Consultation Question 1:

Do biomedical data have special significance?

No. It would not be possible – and of uncertain value - to define biomedical data as a distinct class for data. For example, within a patient's health record there might be demographic information, such as name, address, that would clearly identify an individual, would generally not be regarded as 'biomedical data', but may be important for biomedical care or research (e.g. to allow linkage to other data sets, provide information about potential health hazards associated with the person's location), along with data on diagnoses, blood results, imaging reports that are clearly biomedical in nature. Much medical and health services research and monitoring could be conducted using de-identified data. But, depending how this is achieved, does compromise the ability to inform individuals about their own results, provide updated information about the research results, obtain revised consent, or link to other, relevant, data sets.

Turning to the question of whether particular sub-sets of biomedical data present particular ethical challenges, this is certainly the case when considered from the perspective of the protection of confidentiality (for example, there is a theoretical possibility at least that genomic analysis could result in the identification of individuals from their sequence data, but properly constructed access agreements prevent researchers from identifying individuals). Another factor that needs to be considered is the importance of the data to the future health of the individual research participant: For example, genetic analyses (as well as other research) will identify differences in a variety of factors between individuals. While some could have clear importance for the individual's health (e.g. risk of breast cancer, response to warfarin), the importance of others may be much less clear (particularly at the time the finding is identified), either because the effect size is small, or the association with disease depends on interactions with other genetic or environmental factors.

The key is to ensure that the individual understands what feedback they can expect from those collecting and using the data. This may range from (a) a complete read out of the data (e.g. whole genome sequence) being returned to the individual and/or their doctor; (b) only findings that are likely to significantly increase the risk of serious morbidity (where "significantly increased risk" and "serious morbidity" will need to be quantified); to (c) no feedback (as is the case with genetic analyses in the UK Biobank study).

Consultation Question 2:

What are the new privacy issues?

Big-data does not raise entirely new issues with regard to privacy. The key principles that govern using individual-level data for research would still apply, namely the ethical – particularly the basis of consent - and legal right of the researcher to use the data for a specified research purpose and also how the data might or be might not be shared with third parties.

In a conventional research protocol, where the data to be collected are known and can be fully anonymized, a guarantee could be given to participants that the research would not result in them being identified, but in an era of big-data, there are likely to be future linkage to health and other record systems that were not specified (or even existed) at the time of consent. There is also the inadvertent risk of identification by analyzing multiple datasets. This means that the researcher can certainly not give a guarantee to the research participant that there is no possibility that they might be identified, but the initial consent process should include a clear description of how privacy will be protected.

De-identification is unlikely to guarantee anonymity in all circumstances but substantially reduces the risk. In considering this issue, it is helpful to consider the likelihood that somebody would attempt to re-identify the person, how likely they are to be successful (for example, in isolation a genome cannot be used to identify a person unless there is access to another source of information that contains that person's genetic data in combination with other identifiable data), and the potential consequence of successful re-identification.

See also Any Other Comments

Many of these issues apply to current Research Tissue Banks (RTBs), such as UK Biobank, where a broad consent for future health-related research was obtained, as well as linking to a wide range of future medical and other health-related records. The National Research Ethics Service (NRES) have produced guidance and a specific application process to assist RECs with the ethical review of RTBs and this covers many of the specific issues that need to be considered with respect to privacy and confidentiality.

Consultation Question 3:

What is the impact of developments in data science & information technology?

'Big-data' is an imprecise term and for a specific research study participants should be informed at the outset what data in a general sense will be collected about them, how that data might be used and who can carry out research using that data. For long-term longitudinal studies there is considerable scientific value in linking to other datasets not envisaged at the initial recruitment phase and in these circumstances ongoing communication with study participants is essential – via the study website or the sending of a periodic newsletter – with participants always retaining the option to withdraw if they so wish.

Significant progress in biomedical research should be made in the next few years in making greater use of NHS electronic health records in research.

Vast quantities of data that may be of relevance to health care and research are being collected. Big Data offers methods to analyse these complex and heterogeneous datasets in order to generate new knowledge. This might be new understanding of current disease concepts (e.g. causes and treatment of breast cancer), repurposing of old drugs for a new use (e.g. use of imipramine [an antidepressant] to slow progression of lung cancer) or discover of entirely new disease concepts (i.e. previously unrecognized diseases).

Consultation Question 4:

What are the opportunities for, and the impacts of, use of linked biomedical data in research?

The ability to collect and analyse increasingly large and complex datasets offers enormous potential to better understand the determinants of a wide variety of diseases. As we highlighted in our response to Question 3, large population based biobanks, such as UK Biobank, have already addressed many of the ethical and social issues that have been raised, such as the need to have clear policies on data access, data sharing and collaboration. The REC review of RTBs ensures that the participant information makes it clear who is the custodian of the resource and how data might or might not be shared with other researchers. Large-scale collaborations, including international ones, do not raise any new ethical issues, but again, the basis for data sharing needs to be clearly laid out in a Materials Transfer Agreement. CTSU leads a number of such collaborations, including a number of long-standing systematic overview collaborations (for example, the Early Breast Cancer Trialists' Collaborative Group involves over 600 trialists from around the world)

It should be recognized that access to data (e.g. data sharing between researchers) is necessary but, by itself, insufficient for generating new knowledge. In many cases it is necessary to understand how those data have been collected and processed, and to have the appropriate skills, technology and resource to analyse and interpret the results.

Consultation Question 5:

What are the opportunities for, and the impacts of, data linking in medical practice?

We will limit our response to the potential for linked routine data to be of value for research. The UK is one of only a very few developed countries that has a single universal healthcare system. There are many examples – with public support – where routine NHS data has been used to support research that would not otherwise have been possible to undertake. An example is UK Biobank, which used NHS records to identify and invite over 9m people to take part – with very few concerns raised - and is currently also using routine records for death, cancer and hospital admissions to follow the health of the cohort of 500,000 participants.

While much greater use should be made of routine health records for research, the public should be kept informed how their data is being used and retain the right to opt-out of this use if they so wish. For example, the care.data programme being created by the Health & Social Care Information Centre (HSCIC) will link primary and secondary care data to support clinical commissioning, but also includes and informs patients of the use of this data for research, with patients retaining the right to opt-out of their data being included if they so wish.

It should be noted that such model of opt-out consent is not appropriate in all circumstances. For example, tracking outbreaks of some of the most serious infectious diseases requires comprehensive data collection (without consent). Similarly, to be reliable, monitoring of surgical results requires information about all patients that have been treated (and may require linkage to other data sources, e.g. to obtain details of subsequent hospitalization or death). In both cases, omission of some individuals could result in seriously misleading results.

Consultation Question 6:

What are the opportunities for, and the impacts of, using biomedical data outside biomedical research and health care?

Patients and the public agree to the use of their data on the basis of specific future uses, which for biomedical data is for the purposes of their clinical care and/or to support research. While using such data to support related activities, such as planning healthcare provision and public health, has the general support of the public, using their data for unrelated activities may not generally be supported and would undermine the relationship of trust.

Consultation Question 7:

What legal and governance mechanisms might support the ethical linking and use of biomedical data?

The current requirement for ethical review and individual participant consent with the option to withdraw is adequate for current and anticipated future research. That said, greater emphasis needs to be placed on consent not being presented as a one-time event, but rather an ongoing relationship between the researcher and the research participant. Opportunities will certainly arise in future to greatly enhance existing health resources using additional new sources and types of data not currently used or even in existence at the time of recruitment. Research participants will expect researchers to exploit the potential of these new opportunities. Ongoing communication with participants, using informative and regularly updated websites as well as social media should be more widely used to maintain and enhance the ongoing relationship between the researcher and participants.

Our experience is that participants are often very interested in the results and new knowledge that is being generated from their contribution.

[It is worth noting, that in many circumstances it is not practical to send such details in all participants. For example, many participants move

Consultation Question 8:

Any other comments

High quality health care is critically dependent on the thorough, reliable and timely analysis of multiple sources of data on large numbers of individuals. New advances in methods to collage and analyse biological and health data offer the potential for a step-change in the way that healthcare is practised. Such an Electronic Learning Healthcare System is able to span research & practice, providing facilities for rapid evaluation, analysis, decision making and knowledge distribution to improve the quality of healthcare for both individuals and whole populations.

Such an approach could address a range of important questions, For example:
- Is this person at high risk of diabetes?
- Does this treatment work? Is it safe? Is it better than the alternative?
- What would be the best treatment for this particular person?
- Is this surgeon competent to perform this operation?
- How many nurses should there be on this ward?
- Is funding being spent on diseases or populations that are most needy?

Each of these requires data about the individual combined with knowledge from previous experience (clinical trials, observational studies, audits, etc) and modeling of the information.

It is possible to impose checks at each stage of the information pipeline:
- Approval of the study
- Individual consent by participants
- Agreement to allow linkage to a new data source
- Permission to access the accumulated data (data sharing)
- Dissemination of the results (knowledge sharing)
Which of these is (are) appropriate for a particular activity depends on the nature of activity.