Seminar Series:

Facts and pitfalls of observational studies - How to plan and conduct HRO projects

Q&A from the session

"Data governance and protection: How to navigate the regulatory jungle"

- What regulations/laws apply when a research project is relying on anonymized data owned (with mandatory in-house analysis) by a private, for-profit company, keeping in mind that the only deliverable promised by that private company is a series of 'global tables' (no individual data)?
    - If the data is purely anonymous, the HRA does not apply.

- Where Must the Deanonymization key be stored and how? Is the Project leader allowed to Store it?
    - There is no such thing as a deanonymisation key – if there would be a key to re-identify the involved persons, the data have only been pseudonymised, not anonymised.
    - The pseudonymisation key must be stored separately from the biological material or personal data and in accordance with the principles of article 5 par. 1 HRO, by a person to be designated in the application who is not involved in the research project (article 26 par. 2 HRO).
    - According to article 5 par. 1 HRO, any person who stores health-related personal data for research must take appropriate operational and organisational measures to protect it, and in particular:
        - restrict the handling of the health-related personal data to those persons who require this data to fulfil their duties;
        - prevent unauthorised or accidental disclosure, alteration, deletion and copying of the health-related personal data;
        - document all processing operations which are essential to ensure traceability.

- Question concerning SCTO Guidance on new Federal Act on Data Protection: «Keeping a register of data processing activities is now mandatory (Art. 12)» -> How should such a register look like and what has to be included? What does that mean for small academic groups?
    - Preliminary remark: The federal Act on Data Protection (LADP) is applicable for *private* hospitals and *private* research institutions. I suppose that these entities must indeed maintain such a register according to article 12 LADP.
    - However, for *public* hospitals and *public* research institutions, the *cantonal* Data Protection Act is applicable. E.g. for the USB, this is the Information and Data Protection Act of the Canton Basel-City (IDG BS) and its corresponding ordinance. The IDG BS contains a Clause similar to article 12 LADP (§ 24). USB has, in compliance with this Clause, created and published a list of procedures involving personal data (you can find the relevant passage in regard to research on page 4):

https://www.unispital-basel.ch/dam/jcr:16419282-fdaf-4eab-91b9-c6a95a5283e1/USB%20Verfahrensverzeichnis%20mit%20Personendaten.pdf

- Should data governance be separated form data protection? meaning managed by 2 different departments?
    - Data Governance defines the overall strategy for data (lifecycle) management, data availability, data quality and data protection. It controls how data are collected, stored, used, and deleted. Data protection is a subset of Data governance, responsible for technical and organizational measures to protect data privacy and data integrity.
    - Ob ein eigenes Department für Data Protection zuständig sein soll, oder ob Data Protection mit Data Governance in einem Department abgedeckt werden soll, hängt von der Grösse des Spitals/Unternehmens und den konkreten Zielen ab. Mit HRO hat diese Frage aber eigentlich nichts zu tun.

- How to deal with registry data governance?
    - The requirements are the same as for every other project: restrict access to data, ensure safety of data, document processing activities

- Who owns the data in studies with healthy participants? There are not patients (no health data collected but biological material such as blood/urine/exhaled breath air)
    - Data is owned by healthy participants -> same as with patients, data is owned by person

- How is 'genetic data' defined? Is the name of a rare genetic disease already 'genetic data'?
    - HRA Art 3 g) *Genetic data* means information on hereditary properties or properties acquired during the embryonic phase obtained by genetic testing

- Can the GC only be used for retrospective projects or is it also valuable for prospective Projects that do not have specific study-related Appointments/exams?
    - Refers to big cohort study, where routine data are used, and the discussion point is whether it is prospective or retrospective and if GC is enough --> if only routine data are used, GC is enough (if that GC includes the reuse of future samples and data)

- If samples/data from a Patient that initially signed GC were not used before Patient withdraws GC, could those older samples be used after withdrawal?
    - no, from the point of withdrawing on, no data or samples can be used anymore, no matter if they already have been used or not
    - if they have been used before, those data can stay in that analysis

- Can Coded data be considered anonymised when a research team has given away the code list? Or when the code list has been lawfully deleted?

    - Data is only anonymized if the re-identification is not possible without disproportionate effort. Since the code list still exists, it is easy to restore all identities – it is thus not anonymous, it is still only pseudonymized
    - Even when the code list has been deleted, other so-called "key variables" could be used to link the data to external sources for a re-identification. In order to consider data anonym, the risk or re-identification needs to be determined and below an acceptable threshold.

- How about if patients died without signing GC? could their data still be used?
    - If the person could not consent before death, a close relative or trusted person has to be informed and can consent
    - If the death of the person concerned occured more than 70 years previously, research may be carried out without informed consent, but only if closest relatives do not object.
    - For deceased persons undergoing artificial respiration, research may only be carried out if results cannot be obtained with persons not undergoing artificial respiration, and if the death has been determined by an authorised person who is independent from the research project.

- Can unencrypted sensitive data be shared using HIN-secured email?
    - In theory, if it is made sure that all security measures are still used as intended, it would be possible – in practice however not recommended, since already a forwarding of an email would breach the security measures

- Since anonymised health-related data doesn't fall under the HRA nor the DPL, and since full/proper data anonymisation (i.e de-identification) has become increasingly challenging, shouldn't regulatory bodies or institutions discourage the use of anonymised data for research purposes?
    - Question of opinion: anonymisation is a tool that can be helpful and properly used and if DTUAs etc are in place; but definitely discourages use of anonymised open data -> never sure if safe against reidentification, and data out there without any control
    - Since anonymisation always comes with information loss, pseudonymisation should be preferred over anonymisation

- Are pseudomized Health data still considered as personal data?
    - Clear yes -> key there, reidentification possible -> still personal data as before

- A question about research on unstructured data analysis with Large Language Models (e.g. text entries about patients): they are considered non-identifiable data, right? (might contain names, personal informations etc.) Is there a way to use this data retrospectively via the GC?
    - Very hard to answer without knowing specifics. Strongly suggest to discuss this with the DPO or legal expert at the institution.

    - Free text is almost impossible to anonymise in an automated way, so unless one manually reads through the text and blacks all identifiers, data is personal
    - Note that LLMs can have other potential ethical or legal problems, e.g. consent for research may not include consent to use data for modell training, and all data protection measures need to be guaranteed if data are transferred to an external LLM server

- To whom does anonymized data belong to?
    - Dataset of aggregated data would belong to the institution where this intellectual property has been generated

- To whom belong data of deceased Person/fetuses?
    - Preliminary remark: «Belonging» is not a precise term in this context; the question should rather be: Is research which involves deceased persons legal and under which conditions? These conditions are described in Clauses 36 – 38 HRA.
    - Fetuses: The conditions are described in Clauses 39 and 48 HRA.

- You say that identifiers should not be sent by Email. However in our institution, we need to send a list of patients (PID, name and birth date) to get information on whether patients have sighned the GC or not. (Pre-screening before handing in an ethical application as well as to know, once approved, which data is allowed to be used) The mail is sent from the organisation/institute mail to the receiver organisation/institute mail add. Is this not okay?
    - If the e-mail is properly encrypted (e.g. by HIN) that would be OK, although I would not recommend it. One solution would be to simply password protect the list of patients (and exchange the password via phone). Or you could use a secure cloud provided by your institution, to which you and your partner have access.

- Not all data can be anonymized, right? For example, imaging data.
    - In general, imaging data is hard to anonymise. One has to go through the process of defining a "disclosure scenario": Who and how could a re-identification happen. Than one needs to determine the re-identification risk and decide whether its acceptably low.
    - Blacking or blurring of direct identifying features (like eyes) in an image is probably no longer acceptable, as with Deep-Learning other features (like facial bone structure) may suffice to re-identify a person (especially if there is a public picture somewhere, e.g. in social media or company webpage).
    - (example: In MRI scans, for example, persons can be identified using this information, or from the shape of teeth and ears --> all these parts of the image would have to be blackened for anonymisation --> questionable if the image is still of use)
    - See e.g. Abramian, D., Eklund, A., (2019), REFACING: RECONSTRUCTING ANONYMIZED FACIAL FEATURES USING GANS, 2019 IEEE 16TH INTERNATIONAL SYMPOSIUM ON BIOMEDICAL IMAGING (ISBI 2019), 1104-1108. https://doi.org/10.1109/ISBI.2019.8759515