DIGICORE: toward a European Digital Institute for Cancer Outcomes Research, and a practical answer to RWD studies

Giovanni Apolone[1,2], ***Gennaro Ciliberto***[3], ***Xosé Fernández***[4], ***Claudio Lombardo***[2,5], ***Piers Mahon***[6] *and* ***Ashley Woolmore***[6]

1. Fondazione Istituto Nazionale Tumori, Milan – Italy
2. Organisation of European Cancer Institutes - EEIG, Brussels – Belgium
3. Istituto Nazionale Tumori Regina Elena, Roma – Italy
4. Institut Curie, Paris – France
5. Sustaining Oncology Studies Europe Srl, Genoa – Italy
6. IQVIA Ltd, London – UK

Real World Evidence can be a powerful complement to traditional trials that allows the clinical research community to tackle certain important research topics. These vary from outcomes research to establish the true efficacy of treatment on real populations, through to improving evidence around clinical decision making (especially based on clinical biomarkers) to health systems research to optimise entire care pathways or understand the dynamic nature of care – for instance in different cancer care system responses to COVID-19.

Making high quality real world research takes effort just like high quality trials. IQVIA (the contract research organisation) with UNICANCER in France, Alleanza Contro il Cancro in Italy, and other cancer centres/institutes already certified by the Organisation of European Cancer Institutes (OECI), have been developing a large scale real world research alliance called the "DIGital Institute for Cancer Outcome REsearch" (DIGICORE).

This article lays out the opportunity and the challenges ahead in driving that joint Real World Research programme, based on the partnership's collective experiences to date. We do this to extend a warm welcome to other similar existing networks and cancer centres to join us in our mission to "make every willing cancer patient a research patient and so transform cancer care".

For formal proof of comparative efficacy, the well designed, appropriately powered randomized controlled trial is the gold standard. However, it is costly, often slow and while internally consistent may not generate results that are reflected in the co-morbid, complex patient pools of clinical reality. We also need to recognize that many innovations come to market on proxy endpoints such as Progression Free Survival, which does not correlate well with Overall Survival (and rarely establish that survival benefit once on market).

There are also certain research questions trials cannot easily tackle, for instance to understand how patients are being treated today and the impact of that variation in care on care quality and cost. Trials also struggle to be powered to understand rare events or rare sub-groups. For these topics, observational research may be a useful and complementary approach to randomized trials.

Traditionally, the research community has tackled such questions with consented prospective or retrospective observation studies, as typically used for safety studies. These rely on manual retype from the medical record to an electronic clinical research record (eCRF). eCRF studies have the advantage that they can work with records in any format, including paper. But their manual approach is high cost, creates case selection bias and requires costly supervision and project management to drive appropriate data quality and timely data capture.

However, inside today's electronic medical records (EMR) lies a wealth of information that could create a faster and more efficient research solution for many important research topics, especially in precision oncology given its rare patient groups. The challenge is to make the diverse information collected in the delivery of routine clinical care "ready for research".

What does it take to make electronic medical records research ready? In short: some technology and a lot of research process and methods innovation (not the other way around). Done this way, high quality research is approachable at reasonable cost by most comprehensive cancer centres, clinical centres, and national cancer associations who can comply with the principles within the OECI. Not surprisingly, many OECI Members have already experimented individually with local EMR based research. But to get scale and representativity, we need a large highly interoperable multi-centre, international network – which is much more technically challenging.

The technical challenges for creating network interoperability grow with scale, with these common:

i) The broad range of data definitions, languages and IT systems available
ii) Variation in the practice of medicine, especially internationally
iii) The need to solve for both unstructured and missing data
iv) Appropriate GDPR compliant privacy solutions
v) Internal capability of each cancer centre to put forward skilled human resources

These are now solvable at scale, and there are many examples of those solutions up and running across Europe. The PIONEER programmes have shown that OHDSI's OMOP[1] common data model is extendable to cancer and that a single common data model can be implemented over multiple European countries and EMRs, as well as in elite US cancer centres.

UNICANCER has developed Consore[2], a federated search engine empowering fast data queries across EMRs at national scale. Multi-centre patient cohorts can be identified with this tool in a matter of minutes, instead of weeks. Consore annotates and standardises medical records relying on key medical references (ICD-10, SNOMED, etc.) to structure all the patient files. To illustrate, Institut Curie needs +3000 new documents to be added to EMR system on a daily basis – these become searchable. The system can also infer patient disease history, with machine learning approaches helping to improve the quality of the inference made on the highly heterogeneous underlying data. Consore's underlying data model is highly aligned to OMOP.

Millions of patient descriptions, clinical narrative reports, chemotherapy protocols, administration data, tumour characteristics… are indexed over hundreds of millions of documents across the network nodes. The search engine delivers timely responses across the various comprehensive cancer centres in the network without centralising data in a single location. Each Consore node is deployed in a dedicated environment and can only be queried by other authorised nodes. Only the number of matches corresponding to a given question is provided, no other data items are shared in this process.

A couple of examples can illustrate how Consore empowers clinical teams at Institut Curie for academic research. Mining EMR to identify pregnancy cases after breast cancer was hugely simplified. Likewise, it is possible to analyse and reveal the importance of comedications (and comorbidities) in a multi-centre breast cancer cohort by analyzing the influence on immune infiltration and pathological response to neoadjuvant chemotherapy. However, by being limited to counts, Consore can only identify the members of a research cohort, not drive a full protocol to results.

IQVIA's Oncology Evidence Network (OEN) focused more on solving for transforming all forms of local data into research quality data and the end-to-end delivery of a specific multi-centre protocol. Member sites today have on-site teams and data tools ready to curate and enhance records under hospital

[1] The Observational Health Data Sciences and Informatics (or OHDSI, pronounced "Odyssey") program is a multi-stakeholder, interdisciplinary collaborative to bring out the value of health data through large-scale analytics. It's Observational Medical Outcomes Partnership (OMOP) common data model allows the harmonization of disparate clinical coding systems - with minimal information loss - to a standardized common vocabulary.

[2] ConSoRe (Continuum Soins Researche or continuum of care research) is an evolved tool for semantic search, associated with a Clinical Data Warehouse (CDW), enhancing the use of patients' data in oncology research.

control for ethics approved, protocolised research. IQVIA developed comprehensive end-to-end approaches to privacy with pseudonymisation and risk assessment technologies. With that approach it is possible to reconcile the data diversity over international network to a research fit, protocol specific, common data model and then use data science to drive the protocol at each centre. To give a sense of the power of this approache, the record to date from protocol acceptance to research insights on a large cohort that can be released to a study sponsor is 10 working days. Congratulations to fellow OECI member, the Frankfurt University Cancer Centre, for that record!

As we said earlier the technology and data science are the "easy bit" (if also the exciting bit!). Much harder is to systematically understand and tackle the organizational barriers to collaborative research and to develop new and effective ways of working to solve them.

The two most important of these barriers are a lack of trust and a need for control. Cancer centres have a duty of care to make sure their records are used appropriately for research and that their research autonomy is respected. At the same time, these new methods are complex and require new skills, making the assessment of appropriate controls (such as technical privacy standards) challenging for many centres.

For these reasons of trust and control, the above cited partners agreed to create a new European Economic Interest Grouping with several OECI cancer centres. This new organisation is called the DIGital Institute for Cancer Outcome REsearch (DIGICORE), with the objective to become the European Digital Cancer Institute and global destination of choice for high quality real world research. DIGICORE's constitution also puts the cancer centres "in charge" with 1 member, 1 vote. It also enshrines (among other things) the inalienable rights of a cancer centre to their own data, to research autonomy and to clinical decision making, for instance in molecular test choice.

It will catalyse both high quality academic and commercial international collaborative research.

Two areas are of particular focus for the DIGICORE academic programmes.

The first will be joint programmes of work to develop and validate care quality analytics that are fit for the precision era, such as tracking guideline adoption in near real time and in measuring the impact of those analytics on care quality and outcomes. This focus on care quality improvement fits well with the recommendations of the European Cancer Mission and the European Beating Cancer Plan. Quite rightly – and as recognized from inception by the OECI - patients must be of prime concern to healthcare professionals. Patients rightly demand to receive the best personalised care available. It is only through a swift and thorough analysis of real-world data that it is possible to establish if a therapeutic protocol was truly effective or, at least in part, whether it failed.

The second is in joint research methods development and validation to make best use of these new digital research infrastructures for precision oncology care development. As an example, the development of novel semi-automated study designs that could drive predictive biomarker discovery and validation. Like these will use Mendelian randomisation applied for the first time at scale to real world somatic mutation data upstream of current standard of care therapy. Such methods could be applied to a broad range of cancer therapy responses (such as radiotherapy or generic chemotherapy) to work out *"what works, what doesn't and why"* and so improve clinical decision making and the cost effectiveness of cancer care. As a result of such research we will find subgroups of patients which had been historically hypothesized to be similar, but actually yielded different responses to the same therapeutic protocols. Over time, this will improve care outcomes – and care cost effectiveness.

Secondly, via IQVIA, DIGICORE members have access to programmes of work linked to pharma sponsored real world research opportunities. These start from digital trial site selection and recruitment solutions that help connect patients and their clinicians to appropriate trials. But regulators are also innovating on the use of real world evidence to support the introduction of novel agents (especially in narrow indications). To provide three examples, there are now large effect size drugs that have conditional market approval based on single arm trials and real world comparators. There are now drugs that have secured second indications based on off-label real world evidence alone, for instance in male breast cancer. Finally, recent Dutch publications have shown the power of multi-centre off-label case series in de-risking and accelerating next indication development in secondary indications .

In conclusion, DIGICORE is a new solution geared towards implementing an innovative way to support collaboration between clinical data producers. The Grouping will give a voice to the real users of the incredible wealth of knowledge behind clinical data: the patients and catalyze transformative cancer research.

The DIGICORE constitution will be warmly welcomed by the organisations appointed to coordinate national and European initiatives, which can support the establishment of a European health system where citizens are the protagonists of a dynamic dialogue with those who have the power to assess the outcomes of consolidated or experimental protocols.

Cancer institutes at the forefront of cancer care are well aware that their patients' data are a fundamental tool to advance knowledge and therefore a veritable heritage of the entire community. DIGICORE will serve as a meeting point for the cancer community where partners may find a shared solution to all those questions that involve patients, care providers and private collaborators.

We welcome sister cancer centres of the OECI – and more broadly – to join with us to deliver our mission to

"make every willing cancer patient a research patient and so transform cancer care"