I welcome the EDPB Guidelines 1/2026 and their effort to clarify the application of the GDPR to scientific research. From the perspective of a clinician-scientist working in a public hospital and university environment, the document is timely and necessary. Clinical research increasingly depends on the responsible secondary use of routinely collected healthcare data, including laboratory results, operative records, hospital length of stay, reinterventions, imaging studies, and long-term clinical outcomes. These data are essential not only for academic publication, but also for improving patient safety, evaluating real-world effectiveness, reducing unnecessary variation in care, and developing trustworthy clinical decision-support tools.

The guidelines appropriately recognise that scientific research is a societal good and that further processing for scientific research purposes may be presumed compatible with the original purpose of data collection. However, I believe the final version would benefit from more explicit practical guidance for retrospective clinical research and hospital-based research data infrastructures.

1. Clarify the lawful and ethical pathway for retrospective clinical research using routinely collected health data

In many public hospitals, a large proportion of clinically relevant research is retrospective, observational, and non-interventional. It relies on data originally collected for the provision of care. In such cases, obtaining individual consent from every patient may be impracticable, may introduce selection bias, and may seriously impair the scientific validity of the study. This is particularly relevant in vascular surgery and other fields where outcomes such as amputation, mortality, reintervention, complications, or long-term patency require longitudinal datasets.

The guidelines should more clearly state that retrospective clinical research using existing health data may be lawful when it is conducted under an appropriate legal basis, with a valid Article 9 derogation, institutional authorisation, ethics oversight where applicable, data minimisation, pseudonymisation or anonymisation whenever possible, and appropriate technical and organisational safeguards. Consent should remain an important mechanism, but it should not be presented, explicitly or implicitly, as the default or only acceptable pathway for legitimate retrospective health research.

2. Provide a practical example of a hospital clinical research data repository

The guidelines already address research infrastructures, but the final version should include a specific example based on a public hospital or university hospital creating a clinical research data repository. Such a repository could include pseudonymised data extracted from electronic health records, operating theatre records, laboratory systems, administrative data, and imaging archives, for use in approved scientific projects within a defined area of clinical research.

The example should describe a proportionate governance model: institutional controller responsibility; approval by a research governance body or ethics committee; involvement of the data protection officer; project-level access requests; role-based access controls; logging; secure processing environments; restrictions on export of individual-level data; output checking for small-cell or re-identification risk; confidentiality commitments; and publication only of aggregate or sufficiently de-

identified results. This would be extremely useful for hospitals seeking to enable research while protecting patients.

3. Clarify the distinction between anonymisation and pseudonymisation in high-dimensional clinical datasets

The guidelines correctly recommend anonymised data, or alternatively pseudonymised data, whenever the research purpose can be fulfilled in that way. However, in clinical medicine, anonymisation is technically difficult and sometimes misunderstood. Removing name, address, and patient number is not sufficient when datasets include rare diagnoses, exact dates, procedure combinations, imaging studies, small subgroups, or longitudinal trajectories.

The final version should explicitly acknowledge that many clinically useful datasets described as “anonymised” are in reality pseudonymised. This is especially true for medical imaging, genomic data, rare disease registries, vascular anomalies, complex surgical cases, and datasets involving detailed chronology. Clearer practical examples would prevent both unsafe overconfidence and excessive institutional risk aversion.

4. Add specific guidance on medical imaging and DICOM data

Medical imaging deserves specific attention. CT angiography, MRI, ultrasound, fluoroscopy, and DICOM datasets may contain direct identifiers in metadata, embedded annotations, accession numbers, dates, institution identifiers, and sometimes anatomical features that increase re-identification risk. In some cases, imaging data may also be linked to operative reports, laboratory values, outcomes, or device information.

The guidelines should include a dedicated example on imaging research, including de-identification of DICOM metadata, management of dates and accession numbers, segmentation files, derived images, linkage keys, storage of raw and processed images, transfer to external collaborators, and the use of imaging data for artificial intelligence development. This is a rapidly expanding area where many clinicians and hospitals need practical guidance.

5. Strengthen the risk-based and proportionality approach

The final document should avoid creating the impression that all clinical research data processing requires the same level of governance as high-risk, multinational, commercial, or AI-development projects. A small single-centre retrospective audit using pseudonymised data is not equivalent to a large-scale imaging dataset transferred to an external commercial partner for machine-learning development.

A risk-tiered approach would be helpful. For example, the guidelines could distinguish between: low-risk retrospective studies using limited pseudonymised structured data; moderate-risk research databases with longitudinal linkage; high-risk studies involving rare diseases, imaging, genomics, free text, or vulnerable groups; and very high-risk projects involving external commercial partners, AI model training, cross-border transfer, or large-scale linkage. This would allow proportionate safeguards while avoiding unnecessary barriers to clinically valuable research.

6. Clarify transparency obligations where individual information is impracticable

Transparency is essential. However, in retrospective clinical research involving thousands of historical records, individual notification may be impossible or disproportionate, and may itself require additional processing of contact data. The guidelines should provide more practical examples of acceptable transparency mechanisms in this context.

Hospitals should be encouraged to use layered public information notices, research webpages, institutional privacy notices, patient portal information, posters in clinical areas, annual reports of approved research projects, and clear mechanisms for questions or objections where applicable. Such mechanisms may offer meaningful transparency without undermining the feasibility and validity of retrospective research.

7. Clarify attribution of responsibility in clinician-led academic research

Clinical research frequently involves overlapping roles: the hospital as care provider and data holder, the university as academic partner, the clinical department as the research environment, and the individual clinician as principal investigator. The guidelines should include examples clarifying when the hospital is the controller, when the university may be a joint controller, when an external collaborator is an independent controller, and when an individual clinician is acting under the authority of the institution rather than as a separate controller.

This is particularly important because uncertainty about controller responsibility often delays or prevents legitimate research, even when the project has clear scientific merit and appropriate safeguards.

8. Support secure access models rather than unrestricted data export

The guidelines should more strongly promote secure processing environments for clinical research, especially for large-scale datasets. In many cases, the safest model is not to export full datasets to individual researchers, but to allow controlled analysis within a secure environment, with logging, access restrictions, and export only of aggregate outputs or approved derived datasets.

This model would be especially valuable for hospital research repositories, multicentre registries, imaging research, and AI development. It protects patients while allowing clinically important research to proceed.

9. Address AI development from clinical data without discouraging responsible innovation

Artificial intelligence can generate real benefits in clinical medicine, including risk prediction, image analysis, triage, quality improvement, and personalised follow-up. However, AI development using clinical data also creates specific risks: model memorisation, re-identification, dataset shift, bias, opaque secondary use, and uncontrolled downstream deployment.

The guidelines should include a practical clinical AI example distinguishing research and development from deployment in clinical care. They should also encourage safeguards such as ethics review, DPIA where appropriate, data minimisation, federated learning when feasible, secure environments, performance monitoring, bias assessment, model documentation, and restrictions on re-use by commercial partners.

10. Avoid making governance so complex that only large commercial actors can comply

A final concern is that excessive procedural complexity may unintentionally harm public-interest clinical research. Large commercial actors may have the legal and technical resources to comply with complex requirements, whereas public hospitals, academic departments, and clinician-led research groups often do not. The final guidelines should therefore provide simple, practical templates or examples for compliant low- and moderate-risk clinical research.

The objective should be to protect patients while enabling scientifically valid, ethically supervised, public-interest research. Patients benefit when their data are protected, but they also benefit when healthcare systems learn from previous care.

In summary, I strongly support the direction of the guidelines. I respectfully suggest that the final version should include more detailed examples and operational guidance for retrospective clinical research, hospital research data repositories, medical imaging, AI development, transparency in historical datasets, and institutional responsibility in clinician-led academic research. This would make the guidelines more useful for hospitals and universities and would help ensure that data protection and clinical science reinforce, rather than obstruct, each other.