384.

Merrall ELC, BIRD SM, Hutchinson SJ. A record linkage study of hospital episodes for drug treatment clients in Scotland, 1996-2006. *Addiction Research & Theory* 2013; 21: 52 – 61.