



Reflections on the clinical implementation of precision medicine

Experiences from BigMed, a Norwegian ICT Lighthouse project





ISBN 978-82-473-0091-6

Please cite this report as:

Reflections on the clinical implementation of precision medicine – Experiences from BigMed, a Norwegian ICT Lighthouse project. Vibeke Binz Vallevik, Alia Zaka, Bobbie Ray-Sannerud, Erik Fosse, Pål H. Brekke (editors) on behalf of the BigMed consortium. Oslo University Hospital, Oslo, Norway (2021).



Design, layout and print production of this report: ETN Grafisk

BigMed visual identity: Melkeveien

All images from Shutterstock, unless specified.

About this report

BigMed, launched in 2017 and funded by the The Research Council of Norway, was the first major publicly funded precision medicine initiative of its kind in Norway. Hosted by Oslo University Hospital, the project brought together partners from clinical organisations, academia, patient organisations, and industry to address barriers to the implementation of precision medicine and pave the way for big data analytics in healthcare.

The 2018 BigMed position paper, *Big data management for the precise treatment of three patient groups*¹, which identified the initial barriers to clinical implementation of precision medicine, was the starting point for this project. This report summarizes the footprint and knowledge developed during the project, and points to more detailed BigMed resources where available. Participants from this unique cross sector and cross competence project have contributed through discussions, deliverables and knowledge summaries, resulting in a report that includes perspectives, reflections and wisdom from a broad consortium. The content covers both visions and future goals, as well as practical examples of solutions made in the project.

Throughout the project, knowledge has been documented through podcasts episodes, seminars, articles and technical reports. These are referenced where relevant in this report as further materials and are available on:

[BigMed.no](#)



Contributors

Many people contributed to this report through the various forums and groups in and around BigMed. Contributors from within the project are listed with name and affiliation below. Many stakeholders and key resources were part of making this project a success. We would like to thank each and every one of you for your contributions.

Editorial committee

Vibeke Binz Vallevik	DNV
Alia Zaka	Sykehuspartner
Bobbie Ray-Sannerud	DNV
Erik Fosse	Oslo University Hospital
Pål H. Brekke	Oslo University Hospital

Contributing authors

Alia Zaka	Sykehuspartner	Knut Lindås	Sykehuspartner
Andrew Reiner	Oslo University Hospital	Kristina Haugaa	Oslo University Hospital
Anita Moe Larsen	Norway Health Tech	Laura Slaughter	University of Oslo
Anne Jorunn Stokka	DNV	Lilja Øvrelid	University of Oslo
Anne Kjersti Befring	University of Oslo	Liv Bollvåg	DIPS
Arnoldo Frigessi	Oslo University Hospital, University of Oslo	Loek Vredenberg	IBM
Bjørn Næss	DIPS	Monica Gomez	University of Oslo
Bobbie Ray Sannerud	DNV	Morten Eike	Oslo University Hospital
Courtney David Nadeau	DNV	Oda Bakken	Oslo University Hospital
Dag Frode Nilsen	Sykehuspartner	Odd Arild Lehne	Norway Health Tech
Eivind Hovig	Oslo University Hospital, University of Oslo	Oleg Agafonov	DNV
Eline Aas	University of Oslo	Øystein Nytrø	NTNU
Frédéric Courivaud	DNV	Pål H. Brekke	Oslo University Hospital
Fredrik A. Dahl	Akershus University Hospital	Petter Hurlen	Akershus University Hospital
Gard Thomassen	University of Oslo	Randi Borgen	Oslo University Hospital
Gjertrud B. Mageli	Oslo University Hospital	Serena Elizabeth Marshall	DNV
Hariharan Michael Hallock	DNV	Sevald Cirkov	Sykehuspartner
Ildiko Pilan	University of Oslo	Sharmini Alagaratnam	DNV
Inger Johanne Sand	University of Oslo	Stephen McAdam	DNV
Ingrid Stenstavold Ross	The Norwegian Cancer society	Taraka Rama	University of Oslo
Jan F. Nygård	Cancer Registry of Norway	Tony Håndstad	Oslo University Hospital
Karen Irgens	DNV	Vebjørn Arntzen	Oslo University Hospital
Kathrine Myhre	Norway Health Tech	Vegar Dagenborg	Oslo University Hospital
		Vibeke Binz Vallevik	DNV
		Yngve Sejersted	Oslo University Hospital

Hearing and comments

BigMed steering committee

Erik Fosse	Head of Intervention Centre, OUH	Hariharan Michael Hallock	DNV
Giske Ursin	Managing Director, Cancer registry of Norway	Inger Johanne Sand	University of Oslo
Hans Henrik Merckoll	CEO IBM Norway, IBM	Jan Nygård	The Norwegian Cancer Registry
Inger Johanne Sand	Institute for public law, University of Oslo	Jahn-Henry Løvaas	DNV
Kathrine Myhre	CEO (prev), Norway Health Tech	Kenneth Iversen	Sykehuspartner
Kenneth Iversen	Director, Customers & services, Sykehuspartner	Kjersti Flatmark	Oslo University Hospital
Lars Åge Møgster	Head of Analytics, Akershus University Hospital	Knut Lindås	Sykehuspartner
Liv Eli Lundeby	Special advisor, The Norwegian Cancer Society	Kristina Haugaa	Oslo University Hospital
Odd Arild Lehne	Managing Director (prev), PubGene	Lars Åge Møgster	Akershus University Hospital
Olaug Råd	Research Council of Norway (observer role)	Laura Slaughter	University of Oslo
Sigbjørn Smeland	Head of the Cancer clinic, Oslo University Hospital	Lilja Øvreliid	University of Oslo, Inst. of Informatics
Stephen McAdam	Director of Digital Health, DNV	Liv Eli Bollvåg	DIPS
Thomas Smedsrød	Medical director, DIPS	Loek Vredenberg	IBM
Øystein Krüger	Head of Research, South East Health Authority	Mari Eide Frey	Oslo University Hospital
Øyvind Skraastad	Head of ER Clinic, Oslo University Hospital	Monica Gomez	University of Oslo

Project participants, work package leaders highlighted

Alia Zaka	Sykehuspartner	Oda Bakken	Oslo University Hospital
Andrew Reiner	Oslo University Hospital	Odd Arild Lehne	PubGene
Anita Moe Larsen	Norway Health Tech	Oddgeir Hvidsten	Sykehuspartner (prev.)
Anne Jorunn Stokka	DNV	Oleg Agafonov	DNV
Anne Kjersti Befring	University of Oslo	Petter Hurlen	Akershus University Hospital
Arild Jacobsen	Sykehuspartner	Pål H. Brekke	Oslo University Hospital
Arild Waaler	University of Oslo, Inst. of Informatics	Randi Borgen	Oslo University Hospital
Arnoldo Frigessi	Oslo University Hospital, University of Oslo	Roar Jakobsen	Sykehuspartner (prev.)
Arve Kaaresen	Oslo University Hospital	Sen Zhao	Oslo University Hospital
Asbjørn Finstad	Sykehuspartner (prev.)	Serena Elizabeth Marshall	DNV
Bjørn Edwin	Oslo University Hospital	Sevald Cirkov	Sykehuspartner
Bjørn Næss	DIPS	Sharmini Alagaratnam	DNV
Bobbie Ray Sannerud	DNV	Sigve Nakken	Oslo University Hospital
Chiara Rasi	SciLifeLab, Karolinska	Stephen McAdam	DNV
Courtney David Nadeau	DNV	Stig Ronny Kristiansen	Oslo University Hospital - ICT
Dag Are S. Hov	PubGene	Taraka Rama	University of Oslo
Dag Erik Undlien	Oslo University Hospital, Dept. of Medical Genetics	Thomas Brox Røst	NTNU
Dag Frode Nilsen	Sykehuspartner	Thor Stenbæk	DIPS
Dag Hovland	UiO IFI	Tony Håndstad	Oslo University Hospital
Daniel Haugstvedt	IBM	Tor Solli-Nowlan	Oslo University Hospital
Eivind Hovig	Oslo University Hospital, University of Oslo	Tore Gundersen	Ahus
Eline Aas	University of Oslo	Tri Cao Vu	IBM
Erik Kvanli	DIPS	Valtteri Wirta	SciLifeLab, Karolinska
Fatemeh Kaveh	Oslo University Hospital	Vebjørn Arntzen	Oslo University Hospital
Frédéric Courivaud	DNV	Vegar Dagenborg	Oslo University Hospital
Fredrik A. Dahl	Akershus University Hospital	Vibeke Binz Vallevik	DNV
Gard Thomassen	University of Oslo, ICT	Vigdís Nygaard	Oslo University Hospital
Guro Meldre Pedersen	DNV	Yngve Sejersted	Oslo University Hospital
Haldor Husby	Akershus University Hospital	Øystein Nytrø	NTNU
Haneef Awan	University of Oslo, ICT		

Table of contents

1. The paradigm of precision medicine	15
1.1 What is precision medicine?	15
1.2 How to move forward	17
2. BigMed's framework for addressing the barriers	18
2.1 The Core: data and infrastructure	18
2.2 Legal and ethical framework	19
2.3 Organisational framework and governance	19
2.4 Framework development	19
3. Data capture, analysis, and application – systematic use of health data	20
3.1 A need for a strategy FOR data capture and use	22
3.2 Extracting information from clinical text with natural language processing (NLP)	24
3.3 Biases and pitfalls in using real world health data	27
3.4 Access to genomics data through sharing across organisations	29
3.5 Standardisation and quality assurance of molecular diagnostics	32
4. ICT Infrastructure	37
4.1 Infrastructure of South-Eastern Norway Regional Health Authority for implementation of precision medicine in the clinic	39
4.2 Infrastructure needs of the research & innovation communities	41
4.3 Production platforms for data sharing in real time	44
4.4 Enabling open platforms	45
4.5 Dynamic digital consent	47
5. Organisational frameworks and governance	49
5.1 From ad hoc to adoption – a maturity model for precision medicine	51
5.2 Stimulating an innovation ecosystem with industry cooperation	52
5.3 From innovation to implementation	55
5.4 New clinical study models in the framework of precision medicine	58
5.5 Precision medicine in a health economics perspective	63

6. Legal and ethical framework	66
6.1 Legal reflections to facilitate a data driven healthcare	68
6.2 Evaluation of anonymity with the purpose of sharing genomic data	74
6.3 Ensuring safe application of diagnostic tools – computational trust and regulations	79
7. Practical examples from the implementation of precision medicine in three clinical areas	82
7.1 Prevention of sudden cardiac death	86
7.2 Whole genome sequencing for rare diseases	91
7.3 Colorectal cancer	100
8. Final reflections on the BigMed project	103
8.1 Iterative development in cross competence teams	103
8.2 Safeguarding digital healthcare	103
8.3 A new perspective on data use	103
8.4 An organisation designed for change	104
8.5 Building on the BigMed legacy	104
Appendix: Deliverables overview	107
Appendix: Consortium partners	113
Definitions	116
Abbreviations	117
References	118



Four exciting years

After almost four years of development, experimentation, and knowledge building on clinical implementation of precision medicine, we conclude BigMed with this final report, *Reflections on the clinical implementation of precision medicine*. The report summarizes experiences and reflections from the project and is meant to be a guide for subsequent projects on the implementation of precision medicine.

It has been an exciting and rewarding journey from the first visions and thoughts we had about BigMed in 2016 to the results we see materialise today. It has been an honor to work and discuss with leading researchers and centres both nationally and internationally. We wish to thank all the partners in the project: patient organisations, university faculties, industry, and hospital departments for their input and participation.

BigMed has had the privilege of conducting a dialogue with several important stakeholders throughout the project period, including the Norwegian Ministry of Health and Care Services, the Directorate of Health, and Directorate of eHealth.

By the autumn 2020 the legal department of the Ministry of Health and Care Services suggested changes in the Health personnel Act and the Heath records Act, based in large part on BigMed experiences, that will facilitate the use of patient information to develop an infrastructure for precision medicine. During the coming years we believe further revisions of the legal system and digital structures will continue to emerge and make precision medicine a clinical reality.

We look forward to further discussions, and hope you enjoy the read regardless of whether you are a decision maker, practitioner, expert, or simply share our interest in the clinical implementation of precision medicine.



Erik Fosse, Oslo University Hospital
Project owner

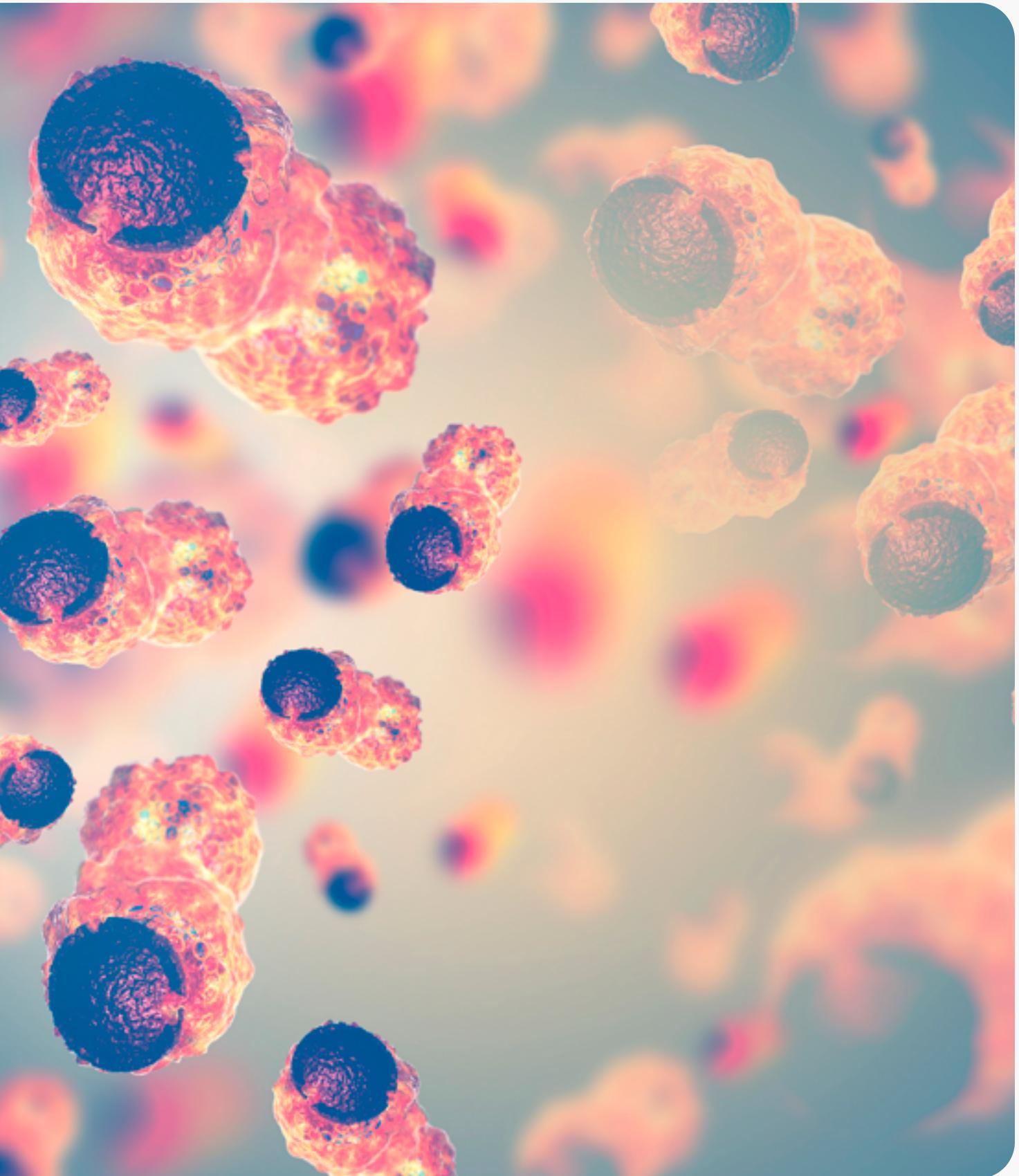


Vibeke Binz Vallevik, DNV
Project manager





Category	Sub-Categories	Value 1	Value 2	Value 3
Genetic Data	Allele Frequency	0.25	0.30	0.20
	Haplotype Block	0.15	0.20	0.10
Environmental Factors	Temperature	25°C	28°C	22°C
	Humidity	45%	50%	40%
Microbiome Analysis	Bacterial Diversity	High	Medium	Low
	Fungal Load	Low	Medium	High
Pathogen Detection	Virus	Present	Present	Present
	Bacteria	Present	Present	Present
Phenotypic Variation	Height	1.75m	1.80m	1.65m
	Weight	70kg	75kg	65kg
Genotype	rs1234567	AA	AB	BB
	rs1234568	CC	CD	DD



The value of health data for the patient

The Norwegian Cancer Society joined the BigMed project to help equip our healthcare system to realize the benefits of personalised medicine.

In cancer, personalised medicine represents a shift in thinking that requires new international cooperation and data sharing to overcome the limits of small databases in Norway. Facilitating this data sharing across national borders, health trusts, and even within levels in the health service, is demanding. We cannot each sit on our separate islands to meet these new needs; infrastructure, regulatory conditions, and financing mechanisms need to change, and we must all work together.

In a world where knowledge develops rapidly, artificial intelligence technology can assist doctors by comparing data from a specific patient to the thousands of other patients with the same profile and guide them towards more effective treatments. This technology provides opportunities that did not exist a few years ago. In the BigMed project, we wanted to find out how we could enable this for routine use. Through the project, we identified the requirements with respect to legislation, funding, and data collection, and were able to give recommendations on what is needed.

Health data privacy is a major concern for us and our patients. Precision medicine requires sharing data but there are concerns about what will happen if a person's personal health data goes astray. It is therefore important that every system that manages personal health data

for precision medicine also takes privacy and personal integrity very seriously. The benefits of using someone's health data must be balanced against their privacy.

The vast majority of cancer patients we have talked to want and expect their data to be helpful to others even if it can't be used to help themselves. We advocate for systems that allow ethical data sharing for this purpose while protecting against private and public players who want to exploit the potential in health data. If we don't enact best practices to protect health data now we risk losing the trust needed to implement precision medicine.

We envision a Norwegian healthcare system that is at the global forefront of treating cancer with precision medicine. We are therefore very proud and happy that the BigMed project is helping to point the way forward for cancer patients, towards a life without cancer.



Ingrid Stenstadvold Ross
Generalsekretær i Kreftforeningen

Executive summary

The BigMed project was funded by The Research Council of Norway (RCN) as an ICT lighthouse project to address the barriers to clinical implementation of precision medicine and pave the way for big data analytics. The consortium, hosted by Oslo University Hospital, was a cooperation of more than 20 partners from industry, academia, and the clinic. In the four years BigMed operated, the multi-disciplinary teams developed solutions on infrastructure, quality assurance, data sharing and clinical decision support based on needs identified through three clinical areas: rare diseases, sudden cardiac death, and colorectal cancer.

In developing a framework for addressing different types of implementation barriers, BigMed considered legal issues, organisational and governance issues, ICT infrastructure, as well as the life cycle of secondary use of data through data capture, analysis, and application.

The project found data is a resource that is seldom reused for clinical decision support. Clinical genomics is an exception and serves as a good example of how data can be reused to support clinical decisions. This can be a model for future development of other areas, to see how data can be used for patient similarity analysis and to speed up knowledge development to the benefit of our patients.

BigMed contributed to reducing the genetic analysis time for critically ill newborns at (OUH) from 8 weeks to 5 days by developing tools to support the implementation and automation of high throughput sequencing pipelines. To monitor quality, BigMed explored

quality control needs and specific tools, and developed and implemented solutions for sharing of genomic data. The legal aspects of sharing genomic data were interpreted and clarified to make implementation possible.

The project developed patient similarity tools for colorectal cancer, building on data from external sources like the Cancer Registry of Norway. These analytic tools supporting clinical decisions were coupled with genomic reporting tools in a dashboard to gather all relevant patient information in a timeline. This clinical case demonstrates the need for data mobility and the benefit of reusing data for clinical decisions.

Based on the case of sudden cardiac death and the need for early identification of patients at risk, we demonstrated the use of Natural language processing (NLP) to pull relevant information from Electronic Health Records (EHR) in order to populate analysis tools and automate risk estimation. Synthetic data proved successful for the first phase of training tools.

Our experiences reveal a need to develop strategies on how best to use data and what infrastructure needs to be in place. Data must be captured in formats that allow for reuse. Raw data should be saved as well, as a lot of value may get lost in the translation from raw data to structured variables. NLP will provide a useful tool to interpret the information hidden in clinical text, yet need to be developed in one area at a time.

Data must be allowed to flow between systems to allow for the successful development and



«Essentially, the main rule is that data should not be reused. The many necessary exceptions to this rule make it a difficult [legal] landscape to operate in...»



implementation of tools and processes that support a precision medicine approach. This flow will be facilitated by defined standards, APIs, and suitable open platforms. There is a significant technology debt in the existing infrastructure that needs to be addressed as we move forward.



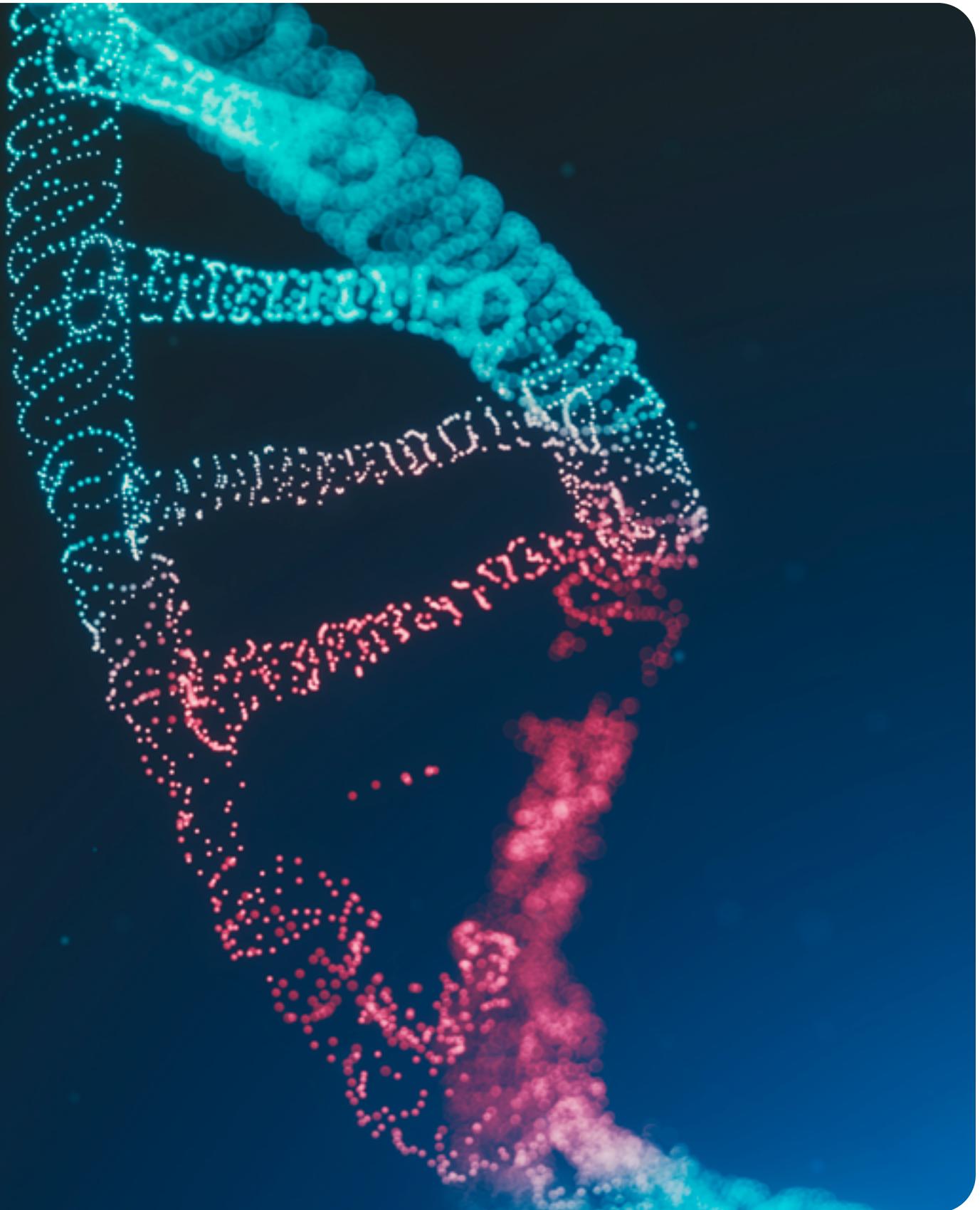
Adoption of precision medicine is best executed with an iterative approach and incremental changes. This is best accomplished through an agile management approach and establishment of a clear path from innovation to implementation, rather than a hierarchical decision structure. This innovation ecosystem requires the hospitals and technology providers to work together towards a common goal.



New models for clinical studies are needed to adapt to the precision medicine paradigm, and health economic models must be revised to reflect the individual rather than a group approach.

The current healthcare regulations are fragmented. Essentially, the main rule is that data should not be reused. The many necessary exceptions to this rule make it a difficult landscape to operate in for both healthcare professionals and researchers. A holistic approach to regulating how data can be reused in a safe and balanced way would support development and benefit the patients rather than create barriers. Clinical decision support (CDS) tools powered by Artificial Intelligence (AI) will be regulated under the EU Medical Device Regulation (MDR) and In-Vitro Diagnostic Regulation (IVDR), yet more supporting systems for safe use are needed. The need for ICT competence near the clinic will increase, as diagnostics move towards more of a multi competence team process.

Moving forward, more specialised initiatives that follow BigMed will continue developing solutions and bringing important discussion topics to the stakeholders and the public. The right setup for allowing incremental changes in our system will allow us to continue working towards our common goal of implementing precision medicine in the healthcare system for the benefit of our patients.



1. The paradigm of precision medicine

Precision medicine promises many opportunities and benefits to healthcare yet clinical implementation has proven challenging. Realising these benefits will require us to simultaneously rethink several areas of existing clinical practice.

Medical doctors have always sought to “*find the right treatment for the right person at the right time*”², however finding the right treatment is often a frustrating process of trial and error. Rapidly increasing knowledge of the biological characteristics of individual people and diseases is making it possible for doctors to use precision medicine to individually tailor treatment and find that right treatment the first time.

The implementation of precision medicine is still in its infancy and the real health economic ratios of different initiatives are still being discussed. More experience and data are needed before coming to a conclusion.

1.1 WHAT IS PRECISION MEDICINE?

Treatment plans have traditionally relied on data and statistics from studies on large clinical trials. Traditional studies are good for recommending treatments that are likely to work for a large part of the population. Precision medicine represents a paradigm shift to more dynamic, individualised treatments and health risk assessments based on a patient's specific clinical and biological information.

Precision medicine

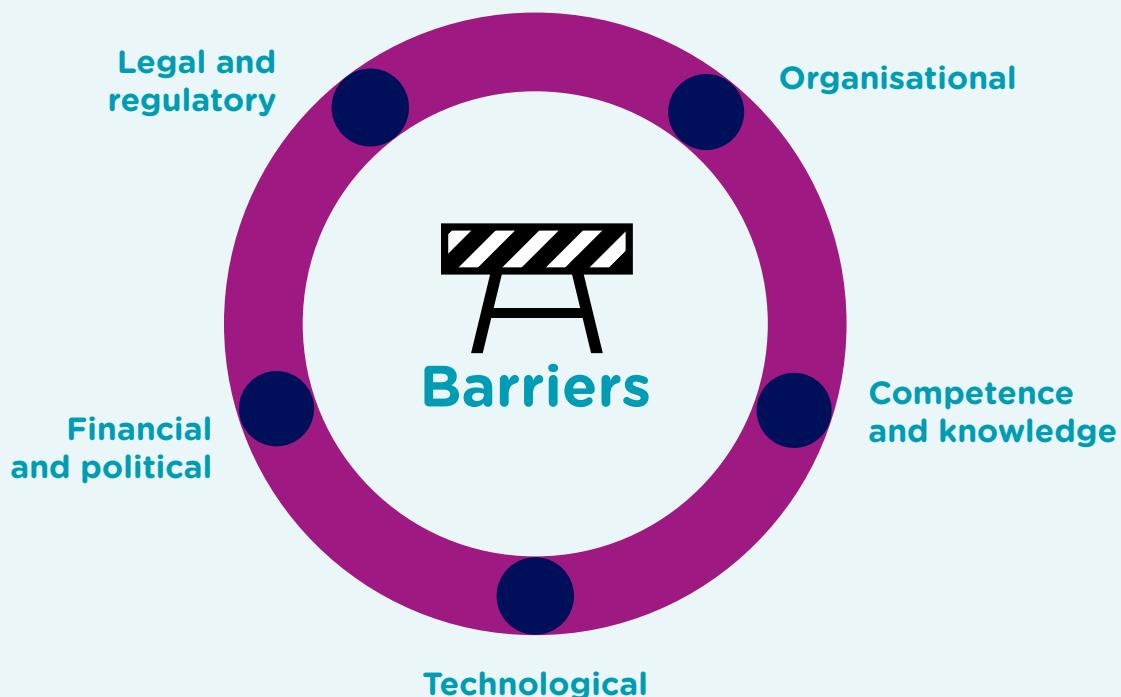
Precision medicine or personalised medicine refers to the tailoring of medical treatment to the individual characteristics of each patient.³ Detailed patient information – from biomarkers like genetic variants, to social attributes – are used to inform customised medical decisions, practices, or products.

In precision medicine, healthcare professionals form multidisciplinary teams to evaluate and analyse large amounts of data from many different sources in order to make the best clinical decision.

The falling cost of diagnostic technologies has made it possible to quickly collect, store, and analyse patient data. Molecular markers and genetic analysis are the most common target for analysis in precision medicine using new technologies such as artificial intelligence. Together, these enable more accurate diagnosis and tailored treatment.

Additional new methods and tools are still urgently needed. BigMed contributes to the development of these methods, software and tools, with an emphasis on efficient reuse of gathered knowledge.

Barriers to clinical implementation



Five major categories of barriers were identified through clinical cases and presented in the 2018 BigMed report: Technological, legal and regulatory, financial and political, organisational, and competence and knowledge.

Technological barriers: There is a need for solutions that support flexible storage, capture, transfer, sharing and use of primary clinical data for analysis, and facilitating direct communication between ICT systems and various sources of data.

Legal and Regulatory barriers: Regulations that address research activities and healthcare services separately, inconsistent privacy rules, and laws that do not facilitate precision medicine research, make it difficult to harness the value of data through sharing and secondary use.

Financial and Political barriers: It is difficult to find consensus when driving disruptive change. Inefficient incentives for innovation and a lack of evidence to support the overall financial benefits are hindering precision medicine implementation.

Organisational barriers: The complex structure of healthcare organisations make them adverse to change, reluctant to share information, and sceptical of industry involvement.

Competence and Knowledge barriers: There is a lack of required expertise (data scientists and bioinformaticians) within the healthcare system, and a lack of cross disciplinary understanding between clinicians, regulators, and data scientists that is necessary to use data for clinical decision making. Patients and clinicians are often unaware of the benefits of precision medicine.

1.2 HOW TO MOVE FORWARD

Several building blocks are still missing before precision medicine can be routinely and meaningfully utilised by the Norwegian healthcare system.

As our understanding of patients become more detailed, and treatment and follow up becomes more individualised, the groups of similar patients become smaller and smaller. Small cohorts require larger data sets for algorithms and clinicians to consider. Local, regional, or even national health databases may not contain sufficient health data on a specific phenotype. This creates a need for broader data exchange between hospitals, regions, and countries.

The algorithms for achieving this exist, but the infrastructure is lacking.

As precision medicine teams grow larger and new competence and technology become part of the diagnostic process, the traditional boundaries between clinical practice and research are shifting. Capturing and analysing data generated in standard clinical practice, outside of clinical trials, is making it feasible for every patient to become their own research project. A broader legal definition of the term “healthcare worker” will be needed to properly understand and regulate this emerging practice.

Precision medicine challenges our existing legislation, medical decision systems, infrastructure, and organisational structures. Our medical system – including methods of prioritising healthcare, technology approval, and drug pricing – will need to be updated.

Experience from the BigMed project show that we will benefit from addressing all of these issues simultaneously by developing iterative solutions to clinical problems as we learn and mature.



2. BigMed's framework for addressing the barriers

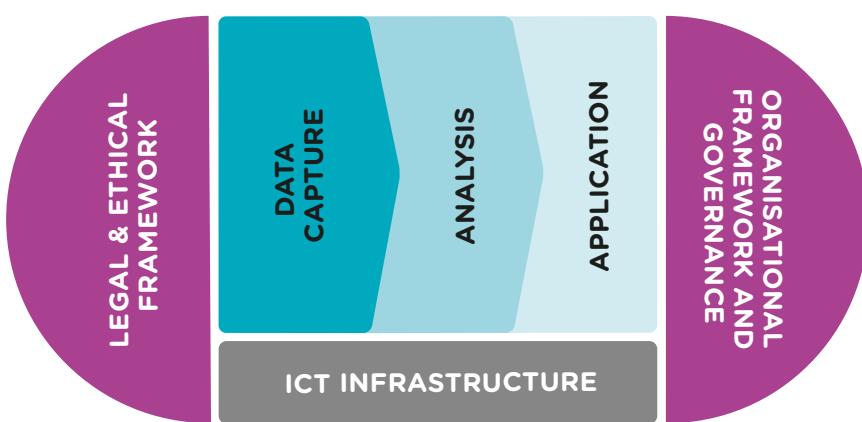


Figure 1. The pill-shaped framework BigMed applied for implementation of precision medicine.

BigMed designed a framework for implementing precision medicine based on a core of clinical data management, supported by information and communications technology (ICT)-infrastructure, and capped at either end by a legal & ethical framework and by an organisational and governance framework. This pill-shaped architecture is designed to overcome the barriers to implementing precision medicine.

Precision medicine will require clinical data to be used in new ways – from capture, through analysis, to the application at the point of care. ICT-infrastructure needs to support this and meet the needs for data flow and sharing of data across institutions. Legal & ethical frameworks need to allow for the new ways of using data while also providing effective regulation and ensuring patient protection. Organisational frameworks and governance, including embedded financial incentives, need to be in place to fully capitalize on the opportunities that arise from precision medicine.

2.1 THE CORE: DATA AND INFRASTRUCTURE

We have developed, tested, and implemented solutions for sharing data between organisations, harmonizing and structuring data, automatic extractions, processing data, and ensuring data quality. We focused our investigation on three model clinical categories: rare disease, sudden cardiac death, and colorectal cancer. The primary barriers to effective

data use were formatting and sharing. We addressed these by improving the data capture and processing system by developing a new identification system using natural language processing algorithms and a new analysis pipeline parameter setup. In addition, we created a digital consent to prioritise variants solution, replacing paper forms with electronic ones to supply structured data that will help quickly prioritise variants in the lab.

As our understanding of the complex nature of data processing matured, new barriers were continuously identified. The initial project plan had to be reworked several times to adjust to a reality where access to data proved even more difficult than expected.

2.2 LEGAL AND ETHICAL FRAMEWORK

BigMed identified an ongoing need for legal and ethical clarifications. Specifically, the definition of anonymity in the GDPR is ambiguous in the context of patient data for precision medicine. The BigMed legal team, with their network of experts from all relevant stakeholders, were involved throughout the development process. They identified and sought to solve these issues as they arose.

Legal barriers and proposed solutions have been lifted to the decision maker level, resulting in proposed changes in the law in 2020⁴ regarding accessing patient data for reuse and sharing of genomic variants.

2.3 ORGANISATIONAL FRAMEWORK AND GOVERNANCE

Working with multi-disciplinary teams and stakeholders, BigMed identified several organisational barriers to precision medicine. Organisational frameworks will need updating to bridge the gap between research and implementation. This includes improving data sharing workflows, financial incentives, and encouraging collaboration between researchers, clinicians, and industry. Through this new framework, BigMed has been able to develop new technologies and proposed several organisational improvements.

2.4 FRAMEWORK DEVELOPMENT

BigMed has addressed each component of this architecture in parallel. The project tested solutions for moving new developments from the lab to clinical practice. The

need for a near-production innovation platform was identified early in the process, and a BigMed innovation zone was created based on these identified needs. This zone enables the providers of infrastructure to test new services before rolling these out on a large scale.

A natural, necessary part of BigMed has been collaboration and the discussion forums. As the project draws to a close, a few areas have been identified as valuable for continued discussions. This includes a necessity for continuous improvement of IT competence and understanding of clinical needs, through discussions between the clinic and the infrastructure and ICT service support, as well as development of the legal competence network for precision medicine, which has been extended to the Nordics and formally established as “Nordic Permed Law”.

The following chapters reflect BigMed's experience in precision medicine. The chapters are organised by the components of this architecture. They cover both practical outputs from the project and visions for future directions and collaborations where concrete solutions are still impeded.

Additional BigMed material

BigMed report

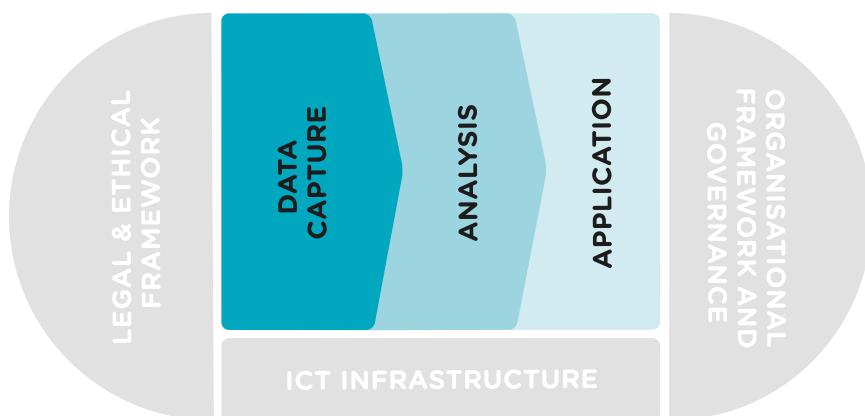
- Big data management for the precise treatment of three patient groups

Recorded webinars

- BigMed-konferansen 2020:
Veien til presisjonsmedisin



3. Data capture, analysis, and application – systematic use of health data



Effective use and analysis of clinical patient data increases our body of medical knowledge and help meet the needs of clinical quality assurance and decision support.

BigMed studied how clinical patient data was used in three model clinical fields: rare disease, sudden cardiac death, and colorectal cancer. We began at the end of the data process by identifying the relevant clinical applications and worked backwards to find the needs of the corresponding supporting analysis. We used these to guide our discussions on best practices of data capture. The analysis revealed several barriers to effective data use.

First, raw data was often either not stored, not accessible, not suitably formatted, or there were no available means to analyse it. Most of the patient data captured in the Electronic Health Record (EHR) is meant for communication between clinicians, billing, and record keeping. Little of it is in a format that can be easily used for other purposes. Clinical data that was reported by clinicians for secondary

use was limited to small sets of structured parameters sent to disease specific registers. The format of definitions and data here was generally selected for tertiary value (national summary statistics) rather than clinical or research purposes.

Second, raw data from medical devices or specialised analysis was often not archived at all. It was stored in formats inappropriate for reuse, or only represented by summary data within proprietary reports.

Finally, there was no ability to analyse data within the boundaries of the clinical systems. Without effective communication tools, data analysis was more of a one-way street rather than a fast feedback loop.

Molecular diagnostics and genomics data are key components of precision medicine and examples of highly sensitive data. After benchmarking exercises showed that differences in pipeline parameter setup could lead to differences in clinical results, BigMed made it a

priority to focus on ensuring that clinical decisions are made on solid foundations. Data sharing is important for quality assurance. Hence, solutions for safe and effective sharing of genomic data were developed in the project.

Before using real world patient data as a foundation for decision support tools, clinicians and developers must understand the bias in health data. While we support structuring certain information for reuse and analysis, we also firmly believe that original clinical text remains the best form for capturing complex and context-dependent information and conveying it between caregivers. NLP can be a useful tool for unlocking this complex information moving forward.

Within this chapter, sections address reflections and learnings from different contributors to the project.

Sections in chapter 3

- 3.1** A need for a strategy on data capture and use
- 3.2** Extracting information from clinical text with natural language processing
- 3.3** Biases and pitfalls in using real world health data
- 3.4** Access to genomics data through sharing across organisations
- 3.5** Standardisation and quality assurance of molecular diagnostics

Related BigMed material

BigMed reports

- Neural classification of Norwegian radiology reports: using NLP to detect findings in CT-scans of children
- Patient similarity networks for precision medicine
- Clinical decision Support Software; Regulatory landscape in Europe from May 26th 2020
- Cancer Predisposition Sequencing Reporter (CPSR): a flexible variant report engine for high-throughput germline screening in cancer
- Accuracy and efficiency of germline variant calling pipelines for human genome data
- Building a Norwegian Lexical Resource for Medical Entity Recognition
- Suggesting Reasonable Phenotypes to Clinicians
- Drivers in rapid genetic diagnostics for rare diseases in infants
- Personal Cancer Genome Reporter: variant interpretation report for precision oncology
- Iterative development of family history annotation guidelines using a synthetic corpus of clinical text
- Clinical sequencing: Regulatory frameworks and quality assurance for NGS-based diagnostics
- Big data management for the precise treatment of three patient groups

Podcasts

- Precision Medicine into Patient Treatment
- Personalised Cancer Treatment
- Genomics and Datassharing
- Machine Learning on Text from Medical Records

Recorded webinars

- NLP in Health – What is Possible, Useful and Allowed?
- BigMed-konferansen 2020: Veien til presisjonsmedisin

An overview of all material from BigMed is available at bigmed.no

3.1 A NEED FOR A STRATEGY FOR DATA CAPTURE AND USE

Pål H. Brekke (Oslo University Hospital), contributions by Serena Elizabeth Marshall (DNV),
Bjørn Næss (DIPS), Vebjørn Arntzen (Oslo University Hospital), Vibeke Binz Vallevik (DNV)

Establishing a focused strategy that can capitalize on the potential value of real-world clinical data is becoming increasingly urgent. Such a strategy should outline how the data value chain, from origin to use, will be supported. The inherent nature of clinical work requires trust in the data used. Therefore, data quality, availability, integrity and interoperability must be defined. Different stakeholder needs for data use and sharing should be covered.

3.1.1 Using real world data in precision medicine

In the data driven future of precision medicine, real world data is a resource for decision support when deciding on treatment for both the current and future patients. Real world data is also helpful for creating new knowledge, improving processes and assuring quality of clinical practice.

Currently, knowledge generated through clinical research projects follows an unappealing timeline. As illustrated in Figure 2, the research project (purple arrow) will take a long time before possibly influencing clinical practice only several years after the original data was produced. In principle, integrating real time data in the clinic (orange arrow) should allow for faster patient benefit.

Analysing the real world data we create every day, and using the results to generate hypotheses or direct decisions at the organisational or clinical level, has the potential to increase new knowledge integration at a much faster pace. Yet, the faster pace requires a different approach to assuring quality. These assurance processes must address, for example, quality, omissions, accessibility, and trust in the origin. These needs will determine the data formats, storage and access requirements and ensure an alignment of priorities.

Real world data

The European Medicines Agency (EMA) defines real world data (RWD) as “routinely collected data relating to a patient’s health status or the delivery of healthcare from a variety of sources other than traditional clinical trials”.

3.1.2 Where and how data should be stored

Many devices used in hospitals, for example blood pressure monitors, spirometers, and portable ultrasound scanners, are not connected to the hospital’s IT network, and therefore do not capture or archive results in a centralised data storage. Data summary reports are more commonly prepared manually or delivered as a printout. Even for the perhaps most ubiquitous medical study, the electrocardiogram (ECG), the recording is printed and then scanned as an image rather than storing the original – and much smaller – digital file. Handling printed reports has a cost in terms of extra administrative work, delayed information flow, and quality and information loss.

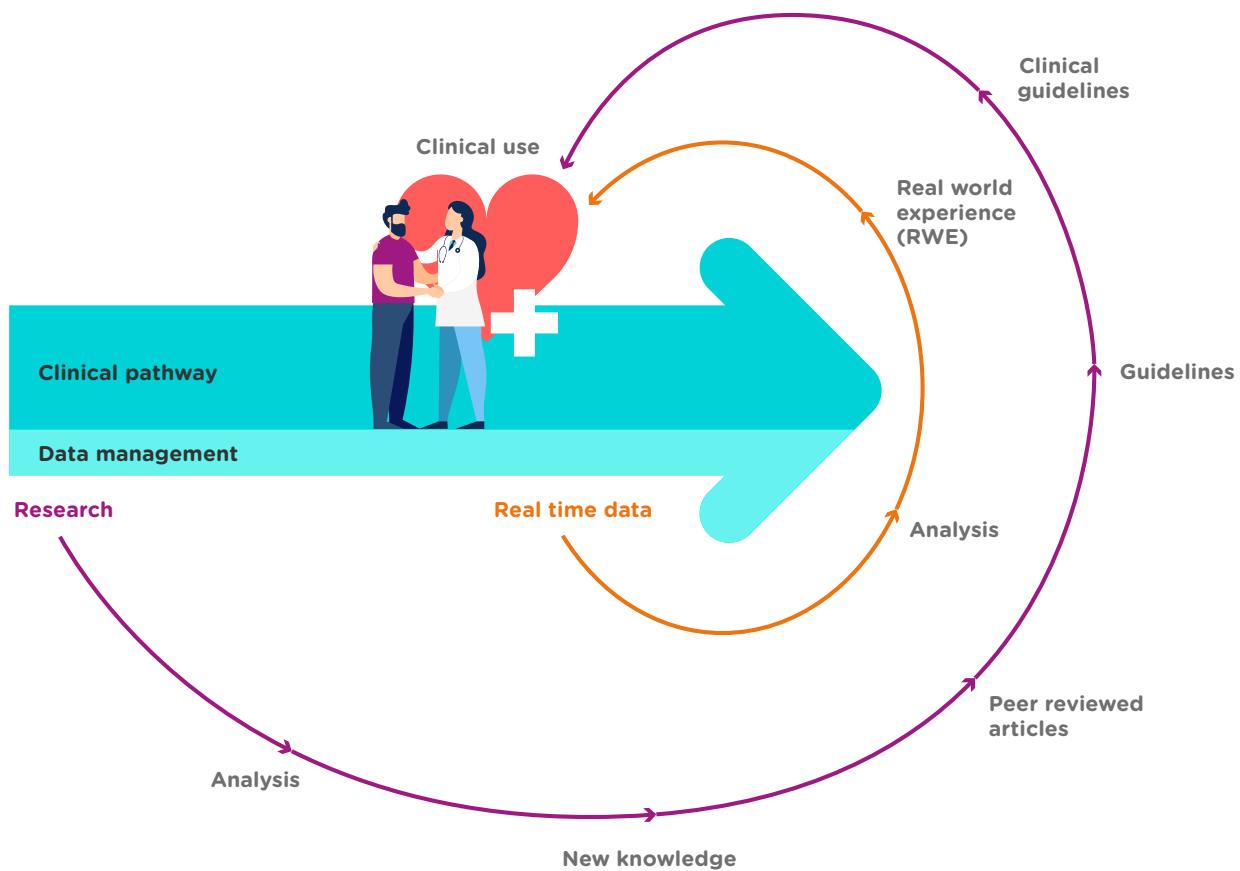


Figure 2. Value for patients from research (purple) vs. real world data (orange).

While preserving every piece of health information in its original format may be excessive, a conservationist approach could enable future diagnosis and treatment technologies. Even if current technology cannot yet analyse all the data flows produced by modern diagnostic equipment in high resolution and high volume, one should consider how the current surge in neural network-based machine learning and AI has facilitated big data processing beyond what was imagined only a decade ago. Paradigm shifts in analytic abilities are generally unpredictable. To paraphrase a well-known ad: Storage is cheap. Knowledge is priceless.

3.1.3 Human or machine readable clinical information

In the past few years, there has been a strong drive towards structuring EHR in order to optimise data for machine interpretation, arguably at the cost of human usability. While describing the complex biological ecosystem of a human being with simple categorical variables is of great use for creating statistical aggregations of larger datasets or for billing purposes, the simplicity of quantified classifications will inevitably result in valuable information being discarded. Clinical texts – the freeform notes clinicians write about their patients – contain not only technical data but also important contextual information describing uncertainty, subjective evaluations, and grey zones.

To avoid a reductionist approach, and cater to unforeseen future data usage, we emphasize the value of human readable text and advocate the use of natural language processing (NLP) tools to bridge the gap between human and computer.

NLP tools could serve as an abstraction layer between clinicians' necessary human-to-human communication and the analytical needs of researchers, hospitals, and organisations. Fit-for-purpose clinical NLP could enrich the human-readable text with a layer of interpreted, analysable data points. This would allow clinicians to communicate and document in their natural language, preserve the reasoning and contextual information in the EHR, and also provide a computable representation of health data in the background. It might even allow the clinician to immediately re-use machine-interpreted data from clinical text as input in risk calculators, registers, or other tools requiring structured data.

3.2 EXTRACTING INFORMATION FROM CLINICAL TEXT WITH NATURAL LANGUAGE PROCESSING (NLP)

Fredrik A. Dahl (Akershus University Hospital), contributions by Lilja Øvrelid (University of Oslo)

Over the last decade, the state of the art for NLP has improved radically due to machine learning (ML) with deep neural networks, which can successfully represent the usage and meaning of single words and sentences. Although such systems have a long way to go before they can reproduce all aspects of human communication, programs such as GLUE and Super-GLUE⁵ are making impressive progress in a wide array of specific language understanding. There is a great need for software that can process clinical text intelligently by finding key pieces of information and summarizing the content of clinical notes, and NLP based on ML fits the bill perfectly.

3.2.1 Data access and legal issues

By definition, medical records contain sensitive information. Privacy issues must be handled carefully. Clinical free-text is even more challenging than structured data in this respect, since each sentence may contain sensitive information that is virtually impossible to anonymize completely for analytical purposes.

From our experience, the Norwegian ethics committee system is not prepared for NLP projects. One example is when we submitted an ambitious project on word vector development to the regional ethics committee (REC) in November 2018. Our plan was to analyse the entire collection of clinical text at Akershus University Hospital (Ahus) with ML, in order to produce clinical word embeddings

– mathematical models of the meaning of individual words. The application was reviewed by several local and national committees over the course of 18 months before final approval. We hope that this exercise has established a precedent for NLP projects within these review committees and that this familiarity will make future reviews faster.

In the group's project on automatic detection of family history of disease, we found a workaround. Rather than using actual clinical records, we developed synthetic clinical text. A clinician produced entirely fictitious journal notes that contained descriptions of family medical history. These were realistic in form and content, but did not contain sensitive information since they referred to non-existing patients. This synthetic data set was used for ML, and the results were published scientifically, which showed that systems trained on synthetic data may also generalize to real, clinical text for the task of family medical history extraction. With a collective effort by clinicians, it may be possible to produce synthetic journal note collections for a wide range of medical contexts – a resource that could be valuable to many NLP researchers.

3.2.2 Text processing pipeline

A well-functioning text processing pipeline is the backbone of any clinical text analysis project. Fortunately, Ahus had most of this in place up front, supplemented by state-of-the-art tools for pre-processing based on Norwegian NLP resources developed by the Language Technology Group at the University of Oslo. The first step is the extraction of the relevant journal notes from the EHR system, where documents are typically stored in RTF or PDF formats, and formatting must be removed. Further processing steps such as sentence splitting, tokenisation, stemming or lemmatisation may be needed. For some cases, a step of provisional parsing and tagging is useful. The pipeline should also have a system for separating out test data in dedicated directories that are accessed only after all the modelling is finished, to ensure valid evaluations of model performances.

3.2.3 Neural network modelling

Deep neural nets are very complex entities, and the algorithms used to train them represent cutting-edge research in informatics, mathematics, and statistics. It is virtually impossible to decide beforehand what kind of model will work best, and which training parameters to use.

However, as an empirical and experimental process, neural network modelling is relatively



easy to perform. There are free online repositories for Python code that implements most state-of-the-art text analysis models. Utilisation of these models does require a competence in general programming skills, but the models themselves are basically plug-and-play and separate neural modules can be combined.

In this project, we have experimented freely with several types of neural networks, including support vector machines (SVM), long-short term memory (LSTM) and convolutional neural network (CNN) models. With the widespread adoption of neural modelling techniques, it is the availability of annotated data that directs research within the field of NLP in general and within BigMed. An important step in advancing clinical NLP for Norwegian language text will be the creation of clinical datasets with expert annotations for use beyond individual research projects.

3.2.4 Supervised learning

A previous case of syncope (loss of consciousness) is a known risk factor for sudden cardiac death. Syncope is frequently described in clinical notes, although its ICD-10 (International Classification of Diseases) code is often not used. Supervised learning in NLP requires annotation by humans. For the task of automatically identifying syncope cases, clinicians read through a collection of clinical texts and identified actual syncope cases. This represented the “ground truth” that we trained neural models to imitate. Having a human expert provide the learning signal by annotating the data set is a typical case of supervised learning.

In another application, we trained neural models to decide if a given CT-scan report describes any abnormal finding. This also required annotations by a human expert, who could provide samples of correct classifications. The effort required for annotation are easily underestimated in ML projects, and the quality and quantity of annotations are often the limiting factor.

3.2.5 Unsupervised learning

Unsupervised machine learning occurs in the absence of expert-generated learning to imitate. Our word vector project referenced in the data access section above is a typical example. By analysing enormous amounts of text, the algorithms can establish a simplified language model, which represents the meaning of a word in an abstract vector space of a few hundred dimensions. Word vectors, or word embeddings, can be useful by themselves, as a means to identify synonyms and contextually related terms, but they are even more valuable as building blocks in supervised learning applications. In the syncope case, the learning task will be easier if the program is fed word vectors that already know that “besvime” (eng: faint) is related to “synkope” (eng: syncope). We tried to use word vectors derived from other text sources than clinical text, with little success. This supports our initial assumption that clinical text is very different from other text sources, with a specialised vocabulary and often simplified syntax. Our clinical word vectors have not yet been completed, but this is an application we are planning.

3.2.6 The road ahead: clinical usefulness

We need to address clinically relevant problems that can be handled technically in a legally and ethically acceptable way. So far, the legal and ethical restrictions may have been the limiting factor. This appears to be changing, in part due to the efforts of BigMed. The technological development has strong momentum, and new language models and training algorithms are developed at an impressive speed all over the world. The critical dimension from now on will be the clinical usefulness. Focus should now be on finding and implementing the applications that give the highest yield in terms of improved clinical decisions and higher clinician efficiency.

There is a problem of information overload for clinicians, but this may be reduced by on-going efforts to improve formal coding systems for diagnoses and treatments and enforcement of common definitions and key terms. However, human language has evolved over thousands of years for the purpose of effective communication, and attempts to replace it with formal systems have had mixed results. The clinical notes that are currently being written in natural language will still be necessary and useful for quite some time.

3.3 BIASES AND PITFALLS IN USING REAL WORLD HEALTH DATA

Pål H. Brekke (Oslo University Hospital), contributions by Arnoldo Frigessi (Oslo University Hospital, University of Oslo), Fredrik A. Dahl (Akershus University Hospital)

Health record data has been described as “the new oil”, as there is enormous potential value in analysing our collective health care experience. To continue the petroleum analogy, there are quite a few steps needed before refining a usable product from EHR data, and there is also a risk of producing something that causes potential harm. As analysis tools become more powerful and easier to use, it will be tempting to use them on large datasets, but an understanding of the processes behind the data is needed in order to avoid drawing the wrong conclusions.

The ways diseases are defined and diagnoses made are continuously changing, and new biomarkers and other diagnostic tools are being developed. Likewise, treatments are changing in step with development of new technology and procedures. For example, dramatic changes have occurred in cardiovascular disease in the past decade or so, where open bypass and aortic valve replacement surgery has shifted towards minimally invasive catheter-based interventions. These changes have implications for ML. For example, if a machine learning-based decision support system was trained on 1990s or 2000s data on aortic valve disease, it would likely recommend open heart surgery, or even no invasive treatment, when current best practice is a preference for catheter-based intervention.

It is likely that both etiological understanding and treatment of several diseases may go through similar paradigm shifts in the future. When using data from clinical practice to inform decision models, one should at least plan to regularly update algorithms in step with changes

in clinical guidelines, lest models “fossilise” outdated thinking. Also, with availability of new procedures and change in associated risk, the populations considered for treatment – and actually treated – change.

A well-known problem for medical researchers, but perhaps new to those approaching medical AI from a data analytics background, is the fact that diagnostic codes (commonly ICD-10) assigned to a patient’s record are significantly influenced by economic considerations. Hospitals are reimbursed for their costs based on reported diagnoses and procedures, and there is even a requirement for hospitals to increase their “coding efficiency” year-on-year. This conflation of billing and medical information potentially leads to apparent rise and fall in disease prevalence in step with reimbursement changes, and also to a disproportional lack of “low value” ICD codes, particularly codes related to symptoms or findings, which in themselves do not lead to any reimbursement. Using diagnostic codes as ground truth labels, then, comes with some challenges.

A more insidious problem in machine learning based on real-world data is the risk of replicating and amplifying human and/or systematic biases. This is highlighted in other fields such as automated loan application processing, predictive policing in the US, and HR department pre-screening of employment candidates. While the Norwegian public healthcare system prides itself on giving equal access, socioeconomic, cultural, and geographical factors influence the use of, and access to, different parts of the healthcare system⁶. Under or over-representation of groups in the data, and potentially different treatment given to different groups, can result in biased predictions or recommendations from AI-supported systems.

Data from clinical practice is inherently less complete than data from a prospective research study in which every person or patient has gone through the same analyses and observations according to a protocol. In real world data, diagnostic tests are performed for a specific reason, which means data will exist – or not exist – in non-random ways which are very hard to reconstruct, model, or correct for. Since most diseases have a time course, tests which are negative or inconclusive at the beginning of a patient trajectory may change to be positive later on – or vice versa. Additionally, any treatment that was started will also affect many of the measurable parameters, further complicating data modelling and analysis.



On a local level, hospitals may cater to very different patient groups according to their available facilities. It is debatable whether patients from a large university hospital and a small hospital are in fact comparable even if they have the same diagnosis. For example, more severe cases requiring specialised care are more likely to be transferred or immediately admitted to larger hospitals with better facilities, which may skew the data. It is a well-known issue that the predictive value of a particular diagnostic test changes with the population in which you apply it. Similarly, AI models trained on data from a university hospital may not be directly applicable to a local hospital without additional refinement.

External validation of ML models is an important consideration as well, and access to a variety of patient data from different settings needs to be considered in project planning. This requirement has implications for laws and regulations in the medical AI space, as increased access to health data – even outside the researcher's own institution – is paramount in order to validate findings and ultimately ensure patient safety if or when the model is deployed clinically.

While the use of clinical data in machine learning or AI applications comes with some caveats, there is enormous potential value in the collective clinical experience archived in our healthcare systems' records. However, in order for AI to be lawful, ethical, and robust, as the guidelines from the European Commission's High-Level Expert Group on AI recommend⁷, it is important to be aware of the potential pitfalls when planning, evaluating, and performing a health AI project using real world clinical data.

3.4 ACCESS TO GENOMICS DATA THROUGH SHARING ACROSS ORGANISATIONS

Tony Håndstad (Oslo University Hospital), contributions by Sharmini Alagaratnam (DNV), Serena Elizabeth Marshall (DNV)

Regardless of the data need – be it sharing in research, inclusion in real time analytics for decision support, in process improvement, or other needs – access to the data at the point of analysis is key. Several standards are ready and in development for using federated computational resources and shared data. These standards will enable clinicians and researchers to leverage the world's genomic and clinical data in a much more automatic and advanced manner than we can imagine today.

3.4.1 Overcoming data silos

Health data silos arise when healthcare institutions collect and store only data generated from their own patients, with limited or no possibility to either share data or access valuable knowledge from other parts of their own institution, similar institutions or organisations. Data silos also occur within organisations when data is produced by different tools or stored in databases that are not networked.

In the field of genomics, sharing of data and knowledge is essential for the quality of healthcare diagnostics and also to advance our understanding of genomics for future patient benefit. DNA sequences can only be understood when viewed in relation to other sequences. For example, by comparing similarities in DNA and phenotype among patients with disease and contrasting that with the characteristics of healthy people, we can better understand the cause of disease.

3.4.2 Quality assurance of variant classifications – making sure our diagnosis is right

Sharing of classifications and supporting evidence between laboratories will lead to harmonisation of classification procedures and standardisation of formats used. This will drive continuous improvement of this emerging area of diagnostics.

In the field of medical genomics, we are typically interested in information about genetic variants and their relation to disease. The distinction between single variant knowledge databases (where an "aggregated" list of individual unrelated variants is accompanied by further information about each variant) and genome databases (where all variants observed within an individual patient are linked together) is an important one as these databases have different use cases and different challenges with respect to data privacy. A single variant database aggregates knowledge about individual variants. In theory, all possible variants across a 4 billion letter reference genome can be included in such a database, but usually only the variants observed in a source patient population or clinical cohort is included.

Box 1: Variant Exchange

A solution developed by DNV in the BigMed project with the aim of facilitating safe and secure sharing of classifications (interpretations) of single variants between different genetics labs. Variant Exchange allows each participating lab to choose which other labs can access its data and automatically notifies them about any classification discrepancies among cooperating labs. The solution is in use today by several labs in the Nordics. Department of Medical Genetics (DMG) at Oslo University Hospital (OUH) and Rigshospitalet in Copenhagen have already successfully shared variant classifications.

3.4.3 Genome databases to increase knowledge

Genome databases contain patient-specific information and typically also phenotype descriptions and other patient details, even familial relationships between genome donors. The information about which variants co-occur in each individual patient can be used to infer haplotypes and is also useful in several other contexts. A genome database with patient specific data naturally contains more privacy-sensitive data and therefore needs better protection than a single variants database.

Table 1: The distinction between single variant databases and genome databases

Single variants database		Genome database
Variant	Frequency and class	Individual genomes
Chr1 35 A>C	2,0%, class 1	Patient A: Chr1 35 A>C, Chr3 535 T>C, ...
Chr1 49 G>T	0,4%, class 3	Patient B: Chr8 4234 G>C, ChrX 1435 delT, ...
ChrX 1435 del T	0,0%, class 5	Patient C: Chr1 35 A>C, Chr5 424011 G>T, ...
...

3.4.4 Federated or centralised sharing of genomics data

Data can be gathered into centralised databases or spread across several de-centralised databases. While the idea of having only a single database can seem attractive, it's not always possible nor desirable to centralise the storage of data into a single database. When it is not, it might still be desirable to standardise the method of access to the data and perhaps even create a way of querying across several databases using a single interface. This is common in federated networks. With such federated databases, communication typically happens according to a defined protocol, and access can also be controlled and restricted by the communication service that acts as an interface between the data and the requestor.

Box 2: Federated data sharing

Systems architecture, data standards, and cybersecurity protocols are often unique to each institution. This makes data sharing challenging. Institutional and individual interpretation of GDPR and national laws leave many clinicians and researchers unsure about the legality of sharing or providing access to patient data for primary or secondary analyses. The necessity for federated databases and federated sharing has arisen due to existing problems with highly limited accessibility to data within silos resulting from interoperability, privacy or organisational proprietary issues.

Although centralised databases with data collected in one common silo have traditionally been proposed as an approach to overcome these barriers, such solutions have been criticised because they create double work and raise sustainability, synchronisation and ownership issues. If unaddressed, limited datasets will hinder creation of the knowledge needed to improve diagnoses and tailor treatments in the precision medicine paradigm.

The proposed solution is federated sharing within federated networks, allowing the user secure and standardised access to data at different locations, without the data having to leave its storage location.

Examples of genomic services that store and share data behind a standard interface include Beacon network⁸ and Matchmaker Exchange⁹. BigMed has implemented and demonstrated the practical use of these solutions.

Beacon is a standard service defined by the Global Alliance for Genomics and Health (GA4GH) that is meant to serve as a simple low-risk platform for data sharing. Organisations can set up a simple web-based service using Beacon that can answer questions like “Do you have any information about variant X?” A beacon response can be as simple as “Yes” (and contact info if further details are wanted) or “No” (in case the variant has not been observed by the organisation). The protocol is evolving to also allow for richer responses, such as information about the observation frequency, any interpretation of the variant, or even info about the patients that had those particular variants. Beacons are typically publicly accessible, but it is also possible to restrict access, partly or fully.

While a beacon is a simple service typically used to query for exact matches against single variants and related data, Matchmaker is a service designed to find similar patient cases. Matchmaker, similar to beacon, is also a web service, but it is only open for communication between trusted partner organisations and allows for more lenient matching. A matchmaker service is queried by presenting a patient case (specifying a suspected gene or variant, and a phenotype description structured according to the Human Phenotype Ontology (HPO)). The queried matchmaker service must then examine its database and evaluate if it has any patient case that is sufficiently similar (e.g. a case with a similar phenotype and another suspicious variant in the same gene). Matchmakers enable the discovery of “the second case”, e.g. evidence of a novel gene-disease relationship. Matchmakers in the Matchmaker Exchange network have already led to the discovery of several new gene-disease relationships and the correct diagnosis of many rare disease cases.

Box 3: Different genomic databases and interfaces with examples

Centralised databases

Single variants

- Frequency databases like *gnomAD* and *Norgen*
- Variant classification databases like *Clinvar*, *CIViC* and *Variant Exchange*

Genomes

- Centralised databases like EGA – European Genome Archive
- *Norvariom*

De-centralised databases and federated data sharing solutions

Single variants

- *Beacon* (sharing information about single observed variants)

Genomes

- *Matchmaker Exchange* (sharing of patient cases to identify patient similarities)
- Federated databases like federated/local EGA

The future of genomics with seamless use of federated computational resources is reliant on more standards development. We need this to successfully enable data sharing and for the use of advanced pattern matching to reach its broadest potential. It is vital that also the legal frameworks are flexible enough to mirror the patient's interests, who's only hope can rely on effective cooperation of labs around the world.

3.5 STANDARDISATION AND QUALITY ASSURANCE OF MOLECULAR DIAGNOSTICS

Eivind Hovig (Oslo University Hospital, University of Oslo), contributions by Oleg Agafonov (DNV), Tony Håndstad (Oslo University Hospital)

Next generation sequencing (NGS)-based testing has transformed molecular diagnostic routines. In some cases it has replaced older technologies, while in other settings – such as single-gene testing, arrayCGH, and karyotyping – it has complemented them.

One of the main benefits of NGS-based testing is the ability to test a high number of loci for various types of disease-causing variants in a diverse genomic context. Nevertheless, despite the benefits of the NGS-based testing, its novelty leads to a lack of harmonised procedures and protocols. This may cause inconsistent results from diagnostics performed at different clinical sites.

NGS-based testing is comprised of multiple steps: test requisition, sample collection, library preparation, sequencing, data processing, variant calling, variant annotation, variant classification and prioritisation, and finally clinical reporting. Figure 3 depicts a germ-line testing procedure. Each of these steps influences the steps downstream, and

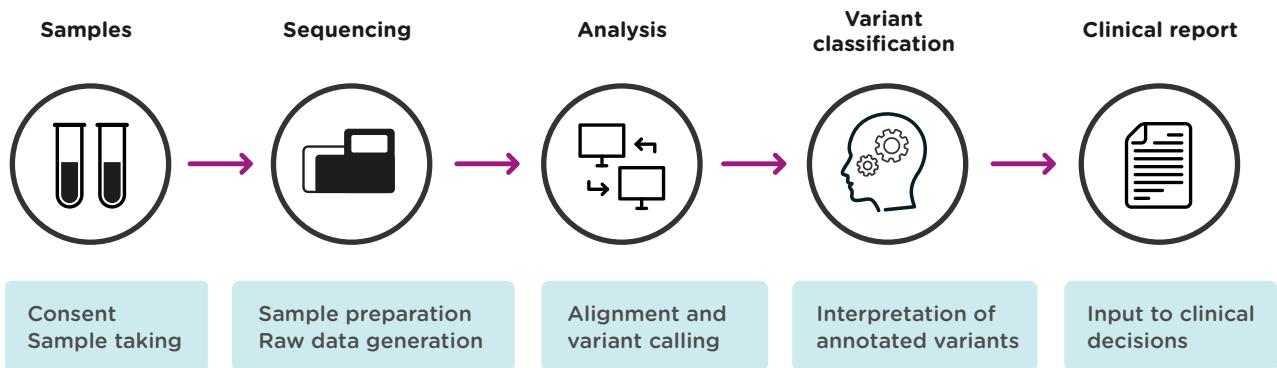


Figure 3. The genetic testing process.

ultimately has the potential to affect patient management and treatment. Therefore, it is paramount to ensure the quality of each step to ensure the quality of the test result.

It is important to note that NGS approaches for clinical diagnostics vary significantly between germ-line sequencing of e.g., rare diseases and cancer, in which the body cells are genetically altered. As cancer cells can harbour thousands of changes, in many dynamically changing ways, it is important to have robust systems to identify the changes in the patient.

This is a difficult challenge, as there are numerous sources of error in sample selection, sample handling, lab methods for sequencing, and the toolkit of bioinformatics methods used to identify the DNA and RNA changes in the tumor. These changes add even more complexity to the task and should be compared to germ-line genomic context of the patient to ensure the best way of identifying the variation that is unique to the tumor cells. Furthermore, there is an array of parameters that may be derived from the raw sequence results that have potential implications for treatment, such as the total number of mutations for a tumor and the mutational load.

High-throughput sequencing is not without technical challenges. In its routine use in diagnostic and research labs, technical artifacts and problems will occasionally occur. Analytical validation and quality control are essential for the safe use of the technology. Quality control (QC) should be implemented both at the individual sample/analysis level as well as regularly monitored on a higher level across samples/analyses. QC per sample should be automated as a part of the variant calling pipeline. An important part of QC is to regularly sequence control samples and/or reference materials. These are samples that typically have a list of known variants (gold standard) confirmed with multiple, orthogonal technologies. Control samples should be automatically compared against gold standard datasets to measure accuracy (sensitivity and precision), lower limit of detection, false positive and negative rate, and other performance metrics. The Oslo University Hospital Department of Medical Genetics (DMG) has implemented trend monitoring of both control sample calling accuracy and several other parameters in an application called MegaQC¹⁰.



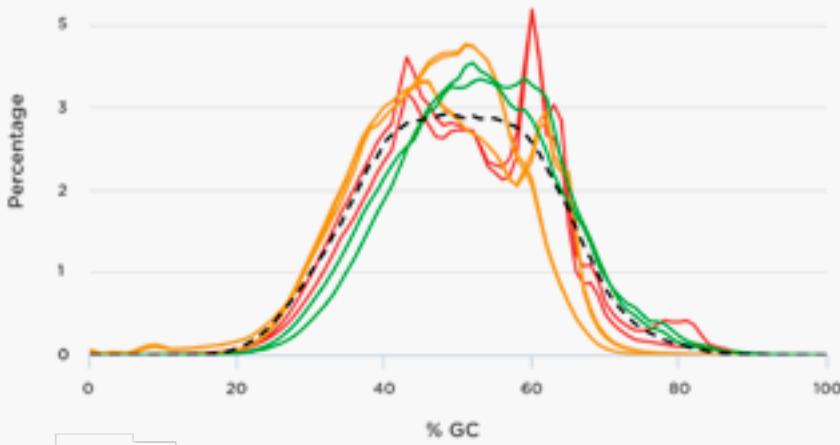


Figure 4. An example from MegaQC dashboard.

Benchmarking is a tool for measurement of the test performance and comparison of the results between laboratories which can be a part of continuous quality improvement and inform test developers on which practices result in the best performance.

Variant calling pipelines are continuously developed as lab protocols, software parameters, and reference data change. It is therefore important to have automated integration and end-to-end tests to validate changes before they are applied to production data. A static (non-changing) control sample data set can then be used for monitoring calling accuracy and successful execution of the pipeline under development. A successful end-to-end test should ideally be logged in the continuous integration system that bioinformaticians use. It can then be referred to in the documentation produced as part of change management procedures.

During the BigMed project, we benchmarked the most popular tools for variant calling of Single-nucleotide polymorphisms (SNPs) and short insertions and deletions (InDels): diagnostic analytical pipelines performance, variant interpretation across Nordic laboratories, and clinical reporting procedures. These exercises provided information which led to modification of analytical pipelines and other routines in the diagnostic laboratories.

By performing benchmarking exercises, we learned that design, administration, coordination, and data analysis takes significant resources and time. Clinical laboratories are unable to provide the same level of resources to these tasks as research laboratories. In some instances, even when the benchmarking was organised by a third party, clinical laboratories were not able to commit to the exercise due to the lack of time. If a laboratory participates in a continuous quality improvement program (e.g., as an activity supporting ISO 15189 accreditation), it is important to ensure that adequate resources are allocated for it.

We have also found that most of the value provided by the benchmarking originates from a comparison of results between laboratories with similar settings. For this purpose, consortia such as ICGC, PCAWG, and NACG have proven to be valuable grounds for national and international cooperation between clinical genetics laboratories. Such consortia allow not only comparisons of performance metrics, but also enable in-depth discussion and the identification of best practices. This leads to improvements of test performance that will ultimately benefit patients.

As cycles of testing continue, standards will eventually emerge where sensitivities and specificities are more robustly defined.

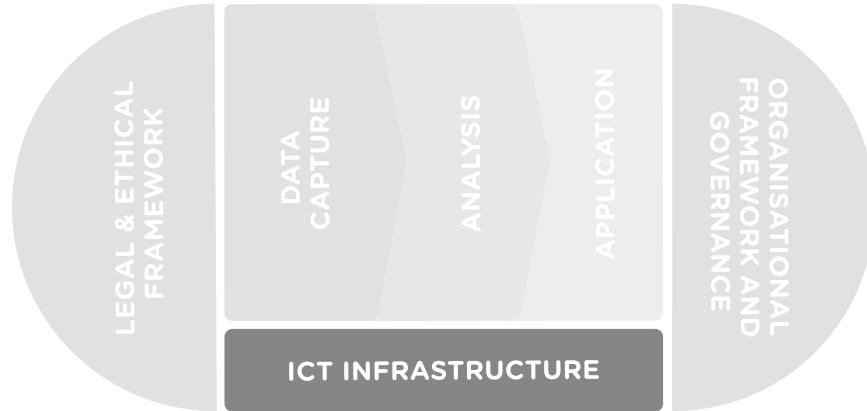
Assay performance should be linked to these. Today, researchers are beginning to define these standards for different end-points, but there is still a long way to go before this work is complete internationally, even for many of the basic assays.

However, developing these standards seems to be a relatively modest problem. Generally, both research and clinical activities utilise the same basic instrumentations and algorithms. Research will inform standards development, and for tests under development, this implies that they may become available as supplementary information to the basic clinical tests being performed in clinical decision making. Quality should be ensured throughout test development, test validation, and ongoing QC. If a laboratory develops its own tests (lab-developed test, LDT), the tasks of both validating the test's analytical and clinical performance, verifying that this performance is met in routine use, and ensuring conformity to the regulatory requirements described in the IVDR fall upon the laboratory. If a laboratory uses a CE-marked IVD test (an off-the-shelf commercial assay), development is performed by a commercial provider. Nevertheless, the diagnostics laboratory still needs to verify that the test performs according to specifications and to perform regular QC.

There are ongoing local, national and international curation efforts to define genetic variants of clinical significance. Ultimately, it seems that the most informed solution will be to assemble information from all patients in the world. There are ongoing efforts to share data across borders to enable such ambitions, and international consortia are already actively engaged in handling knowledge management for selected genes and conditions. Knowledge resources will be constantly enriched over time, and the potential for therapeutic options will grow. Thus, EHR systems should also be handled by automated flagging procedures of previously analyzed patients, in order to ensure that clinicians are kept up to date on new therapies.



4. ICT Infrastructure



A key element in precision medicine is the data and the knowledge that comes with it. In order to utilise this knowledge, we need to establish an ICT architecture suited for precision medicine. Basic infrastructure must be able to facilitate flexibility, storage, processing power, access control, and security requirements for sensitive data. Open and dynamic platforms, established on top of the ICT infrastructure, and services to facilitate data sharing are also crucial for the success of precision medicine.

Prior to BigMed, there were no suitable existing innovation platforms for precision medicine within Oslo University Hospital (OUH). A major part of the infrastructure in the project work has therefore been to design an innovation platform and automate data flows. An innovation zone with the potential to mimic a production environment was established and tested within the hospital by Sykehuspartner.

During the active project phase, BigMed had a testbed on the Services for Sensitive Data (TSD) at the University of Oslo, which has an existing high-performance computer (HPC).

Throughout the whole project, the infrastructure group was involved in all tool development processes – from the concept phase, the mapping of needs, all the way to supporting the work – including designing and building solutions that will be left behind as a permanent feature for the next innovation project.

BigMed has also worked with standards and frameworks that promote interoperability and open platforms. This work has stretched from the use of ontologies such as SNOMED CT and Human Phenotype Ontology (HPO) and standards like HL7 FHIR and openEHR, to structuring of clinical content to technical specifications for metadata and APIs.

While some fundamental building blocks have been established through BigMed, there is still a long way to go before we have a high functioning ICT platform for precision medicine in the clinic and data driven innovation.

Sections in chapter 4

- 4.1** Infrastructure of South-Eastern Norway Regional Health Authority for implementation of precision medicine in the clinic
- 4.2** Infrastructure needs of the research & innovation communities
- 4.3** Production platforms for data sharing in real time
- 4.4** Enabling open platforms
- 4.5** Dynamic digital consent

Additional BigMed material

Bigmed reports

- Implementing NGS-based diagnostics in cancer care:
Technical and organisational factors in the Nordics
- Consent for clinical genetic testing in Norway – Considerations to the development of process and content
- Germline genomic medicine: A BigMed needs analysis
- Big data management for the precise treatment of three patient groups

Podcasts

- Cardiology and Technology
- ICT Infrastructure for Data driven Innovation
- IT Infrastructure needed for Precision Medicine in Norway

Recorded webinars

- BigMed-konferansen 2020: Veien til presisjonsmedisin
- Federated Analytics of Health Data

An overview of all material from BigMed is available at bigmed.no



4.1 INFRASTRUCTURE OF SOUTH-EASTERN NORWAY REGIONAL HEALTH AUTHORITY FOR IMPLEMENTATION OF PRECISION MEDICINE IN THE CLINIC

Alia Zaka (Sykehuspartner), contributions by Knut Lindås (Sykehuspartner), Sevald Cirkov (Sykehuspartner), Vebjørn Arntzen (Oslo University Hospital), Tony Håndstad (Oslo University Hospital)

Below we will briefly highlight the areas of the South-Eastern Norway health region (SEN) infrastructure that we worked on as part of the BigMed project. These areas will require further development to ensure that precision medicine becomes an integrated part of the hospital services. Reflections and opinions are based on our experiences in the BigMed project.

The ICT architecture of the SEN health region is a complex landscape. This, coupled with the lengthy processes for a coordinated implementation of new technology across all hospitals in the region, makes it challenging for clinics aiming to offer services within a fast-moving field such as precision medicine.

4.1.1 Secure innovation platform within the hospital network

The project identified the need for a secure innovation zone within the hospital network due to the vast number of research projects in SEN and the information security requirements and privacy concerns regarding personal information. To address this need, we designed and constructed the BigMed-zone where health personnel and researchers in collaboration with the industry can develop, test, and validate innovative software solutions before commercializing or implementing new tools for clinicians in the clinic.

Box 4: Innovation in the BigMed zone

The BigMed zone is a secure project place for data-driven innovation, established through the RIF-program which is run by SEN regional health authority. The zone is established on the Oslo University Hospital infrastructure and fulfills the same technical and security requirements as for production zones regarding both storage and exchange of patient data. Access can be given to one project at a time.

There are plans to make the process of requesting such an innovation zone and obtaining access for the project members a standardized self-service in SEN, accessible to all hospitals. However, it is important that necessary risk and safety assessments are carried out only once, when the service is established. Which applications are to be tested, which test data is to be used, and the compliance with policies and regulations, should be the requesting project's own responsibility.

4.1.2 Data provisioning for research & development purposes

Access to health data is crucial for the development, testing, and validation of data-driven applications based on ML and AI. The process for both the REC-approval, and the following technical process of data extraction from the hospital's EHRs, is complicated due to strict regulations on patient-sensitive information. Each data extraction must be performed manually by creating a new customized script every time.

Rapid Whole Genome Sequencing Pipeline

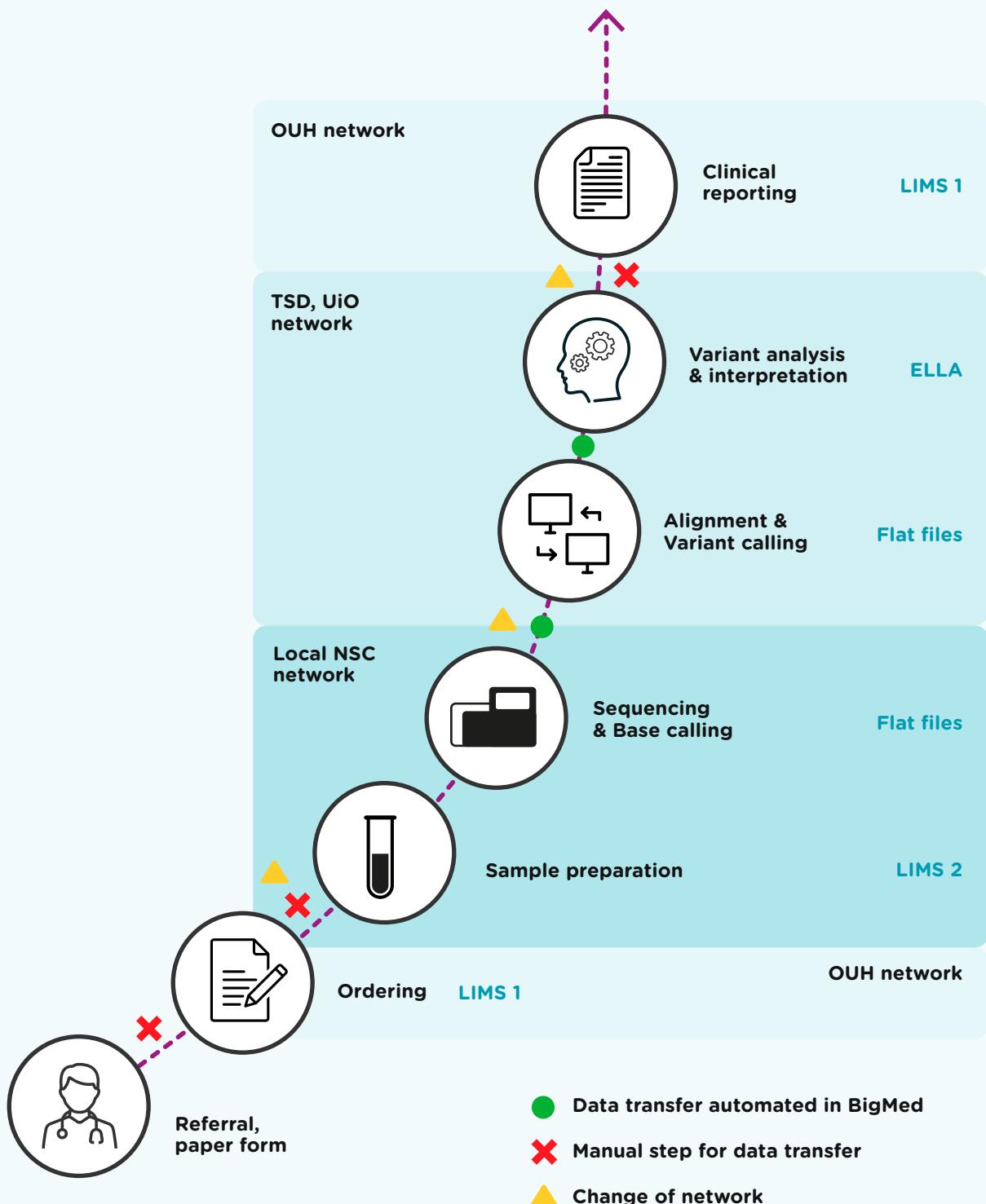


Figure 5. WGS pipeline at the Department of medical genetics at OUH. This pipeline is a prime example of a complicated and vulnerable diagnostic pipeline and inefficient use of the staff's time. Such a pipeline could benefit from further digitalisation and automation as initiated through BigMed, as well as fewer steps of data transfer between different networks both for efficiency and security purposes. As diagnostics move towards more precision by using more data, this pipeline becomes further complicated by the incorporation of phenotype data (sent from LIMS1 (Laboratory Information Management System) located in the OUH-network) in variant analysis (in ELLA located on the TSD-platform in the UiO-network).

Standardised and automated pipelines for efficient extraction of health data from EHRs are needed. De-identification is often a need driven by the data minimisation principle in the privacy regulation.

Such pipeline solutions should be established as self-services based on the REC approval or dynamic patient consents if such a digital solution for patient consent existed on the SEN platforms.

Box 5: De-identification of health data

As we worked towards automating the process for data extraction from the EHR system DIPS, we tested a tool for de-identification. However, the tool removed substantial amounts of contextual information which may make it difficult to find trends and patterns in the data and build AI-models. Commercially available de-identification tools require users to spend a considerable effort on maintaining the libraries on a regular basis and create customized filters for each project. This makes them less suited for automated pipeline services.

4.1.3 Storage and computational power for the clinic

Storage and high-performance computing must be an integrated part of the SEN infrastructure in order to efficiently process and analyze the vast amounts of clinical and genomics data needed for both research and diagnostics. Today, each individual clinic has been left to care for its own needs. As a result, many in the genetics communities are using the TSD storage and computational power services at the University of Oslo (UiO), and exploring other more expensive options in commercial cloud-based platforms.

As clinics have to access storage and computational power outside the hospital network, different parts of diagnostic pipelines run on different platforms. This results in the clinics spending valuable time on creating workarounds for their ICT-solutions. This is time taken away from research and patient diagnostics, and the whole argument for a centralised ICT-service within SEN falls short. The complex pipelines need several checkpoints due to data transfer between different platforms. Additionally, such solutions are not scalable on short notice.

4.1.4 Platform for agile software development and deployment

Today, the rate of adoption of new technology into the clinics is slow. There is a need for ICT platforms and services that will facilitate more agile development and deployment. More frequent releases of specialist applications will allow the clinic to benefit from the latest functionality and drive a continuous digitalisation of their workflows.

The modernisation rate of health services will be greatly improved by also facilitating local efforts for incremental functionality improvement in hospital EHRs in the innovation zone, and by having streamlined decision processes for rapid evaluation and coordinated implementation of the new functionality in hospitals across the SEN health region.

To be able to offer state-of-the-art diagnostics, clinics also rely on in-house development. It is of concern that they use alternative infrastructure not only for development purposes, but also continue to run diagnostic pipelines on infrastructure that is dedicated to research, hence does not deliver on the higher uptime requirements for clinical diagnostics pipelines. A solution must be found to avoid leaving vulnerable patient groups at risk of not receiving the proper medical care at the right time. Support for in-house development and subsequent frequent deployment to a stable and secure production platform will ensure that the clinic can use the latest technology to provide immediate care in a predictable way.

4.2 INFRASTRUCTURE NEEDS OF THE RESEARCH & INNOVATION COMMUNITIES

Gard Thomassen (University of Oslo), contributions by Tony Håndstad (Oslo University Hospital), Loek Vredenberg (IBM)

Through the BigMed project, our UIO ICT partner has identified common needs that must be addressed in order to create a functioning innovation zone for hospital related projects:

- Sufficient and adequate hardware and software
- Better data access
- A test site that does not jeopardize privacy or hospital operations
- Access to international reference data
- Data/result visualisations for clinicians
- A pipeline for moving innovations into hospital production systems

```
x = False  
"MIRROR_X":  
x = False  
y = True  
z = False  
"MIRROR_Z":  
x = False  
y = False  
z = True
```

```
the end -add back the deselected mirror modifier object
```

```
-1  
objects.active = modifier_ob  
str(modifier_ob) # modifier ob is the active ob  
ct = 0  
selected_objects[]  
new_selection.select = 0  
selected_modify the object, then return to the previous state
```

4.2.1 The challenges research & innovation environments face

Today we struggle with the gap between platforms that serve researchers and clinics. As precision medicine converges towards the idea of “one patient equals one research project”, the traditional split between research and the clinic will erode (clinical trials within cancer are relevant examples of this development). Seamless dataflow between the clinic and research will therefore become critical for research in the near future.

The world of research differs greatly from that of clinical diagnostics. It is based on trial, error, and testing of hypothesis. It can be a slow and lengthy process and the future impact on patient treatment is uncertain. It is therefore often thought of as less time critical.

When research and clinical activities are supported by the same ICT provider, it is likely that research and innovation needs must yield to clinical needs. To ensure a steady focus on research and innovation activities, ICT resources should be delivered without impeding clinical deliverables, but while still ensuring an effective flow of data between clinical and research/innovation ICT infrastructures.

Experience from the Nordics, like the Danish Genome Center & Copenhagen University HPC and Karolinska University Hospital & SciLifeLab at the Karolinska University, show that collaboration with an academic partner is one way to help hospitals gain access to specialised IT capacity and competence, and allow mutual sharing of experience and knowledge.

4.2.2 Platform based IT deliverables

Modern IT solutions (AWS, MS Azure, etc.) are appropriate for internal and external use. These platforms are made up of fairly independent components with certain obligations to each other. This implies that components can evolve and change as long as they fulfill their obligations to each other. The platform approach is the most sensible solution when designing the ICT systems of the future, as this will prevent one single system from becoming too complex.

Traditional IT-systems often create “vendor lock-in” at the expense of APIs and collaboration. We postulate the following:

- The hospital IT networks must be agnostic with respect to the payload that is sent (via APIs) over the network (transport layer). By transferring the network into a transport layer, any authorized and authenticated research & innovation centre will be able to access the data.
- Clinical IT systems should strive to enable APIs on all functionality, including large scale data-dumps. The ultimate goal should be to offer (RESTful) API solutions on all IT services in HSØ.
- Establish one HF/company which operates (here “HCN” – “Health Compute Norway”) two similar physical sites for “advanced data processing in the clinic”. HCN should provide services for hospital clinics that require massive storage combined with high performance computing. HCN should also collaborate strategically and competence-wise with Uninett Sigma2 AS, UiT, UiB, NTNU andUiO. This includes lessons learned in the HUNTcloud and TSD infrastructures.

- SEN RHA should establish a new IT entity for research and innovation services in close collaboration with the IT department at UiO, or outsource it to UiO in its entirety. This will enable a setting where HSØ IT research does not compete with HSØ clinical activity, while still keeping synergies of ongoing collaboration.

Changing the architecture of health IT should be done in a step-by-step manner. The goal should be platform-based delivery. This is the most likely way to enable stable operations combined with an opportunity for several third-party clinical IT suppliers, especially from the SME segment.

There is a need for a significant change in how IT services are offered for research and innovation in connection with precision medicine. We require improved data access from and to hospital systems. Research and innovation IT services should be delivered by an entity not responsible for clinical operations. Advanced clinical computing and storage should be handled by a new entity focusing only on this. Both organisational, IT technical, and IT policy changes must be implemented to make this possible.

4.3 PRODUCTION PLATFORMS FOR DATA SHARING IN REAL TIME

Loek Vredenberg (IBM), contributions by Gard Thomassen (University of Oslo)

The drive towards precision medicine has resulted in an overarching technical vision which we have not seen before. BigMed aimed to develop a technology platform that would allow for the management of data and operation of services and applications for precision medicine. In the initial phases of the project, we established a more production-oriented platform to manage this from a clinical production standpoint. In developing this platform, we identified several technological challenges and possible solutions.

4.3.1 Examples of existing platforms

Internationally, there are a few examples of operational platforms performing similar tasks in healthcare akin to what BigMed is working to establish. These platforms have been in production for several years. They are able to manage large amounts of patient data that deliver a longitudinal view of patients and their involvement with their health systems. Updates such as doctor visits, interventions, lab tests and other tasks are mostly performed in near real-time.

The platforms that are in use are clearly distinguished between two types of systems:

- Operational healthcare data systems such as EHR systems, lab systems, imaging systems, and systems that connect with and store data from a diverse range of medical technology equipment
- Analytical systems that are designed to access and analyze large amounts of diverse types of data like HPC for genomic data analysis, research data, statistical data analysis, real world evidence (RWE) systems, and more.

Traditionally, the two systems have served different purposes. The first system addresses the need for speed and direct access, whereas the latter needed capability for large volumes of data and speed of analysis. For that reason, the systems have different architectures driven by their specific needs.

4.3.2 Challenges of developing a new architecture

It is difficult to integrate hundreds of different clinical systems from cohorts of patients into one coherent architecture. Constant innovation and adaptation of new technologies compound this challenge. A good integration of architecture that also provides the flexibility to add new data sources over time is of paramount importance.

To meet these challenges, the new architecture must be able to:

- Anonymize and/or de-identify patient information from unstructured data while maintaining data integrity
- Efficiently integrate many different source systems
- Ensure that correct and timely data is used for precision medicine decisions
- Secure access and manage data in accordance with local laws and regulations

4.3.3 The future is Open Architecture

We drafted a reference architecture based on open-source software with a clear division of responsibilities. This new reference architecture will ensure analysis results have scalability, reliability, and speed. It was developed as an iterative collaboration between technology providers and healthcare providers. Data sharing and system integration are key in the new architecture.

The specifications of the reference architecture are based on integration architectures that support the major Health

Care and Life Sciences (HCLS) standards like HL7 and FHIR. Integration is about connecting different systems with each other, but also includes transformation and routing. These systems often have a variety of data formats. The integration software will therefore need to first transform the data to help the receiving ICT systems “understand” and make use of it.

4.3.4 Key lessons learned from production platforms

In studying existing platforms and their common challenges, we have identified five key lessons that will inform development of new platforms going forward.

- **Access control:** Health data is highly personal and sensitive so access control is important. API Management solutions are addressing these needs.
- **Security:** Sensitive data requires an extra focus on security and protection. All production systems employ encryption for data at rest as well as data in motion. Data access technologies as well as monitoring data usage are crucial. Security information and event management (SIEM) solutions to monitor as well as actively avoid any threats are another layer on top of Identity and Access Management (IAM) types of solutions. In addition, it is important to limit the impact of any data breach that might occur.
- **Capacity planning.** The usage and flow of data is hard to predict, so it is important to create a robust and scalable infrastructure that can react to changing demand. Auto scaling and automated provision of storage and compute make the architecture robust. This is especially important for systems that are not just used for research but also for clinical processes (precision medicine).
- **Data Privacy.** When patient data is used for secondary purposes, the data must be de-identified or anonymized. These techniques need to be applied to all data exchange mechanisms inside and outside the platform.
- **Health Data Model.** Sharing data between many diverse systems is difficult because each system may employ different integration formats and protocols. For data quality purposes, it is important to maintain a common data model within the platform. IBM uses the Universal Data Model for Health (UDMH) as the canonical model, and translates to and from this model, using SNOMED based adaptors, which works very efficiently.

App connect is an IBM platform which can be used in healthcare. App Connect supports healthcare standards like HL7 FHIR, DICOM, CDA & CCD, IHE, and IoT connectivity, in addition to traditional capabilities. The platform allows clinicians to send and receive medical images, recordings, and data from medical journals in real time, as well as to receive data from medical equipment.

4.4 ENABLING OPEN PLATFORMS

Bjørn Næss (DIPS), contributions by Vebjørn Arntzen (Oslo University Hospital), Gard Thomassen (University of Oslo)

Open IT platforms promote development & innovation better than single-supplier systems because they are supplier and technology neutral, eliminate lock-in, facilitate innovation and competition, and force suppliers to compete for quality, value and service. These traits are crucial for a rapid development within precision medicine and the health sector in general.

There is a general agreement today that e-health should be built with interoperability so that the various systems can exchange data with each other. OpenEHR is an open IT platform that meets these demands. It is currently used as a standard for storing data in medical records, but there is still a need for more focus on the definition and representation of clinical content. Clinical concepts are advanced and further work with clinical modeling is required.

In order to have standardized and automated interaction between systems, we depend on developing good standards and protocols for message exchange, as well as well-defined coding systems that allow the registration of data in a uniform manner.

4.4.1 Efficient healthcare and research need a scalable platform for clinical data and metadata

To build good end-user applications and make research possible, we need solutions that can handle the great variety of systems, data sources, data definitions and organisational boundaries. One such solution has already been implemented in Germany. The HIGHmed project (<https://highmed.org>) has developed and deployed an open platform approach to enhance care and research across institutional boundaries. Hospitals

joined forces and transformed their socio-technical infrastructures from an isolated and function-based approach to a collaborative and data-driven one. To make this happen, several properties that are common in open IT platforms have to be in place:

- **Governance needs to be established.** The work must be anchored with all relevant stakeholders. The driving purpose must be to establish good framework conditions for continuous work with the development and management of the systems.
- The data and applications need to be **interoperable** across different platforms and implementations. An open platform achieves this when the requirements for data modelling are implemented and the data is separated from the applications. This will benefit research programs in the ETL process from primary source into the research data model. All structured data can be (semi-)automatically transformed to a variety of downstream formats¹¹. This is a benefit of using archetypes, which allow changes to clinical models while the transformation algorithms rely on the stable openEHR Reference Model.
- **Portability** makes it possible to write software and applications which can be used in other contexts. One such example is the application built by The Cancer Registry of Norway that provides insight based on data in the national registry to suggest and recommend treatment for the patient at hand. It should be possible to develop this kind of applications only once and run in any hospital environment by using standardized data definitions and APIs. An application developed for DIPS Arena at OUH can then later be used within EPIC at Helseplattformen.
- Separating data from applications and a specific storage layer also reduces and **eliminates the risk of vendor lock-in on data**. This is critical since data is the definitive value in an open health platform and an effective system can't risk losing its data if the vendor changes. Separating data from applications require efforts on semantic modelling. This is a critical activity to generate the necessary knowledge to safely access data in a cross-enterprise environment.

The eHealth standards and systems continue to evolve over time. Semantic domain models should be built as completely independent entities, separated from specific software products, solutions, or technologies, run by and for domain experts. One effective example of this is openEHR, which defines a formal specification and methodology for multi-level modelling based on archetypes which defines constraints on a reference model. Once archetypes are defined, they represent definite models of semantics that can be used for multiple purposes, including the generation of data capture forms in EHRs, database schemas, transformations, messages, data validation algorithms, data querying, etc. Archetypes enable both syntactic interoperability and semantic interpretability, which are two necessary components of semantic interoperability.

4.4.2 Open health platform architecture

A platform progress away from being locked-in to a monolith of fixed commitments, toward an open ecosystem (<https://wolandscat.net/2014/05/07/what-is-an-open-platform>) through:

- **Open Service Models:** all specifications of the provided application programming interfaces (APIs) are openly accessible to everybody. Specifications include data security and privacy, electronic health record (EHR) management, and database queries.
- **Open Information Models:** All clinical models are well defined based on established open standards. Data based on these models can be reliably processed and computed in local and distributed environments. In addition, all models that are defined and reused should be made openly available to the community. Examples are SNOMED CT, openEHR, HL7 FHIR and HPO.
- **Open System Specifications:** All system components and protocols are openly specified using licenses feasible for commercial and non-commercial use so that every component in the system can be replaced by software from multiple vendors or open source communities.

Features of an open platform

Data definitions

- Shared and open data definitions
- Transparency on what kind of data is present

Data access

- Shared cross domain query language
- Well defined API
- International standards

Applications

- Data separated from applications
- Build applications faster
 - Applications have short lifetime
 - Data has eternity

4.5 DYNAMIC DIGITAL CONSENT

Bobbie Ray Sannerud, contributions by Sharmini Alagaratnam (DNV)

The increased utilisation of genomic sequencing in Norwegian clinics is creating new practical, legal and ethical issues for the development of content and process of informed consent. To address the issues surrounding clinical genetic testing, DNV led several discussions with Oslo University Hospital's Department for Medical Genetics (OUH, DMG) on the nuanced considerations associated with consent. This partnership published a white paper with a set of recommendations to serve as a starting point for laboratories and clinics developing content and process of consent management in the clinical genomic setting.

Consent is achieved through dynamic interactions, adapting to the needs and situation of the patient. In some cases, this is an ongoing process rather than a one-time informational session. Bridging the gap between the law and clinical practice so that the process of informed consent delivers its value is essential. Interviews with various stakeholders across the clinical genomics value chain suggest that despite this recognition, interpretation of the law for the implementation of consent management in clinical practice can be difficult. As a result, practices for obtaining consent and the content included in consent forms often differ between healthcare institutions across Norway, and even sometimes between clinicians and laboratories in the same hospital.

The white paper advocates for healthcare organisations to develop institutional policies that enable consistent clinical genomics consent practices across the patient and sample analysis journeys. Such policies may also inform the patient about whether or not to go through genetic testing. Additionally, nationally harmonised consent approaches may enable advanced IT tools for sharing genetic information and other health data which could match the patient with other similar patients across not only Norway, but the world.

Further reading:

- Consent for clinical genetic testing in Norway – considerations to the development of process and consent by DNV
- Pan-Nordic clinical consent framework for genetic testing by DNV
- Dynamic consent in clinical genetics: Implementation barriers

Box 6: TSD Consent solution by Gard Thomassen (UiO)

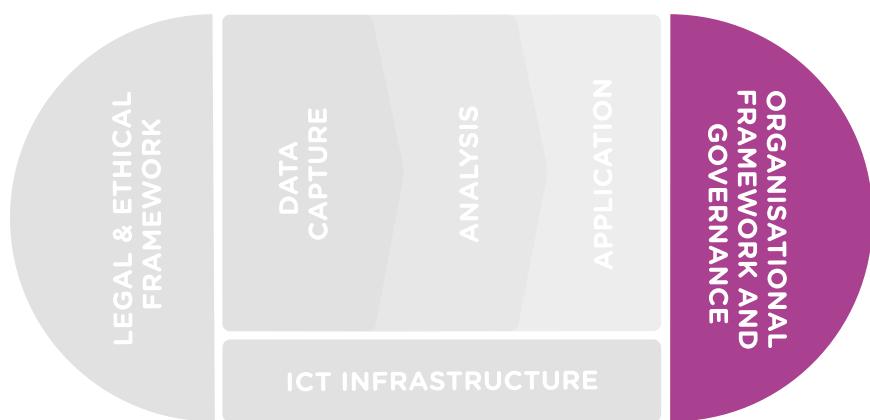
UiO has successfully developed and implemented a dynamic digital consent solution on the TSD platform in cooperation with the Department for Medical Genetics (OUH). The solution grants access by using BankID on the mobile signature solution from Digdir.

The TSD solution uses UiO's Nettskjema (a self-service online questionnaire solution) in such a way that the research project itself creates a consent form. TSD arranges encryption and connection to BankID, as well as connection to TSD's consent solution. This is a complete self-service process for the researcher, and the time spent depends only on the complexity of the consent, which means that the consents can include several questions, which are treated as separate consents in the further process.

With the design ready in 2019, the solution was developed through BigMed in 6 months as a result of an IT platform architecture already set up with required authorisation, authentication and security mechanisms. This, combined with loose connections between the components of the IT platform, allowed TSD to rapidly develop a functional solution to a problem other more monolithic systems have struggled to develop.



5. Organisational frameworks and governance



Organisational issues have been a common denominator influencing all parts of the project and remain the major area that needs to be addressed moving forward. We have gained valuable lessons and experience in organising cooperation between partners with different cultures and expectations.

BigMed created a well-functioning ecosystem consisting of different stakeholders with separate individual agendas from both public and private sectors with different roles and responsibilities working towards the same goal: to address the barriers of precision medicine to the benefit of patients.

Experiences from the project show moving innovation and new technology from research into clinical practice is often problematic.

A clear strategic goal could help all stakeholders move in the same direction. The simultaneous and gradual maturing in

the different areas promote an iterative approach of testing and experimenting.

Following this, we believe that rapid development of precision medicine in Norway could be supported by processes that facilitate both iterative innovation and by establishing organisational clear roles and responsibilities for moving new technology from research to the clinic.

This chapter contains reflections on lessons learned in the project focusing on the topics of organisation and governance. First, we investigate the implications needed to be taken into consideration, then observations on the micro level of innovation at the hospital and then zoom out to the macro level in decision methodology and societal priorities.

Sections in chapter 5

- 5.1** From ad hoc to adoption – a maturity model for precision medicine
- 5.2** Stimulating an innovation ecosystem with industry cooperation
- 5.3** From innovation to implementation
- 5.4** New clinical study models in the framework of precision medicine
- 5.5** Precision medicine in a health economics perspective

Additional BigMed material

BigMed reports

- Implementing NGS-based diagnostics in cancer care: Technical and organisational factors in the Nordics
- Consent for clinical genetic testing in Norway – Considerations to the development of process and content
- Germline genomic medicine: A BigMed needs analysis
- Drivers in rapid genetic diagnostics for rare diseases in infants
- Clinical reporting of NGS data: A systematic Nordic collaborative, peer-reviewed benchmarking
- Big data management for the precise treatment of three patient groups

Podcasts

- Precision Medicine into Patient Treatment
- Health Benefits and Resource Utilisation
- The Key to Precision Medicine

Recorded webinars

- Precision Medicine: A Health Economics Perspective
- BigMed-konferansen 2020: Veien til presisjonsmedisin
- Real-World Data, digitalisation and decentralisation of future clinical trials

An overview of all material from BigMed is available at bigmed.no

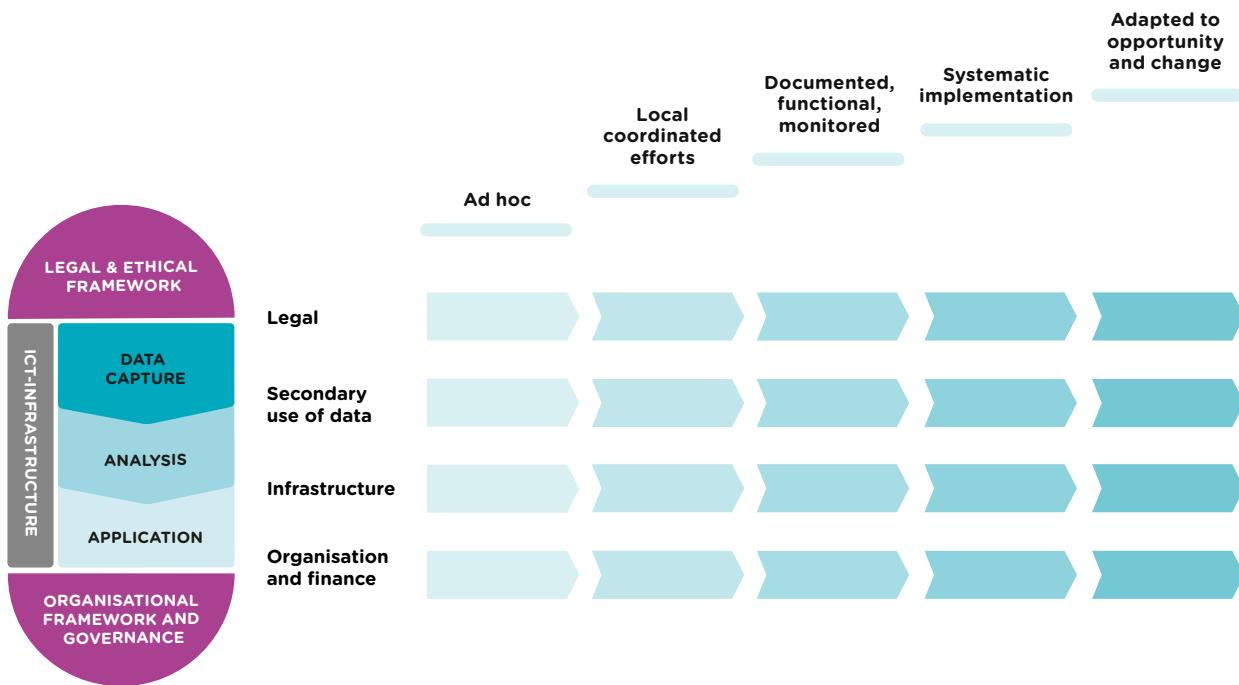


Figure 6. Maturity model for precision medicine implementation.

5.1 FROM AD HOC TO ADOPTION – A MATURITY MODEL FOR PRECISION MEDICINE

Vibeke Binz Vallevik (DNV), Alia Zaka (Sykehuspartner)

The history of industrialisation is full of examples of new technologies that initially do not deliver on their promise. When we invent new technologies, but continue to work in the old way, we cannot expect the innovations to have the exponential effects that we may have hoped for.

Unfortunately, this is the pattern we observe today within precision medicine. In order to harness the full potential, we need to grow our understanding of the opportunities the new technology gives, and incorporate it into our system of care. Until we fully integrate precision medicine into our systems, the health economic benefits can be expected to lag.

Pharmacogenetics is a good example of a “new technology” that has been integrated into an existing process. We have knowledge of several genomic variants and a simple PCR test that, administered before treatment, will reveal whether a particular drug will give low or no effect, or even adverse effects. An alternative proposed¹² way of integrating pharmacogenomics is to store our patient’s genomic variants in a register and allow an automatic cross check of relevant variants with the record before certain medications

can be prescribed. This would allow a more efficient reuse of resources, also ensuring safe treatment of patients.

From our experiences in BigMed we have drafted a model for precision medicine (Figure 6) that describes the different maturity levels within each area of our architecture: secondary use of data (data capture, analysis and application), infrastructure, legal, and organisational framework & governance.

For precision medicine to reach its full potential, we need supporting infrastructure to facilitate the flow and use of data. There is also a need for corresponding adjustments to our legal framework. Some actual changes to the law are needed, while some issues can be handled through clarifications of how our regulations should be interpreted. This is needed to align practice nationally. We need to redesign our decision processes to facilitate the uptake of new technology and encourage patient care based on individual patient characteristics. Most importantly, we need to understand where and how this new technology can allow us to deliver care in a different and better way.

At the first step of the model, we observed mostly **ad hoc initiatives**, locally driven by dedicated individuals. The next step has been **locally coordinated efforts**. As these efforts

mature and become fully **functional**, they will be documented and monitored. Documented effects are a natural prerequisite before a wider **systematic implementation**.

In the scenario of a **systematic approach**, we can expect clinical data to be captured and stored for systematic use based on the current data strategy. Data sharing is a good example of a systemic approach: there are available tools for sharing, the legal boundaries for use are clear, and supporting standards and ontologies are commonly used.

In this scenario, clinicians will have access to essential data from which they can evaluate the statistics of outcomes for different patient segments that they in turn can use as a basis for decisions on treating the next patient. For example, they may be able to search through molecular tumor profiles from previous patients to find statistics for a new patient with similar characteristics. A process and digital tools for consent will facilitate the secondary reuse of data. A harmonisation between sites encompasses standardisation of data generation and storage; to ensure quality and similar outcomes regardless of geography.

In a system **adapted to opportunity and change**, we expect the gap between research and implementation to be bridged by a process of facilitated integration to ensure iterative improvements. Roles and mandates for this activity should be clear, and a supporting infrastructure made available. Funding mechanisms would reflect efforts needed to test and validate new technologies and methods before moving these into clinical use. The decision system for approving new technology will be balanced to ensure sufficient safety and a fair use of our common resources on one side, while on the other side not bottlenecking access to the care our patients deserve.

The Norwegian strategy for Personalised medicine 2017–2021 had an “overall aim to ensure coordinated building of expertise and coordinated knowledge-based developments in the field of personalised medicine, and to pave the way for further research and innovation.”¹³ An updated strategy for the next period should address the implementation of innovations in precision medicine. A gap analysis based on the maturity framework could be used as a guide to set the new ambitions.

5.2 STIMULATING AN INNOVATION ECOSYSTEM WITH INDUSTRY COOPERATION

Liv Bollvåg (DIPS), contributions by Bobbie-Ray Sannerud (DNV), Anita Moe Larsen (Norway Health Tech), Stephen McAdam, (DNV), Odd Arild Lehne (Norway Health Tech), Kathrine Myhre (Norway Health Tech), Loek Vredenberg (IBM)

The BigMed project has fostered promising new tools and innovations, thanks to a tight collaboration between industry partners, clinicians, and researchers.

Precision medicine is not just a technical venture, it also challenges traditional workflows in healthcare. Smart organisation of knowledge and data, and an open-minded approach free from skepticism towards public-private collaboration, are necessary for providing effective and safe healthcare services as the amount of new knowledge exceeds the

«The ideal set-up for innovation activity and solving challenges for better healthcare is a close collaboration between industry, researchers, and clinicians»

individual capacity of each party. Precision medicine requires us to acknowledge the need for cross-disciplinary knowledge and competence from both sectors.

5.2.1 Guidelines for innovation

The ideal set-up for innovation activity and solving challenges for better healthcare is a close collaboration between industry, researchers, and clinicians. Working within a host organisation (hospital) with a complex landscape of IT solutions makes it challenging to provide staff that can facilitate software upgrades and test-facilities. To work as a team, all parties contribute to build bridges between real world problems and possible solutions.

The first BigMed report¹⁴ identified some important guidelines that needed to be in place for the successful development of a well-functioning system of public-private collaborations. These were:

- Focus on innovation in areas where good solutions are not commercially available.
- Address current culture of industry skepticism from a political and top management level.
- Acquire and test available solutions and technologies to a larger extent than currently practiced.

The BigMed project has been successful in creating forms of cooperation where industry and clinicians work together towards a common goal. Moreover, it offered the economic incentive needed for industry to be able to invest in developing tools with an uncertain future benefit. This allowed development of a range of new technologies (see Box 7). This supported a collaborative environment in which both industry and clinicians could focus jointly on the clinical needs throughout the developmental phase of the project.



Box 7: New tools developed in the project through industry cooperation

Variant Exchange – Sharing of interpreted genomic variants	DNV, Oslo University Hospital (OUH), Scilife Lab/Karolinska
Data provisioning OUH – pipeline for text extraction from EHR	OUH ICT, DIPS, University of Oslo (UiO), Sykehuspartner (SP)
Data provisioning Akershus University Hospital (Ahus) – pipeline for text extraction for EHR	OUH ICT, DIPS, UiO, SP
BigMed zone innovation platform with pipelines for data extraction (on RIF)	SP, OUH ICT, IBM, DNV, DIPS
MaxManus – automatic anonymisation for free text	OUH ICT, DIPS, UiO, SP
Digital consent solution in TSD	OUH ICT, UiO, SP
Risk calculator for sudden cardiac death	OUH, DIPS
ML: Automatic echo measurements for input to calculator (machine learning)	Inmeta, OUH
NLP: Identification of patients at risk for SCD from EHR, identification of "syncope" to populate risk calculator.	Akershus University Hospital (Ahus) & UiO Institute of Informatics, Language technology (IFI LTG)
NLP: Pedigree tool – Extraction from free text – family relations relevant for medical condition	Ahus & IFI LTG, OUH
NLP: Interpretation of CT descriptions (validated with MR and CT caput-descriptions)	Ahus & IFI LTG
Dashboard with timeline in the EHR, DIPS arena	DIPS, OUH
Patient similarity classifier predictor (Netdx)	OUH Oslo Centre for Biostatistics and Epidemiology (OCBE)
Automatic reporting tool from EHR to the cancer registry	DIPS, OUH, The Norwegian Cancer registry
Boolean search for research articles	OUH, PubGene
English dictionary for search of research articles	PubGene, OUH
Research data capture directly in EHR – (ProtheCT)	DIPS, OUH
Text mining from EHR for automatic population of Dashboard – NLP testcase for WatsonExplorer	IBM, OUH

For a complete list of industry cooperation results in the BigMed project, see appendix.

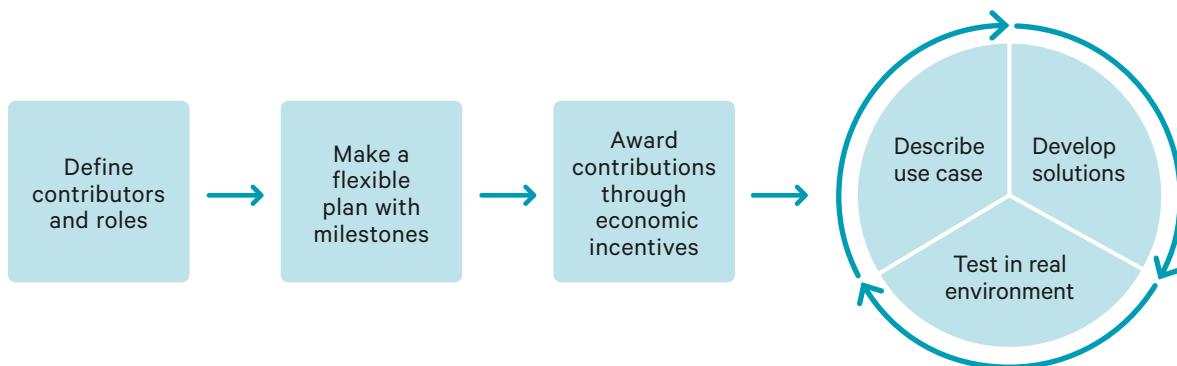


Figure 7. Industry-hospital collaboration process where all parties are involved from the start.

5.2.2 Overcoming the barriers to implementation

A major limitation has been the lack of mutual understanding of needs, tech readiness, and process readiness. Overcoming this limitation represents a major success of the BigMed project.

One of the biggest organisational barriers to clinical implementation of precision medicine identified in the 2018 BigMed report¹⁵ was skepticism towards industry involvement. The report pointed to, among other things, a lack of tradition for cooperation and missing incentives. The setup of the BigMed project with a cooperation between academia, hospitals, and industry partners has been an exploration of how cooperation can be organized between partners with different cultures and expectations. Industry was involved to develop the ideas into sustainable commercial solutions and ensure an opportunity for clinical implementation.

The process has provided many valuable lessons. Transparency and an understanding of each other's ambition proved a key starting point. Coming together as one group and working together towards one common goal over time has built relationships on a peer-to-peer level.

While the collaboration across disciplines and sectors on a research level is showing results, the real challenge of bringing innovation into operational practice is still present and a hinderance to further value creation. With an understanding and acceptance of the different motivations of the stakeholders, a well-functioning ecosystem of trust could emerge over time.

5.3 FROM INNOVATION TO IMPLEMENTATION

5.3.1 The implementation gap between research and clinical environments

Courtney David Nadeau (DNV), contributions by Eivind Hovig (Oslo University Hospital, University of Oslo), Anne Jorunn Stokka (DNV), Vibeke Binz Vallevik (DNV)

Technologies develop on a continuum from initial high-risk, conceptual academic research to routine and appropriate use in the health system. While investigating NGS-based diagnostics and machine learning in hospitals, BigMed found major gaps between the possibilities of research and implementation in the clinic¹⁶.

This gap exists in large part due to differing incentivisation structures for research and for clinical work. Researchers in a university or translational setting are generally incentivized to publish in academic journals that expand scientific knowledge and to leverage their cutting-edge work to fund future research. In contrast, clinical diagnostics units are tasked with developing and providing robust, quality-assured diagnostics within a highly regulated funding and regulatory framework. In many cases, these disparate incentives create separate research and clinical infrastructures.

Much of the work needed to bring the research from the lab to the clinic is not directly incentivized for either academic or clinical stakeholders. Re-writing software packages, quality-vetting supply chains, risk management, process change management, and regulatory compliance efforts in most cases neither lead to high impact-factor publications, nor are they directly reimbursable activities for clinicians.

«An agile project management style is well suited to paving the way for precision medicine with the ultimate goal to benefit future patients»

When exploring the implementation of NGS-based cancer diagnostics in the Nordics, BigMed found that many clinical tools were prototyped in a research setting, but were rarely adapted for use in a regulated hospital setting. Even in cases where clinical stakeholders could see the benefits of using some of these tools, they were not implemented because of the high cost of development and validation efforts.

Creating specific funding opportunities for operationalizing research tools (without the expectation of peer-reviewed publication) and promoting cooperation with industry partners are two mechanisms to incentivise the development of medical technologies. Greater clarity on regulatory and quality requirements for both clinical and research stakeholders could also alleviate these implementation hurdles.

5.3.2 An agile organisational path to implementation

Alia Zaka (Sykehuspartner), Vebjørn Arntzen (Oslo University Hospital), Vibeke Binz Vallevik (DNV)

BigMed found that organisations with the competence and organisational structure to develop and maintain their own IT tools for clinical use were better able to move research from innovation to implementation, like the case of genomics laboratories with bioinformatic competence.

In one case, the BigMed project team worked in close cooperation with all stakeholders to develop a solution for automatically generating colorectal cancer reports to the cancer registry directly from the EHR system (description in Box 8). The solution was developed quickly within the research project, but there was no clear pathway for implementing it in the hospital. It was not clear what the decision gate was nor who had the authority to advance the project from innovation to implementation. This resulted in unnecessary delays.

There are many stakeholders in the process of implementing new tools: clinical and ICT departments at the hospital, the technology provider, and the regional health authority. Clear organisational processes and mandates, including a budget for testing and validating new tools, are needed to bridge the implementation gap.

Allowing different parts of the organisation to autonomously develop, test, validate and implement incremental improvements within a set of frames – instead of a hierarchical model – could shorten the time spent to move from innovation to implementation for many projects. Teams need to be able to pivot as they obtain new knowledge and collaborate to overcome new barriers. An agile project management style is well suited to pave the way for precision medicine with the ultimate goal to benefit future patients.

BigMed use case

A solution for automated reporting from the hospitals' EHR to the Norwegian Cancer Registry

Almost 100,000 cancer reports are submitted to the Norwegian Cancer Registry every year. On average, healthcare professionals spend 10 minutes on each case completing a form of structured variables. The clinician needs to open an external web portal service operated by the Norwegian Cancer Registry and manually enter the data. What if we could retrieve such variables directly from the EHR, rather than having to repeatedly enter all the data by hand each time it needs reporting?

BigMed tested NLP solutions for automatically extracting structured parameters on tumor characteristics from clinical text, in addition to creating a solution for automatically sending a structured form from the EHR to the cancer registry. This will cut the clinicians' time spent on cancer reporting in half. It will also allow for re-use of the parameters and avoid extra barriers logging into different systems. Reducing the number of duplicated data from manual registration will in turn reduce the risk of errors.

A solution was designed based on open EHR archetypes, allowing for scalable implementation process starting with the colorectal cancer module while continuing to develop a data structure for other types of cancer.

Box 8

Box 9: Three principles supporting agile management

In the Age of Agile, Denning summarizes the principles for agile management through three “laws”: The law of the customer, the law of the team and of the law of the networks¹⁷.

Patient needs

The law of the customer states that the goal must always be to increase customer value. In a state funded hospital setting, this could translate to increasing patient health value rather than having a department budget focus. Similarly, value-based healthcare¹⁸ has been introduced as a concept to steer organisations in this direction through financial incentives.

Autonomous teams

The law of the team emphasizes the need to allow for autonomous teams that operate within a certain mandate to optimize customer – or patient – value. The management culture needs to allow for the power of decision to lie with the people who have the best prerequisite to make that decision, no matter what level in the organisation they are.

Non-hierarchical cooperation

The law of the network points to the need for people to interact horizontally on a needs-based principle rather than follow strict hierarchical lines of command. We need to develop our organisational culture moving from the old fashioned bureaucracy where decisions are made by persons high up in the hierarchy, into an organisation with autonomous and efficient teams that allow horizontal networking of decision making to happen in the level of most competence.

5.4 NEW CLINICAL STUDY MODELS IN THE FRAMEWORK OF PRECISION MEDICINE

Anne Jorunn Stokka (DNV), Karen Irgens (DNV)

Randomised clinical trials (RCTs) where patients are recruited and monitored under strictly controlled conditions are considered the gold standard for clinical research, and are used as evidence for regulatory decisions. In this design, subjects are randomly assigned to either an experimental group receiving the study intervention, or a control group receiving placebo or standard care. Although the RCT probably represents the best available standard to generate evidence, this design is challenged by new innovative study models incorporating biomarker expression, genetic profiles, technology and digital tools.

A challenge with biomarker and genetic trial stratification could be small sample populations in the study. RCT can also be limited by a recruited study population that might not represent the real-world setting. These elements all make downstream health technology assessments (HTA) more complex.

Precision medicine, decentralisation, and the digitalisation of health care, together with an increased focus on patient-centered care represents a new era of clinical research and contribute to the ongoing paradigm shift in clinical trials. In this section we will focus on the contemporary drug development landscape. We will discuss innovative trial designs, digital platforms and the magnitude of real-world health data (RWD) that is analyzed to generate real-world evidence (RWE) to be used for decision making. This report does not attempt to present an exhaustive list of different clinical trial models, but rather focus on a few models and elements that are believed to have a strong impact on shaping current and future trial design.

5.4.1 Precision medicine in cancer – innovative study models

Today, the majority of oncology trials are designed to enroll eligible patients with specific biomarkers or genetic aberrations. Development of precise diagnostic tools allows for recruiting a more tailored patient group, moving away from the “one size fits all” thinking.

The main challenge with these trials is small study cohorts that challenge statistical power and limits robust clinical evidence. On top of that, many trials do not contain an active control arm in the study, which makes it even more complex to assess clinical outcomes.

Basket and umbrella trials are innovative study models that have been developed under a master protocol framework. So far, such master protocols have not been well established in fields outside of oncology^{19,20}.

In a basket trial, a single drug is tested in different cancers that all express the same biomarker or have the same genetic aberration, independent of histology location. In contrast, an umbrella trial is designed to test several different drugs in one type of cancer with multiple molecular sub-groups. These designs allow broad evidence generation across multiple cancer types and molecular aberrations due to extensive diagnostic testing. A few case examples of ongoing trends in the oncology precision medicine clinical trial landscape are discussed below.

5.4.2 Histology agnostic indications – basket trial design

During the last year, two Neurotrophic Tyrosine Receptor Kinase (NTRK) gene fusion inhibitors were approved in the EU with histology independent indications, i.e. their use is based solely on a genetic aberration and the tumor site of origin in the body is not taken into consideration^{21,22}.

The efficacy and safety of these two medicines were studied using a basket trial design and included several different tumor sites. This was a small revolution in regulatory drug approval as it was the first time in Europe a medicine was granted market authorisation based only on the presence of a biomarker. Clinicaltrials.gov lists several similar ongoing basket studies in oncology, particularly solid cancer, so it is expected that more drugs with histology agnostic indications will reach regulatory submission soon.

These NTRK inhibitors have paved the regulatory way for this type of clinical registration study and represent a new paradigm for national HTA and reimbursement discussions because available efficacy and safety data at the time of approval is very sparse.

5.4.3 The drug rediscovery protocol (DRUP) – combined basket and umbrella design

DRUP is an innovative, combined basket and umbrella design to facilitate the expanded use of existing anticancer drugs outside their approved indication²³.

The combination of umbrella and basket design creates opportunities for including many different mutations and approved medicines in the study. The medicines can then be used in a range of combinations outside their regulatory approved indications to allow for individualised patient treatment. In the DRUP study, 34% of included patients reported clinical benefit by defined criteria²⁴.

There are several similar initiatives in the EU, including IMPRESS Norway (IMproving public cancer care by implementing PREciSion medicine in Norway). Impress is coordinated from Oslo University Hospital and aims to reach out to all eligible cancer patients in Norway through a national tumor board where all hospitals can refer patients²⁵. Generally, these studies enroll patients with advanced cancer who have exhausted all available treatment options.

5.4.4 Technology platforms and decentralised clinical trials

In a decentralised clinical trial, study participation is facilitated from the patient's home or in their local communities.

Physical study site visits are replaced by telemedicine or mobile/local healthcare providers (HCPs), and researchers capture data remotely through use of mobile technologies and wearables. A fully virtual (site-less) trial allows the patient to be home-based at every stage of the clinical trial²⁶.

Such trials are not feasible for all types of studies or disease areas and will in general be most suitable for diseases that are not life-threatening. Many trials need to perform on-site visits like MRIs, biopsies, and other assessments that cannot take place at a patient's home or local community. For these latter trials it may be possible to exchange some physical visits for virtual appointments, referred to as hybrid trials or approaches. A hybrid trial is traditionally defined as a clinical trial that includes both traditional and pragmatic clinical trial elements²⁷. In contrast to a randomised clinical trial, a pragmatic trial is designed to show the real-world effectiveness of the intervention in broad patient group²⁸.

Typical elements that are part of decentralised trials are remote trial recruitment through social media or use of digital tools, nurse visits at home, use of telemedicine/video consultation, digital data capture from devices or wearables, and electronic capture of Patient Reported Outcome Measures (PROMs)/ Patient Reported Experience Measures (PREMS) data. Using technology for data collection can provide more granular day-to-day details compared to traditional trials where data are collected at specific and selected time points.

Mapping during this project period shows that decentralised and virtual trials in general have higher recruitment rates, better compliance, lower drop-out rates, and are conducted faster than traditional clinical trials. Manual processing at the study sites or at trial sponsor's end is reduced as the data is collected directly from the digital devices.

Another important element of decentralisation is the possibility to include a more diverse patient population as barriers such as long-distance travel, physical impairment, age or other limitations is no longer preventing trial participation. In conclusion, decentralisation is about moving towards a patient centric trial design where technology is designed around the patient.



5.4.5 Use of real-world data (in clinical research) to generate real-world evidence

The European Medicines Agency (EMA) defines RWD as “routinely collected data relating to a patient’s health status or the delivery of health care from a variety of sources other than traditional clinical trials²⁹”. RWE, on the other hand, is the clinical evidence that can be derived from analysis of RWD.

In many cases a market authorisation for a new medicine in the EU is granted with limited clinical evidence due to small patient groups in the study, particular in cases where the new intervention targets a rare disease with an unmet medical need and/or where the intervention is linked to expression of a biomarker. In such cases, regulatory approval may be granted with less outcome data, and more flexibility is needed to support interim funding to ensure early access to the new intervention while further RWE is collected in the post-approval setting³⁰.

Data analytics, digital tools and digital transformation is one of six strategic focus areas in the EMA network strategy to 2025. Here, EMA emphasizes the importance of adapting to the rapid global evolution of digital healthcare systems and the use of digital technologies to generate RWE to fill the evidence gaps that exist between outcome data emerging from a clinical trial designed for regulatory approval and clinical evidence needed by downstream stakeholders including HTAs, payers, and ultimately clinicians and patients³¹.

Feedback from stakeholders also raised the issue related to lack of local regulatory guidance on how RWD can be utilised as decision basis for health economic evaluations and reimbursement decisions. One step towards this is a draft guideline on registry-based studies, published in September 2020, by the EMA. The draft guideline aims to optimize the use of registry-based studies as a source of real-world evidence in the context of benefit-risk evaluation of medicinal products in Europe³².

5.4.6 Status in Norway

Norway is in a unique position with many and data-rich health registers. Unfortunately, extracting data from these registries can be a cumbersome and timely process. In addition, these registers have varying data quality and, in some cases, incomplete records. The Directorate for e-health established the Helsedataprogrammet in 2017 with the goal to simplify and improve the use of Norwegian health data. Helseanalyseplattformen (HAP) as an ecosystem for analysis of health data will be established as part of this program. The goal of the platform is to simplify access to and enable analysis across the different health registries, without compromising on data security.

The number of industry-sponsored clinical trials has dropped significantly in Norway over the past few years³³. As a measure to reverse this negative development, the Ministry of Health and Care Services (HOD) launched an action plan for clinical trials in 2020 (Handlingsplan for kliniske studier). In addition, more than 60M NOK was granted in the State budget for 2021 to back this plan and to increase the focus on precision medicine and innovation. This included establishing NorTrials, which is a public-private partnership for clinical trials. National initiatives such as Impress and the diagnostic platform InPred (Infrastructure for precision diagnostics) are important to keep up with the rapid developments in precision medicine, and digital elements are gradually being implemented in patient care and clinical research in Norway.

As mentioned earlier, even with limited data available at submission, new technologies within precision medicine (such as NTRKs inhibitors) have received regulatory approval in the EU. However, the following local health economic evaluations and reimbursement discussions proved difficult. The Norwegian reimbursement system for new health technologies is not set up to handle treatments within precision medicine.

To meet this new era of highly specialised treatments, in 2020 the Ministry of Health and Care Services directed the hospitals to “*under the leadership of Helse Midt-Norge RHF, study and implement schemes for temporary introduction and reassessment of new methods in Nye Metoder to facilitate the introduction of personalized medicine in the service*”.

5.4.7 Covid-19 – a catalyst for decentralised elements

The trial models discussed above, with focus on precision medicine and decentralisation, will all be part of shaping the future of clinical research. Technology is enabling decentralisation of trials and these will play an increasing part of the future alongside traditional site-based clinical trials. The ongoing covid-19 outbreak has been a catalyst for decentralised elements, and it is expected that, post-pandemic, trial innovation will continue along this path of innovation.

5.5 PRECISION MEDICINE IN A HEALTH ECONOMICS PERSPECTIVE

Eline Aas (University of Oslo), contributions by Monica Gomez (University of Oslo)

Worldwide, Norway has been a pioneer in explicit guidelines on priority settings in healthcare. The process started in 1987 and has been continuously revisited. The introduction of precision medicines challenges current guidelines.

5.5.1 Priority settings in healthcare

The first priority settings in 1987 were based on two pillars: severity and *health effect* of new treatments³⁴. In 1997, *use of resources* was introduced as a third pillar indicating cost-effectiveness (balancing costs and health outcome) as a guiding principle³⁵. In a revision of the priority settings in 2014, which is the current guidelines, *health effect* and *resource use* (cost-effectiveness analysis) were set as the two first pillars, moving severity to be the third, measured by absolute shortfall and linked to the threshold value for an incremental health effect^{36,37,38}.

In 2007, the Norwegian Medicines Agency (NOMA) was the first institution in Norway to apply the explicit priority settings for drugs included in the National Insurance Scheme³⁹. Decisions related to reimbursement of medicines in healthcare institutions (hospitals and nursing homes), medical equipment, new surgical procedures, and organisational changes were not subject to guidelines. In the past decades, several organisational changes have been initiated to ensure that decisions are based on the same principles for the entire health sector^{40, 41}.

The National System for Managed Introduction of New Health Technologies within the Specialist Health Service in Norway (Nye Metoder) was established in 2013 to: ensure that harmful and ineffective treatments do not enter the market, establish a knowledge platform based on health technology assessments (HTAs), take resources into considerations, and finally establish systems for implementation⁴². As a result, NOMA started to evaluate reimbursement decision for drugs in hospitals. The final decision on whether the new drug used in hospital should be financed is taken by a board (Beslutningsforum) consisting of the four CEOs from the regional health authorities.

The latest discussion on priority setting is related to interventions initiated in the primary health service. The principles from the previous report (health outcome, resource use, and severity) were suggested to be continued, but with a proposal to include coping as an additional outcome measure to be taken into consideration.

5.5.2 Cost-effectiveness analysis in precision medicine

In economic evaluations, the net present value (NPV) of all relevant (often long-term) health outcomes and costs of alternative strategies are calculated^{43,44}.

The two most common types of evaluations are cost-effectiveness analysis (CEA) and cost-utility analysis (CUA). Both measure the health outcomes in natural units, the latter in quality-adjusted life-years (QALYs). CEAs are used to inform population-level decisions, but are not aimed at directly informing individual-level decisions. The evaluation of each strategy is represented by the incremental cost-effectiveness ratio (ICER). When only two alternatives are compared, the ICER is defined by the incremental costs (differences between the new treatment and standard of care) relative to incremental health outcomes (differences in health outcomes between the two strategies).



«If the incremental cost is negative and yields greater health outcome, the new treatment is considered a dominant strategy»

In a CUA, the ICER expresses cost per QALY gained. If the incremental cost is negative (new treatment cost saving) and yields greater health outcome, the new treatment is considered a dominant strategy. If, however, the cost of the new treatment is higher than standard of care and yields smaller health outcome, standard of care is the preferred treatment option. In a situation where the new treatment is more costly and yields higher health outcome than standard of care, the preferable strategy will depend on how much a QALY gained is valued by the decision-maker, often referred to as the threshold value.

In the context of evaluating costs and health outcomes, decision-analytic modelling (simulation or mathematical models) is the preferred approach and the standard method in application to NOMA. An advantage of modelling is the ability to synthesize available evidence (e.g., RCTs, observational studies, registry data) from multiple sources and extrapolate data beyond the time horizon of studies. As part of an economic evaluation, both parameters and structural uncertainty have to be addressed. Identifying uncertainty, expressed as the likelihood of a new treatment being cost-effective, is important to inform reimbursement decisions. Decisions made by NOMA and Beslutningsforum reveal that uncertainty is an important argument for not reimbursing a new drug.

Parameter uncertainty stems from estimation of model parameters that are inherently uncertain, such as the probability of experiencing a specific event (such as disease progression and death) or the accuracy of a diagnostic test. When incorporating multiple sources of data, these parameters may take a range of different values, which should be accounted for when evaluating strategies.

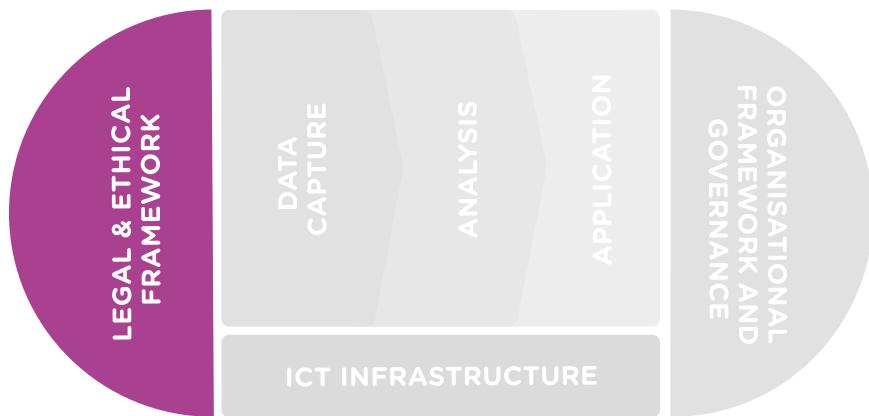
To assess the impact of parameter uncertainty, all parameters are varied simultaneously by assigning predefined probability distributions (e.g., beta, gamma) to each parameter and multiple sets of parameter values (referred to as probabilistic sensitivity analysis, PSA) are

sampled. While the impact of structural uncertainty typically is tested by different model assumptions, which often is related to applying different parametric specifications on progression free survival and overall survival (such as Weibull, log-normal and log-logistic). These two types of uncertainty are important to show decision uncertainty and to guide reimbursement decisions.

With precision medicine, the standard framework described above has two main shortcomings that needs to be addressed in the future before it can properly guide reimbursement decisions. Firstly, current guidelines are based on effects and costs for a given patient population, founded on evidence from randomised trials. As treatments are becoming more and more tailored, efficacy data could be derived from non-randomised controlled trials. Hence, new guidelines need to include how efficacy should be measured to adjust for potential selection bias, where synthetic control groups is one example of a possible method. Second, new methods on how to handle uncertainty around estimates in small cohorts of patients have to be developed. When small numbers of observations mean that there is not sufficient evidence to assign distributions to important input parameters, the standard methods of probabilistic sensitivity analysis must be combined with additional methods to account for the uncertainty around efficacy and safety.

In addition to the challenges related to estimating cost-effectiveness in precision medicine, a review of alternative financing systems is needed because the current system only has two decision options: yes or no. Managed entry agreements should be considered, not only in precision medicine, but in general to better combine the information from the uncertainty analyses with the decision. For instance, a conditional decision option (for instance to wait for additional information on the efficacy from an ongoing trial) should be explored. Further, a discussion of public-private partnership in testing new drugs is also warranted.

6. Legal and ethical framework



BigMed's legal team has been involved in all stages of the development of precision medicine tools, from idea to deployment. The early and consistent involvement of legal resources has enabled us to design solutions that fit the regulatory requirements of today and allowed for early identification of barriers to implementation.

The work has been carried out in three major areas:

- Navigating within the regulatory landscape of today
- Identifying where regulations need modification
- Suggesting new forms of regulation
- Establishing legal networks and the Nordic Permed Law (NPL)

The BigMed legal team established a network of legal practitioners from all relevant stakeholders. This diverse team ensured an increased awareness of the issues, an increased consensus of interpretations, and initiated some of the necessary changes to regulations.

The emphasis of the legal work in BigMed was on cooperation and involvement. This culture was open about new legal issues that are arising along with changes in medical practice and the field of healthcare. The legal work continued after the conclusion of BigMed through the established network Nordic Permed Law. Section 6.1 in this chapter summarizes this work.

After the General Data Protection Regulation (GDPR) was implemented, many discussions arose around privacy and the use of healthcare data. There is a need for clarity in interpreting existing laws. For instance, how can we balance the obligation

to provide state-of-the-art healthcare services while safeguarding patient privacy? A particularly important discussion in the BigMed project was about sharing of genomic data. This chapter includes a section evaluating the anonymity of interpreted genomic variants.

Computational risk and trust in AI are major issues when applying machine learning in a clinical context. The last section of this chapter focuses on the regulation of clinical decision support software and lab developed tests through the coming Medical Devices Regulation and the In-Vitro Diagnostic Medical Devices Regulation in the EU, including reflecting on ethical guidelines for trustworthy AI.

Box 10: State of the art healthcare

In this chapter the phrase “State-of-the-art healthcare” or “sound and proper healthcare” refers to what is known as forsvarlighetskravet or forsvarlighetsplikten in Norwegian. This is a fundamental principle that permeates Norwegian health laws (see e.g. the Health Personnel Act §§ 4 and 16 and the Specialist Health Services Act § 2-2).

Sections in chapter 6

- 6.1 Legal reflections to facilitate a data driven healthcare
- 6.2 Evaluation of anonymity with the purpose of sharing genomic data
- 6.3 Ensuring safe application of diagnostic tools – computational trust and regulations



Additional BigMed material

Relevant BigMed reports

- Befring: Persontilpasset medisin
- Befring and Sand: Kunstig intelligens og big data i helsesektoren
- Consent for clinical genetic testing in Norway – Considerations to the development of process and content
- Big data management for the precise treatment of three patient groups
- Report: BigMed-konferansen 2019: Rettslige reguleringer av persontilpasset medisin
- Clinical Decision Support Software Regulatory landscape in Europe from May 26th 2020
- Clinical Sequencing, Regulatory Frameworks and Quality Assurance for NGS-based Diagnostics

Podcasts

- Legal issues

Recorded webinars

- BigMed-konferansen 2020: Veien til presisjonsmedisin

An overview of all material from BigMed is available at bigmed.no

6.1 LEGAL REFLECTIONS TO FACILITATE A DATA DRIVEN HEALTHCARE

Anne Kjersti Befring (University of Oslo), contributions by Randi Borgen (University of Oslo), Inger Johanne Sand (University of Oslo), Gjertrud B. Mageli (Oslo University Hospital) and Oda Bakken (Oslo University Hospital)

6.1.1 Method and key findings

In order to maintain a high level of involvement of important stakeholders that influence both the regulation and how it is interpreted, BigMed's legal team arranged seminars, conferences, and other discussion arenas, with the following specific desired outcomes:

- Close collaboration between hospitals and academia
- Broad involvement of health law professionals
- Establishment of national and international networks
- Cross-discipline cooperation

Several new legal issues were identified through these interactions. These include consent in the contexts of healthcare, research and data processing, the storage of health data for use in healthcare and research, the use of artificial intelligence and big data, and the access to and ownership of health data.

Clarification of facts

Identifying and analysing legal issues require understanding of both the context that current legislation was designed for, and the new reality that the legislation must take into account going forward.

To ensure that current laws are correctly applied and future laws are properly developed, it is crucial to clarify the potential benefits and the consequences of using new technology. This includes not only to new knowledge and technology, but also the consequences of new events such as the global Covid-19 pandemic. The way that facts are evaluated in a legal context is a core aspect of legal methodology, as Befring discusses in her doctoral dissertation⁴⁵.

Ambiguities can arise in the legal framework when the context has changed radically since the law was written. For instance, new possibilities that come with technological advancement can lead to new questions related to values and ethics which the legislature never considered. Describing and analysing changed conditions and preconditions is of great significance for legal analysis and includes both the application of the legal dogmatic method and the judicial/legal policy analysis. Another important issue in health law is how EEA regulations and other international laws interact with national laws within areas of law where the context changes rapidly on a global level.

The development of the legal dogmatic method is described in chapter 2 of "Personalised Medicine – Legal Perspectives"⁴⁶ and in chapters 1 and 2 of the book "Artificial Intelligence and Big Data in the Health Sector – Legal Perspectives"⁴⁷.

Identification and analysis of “common denominators” in health law

BigMed sought to identify “common denominators” and fundamental principles and values which influence legal questions in modern digital medicine. These considerations

are particularly relevant in the application of new technologies and identifying them is necessary in order to raise and analyse new legal issues.

The following reflections must be considered:

- Changes to conditions from the time the law was designed must be clarified as a starting point for application and assessment of health law and policy.
- How changes to conditions might change the premises that current legislation was based on, and may reveal a need for new or additional legislation.
- How we interpret and develop laws to regulate artificial intelligence and big data within the health sector by identifying, developing and reflecting on the basic principles, values and considerations that can be defined as “common denominators”.

6.1.2 Changes in the understanding of legal sources precipitated by digitalisation

Medical and technological development requires extensive use of patient data in completely new ways. This requires looking at the health legislation with new eyes. Norwegian health legislation is a fragmented landscape. Changes in the actual circumstances or premises for providing modern medicine require corresponding changes in the regulation of health care, health data and new technology.

The division between health research and clinical practice makes it challenging to introduce certain elements of precision medicine. As healthcare and health research are increasingly digitised, the ability to maintain a high quality of healthcare services, both when treating each patient and when developing new treatment methods, is largely becoming dependent on the ability to store and utilise patient data. Health law must also take into account aspects of human rights and international obligations.

Using patient information and the patient record as a basis for knowledge

When a patient's health information and medical records are used as the basis to support decisions of diagnosis or treatment for other patients, they change into a more generic source of medical knowledge.

Patient data becomes the basis for the patient's own treatment, and is also used for developing treatment for

others. An “exchange relationship” is formed between the public and the health services, i.e. medical knowledge and knowledge used by the health sector are drawn from patient journals. One patient's data points will benefit the next, and so on. Data is constantly being generated, and these are of great scientific and commercial value. The ability to compare, analyse and draw conclusions from a large dataset, drawn from multiple patients' experiences (often massive numbers, in the case of big data) can be critical to providing “state of the art” healthcare.

Using the patient records as a basis for the development of algorithms can lead to more precise diagnoses and better care. At the same time, a fundamental prerequisite for providing sound and proper health services is that the public can trust that their medical information will be kept confidential. Health services must ensure that the patients' confidentiality is maintained when patient records are used in new ways in the health service. In cases where the patient data cannot be anonymised in the GDPR sense, the identity of the person must be protected in a different way.

Unclear regulations regarding healthcare, processing of health data and supervisory authorities

There is some ambiguity in the current legislation that regulates processing of health data. The legal regulations are fragmented on international, European and national levels. Many of the regulations Norway operates under today were adopted in a different environment. The arrival of new technologies has changed the function of data and turned the course of treatment into elements of research.

The two fundamental principles: State-of-the-art healthcare (see Box 10) and confidentiality, are of great importance as the basis for health regulations in a situation with increasing degree of fragmentation of the legislation and of supervisory responsibility. For example, determining whether the Norwegian Board of Health Supervision (Helsetilsynet) or the Data Protection Authority (Datatilsynet) should decide whether a kind of processing of patient data is necessary to fulfill the duty of providing sound and proper healthcare. The requirement for best medical practice is of utmost importance when determining what data is collected and how it is processed, stored, documented, and used.

There is a need for legal clarification of who has authority for supervision and enforcement of these regulations. Currently this is just as fragmented as the legislation and split between the Norwegian Board of Health Supervision

and the Norwegian Data Protection Authority. If the Data Protection Authority is to supervise health services based solely on the general provisions of the GDPR and the Patient Records Act, without also considering the specific responsibilities that pertain to health law, or the connections between these regulations, errors may occur during treatment that present high risks to patient safety.

There needs to be a greater understanding around how the legal duties across legal disciplines are tied together, and how they can be in conflict with one another in the processing of personal data. It must be clear that establishing costly and bureaucratic procedures meant to provide oversight and control, may also weaken the ability to share patient data quickly when necessary, and increase the risk associated with some areas of the patient care.

Patients as well as research subjects: the processing of personal health information is inherent to clinical research

Because modern medicine is under constant development, in many cases patients also become subjects for clinical research. This raises questions of patient rights and clinicians' obligations toward the patient. The regulation of healthcare, medical research, and processing of personal data are oftentimes primarily regulated within different sets of laws which can overlap and interfere with each other.

For example, when algorithms are developed in the course of a clinical trial or other research based on big data, they should also become available for use by the healthcare system. The same applies to storing and processing of genetic variants that are collected for medical science purposes and can be needed to diagnose individuals. Interpretation and sharing of genetic variants not only contribute to better and more precise diagnosis and treatments, but also to building new knowledge about the human genome outside the scope of research. Due to technological advancements and new possibilities to process and share genetic data, building knowledge about our genes has become an integrated part of the everyday life of laboratories and a part of how they provide healthcare and perform quality assurance.

Unnuanced categorisations of data are not sufficient

Evaluation of sensitive data will vary based on the context in which it is processed. The duty of care principles include both protecting human beings, the uses of personal data and biological material.

The patient must be sure that confidentiality will be maintained when their personal information is handled.

Within health law, the division between "personal data", "health data", "genetic data", and "anonymous data" are not enough to capture all the nuances needed to process data in the precision medicine era. There should be more nuanced and specific regulation for how to use different types of health data within different contexts.

After further considerations, one conclusion is that medical information can be handled differently in different situations, depending on the reason it is being used. Roughly separating different types of sensitive data may be insufficient according to the standards of State-of-the-art healthcare and confidentiality. It must be possible to block some information while immediately sharing other information. This highlights the need for a more nuanced and specific regulation of how health data can be used and processed within different contexts.

A good example of this is the discussion surrounding different types of *genomic data*. Genetic analysis and the interpretation of genetic variants are increasingly used in medical diagnostics and treatment because of precision medicine. Individual genetic variants and their interpretations are not only relevant to the individuals who have it in their genome, they also provide important knowledge about human genetics in general. Hence, health law must be designed in a way that regards genetic data as both. This shows how the nature of information gained from providing health services can be changed from personal data to data used in medical science, and how the knowledge gained by interpreting genetic variants must be shareable in order to make correct diagnoses.

6.1.3 Requirements for legal consent and the standard of equality

A need for new consent schemes to facilitate precision medicine

A key finding of BigMed is that emphasizing the patient's ability to provide valid consent can lead to complications, considering that the patient has a right both to protect oneself from certain information (the right not to know) and to receive healthcare. This is particularly important in relation to genetic screening, since such testing can reveal a lot of information that the patient may not want to know. Thus, a requirement to receive all the information in order to provide a valid consent creates a conflict.



This may lead to situations where patients who cannot provide valid consent receive a lower standard of medical care than those who can. These patients may also miss out on opportunities for inclusion in clinical trials.

Another key finding is the complications resulting from consent being regulated differently for providing health care, health research, and to processing of personal data. As health care becomes more and more data-driven and increasingly relies on technological tools that require extensive use of personal data, it is crucial to ensure that healthcare providers have sufficient legal basis to process the data that is necessary to provide State-of-the-art healthcare. In order to comply with the GDPR, health institutions have to explain a legal basis for their processing. If they do not find a sufficient legal basis in national health law for the processing in question, they must turn to the GDPR, which leaves them with no other option than to obtain valid consent, which may be difficult.⁴⁸

There are challenges to detailing broad consent for research while at the same time fulfilling adequate guarantees for healthcare. Furthermore, there is a discussion about distinguishing between valid consent in a research-ethics sense, as defined by the Health Research Act and the Declaration of Helsinki, and consent as a condition that must be in place in order to process personal data in accordance with the GDPR. Even if consent is granted to use the data in research, as required by the Health Research Act and the Declaration of Helsinki, questions arise about whether the same health data can be used for treatment.

The interpretation of the GDPR has been that it supersedes all other standards, even though its relevance is limited in the context of healthcare and medical research⁴⁹. This does not conflict with the national legal requirement that both healthcare and participation in medical research or trials should be voluntary.

The criteria for prioritisation need to be revised

We have seen that the development of technology may lead to more risk and imbalances in patient care, despite the norm of equality. This affects the criteria on which medical decisions are made, how they are used in laws for prioritisation, and the clear divide between healthcare and research. Changes to the factual basis and fundamental principles can justify developing rules regarding changes to the delineation between healthcare and research.

The right to healthcare may increasingly include preventive treatment and a proactive process in providing healthcare as genetic science is developed and applied in practice. Healthy patients may have a significant need to act to prevent serious future illness⁵⁰. Preventative care can also lead to a more efficient healthcare system. New understanding of genetics, along with cheaper and more widely available methods and tools will make genetic screening much more widespread. The healthcare can become “circular” when advances in genetic knowledge provide increased opportunities to correct previous diagnoses and change the course of treatment.

New knowledge in the field of genetics may identify pathogenic variants that were once thought to be harmless. This raises the question of how often laboratories will, or should, re-analyse a patient's sequential data and report on new results. The right not to know may also be significant for many people, presenting difficult ethical dilemmas for healthcare professionals.

«There are challenges to detailing broad consent for research while at the same time fulfilling adequate guarantees for healthcare»

Knowledge of genetic predispositions creates new issues connected to the equality principle in health services, as does the way costs are used as an instrument for prioritisation. New questions arise around the requirements of informed consent as this contradicts the patient's right to choose "not to know", and questions arise in relation to how genetic variants of uncertain significance should be used and stored.⁵¹

Artificial intelligence, machine learning, and clinical decision support tools

Machine learning and the use of clinical decision support tools in patient care require structural changes within the health sector and adaptations in the law.

Artificial intelligence and machine learning can use health data to provide new services and improve the existing ones, but their use and development is completely dependent on access to data. As the need for both primary and secondary use of data grows, healthcare and health research will become more and more intertwined. Furthermore, with digital tools and decision support tools, health data will serve additional purposes beyond providing healthcare.

Artificial intelligence requires a new foundation for viewing the structural and consensual nature of health data, and the many factors connected to this must be considered in any new legislation. Large amounts of data are used in medical genomics that bioinformatics tools (more or less based on AI-methods) can use to make diagnoses quickly and correctly. These may be linked to global search engines on the internet and may process health information or collect metadata in such a way that they act as a register of personal data.

The challenge is categorizing this type of register in light of the fact that there is a need to view healthcare and medical research within the same

context while maintaining the principles of privacy, confidentiality, and State-of-the-art healthcare.

There is a need for a legal definition of decision support tools that make fully automated individual decisions. This is in part because it can be difficult to distinguish between automated processing of a doctor's decision-making basis and fully automated decisions. Development and training of machine learning algorithms in health research must be seen in connection with their implementation in the clinical setting, and the dynamic development of the technology.

Liability for the production, procurement, development, and ownership of artificial intelligence must be analysed in more detail. This includes, responsibility of the manufacturer and the health personnel who feed the tool with data on a daily basis, responsibility for the data set (risk of sample bias), and whether health personnel can override conclusions made by algorithms.

6.1.4 Conclusions and continuations

The interdisciplinary approach was key to success in this project. It included legal professionals from governmental bodies such as the Ministry of Health and Care Services, the Norwegian Directorate of Health and the Directorate for eHealth, as well as supervisory bodies such as the Norwegian Board of Health Supervision, the county governors, the Norwegian Medicines Agency and the Norwegian Data Protection Authority, lawyers from the Norwegian Medical Association, patient advocacy groups, the Norwegian Cancer Society, and lawyers and attorneys specializing in health law, personal data, and medical technology.

Seminars and conferences were held, including two Nordic conferences with over 100 participants, one in June 2018 and the other in November 2019. The legal team was generally involved in all the work packages,

which created a shared understanding of the actual current conditions and needs. This has provided a foundation for defining and increasing knowledge about legal issues. Together we generated professional legal articles, studies, contributions, doctoral dissertations, and blog posts. See appendix for complete list of publications.

The network will continue collaborating on studies and research after BigMed under the umbrella of a new Nordic legal network, Nordic Permed Law. It will be an arena for research collaboration and discussion. The ambition is to create awareness of legal issues, contribute to enlightenment, reflection and debate, and be able to begin the work of making needed changes and adaptations to health law and regulation.

6.2 EVALUATION OF ANONYMITY WITH THE PURPOSE OF SHARING GENOMIC DATA

Anne Kjersti Befring (University of Oslo) and Oda Bakken (Oslo University Hospital)

During the project, a recurring question has been whether or not data is anonymous. Several of the other work packages in the project have worked on developing tools for sharing and processing different types of health data to be used both in research and clinical settings, to generate new and general medical knowledge. Some of these tools are so-called bioinformatic tools that streamline and facilitate increased sharing and processing of genetic data in healthcare.

Norwegian law regulates 1) information about someone's personal circumstances, cf. the duty of confidentiality in the Health Personnel Act Ch. 5, and 2) "personal data", "health data" and "genetic data", cf. e.g. the Patient Records Act § 2 letter b, the Health Register Act § 2 letter a, and the GDPR art. 4 no. 1, 13 and 15, cf. the Personal Data Act § 1. Together, these definitions constitute the scope of several regulations of data generated in the health service. Thus, how the boundaries between personal data and anonymous data are drawn will have a crucial impact on how the tools developed in the project can be implemented in different settings.

Precision medicine will increasingly make genetic screening and interpretations of genetic data a larger part of medical diagnostics and treatment. Interpretation of genetic variants is an area that is constantly evolving. To understand whether genetic variants are causing disease, they must be assessed based on other information about patients' disease development and compared across both health institutions and geographical borders to interpretations made by other laboratories. A prerequisite to successful implementation of precision medicine, is increased processing of genetic data in order to provide State-of-the-art healthcare and increase knowledge about the genetics of humans, cancer tumors, bacteria and viruses.⁵² Thus, clarifying the way different types of data – and especially genetic data – are regulated becomes crucial.

The BigMed project has looked further into whether interpreted genetic variants can be anonymous and how the boundaries between anonymous and personal data can be drawn.⁵³ This section provides a brief summary.

Box 11: What is an interpreted genetic variant and why do we need these?

The genome is the entire inheritance facility (DNA) and consists of approximately 3 billion bases (nucleotides) that make up the chemical basic units of the DNA molecules. There are four different types of bases, described by the letters A, G, C and T.⁵⁴ The composition and order of the bases determine the hereditary characteristics of humans.*

An interpreted (classified) genetic variant is a description of a genetic variant together with an interpretation of what medical consequences the variant may have for a human being.

Interpreting a genetic variant means that the variant is classified as pathogenic or not on a scale of 1 to 5, where 1 = benign and 5 = pathogenic.

A genetic variant is identified by comparing a patient's genome with the reference genome, that is, by referring to similarities and differences.⁵⁵

The reference genome is essentially identical to the genome mapped in the human genome project⁵⁶, and this effectively constitutes a standardized coordinate system for the description of genetic variants.

It is an acknowledged challenge to patient safety that different laboratories sometimes classify variants differently and that this can have serious consequences for patients through incorrect diagnoses and medical care.⁵⁷ Thus, comparison with other laboratories' classification of variants is important in order to be able to quality assure the variant classification.

The medical knowledge will become more and more reliable as the evidence material increases for each variant. Increased genetic knowledge will contribute to faster and more accurate diagnoses and treatment in the long term.

6.2.1 The challenges of protecting genetic data

An assessment of how genetic variants are regulated must be based on what (interpreted) genetic variants really are. Simply, an interpreted genetic variant is a description of the genetic variant together with an interpretation of what medical consequences the variant may have for a human being. There are many examples of diseases caused by genetic variants.

A challenge with genetic data is that it in general is inherently sensitive with respect to privacy. The genome of any human being is unique to that individual. In addition, genetic data can contain a lot of information about the individual and their relatives, sometimes even a whole group and extend generations. Furthermore, we may not yet know the significance of all of the data because knowledge of human genetics is under continuous and rapid development. Due to these characteristics, genetic data is given a so-called "special status" in UNESCO's International Declaration on Human Genetic Data.⁵⁸

There is also variation in the methods of identification and the sensitivity of genetic data depending on its characteristics. In order to determine whether an interpreted genetic variant is anonymous and how the boundary between anonymous and personal genetic data can be drawn, the current law and relevant legal sources and arguments must be considered.

The first question is whether it is possible to trace the interpreted genetic variants back to an identified or identifiable natural person or whether they are guaranteed anonymous (referred to

*The most common type of genetic variation among humans is called single nucleotide polymorphisms, SNPs ("snips"), where one of the bases (A, G, C or T) is replaced with another when compared to the reference genome.

here as objective anonymity). If the answer is that the variant cannot be linked to an identified or identifiable natural person, the question becomes whether it can still identify a person if the number of occurrences (frequency) of the genetic variant in a population is low (here referred to as “occurrence identification”). If it is possible to link the genetic variant to an identified or identifiable natural person, it must be considered whether they are anonymous as a result of a low risk of identification (risk-based anonymity). Risk-based anonymity is based on a dynamic assessment of the likelihood of the data being linked to a person based on both the quantity of the data and available means. Relevant factors include the availability of technical equipment, costs, characteristics of the data being processed, how they are filtered and processed (context and quantity), and access control.⁵⁹

Furthermore, it is necessary to consider to what extent precautionary considerations (the precautionary principle) can be emphasised when assessing how genetic variants (and also other categories of genetic data) can be processed when the purpose is to provide healthcare and underpin new knowledge and research. Uncertainty about the possibility of identification and how future technology will develop may be an argument for obtaining consent to process genetic data.

The conclusion is that a single, interpreted genetic variant is anonymous provided that it is not connected to other personally identifiable information. The number of occurrences of genetic variants in a population is not essential for how the boundary between anonymous data and personal data is assessed, or for how genetic and medical knowledge can be shared. The prevalence of genetic variants does not provide sufficient risk of identification.

6.2.2 Legal review

An interpreted genetic variant can be described as both a genetic finding and as new medical knowledge. This knowledge is equivalent to understanding other disease causes and is not itself sensitive information:

“A link between a particular genetic variant and clinical features of a disease is not personal information any more than the link between high blood cholesterol and heart disease.”⁶⁰

Biological prerequisites and symptoms of disease, similar to genetic predispositions, are described as “medical knowledge” in scientific articles, textbooks and on the Internet. Medical knowledge shall be shared so that humanity as a whole

can benefit from this knowledge and scientific progress. The legal challenges arise when genetic data can be regarded as both medical knowledge and personal data.

In the legal proposition of the Personal Data Act duty of confidentiality is referred as a “guarantee” to protect patient information and research subjects’ integrity.⁶¹ The duty of confidentiality provides the basis for the protection of both the information and its source, but is adapted to the need to share information for State-of-the-art healthcare and health research.⁶² This duty does not apply to those who already have knowledge of to whom the genetic variants belong or when the person’s identity is adequately protected.

The GDPR is incorporated into § 1 of the Personal Data Act.⁶³ The boundary between anonymous and personal genetic data is mainly drawn through the legal definitions of “personal data”, “health data” and “genetic data” in the GDPR. The legal definition of “personal data” in GDPR art. 4 (1) and interpretations of this provision form a common core for the definitions of “genetic data” and “health data” in Art. 4 (13) and (15). In national health laws, reference is made to the definition in GDPR art. 4 (15).⁶⁴

One consequence of the fact that genetic data is covered by these legal definitions is that at least one of the terms of GDPR art. 6 (1) and art. 9 (2) must be fulfilled, though there are exceptions in GDPR art. 9 (2) letter h and (3) for processing genetic data in health care. In both of these provisions, reference is made to national law. This means that the national health laws will apply, e.g. the Health Personnel Act, the Patient Rights Act, the Patient Record Act, the Specialist Health Services Act, the Health Research Act and the Health Register Act.⁶⁵

An overarching requirement of both national legislation and the GDPR is that human dignity shall be safeguarded, i.e. human rights, which also form the basis for the protection of integrity. Human rights are regulated by the Norwegian Constitution, the legislation in general, and international conventions. They include a protection against discrimination and against physical violations, interventions in family and private life, and the safeguarding of volunteerism (autonomy).⁶⁶ At the same time, the respect for human dignity and the duty to provide State-of-the-art healthcare are arguments for accessing data, both anonymous and personal, and for the storage of relevant and necessary data.⁶⁷ This obliges risk assessments to include different consequences depending on the processing of the genetic data.⁶⁸

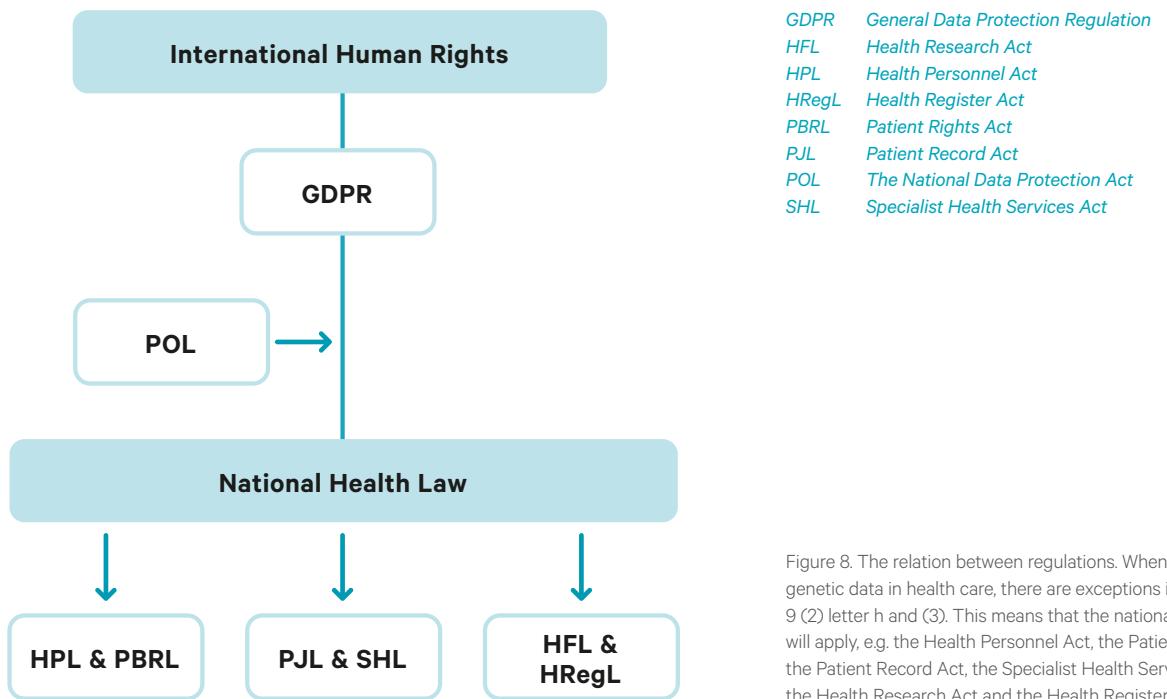


Figure 8. The relation between regulations. When processing genetic data in health care, there are exceptions in GDPR art. 9 (2) letter h and (3). This means that the national health laws will apply, e.g. the Health Personnel Act, the Patient Rights Act, the Patient Record Act, the Specialist Health Services Act, the Health Research Act and the Health Register Act. (16).

6.2.3 Summary of relevant laws

The European Convention on Human Rights of 1950 (ECHR) and the UN's International Covenant on Economic, Social and Cultural Rights of 1966 (ICESCR) contain several articles relevant in the processing of genetic data: ECHR Art. 2, 3, 8 and 15 and ICESCR art. 12. The UN's International Covenant on Civil and Political Rights of 1966 contains rules on health research and the processing of data, cf. Art. 7 and 17. These conventions are implemented by the national Human Rights Act which determines that they precede Norwegian law in the event of conflict.⁶⁹

The Biomedicine Convention of 1997⁷⁰ is not covered by the Human Rights Act, but contains obligations that Norway has committed to.

The relationship between international human rights, the GDPR, the national Personal Data Act and the national health laws can be illustrated as shown in Figure 8.⁷¹

The International human rights are fundamental and superior to both European and national legislation. These are fundamental and supranational rights that all people have by virtue of being human. The GDPR, on the other hand, applies only to countries within the EU/EEA and is presumed to comply with human rights. Furthermore, the GDPR is a general regulation for the processing of personal data. This general regulation allows member states to regulate the processing of personal data more specifically, for example, when the purpose is to provide healthcare. Norwegian health legislation constitutes such national legislation and is assumed to comply with both the GDPR and international human rights.

6.2.4 Anonymising unique genetic codes

When the implementation of broad genetic screening is discussed, it is often pointed out that it is difficult to guarantee the anonymity of research participants because each individual's genetics are unique.⁷² Genetic data and biological material, with advanced technology, can be a source of identification and comprehensive information, but possibilities for identification vary depending on scope, expression, and sensitivity. Genetic data cannot always be linked to a person.

When assessing genetic variants, an important factor is that many genetic variants are common between humans. When an interpreted genetic variant is detected in several different human beings, it cannot necessarily be traced back to only one person. Based on an objective approach, a single interpreted genetic variant without any other identifying information, is anonymous.

Actual and judicial assessments conclude that the number of occurrences of a genetic variant (frequency) will not, as a general rule, be decisive in determining whether a person can be identified. The number of times an interpreted genetic variant is observed does not provide information that can be linked to a particular human. It is not possible to identify a person without access to an overview of all genetic variants and how these are distributed between human beings or a population.

However, a distinction must be made between identification based on *frequency* (occurrence identification) and identification based on a larger *amount* of data that may include several genetic variants and other information that makes it possible to identify a person. In this situation, the information may remain anonymous even if it is not objectively anonymous, on the basis that there is a low risk of identification (risk-based anonymity), depending on the context in which the data is processed. It must be considered at what point the amount of genetic variants and other information or data increases the likelihood of that a person is identified.

One starting point is that fewer than 100 single nucleotide polymorphisms (SNPs) are needed to distinguish between two individuals' DNA profile.⁷³ The sum of many variants from the same individual in one limited database can lead to an identifying "fingerprint". This form of identification requires access to advanced means that might be unlikely to be applied, and must be viewed in the context of how the data is processed.

Assessing the risk based on the amount of data must be carried out on the basis of how the data shall be processed and the data's characteristics.⁷⁴ The purpose of the database, access to the database and access control, how information is compiled, and the range of data can also affect whether it is anonymous.⁷⁵ The possibility of identification increases when a dataset contains other personal data and health data in addition to genetic data.⁷⁶ Furthermore, it must be considered what is required of technology in order to identify genetic data, for example whether it is generally available or whether relevant technology is unlikely to be used.

If several genetic variants are stored from multiple people, anonymity may be maintained if the relationship between the data is adequately protected. Although there is a theoretical possibility of re-identification, genetic data filtered in such a way that the variants cannot be associated with an individual can still be deemed anonymous.^{77,78} Large databases, such as in the Beacon network, can be aligned so that the risk of re-identification is low by controlling how genetic variants and other data are made available. The boundary between anonymous and identifiable genetic data should be treated as dynamic, not static.

With the technological possibilities of the future, the options for defining health data and genetic data as anonymous may become more limited. In an anonymity assessment, it is merely possible to assess the current risks, although it is a requirement to also consider any future technological developments. Future technology must be assessed continuously in light of how data is stored over time and probable threats to the stored data.⁷⁹

In case of uncertainty regarding the risk of re-identification, a relevant precautionary consideration can be to consider the consequences of the data being defined as personal data and of the data being defined as anonymous. Uncertainty related to anonymity can lead to unfortunate and unintended restrictions on the sharing of medical knowledge. As a result, health institutions and healthcare professionals may fail to use modern, necessary and accessible tools to share interpretations of genetic variants, even though they know it can impact diagnostics and treatment methods.

Furthermore, uncertainty can lead to increased use of resources and implementation of measures that may not be necessary, or in the worst case, to the detriment of patient safety.

Thus, in legal “gray zones”, it is necessary to consider risk in a broader perspective. The risk of identification must be balanced with the risk of not sharing or processing the data.* This includes assessments of proportionality and prudence of sharing genetic data (e.g. in order to provide medical treatment to a patient with a rare condition).⁸⁰

6.2.5 Looking forward

National authorities have not yet communicated a clear strategy for how genetic data should be processed in Norway. There are many different types of data that will all qualify as genetic data, but not all genetic data will be categorized as “genetic data” according to the definition in the GDPR. As increased processing of genetic data lays the ground for State-of-the-art health services in Norway, it is crucial to clarify the regulations for genetic data. For now, it is up to the data controllers to make independent assessments. This can lead to different practices across health institutions and contribute to unequal health services.

There is a need for a more nuanced regulation of data. Current consent schemes are not adapted to the needs of data in health care, or to fulfil other patient rights**. General regulations of genetic data both within and outside the healthcare sector contribute to uncertainty and unpredictability. Relying on the health institutions to set the course may have an unfortunate impact on patient safety, and furthermore have a significant impact on the implementation of precision medicine. It is a state duty to make sure that citizens can enjoy the benefits of scientific advancements. The introduction of precision medicine, artificial intelligence and big data should lead to the development of new statutory data sharing and consent schemes. The technological development and the prohibition in GDPR art. 9 (1) presume a wider range of legal bases in national law for the processing of data covered by the legal definitions.

Storing, sharing, and comparing genetic data is essential in order to implement precision medicine and new technology in accordance with national strategies. This justifies that the health authorities should work with the stakeholders in the health sector to set the course for regulations.

*Both proportionality assessments and prudence assessments can have an impact in this situation. In many situations, the consent scheme may be too limited, which can be illustrated by the fact that requiring a consent may limit other fundamental patient rights (e.g. the right to receive health care regardless of the ability to give consent and the right to protect themselves from information about their health). Requiring consent may also limit the possibilities to use other patients' data in diagnostics even though the patients' identities are adequately protected.

**For example, the consent scheme is not adapted to patients and research subjects who wish to contribute with their health data without being in interaction, which a dynamic consent adds up to.

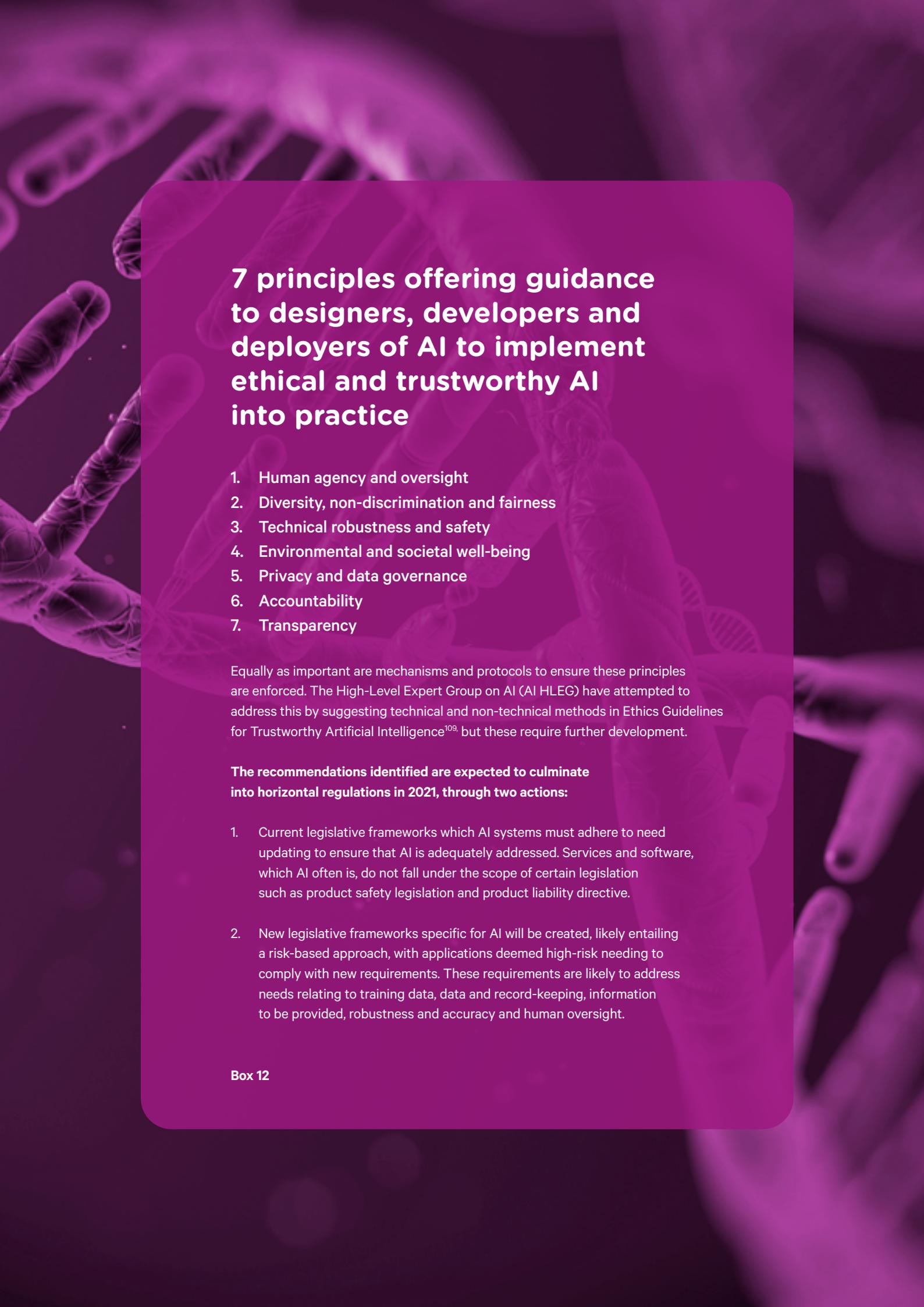
6.3 ENSURING SAFE APPLICATION OF DIAGNOSTIC TOOLS – COMPUTATIONAL TRUST AND REGULATIONS

Hariharan Michael Hallock (DNV), contributions by Frédéric Courivaud (DNV), Courtney David Nadeau (DNV), Dag Frode Nilsen (Sykehuspartner)

Since 1993, Europe has required that all commercial products have a CE marking to demonstrate their compliance with health, safety and environmental protection regulations. Medical devices are assessed to receive CE marking in one of three categories: active implantable medical devices (AIMD), medical devices (MDD), and in-vitro diagnostic medical devices (IVDD). In response to significant technological and computational advances and patient safety breaches, these three categories, will be replaced by new regulations (MDR and IVDR) which come into effect in May 2021 and May 2022. While the MDR and IVDR cover different subsets of devices, the MDR is the ruling regulation, with the IVDR aligning on MDR principles for the regulation of in-vitro diagnostics devices (including software).

The MDR/IVDR are more stringent than previous directives, with a wider definition of medical devices. It includes all devices which provide information used in medical diagnosis or prognosis of human beings (article 2(1)), specific caveats addressing technologies created within health institutions ‘in-house’ (article 5(5)), and re-classification of software as higher risk (according to rule 11 – Annex VIII). Therefore, software will primarily be classified as class IIa, IIb and III, and will require involvement of a Notified Body in the certification process. (see – BigMed’s CDS software whitepaper)⁸¹

These new regulations aim to ensure proper documentation about the safety and performance of the device during its life-cycle and align with European single market principles. Although AI is not explicitly mentioned in the regulations, these solutions may be classified as medical devices and will require manufacturers and economic operators obtain MDR/IVDR certification. Short comings in these current



7 principles offering guidance to designers, developers and deployers of AI to implement ethical and trustworthy AI into practice

1. Human agency and oversight
2. Diversity, non-discrimination and fairness
3. Technical robustness and safety
4. Environmental and societal well-being
5. Privacy and data governance
6. Accountability
7. Transparency

Equally as important are mechanisms and protocols to ensure these principles are enforced. The High-Level Expert Group on AI (AI HLEG) have attempted to address this by suggesting technical and non-technical methods in Ethics Guidelines for Trustworthy Artificial Intelligence¹⁰⁹, but these require further development.

The recommendations identified are expected to culminate into horizontal regulations in 2021, through two actions:

1. Current legislative frameworks which AI systems must adhere to need updating to ensure that AI is adequately addressed. Services and software, which AI often is, do not fall under the scope of certain legislation such as product safety legislation and product liability directive.
2. New legislative frameworks specific for AI will be created, likely entailing a risk-based approach, with applications deemed high-risk needing to comply with new requirements. These requirements are likely to address needs relating to training data, data and record-keeping, information to be provided, robustness and accuracy and human oversight.

Box 12

regulations relating to AI will be addressed through the EU Medical Device Coordination Group (MDCG), with specific guidance due in the near future. In preparation for this, manufacturers and stakeholders can begin (or continue) to familiarize themselves with recommendations for trustworthy and ethical AI (e.g. ISO/IEC TR 24028:2020⁸², *Assessment List for Trustworthy Artificial Intelligence (ALTAI)* self-assessment tool⁸³) and build their conformity documentation in line with the requirements that correspond to their device intended purpose, classification, and conformity assessment route.

6.3.1 Regulatory compliance for CDS

Clinical decision support (CDS) software helps clinicians to “retrieve, filter and/or analyse patient data and assist them in the decision-making process”⁸⁴.

CDS software has the potential to aid delivery of safer, quicker and higher-quality care through diagnostics support, prognosis of treatment and prediction and monitoring of illness outcomes. As such, it is likely that CDS software will need to adhere to MDR/IVDR. Manufacturers of such software, whether it be developed by industry or by health institutions ‘in-house’, should consult MDR/IVDR themselves to assess if compliance is required.

A growth in CDS software in the healthcare market coincides with the digitalisation of healthcare globally, and the resultant increase in health data available. However, healthcare stakeholders should not mistake the publication of academic papers demonstrating CDS technologies on isolated data sets and patient populations as evidence that the CDS necessarily is safe and effective. MDR compliance is required for software manufacturers to ensure that their tools are safe, effective and used responsibly.

Discussions enabled by BigMed generally indicate that stakeholders are unaware of the details of these regulations and their regulatory responsibilities. Many underestimate the scope of MDR, and the quality, engineering, regulatory compliance and clinical investigations needed before these tools can be made available for physician or patient use.

Both regulatory and manufacturing challenges remain for AI based CDS software wanting to demonstrate compliance to directives, especially while standards for consistent assessment are not yet available.

6.3.2 Regulatory compliance for lab developed tests

Sequencing-based genetic tests, both for cancer and for rare diseases, are part of standard diagnostic care in Norway. To compliment these, multiple use cases and pilots for AI and ML technologies have been developed. These have unveiled complex safety, quality and regulatory considerations.

From May 2022 NGS-based genetic tests will require IVDR-compliance. Today in the Nordics, most of these diagnostics are provided as lab-developed tests by individual hospitals, where these tests are developed, validated, verified and deployed. Under the new regulations, hospitals face more robust regulatory requirements designed to ensure patient safety and test performance. Additionally, the indications for which health institutions can develop their own tests is greatly reduced in scope, and a broader shift to certified solutions is expected.

6.3.3 AI in precision medicine

AI is increasingly being used in medical devices, and there are now several examples of “AI as medical device software” (AI-MDSW) used in radiology and image analysis. These are CE marked, mostly AI-enabled CDS software, certified under the (less stringent) MDD (not the MDR). EU-MDR applies to any software qualified as a medical device regardless of the presence of artificial intelligence. However, the interpretation of specific regulatory requirements for AI driven software needs some clarification. Specific guidance on that matter is expected to be published soon by the MDCG.

Discussions enabled by BigMed have highlighted mixed perspectives when it comes to implementing AI within healthcare. While many stakeholders see the potential of AI to improve accuracy of diagnosis and success of treatment, many also fear that the understanding of AI is still immature and that mechanisms to quality-assure AI algorithms and models are inadequate. These feelings are common across several industries and are addressed in recent EU publications on AI. The High-Level Expert Group on AI (AI HLEG) *Ethics Guidelines for Trustworthy Artificial Intelligence*⁸⁵, and more recently the AI HLEG’s *Assessment List for Trustworthy Artificial Intelligence (ALTAI)*⁸⁶ focuses on ethics and trustworthiness; the EU White Paper on Artificial Intelligence: a European approach to excellence and trust⁸⁷, focuses on potential regulation.

7. Practical examples from the implementation of precision medicine in three clinical areas

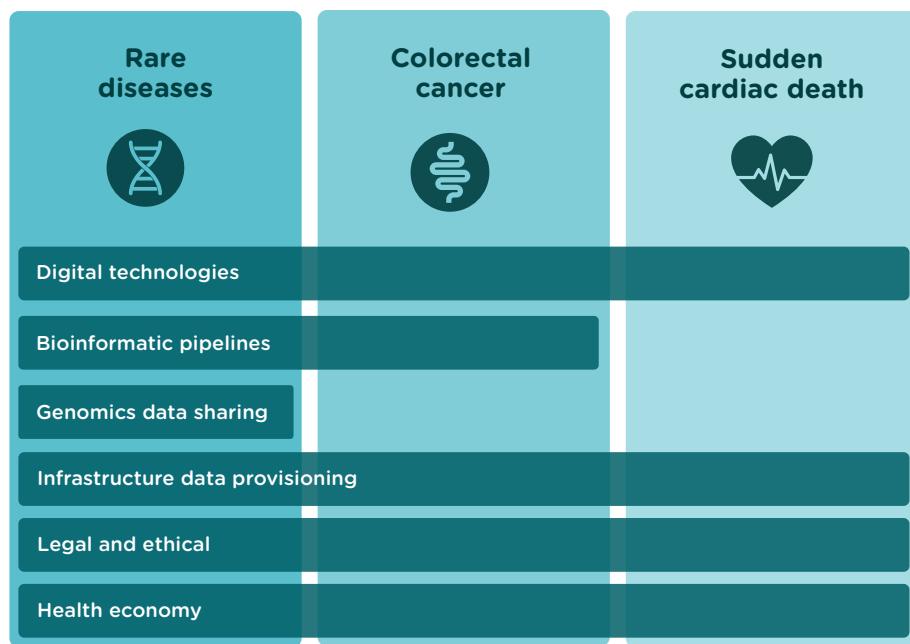


Figure 9. Matrix organisation of BigMed.

At the outset of the BigMed project, it became clear that the clinical fields and work packages would intersect with technology, data protection, legal issues and infrastructure. This was reflected in a two-dimensional, interwoven organisation of the project (Figure 9). The expectation was that the clinical research work would identify and challenge barriers, and that the project would use these experiences to identify and hopefully address general obstacles.

In the cardiology work package, the focus was on making clinical text analysable using natural language processing (NLP) technology. The work highlighted the immaturity of the current regulatory and infrastructure paradigm when it comes to enabling efficient re-use of clinical data. These experiences are generalizable to most large-scale clinical data use cases. It also demonstrated the usability of NLP in extracting analysable data from clinical notes, and the value of synthetic data for research and development purposes.



«The overall aim of the genetics work was to increase the speed of genetic diagnosis in critically ill children with rare diseases»

The work underscored the need for Norwegian clinical word embeddings, giving impetus to a wider collaboration on NLP, and the inclusion of Akershus University hospital in the project (Section 3.2: Extracting information from clinical text with natural language processing).

The overall aim of the genetics work was to increase the speed of genetic diagnosis in critically ill children with suspected rare diseases. This entailed establishing new pipelines for genetic analysis using new infrastructure, improving the precision and speed of clinician feedback to geneticists employing ontologies, and making a new variant calling software application called ELLA. Sharing information on genetic findings and related phenotypes between laboratories is essential, and the current work highlighted and challenged regulatory approaches to genetic data, spurring on suggested changes in Norwegian law.

The colorectal work package took on the challenge of presenting large and complex data sets to clinicians in order to facilitate important treatment decisions. The work also demonstrated a direct interaction with national cancer registry data and the use of patient similarity networks in comparing the current patient's data to similar patients' disease trajectories and treatment responses, approaches that could aid both clinical decision making and patient involvement in therapy choices. Reusing previous patients' data from registers – or directly from the EHR – challenges the current partitioning of clinical care, quality assurance and research, and emphasises the need for rethinking health data regulations.

The clinical research work in BigMed grew out of established and on-going projects. The following sections, written by the clinical researchers supported by their teams, highlight some of the practical challenges encountered when trying to undertake precision medicine-related research under the current technological and regulatory paradigm. Both practical and pragmatic workarounds had to be applied, while certain achievements were possible due to the environmental changes achieved in the course of the BigMed project. A vision for a more streamlined big data work flow in the respective fields is presented.

Sections in chapter 7

- 7.1 Prevention of sudden cardiac death
- 7.2 Whole genome sequencing for rare diseases
- 7.3 Colorectal cancer

Additional BigMed material

Relevant BigMed reports

- Patient similarity networks for precision medicine
- Implementing NGS-based diagnostics in cancer care: Technical and organisational factors in the Nordics
- Consent for clinical genetic testing in Norway – Considerations to the development of process and content
- Germline genomic medicine: A BigMed needs analysis
- Suggesting Reasonable Phenotypes to Clinicians
- Drivers in rapid genetic diagnostics for rare diseases in infants
- Clinical reporting of NGS data: A systematic Nordic collaborative, peer-reviewed benchmarking
- Big data management for the precise treatment of three patient groups

Podcasts

- Personalised Cancer Treatment
- Genomics and Datasharing
- Cardiology and Technology
- The Key to Precision Medicine

Recorded webinars

- BigMed at EHiN 2020
- BigMed-konferansen 2020: Veien til presisjonsmedisin

An overview of all material from BigMed is available at bigmed.no



7.1 PREVENTION OF SUDDEN CARDIAC DEATH

Pål H. Brekke (Oslo University Hospital), contributions by Kristina Haugaa (Oslo University Hospital), Lilja Øvreliid (University of Oslo), Fredrik A. Dahl (Akershus University Hospital), Petter Hurlen (Akershus University Hospital), Ildiko Pilán (University of Oslo), Øystein Nytrø (NTNU), Taraka Rama (University of Oslo).

Sudden cardiac death (SCD) and sudden cardiac arrest (SCA) are dramatic events for the individual patient and their families. While exact data are lacking, national and international statistics suggest there are 5000-6000 cases of SCD per year in Norway, and survival of out-of-hospital cardiac arrest is only 5-10%. The incidence of SCD increases with age, but in younger age groups, the proportion of SCD as a cause of death is larger than in older age groups. Exploring new and more effective ways to prevent SCD and SCA is of considerable medical and socioeconomic importance.

SCD is generally defined as an unexpected death without an obvious non-cardiac cause. Indeed, at the turn of the 21st century, current knowledge indicated that for nearly half of SCD cases, cardiac arrest was the first symptom of cardiovascular disease⁸⁸. In children and younger adults, genetic heart disease such as arrhythmia syndromes and cardiomyopathies are the leading causes of SCA/SCD. Coronary artery disease increases rapidly from middle age⁸⁹.

The healthcare system strives to identify individuals who are predisposed for such serious conditions, and to provide individually tailored prevention and/or treatment. On the one hand, there is continuous effort in improving SCD/SCA risk assessment in established cardiac disease in order to more accurately identify those patients requiring more intensive follow-up, and letting patients with very low risk lead normal lives with as little intervention from healthcare systems as necessary. On the other hand, the challenge for preventative care is identifying people at risk even before they have received a medical diagnosis.

Box 13: Hypertrophic cardiomyopathy (HCM)

In BigMed, we have used the most common genetic heart disease, hypertrophic cardiomyopathy (HCM), as an example condition where big data assisted diagnosis and risk stratification may improve prognosis and prevention of SCD. Our approach, however, has been to highlight areas where the methodology can be generalised to other conditions.

HCM has a prevalence of 0.2–0.6%. A pathogenic mutation is found in approximately 60% of patients. In HCM, the heart wall muscle can increase to several times normal thickness, which affects heart muscle contraction, increases risk of ventricular arrhythmias, may obstruct blood flow through the heart, can impair oxygen delivery to the heart muscle itself, and cause formation of scar tissue inside the heart walls. For patients with HCM, possible symptoms are chest pain during exercise, breathlessness, palpitations, dizziness, loss of consciousness (syncope), fluid retention, and fatigue. Studies have shown that patients with HCM have a risk of ventricular arrhythmia, cardiac arrest, or SCD of nearly 1% per year.

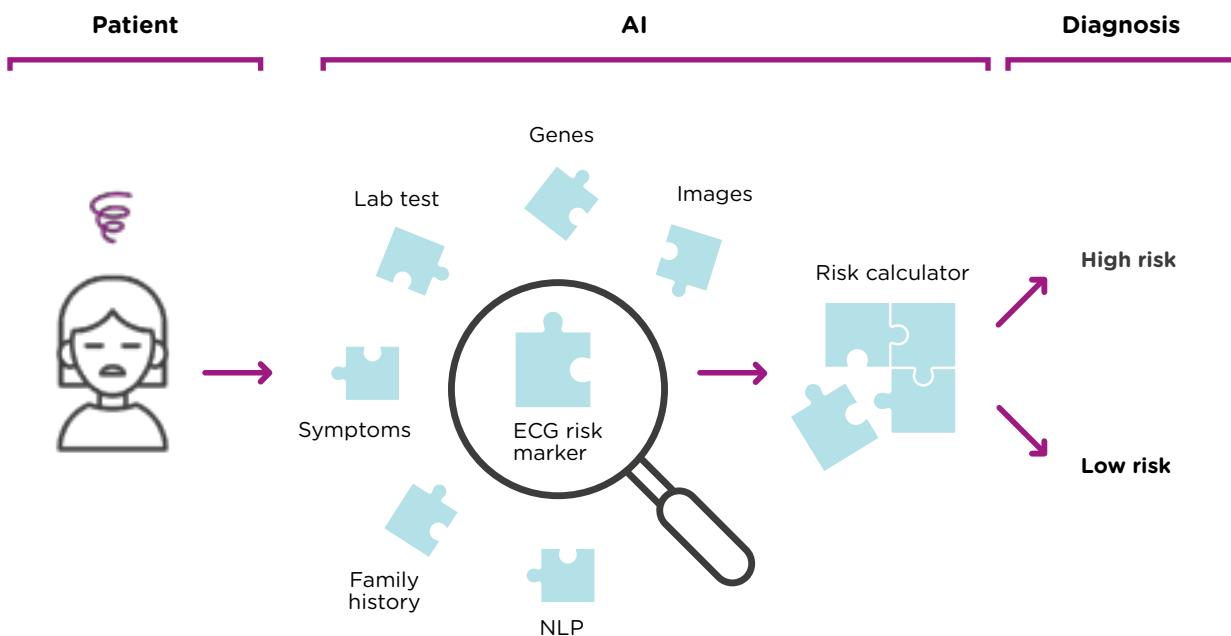


Figure 10. AI tools can potentially detect features not readily visible to clinicians, and support tasks such as risk stratification or diagnosis.

While some factors contributing to individual risk – such as lab test results and electrocardiographic abnormalities – may be easily measurable and available for analysis, there is a lot of information in the textual part of the electronic health record that has potential to add substantially to future risk models. Most people don't live their lives under constant medical surveillance (future smart devices may certainly invalidate this statement). Therefore, the patient's narrative of symptoms and event, and the clinician's observations and reasoning, are important sources of health data.

The BigMed project has highlighted the need for language and domain specific basic NLP tools (see section 3.2). By employing this technology, it is possible to extract symptoms, findings, measurements and other data points from clinical text and store these as analysable values. This will allow the preservation of the expressive and nuanced human language in clinical notes, essential for clinician-to-clinician communication about the patient's condition and treatment considerations, and obviate the need for time-consuming data structuring – and inherent information reduction – on the input side.

The BigMed project set out to pave the way for NLP use on Norwegian language health records. However, due to regulators' privacy concerns, limited access to medical records forced a detour via *synthetic* (made up, but realistic) clinical text data, which provided some important learning points. We were able to show that in cases where data access is difficult or impossible, synthetic data may be a useful starting point for developing NLP⁹⁰. We also showed that data models based on synthetic data work quite well when transferred to real clinical data. The clinical use case for this was family history⁹¹, which when structured, can assist the clinician in ascertaining hereditary disease patterns, and when linked to clinical events in the family, can support individual risk assessment.

The project also demonstrated the effectiveness of NLP methods more commonly used in sentiment analysis – rather than named entity recognition – to identify patients who have had a loss of consciousness (*syncope*)⁹² which is a known risk factor for sudden death in HCM. Syncope is a commonly occurring clinical phenomenon which in most cases is harmless, but in specific cases and conditions is associated with serious incidents. However, since the ICD-10



symptom code for syncope is not associated with any economic reimbursement for hospitals, the diagnosis is often missing from hospital diagnostic registries. A Danish study has shown that 1/3 of syncope cases admitted to hospitals are missing the ICD diagnostic code. This has obvious implications for the use and accuracy of risk models based on registry data⁹³.

Clinical utility of such tools could be NLP systems that recognise that there has been a likely sudden death in the patient's family history, or that the patients' previous records contain an episode of loss of consciousness (syncope), which could direct the clinician to order more extensive diagnostic tests, such as genetic analysis.

Norwegian clinical text from hospital records is sparse, highly context dependent, rich in domain-specific terminology and abbreviations, and sentences are often incomplete. All of these characteristics are challenging general NLP tools. Fundamental building blocks in NLP such as word embeddings – terms with similar meanings will have similar positions in a vector space that aid in generalising a machine learning model – are non-existent for Norwegian clinical text. BigMed has aimed to address this deficiency by building word embeddings from hospital records from Ahus, but the analysis was only able to start in late 2020. The project faced a near two-year delay in getting data access due to regulatory bodies' unfamiliarity with big data subjects and machine learning/AI, and inconsistent interpretations of regulations in local, regional and national bodies. This again illustrates the importance of the work BigMed has done regarding legal interpretation and simplification (see chapter 6).



AI in medical imaging and biological sensor data

As general image analysis has been at the forefront of AI and neural network technology evolution, it is natural to assume that medical imaging will be among the first areas of medicine impacted by AI/ML solutions. Indeed, several AI based tools for tasks such as breast cancer diagnosis, stroke detection and image segmentation are already being marketed to radiology departments. In cardiology, related to the SCD/SCA problem, the expectation is that AI-supported analysis of echocardiograms, angiographies and MRI images will provide novel markers of disease and progression.

Very likely, subtle patterns in large imaging datasets contain important distinguishing phenotypical features which will allow more personalised prevention and treatment. Automated or AI-supported image analysis could alert the clinician to a high risk of SCD in a patient independent of disease history or genetic data due to specific but subclinical patterns. Additionally, in patients with established disease, AI-enabled imaging studies will more precisely classify patients as having high or low SCD risk depending on the context.

Research from OUH and other BigMed partners in the cardiological imaging field has shown that manual analysis of speckle tracking echocardiography and changes in segmental movement expressed as global strain, mechanical dispersion, and myocardial work have significant diagnostic and prognostic importance in several cardiac conditions^{94-95, 96-97}. It is likely that AI-supported cardiac image analysis can reveal further measures of clinical importance.

While medical imaging was not part of the BigMed charter, similarities in big data approaches between NLP and image processing lead to a collaboration with the ongoing NordForsk funded project PM Heart⁹⁸, which aims to combine genetic, imaging, and health record data to achieve more targeted treatment of ischemic heart disease in order to prevent both under- and overtreatment.

The most commonly used diagnostic tool in cardiology is the electrocardiogram (ECG), and rule-based expert systems have been built into ECG machines for decades. With the expansion of ML/AI-based diagnostics, it is probable that future ECG devices will have much better diagnostic algorithms. An interesting example is the ability of an AI algorithm to identify patients with paroxysmal (periodic) atrial fibrillation from ECG recordings of sinus rhythm; ie. the model was able to recognise changes in the ECG that are not visible to human readers. Future ECG devices, including home use and/or wrist-worn personal devices, will perhaps (or even probably) be able to recognise pre-clinical disease and alert the wearer or his/her clinician to these findings. The growth of other personal sensors and behavioural data available from personal smart devices will likely give rise to new diagnostic, prognostic and therapeutic tools not yet envisaged.

Box 14

The expectation is that genetic analyses will become even more widely available. There is also an assumption that the interpretation of single and multiple gene variants and associated epigenetic changes will allow us to better understand disease mechanisms, and thus treatment and prevention options, and provide more individualised risk estimates.

There is also great potential in environmental, personal sensor and behavioural data that is not currently being collected or analysed at scale by healthcare organisations. Supporting this are findings such as a recent registry study from Denmark showing that a large percentage of patients with SCD contacted the healthcare system in the days immediately prior to their event⁹⁹, which is highly suggestive of interpretable symptoms and signs existing as a precursor.

Expanded access to data sets consisting of genetic and out-of-hospital data combined with AI-assisted re-interpretation of contemporary and historic data from medical imaging, heart rhythm recordings, lab data and text records will likely add significantly to our understanding of disease progression and adverse event risk in patients with established heart disease. Most clinicians and researchers believe a big data approach has the potential to reduce the proportion of “blue sky” SCD and SCA events significantly.

7.2 WHOLE GENOME SEQUENCING FOR RARE DISEASES

Yngve Sejersted (Oslo University Hospital), contributions by Laura Slaughter (University of Oslo), Tony Håndstad (Oslo University Hospital)

Rare diseases are conditions that affect less than 1/2000 people in the population. Although each disease is rare, the prevalence of rare diseases is quite common. Roughly 30.000 to 100.000 Norwegians have one or more rare diseases¹⁰⁰. Approximately 80% of rare diseases have an identified genetic origin¹⁰¹ and their effects can include life-threatening conditions in the newborn and chronically debilitating or late-onset disorders.

Box 15: Monogenic disorders in critically ill newborns

Genetic conditions cause more than 20% of all infant deaths in developed countries. Identification of genetic disorders in the neonatal intensive care unit may be a challenge if the phenotype is not easily recognizable, as the clinical presentation may be atypical at an early stage, the full phenotype may not have evolved, and prematurity or other complications may mask diagnostic clues. It is therefore important to establish an iterative process involving the paediatrician and the laboratory performing the genetic analyses for the critically ill newborn, as new information about the patient can have a heavy impact on the interpretation of genetic variants.

In BigMed, we have focused on this patient group to address issues hampering an efficient workflow for genome diagnostics, mapping out efficient ways for flow of information to increase precision, speed and patient/family empowerment. Critically ill newborns may benefit from reduced turnaround time for genetic analyses, increased data sharing and reiterations of data analysis (structured phenotyping, communication solutions and multidisciplinary team evaluations). We have highlighted certain themes due to their generalizability to diagnostics of other genetic disorders.

«The Department of Medical Genetics at Oslo University Hospital (DMG) is the first diagnostic lab in the Norwegian national healthcare system to offer Whole Genome Sequencing (WGS) as part of routine care»

When newborns are admitted to a neonatal intensive care unit with a suspected genetic disorder, the traditional approach is to reach a genetic diagnosis through a prolonged and costly diagnostic workup including candidate gene analysis (single-gene testing). A faster and more accurate diagnosis may be highly beneficial for precision intervention, tailored care, accurate genetic counselling of the family, and result in a reduced workup burden.

The Department of Medical Genetics at Oslo University Hospital (DMG) is the first diagnostic lab in the Norwegian national healthcare system to offer Whole Genome Sequencing (WGS) as part of routine care. The implementation of new techniques comes with an obligation to build and share knowledge. To maintain public trust in genomic medicine for years to come, there will be a high demand for transparency, appropriate regulation, quality assurance and ethical frameworks. Robust technical infrastructure is mandatory to ensure data availability while maintaining high security. Patient empowerment should be a priority for all stakeholders.

These goals can be achieved through public/patient education and systems for dynamic consenting that will allow a tailored continuum of care. DMG is ready to pilot a secure, dynamic consent system developed by BigMed-partner Tjenester for Sensitive Data (TSD at USIT), which would allow annotation of patient data sets already stored in TSD.

7.2.1 The importance of structured phenotypic data

Vast amounts of sequence data are generated and processed in bioinformatics pipelines. Variation from the

reference genome is filtered algorithmically, but the results are manually interpreted. This process of distinguishing benign from pathogenic variations is highly labour intensive. The amount of work required for interpretation and reporting can be somewhat reduced by limiting the analysis to immediately relevant genetic findings. For this reason, routine diagnostics using whole exome and whole genome sequencing techniques are often performed with predefined gene panels to limit the amount of data for manual interpretation and reduce the number of irrelevant genetic findings. An alternative approach is using phenotypic data to guide the interpretation of a whole exome/genome through prioritising/ranking of variants within genes fitting the patient phenotype. For this to be effective and objective, structured patient phenotypic data is a prerequisite.

In addition to **objectivity** when evaluating the relevance of genetic findings in patients, structured phenotypic data are important for other reasons:

- 1) **Reproducibility** of results, quality, and patient safety.
- 2) **Sharing** (Matchmaking of variants and phenotype to identify other patients worldwide, which may provide evidence for reclassification of a variant of uncertain significance).
- 3) **Equality** (patients with similar phenotypes receive the same analysis and interpretation)
- 4) **Knowledge** (national database/variant registry).

Box 16: BigMed use case HPO through requisition form

Labs and medical device manufacturers need to examine carefully how and what data they collect, how this is used in practice, and how this integrates with the IT of the wider health system. The electronic health record (EHR) is a reputable source of information for phenotypic data. In BigMed, we have developed a proof-of-concept web application using semantic technologies to create a phenotype suggestion service for paediatricians. Using the now well-recognized Human Phenotype Ontology (HPO), biomedical ontologies, and other knowledge resources to provide additional information to clinicians to suggest related phenotypes for further work-up¹⁰². The tool prompts the clinician to be thorough in describing patient phenotypes and increases the quality of information communicated further to the laboratory. The work addressed methods to provide suggestions and user “information interaction” issues that are connected to the user interface.

We provide suggested diseases and HPO codes to clinicians based on a lookup mechanism with information about disease-HPO code associations. These are found in the HPO ontology annotation file/footnote¹⁰³. The tool under construction will provide HPO codes that point to further workup needed in the diagnostic process, however it is not a diagnosis “clinician-replacement” machine, instead aiming to augment clinician knowledge with suggestions for diseases that are very specific and/or rare. The Norwegian EHR provider DIPS has implemented the suggestion service and based their requisition form interaction on the user interface in the proof-of-concept tool.

7.2.2 A high-speed pipeline for WGS

A short turnaround time from sample submission to a complete genomic report is essential for making genomic medicine relevant to critically ill infants¹⁰⁴. Three main bottlenecks determine turnaround time of rapid WGS pipelines: 1) Library prep and sequencing, 2) Upstream bioinformatics, and 3) Prioritisation and interpretation (downstream).

DNV has addressed 1 in other projects. DMG found solutions for 2 & 3:

The upstream bioinformatic processing time of a WGS trio can be reduced from four days to three hours by using Dragen, an FPGA-based system. We automated the flow of data between different infrastructures. In addition, we developed a pipeline to identify structural variants from WGS data. Manual labour does not scale well. To increase quality and speed, DMG has implemented an in-house developed analysis tool (ELLA, allel.es) for supporting interpretation of genetic variants. ELLA is based on internationally accepted standards for variant assessment, as specified in the ACMG-AMP guidelines.

We implemented “Variant Exchange”, a tool for sharing variant interpretations between different labs. We also developed “Beacon” and “Matchmaker” services to quickly locate similar patient cases. Deployment in routine diagnostics rests on the dynamic consent system previously mentioned.



In total, this has significantly reduced the turnaround time from several months down to around 5 working days, increasing the clinical utility of the analysis when time is critical.

Currently, the time sink is waiting for the weekly sequencing run, and for missing parental samples necessary for performing trio analyses. Counterintuitive as it may seem, increasing sample volumes for whole genome sequencing will further reduce turnaround time due to more frequent runs.

Box 17: BigMed Use case ELLA

ELLA is a software tool for clinical interpretation of genetic variants that is developed and in use at DMG. Within BigMed, two projects have aimed to expand the existing capabilities of ELLA, for the benefit of our patients.

The first project involves expanding the use of high-throughput DNA sequencing technologies from the current, short genetic variants to also include larger, structural variants. This means that a near-complete picture of a patient's genetic makeup can be produced in a single sequencing run, both increasing the positive detection rate and reducing the time and cost required to reach an answer. In BigMed, we have developed a bioinformatics variant calling pipeline that now includes structural variants, and thoroughly validated it for clinical use. In addition, we have made significant strides towards presenting the results in ELLA alongside shorter variants, with the first implementation expected in coming months.

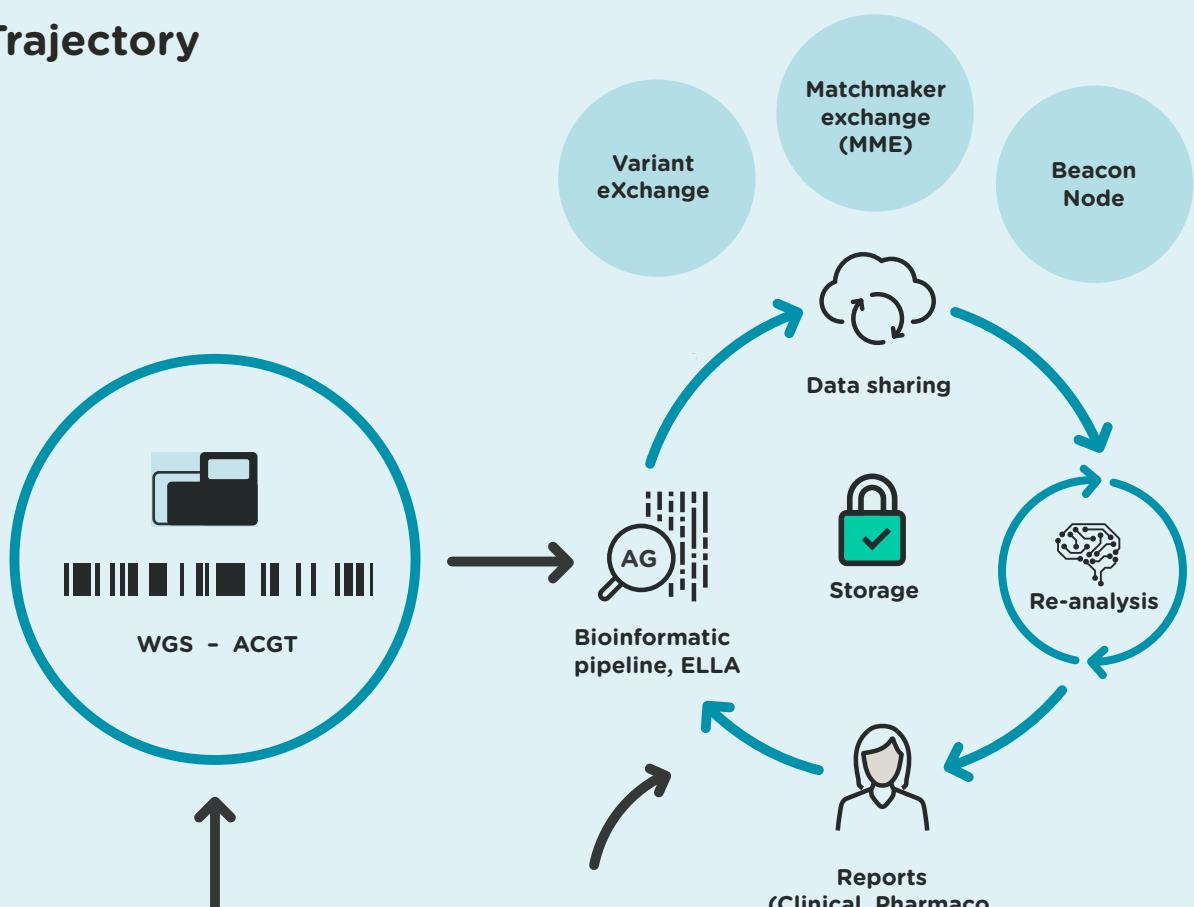
The second project involves using structured phenotypes gathered from patients for more efficient interpretation of genetic variants. In BigMed, we have identified ways of solving this problem using either simple methods that narrow down the list candidate genes and therefore the number of variants to interpret, or more advanced approaches that allow sorting of variants according to likely pathogenicity. This project will need input from phenotyping tools and the functionality is expected to result in significant efficiency gains for the high speed high throughput sequencing (HTS) pipeline.

The second project involves using structured phenotypes gathered from patients for more efficient interpretation of genetic variants. In BigMed, we have identified ways of solving this problem using either simple methods that narrow down the list candidate genes and therefore the number of variants to interpret, or more advanced approaches that allow sorting of variants according to likely pathogenicity. This project will need input from phenotyping tools and the functionality is expected to result in significant efficiency gains for the high speed HTS pipeline.

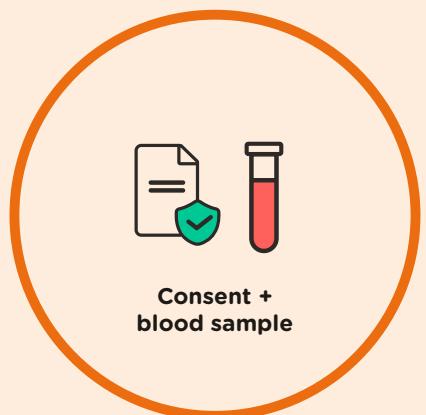
7.2.3 The future vision of rare disease care in NICU

Medical genetics services should in a few years be able to provide over-night diagnostic services for critically ill patients in NICUs and other ICUs. This puts extreme demands on performance of ICT infrastructure in public hospitals. Future diagnosis and treatment of rare disease patients is moving towards a continuum-of-care model. Using NLP to interpret the EHR free text and identify relevant genetic conditions can greatly speed this process. Clinicians will need to approve NLP suggestions for structured phenotypes. These

Lab Trajectory



Phenotype info
e-requisition



Clinical Trajectory

Figure 11: The data flow in whole genome sequencing and continuous reuse of the data.

«Patients can opt in for research, which may provide easy access for patients and families to relevant research projects and clinical trials».

NLP-suggested phenotypes can predict diagnosis before submission to a laboratory, and guide patient-specific bioinformatics and variant interpretations from patient genomes in critical disease, and also later in life.

Patient portals can support informed preference capture, gain collaborative feedback and self-reporting by patients, and enable information and communication about reanalysis and research to be conveyed. The patient/guardian is notified through a patient portal (helsenorge.no or a local/national electronic consent solution, i.e. UiO/TSD) that their physician has placed an order for genomic testing, and they are encouraged to provide their preferences (consents).

This information is used in the bioinformatics pipelines to automatically build a relevant gene panel and predict the post-test probability of candidate diagnosis from the sequencing results prior to manual interpretation of the identified genetic variants. When result interpretation is inconclusive, the paediatrician receives notification in EHR asking for supplementary information. Patients are invited to actively collaborate through self-reporting (health issues, data from wearables etc.), and are given the option to consent to the sharing of data with other collaborating laboratories to increase the probability of reaching a correct diagnosis. Genomic data is stored for documentation which also enables later reanalysis and return of new results according to the patient preferences (the patient must have opted in for recontact).

Patients can opt in for research, which may provide easy access for patients and families to relevant research projects and clinical trials. After genetic counseling (online certification/exam), patients/parents may also choose to have their data analysed

for secondary findings (ref. ACMG-list), including pharmacogenomics and information relevant for preventive healthcare, or for family planning purposes.

(Interpreted) genomic variants and structured phenotype data are stored in a searchable format in the EHR. This enables rapid identification of patients with comparable genetic variation and identification of patients with a high likelihood of having an undiagnosed genetic disorder. This requires the EHR to be searchable for similar patients and the legal grounds for such a search.

With increasing complexity and dissemination of genome analyses in medical genetics, the need for quality assured information and communication between the patient and doctor, caregiver, and lab has emerged as an important premise for state-of-the-art medical genetic services. There is a substantial patient benefit in a systematic approach to data sharing (read more in chapter 3.4) and reinterpretation. Genomic medicine is moving towards a “continuum of care model”, where surveillance and reanalysis of genomic information, and reinterpretation and reclassification of results will inform disease management, reproductive health, family planning, and preventive medicine throughout life. The individual needs and expectations will be very different depending on the context in which genomic data is produced. This context can range from predictive prenatal testing to diagnostic testing for acutely ill newborns, late onset disorders, or carrier screening in family planning. To meet these individual expectations, a transparent solution for patient information and consents should be in place. There needs to be a clear and concise contract between the patient and the lab, enabling the patient to make informed choices, and change these preferences throughout their lifespan.¹⁰⁵

Box 18: Bioinformatics

Bioinformatics is an interdisciplinary field that develops methods, analyses and software tools for understanding biological data.

In the context of HTS-based diagnostics, bioinformatics can be thought of as consisting of two main parts, variant calling and variant interpretation.

Variant calling

This is the process of identifying all the genetic variants found in a patient sample.

- / Mapping of sequencing reads to a reference genome
- / Identification of “variants” – i.e. positions where the sample deviates from the reference genome
- / Quality control to decide if sequencing has been successful

Variant calling is done by connecting several different tools in a pipeline. The execution of these pipelines requires high-performance computing. A typical genome consists of 70 GB of raw data and the processing can take up to several days depending on method. To reduce the processing time, it must be parallelised, which increases the demand for computational power, or alternatively specialised algorithms and hardware (e.g. FPGA, GPU) can be used. Fully automating all data transfer and processing is necessary to minimise turnaround time.

Assuring the quality and improving the sensitivity and precision of analyses is a continuous challenge as new software, lab protocols and reference data are developed.

Variant interpretation

This is the process of providing computational support to identify the relevance of each variant and selecting a subset for further manual evaluation.

- / Variant annotation – looking up what is known about each variant in different databases
- / Filtering and prioritisation of variants – automatically identifying the most relevant variants for further manual evaluation
- / Providing decision support for manual interpretation
- / Sharing of knowledge and data, and finding other similar patient cases

Variant interpretation is often the bottleneck of the diagnostic process due to the manual work required. Computational support is essential to identify the few relevant variants for manual evaluation from the over 4 million candidate variants in a patient genome. To effectively filter variants, information about each variant must be looked up in several databases and reference datasets. Combined with input from bioinformatic prediction tools and information about the patient’s phenotype, filtering algorithms and prioritisation tools can then reduce the number of variants for manual interpretation. The interpretation process itself is highly complex and requires decision support software. Tools for data sharing and finding of similar patient cases can increase the efficiency, quality and sensitivity of the diagnostic process.



7.3 COLORECTAL CANCER

Vegar Dagenborg (Oslo University Hospital), contributions by Jan F Nygård (The Norwegian Cancer Registry), Andrew Reiner (Oslo University Hospital), Vebjørn Arntzen (Oslo University Hospital), Vibeke Binz Vallevik (DNV), Bjørn Næss (DIPS)

Colorectal cancer (CRC) is one of the most prevalent cancers in the developed world. Many patients have metastatic disease (metastatic CRC: mCRC) requiring multiple hospital visits and treatment options, and often longer periods of surveillance. Like with many other chronic disease states, the health records of mCRC patients can contain substantial amounts of data covering a considerable time span. Efforts are being made to set up high-throughput somatic gene sequencing in Norwegian precision medicine clinical trials, building on infrastructure developed in BigMed and the number of clinically important genomic biomarkers is expected to grow in the next few years.

Box 19: Colorectal cancer (CRC)

Treatment of CRC and subsequent mCRC frequently requires multiple hospital visits, varying therapy options, increased surveillance and genetic screening.

This leads to an accumulation of a substantial amount of data. Using and understanding all of the data that is created by each patient, and accessing previous knowledge about similar patients to guide treatment can be challenging.

In BigMed we have focused on this patient group to create a dashboard that can combine each patients EHR data, and directly input relevant data from the Cancer Registry, to support clinicians in diagnostic and treatment strategies. The approximately 4500-yearly new patients in Norway suffering from CRC and mCRC may benefit from this merged-data approach, as new treatment strategies that are mapped to genetic profiles are explored and outcomes reported.

Additionally, the methodology for combining data sources, and approaches for overcoming associated challenges, can be used in clinical practice for the treatment of other diseases where combining multiple personal and public data sources could be beneficial.

With the advent of genetic analyses – both patient and tumor specific – there are opportunities for finding more optimal and individualised treatment options. However, the amount and complexity of data available to clinicians when making therapeutic decisions can result in confusion rather than clarity.

7.3.1 The future of molecular diagnostics in colorectal cancer

In BigMed, we envisioned an efficient future workflow where pertinent clinical information was easily accessible to clinicians, with particular focus on the multi-disciplinary team (MDT) setting, and where all the specialists and caregivers involved in patient treatment meet to decide on treatment options. In such a setting, there is a need for easy visualisation of complex patient cases. This involves an at-a-glance overview of previous treatments and important events in the patient history.

With genomic information, advanced analyses of both radiological and histological images, and up-to-date patient data available, each patient would be compared to similar patients and relevant guidelines presented in order to find the most optimal treatment plan. Better visualisations of treatment options would be helpful both for clinicians and in improving patient participation in care decisions.

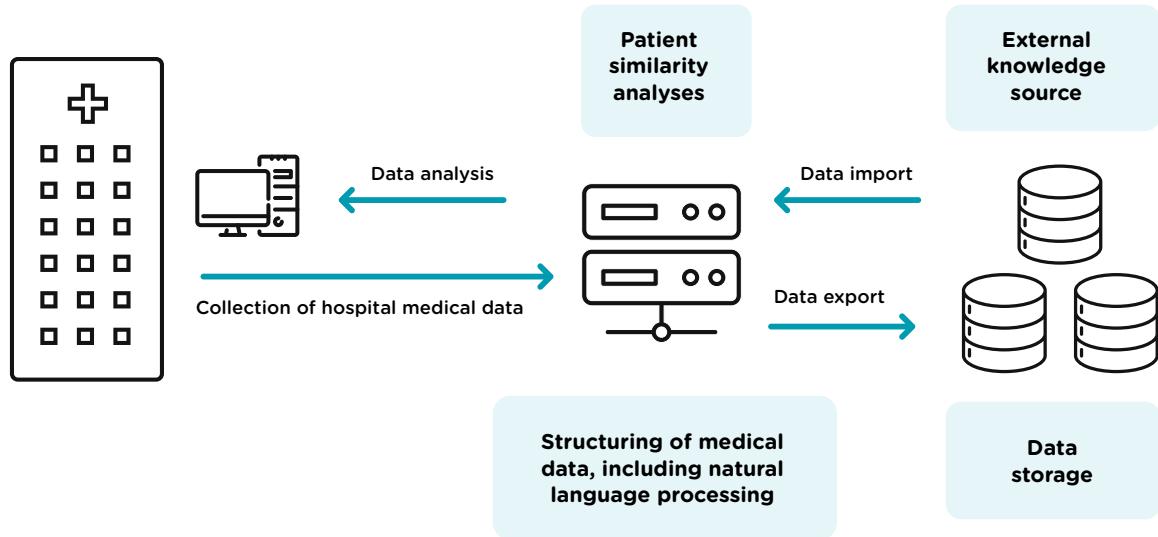


Figure 12. Data flow to allow for clinical decision support.

The mCRC work package focused on three elements of the future vision: data visualisation, improved bidirectional flow of cancer registry data, and a data model for patient similarity developed in collaboration with Oslo Centre for Biostatistics & Epidemiology.

7.3.2 Developing and deploying solutions

In BigMed, we developed a prototype clinical dashboard in the EHR, which draws up a timeline of previous treatments and hospital visits, and enables an indexing of important medical notes needed for the planned patient visit or meeting discussion. The clinical dashboard is also able to visualise data from patient similarity analyses or show important genomic variants.

This project also addressed the need for clinicians to be able to explore relevant medical literature regarding the patient's clinical and molecular characteristics, with the option of referencing and storing the source from the literature. This is a valuable way to document the body of knowledge available when the treatment plan is formed.

Large patient data registers hold considerable potential value in cancer patient treatment decisions. At present, register data is slowly accumulated through a semi-manual process, and data is analysed and published in years-long cycles. In BigMed, we hypothesised that more efficient reporting could increase register data quality and

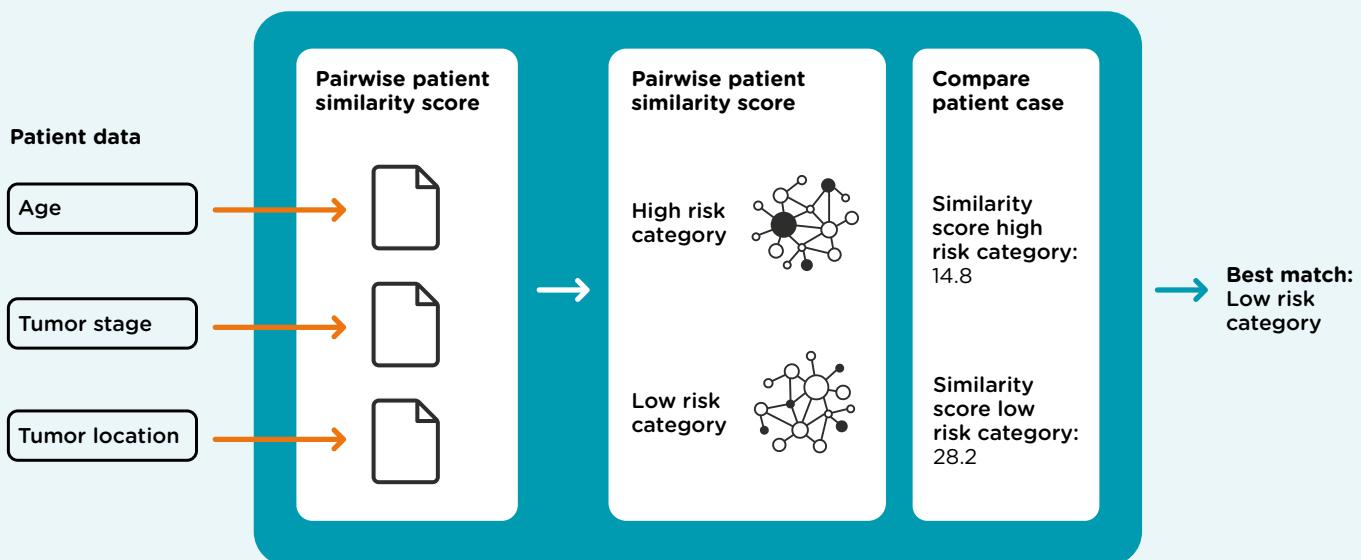
coverage, and that high quality cancer registry data could be used in real time to compare the current patient's data with similar patients in the registry, informing treatment choices (see page 102, Patient Similarity Networks).

Clinical utilisation of register data puts high demands on their accuracy. At present, the inefficient flow of clinical data from hospitals treating cancer patients is a bottle neck for registers like the Norwegian Cancer Registry. BigMed created a solution for this challenge. By harvesting data from the EHR with direct transfer from the clinical system to the Norwegian Cancer Registry, the data is more readily available. Bidirectional data exchange with registers requires communication protocols and parameter mapping and/or common information models. We made the process more efficient by defining variables as openEHR archetypes and by creating structured data in the EHR. We have also shown how algorithms can automatically retrieve cancer data from unstructured medical notes to populate structured forms.

Clinicians need both human readable text and structured data for reuse. An automatic structuring of data with natural language processing (NLP) with direct transfer to research databases, registers, or other repositories, will facilitate the implementation of the other clinical decision support tools that are part of the future paradigm of precision medicine.

BigMed use case

Patient Similarity Networks



Through BigMed, Oslo Centre for Biostatistics & Epidemiology (OCBE) developed a proof-of-concept computational tool to predict the risk of relapse in patients with colorectal cancer. This tool used a novel technique called *Patient Similarity Networks* (PSNs), which utilises known data and outcomes from treated cases to predict the outcome of new cases. The PSN approach measures pairwise similarity between all cases in the training data set on a variable-by-variable basis, and then uses this information to create integrated similarity networks, with each patient pair given an integrated similarity score. This score combines the similarity scores from the separate variables and is a measure of similarity based on input from all the modeled variables. The method then uses the integrated networks to categorize the treated cases into low-risk or high-risk groups. An optimized model is then built for each category, and a new case can then be compared to the two groups. The new case is assigned to the group that best matches it.

The model was built using netDx, a recently developed framework implemented in the R statistical language¹⁰⁶. The model data was drawn from the Oslo-Comet trial¹⁰⁷. The resulting classification system used a subset of the Comet data variables, which were selected through an iterative evaluation process that measured prediction performance. The variables selected are listed below; the classification accuracy was 76%:

- Age at liver resection
- Site of primary tumor
- Tumor stage
- Lymph node stage
- Gender
- CEA level at liver resection
- Chemotherapy prior to liver resection

The PSN approach to patient classification supports many of the principles of trust and transparency for automated systems. The model builders provide control and oversight over the data variables used to construct the model. While the model optimisation process is automated, the model builder chooses

which data variables are used, and how the variables are encoded in the model.

Data privacy can be enforced by using appropriate methods to encode the data. Using genomic data as an example, the model builder can avoid the risk of using potentially identifiable genomic profiles by aggregating the data (by using pathway data instead of gene-level data, for example) in a non-identifiable way. Likewise, patient age need not be explicitly represented in the data. In general, then, the resulting model contains encoded similarity scores, and not the primary, identifiable patient data.

PSN-based models have a high degree of interpretability and transparency. The process is visible to the model builder, who can examine which variables were selected for the model, and which weights were assigned. Additional parameters can be examined and set to affect the outcome.

More info: BigMed/Reiner 2021, Patient Similarity Networks for Precision Medicine.

8. Final reflections on the BigMed project

In 2016 the BigMed project set out with the ambitious goal of addressing the barriers to clinical implementation of precision medicine and to pave the way for big data analysis in healthcare. The project was made up of a diverse group of stakeholders from the clinic, industry, and academia. There were as many motivations and aspirations as there were people engaged in the project, yet a cooperative focus on the common vision helped move us forward in paving the way for those following us. The ambitions were met through achieving the key elements described in the following section.

8.1 ITERATIVE DEVELOPMENT IN CROSS COMPETENCE TEAMS

The work in BigMed used three model clinical areas – colorectal cancer, rare diseases and sudden cardiac death – to identify needs, develop solutions, and address issues. Although several IT solutions have been created, commercialized, and/or implemented at the end of the project, the biggest advancement has been knowledge development enabled through a meeting place for debate.

As we solved one bottleneck in the project, we would uncover several new bottlenecks. Solutions had to be constantly modified and adapted, underlining the need for iterative development and organic growth.

Developing specific solutions to barriers allowed our discussions to be more detailed and concrete. The peer-to-peer connections have been key, as has the experimentation format. Initially, we would often underestimate the barriers and misidentify the real issues. By working through the concrete cases in cross competence teams of technologists, biologists, informaticians, clinicians, geneticists, engineers, lawyers, and economists we overcame these hurdles. Some were technical and could be solved within the group, others were related to frame conditions and were elevated to decision makers or other stakeholders for

debate. The legal team established a network for lifting legal issues that needed clarifications or changes in the law, where the technical experts could be invited to share their knowledge and to highlight the issues. By Q1 2021, the effort of the legal team had already resulted in several proposed changes to the Norwegian law.

8.2 SAFEGUARDING DIGITAL HEALTHCARE

In a data driven healthcare ecosystem, different partners need to work together. In the interfaces between partners and processes there is a need for trust mechanisms on many levels, from standardized data formats and APIs, qualification of digital tools for a specific scope of use, processes to define and assure data quality and safety, and robust governance systems.

The new medical device regulations (MDR and IVDR) that is expanded to envelope software, have the intention of safeguarding patient safety through setting requirements for professional manufacturers and in-house developers. In addition to this, we need to address the safe application of new tools when they are used “off label”, through support for proper validation and testing, understanding the uncertainty in results, and the bias in data.

8.3 A NEW PERSPECTIVE ON DATA USE

Implementation of NGS gave us the opportunity to investigate an interesting data-intensive case that exemplifies the needs of a data-driven healthcare future that is increasingly tailored to the individual patient. We see genomics medicine as a model for providing insight on how data can be used to directly benefit the patient. Analysis of large data is a necessity which fuels a higher demand for IT competence integrated in clinical practice.

Today, this is a constant challenge in a system where health data, as a rule, should not be used for other purposes

than documentation of one pertaining patient. Protecting privacy is key to maintaining trust between patient and healthcare provider. Still, we need to make exceptions to this rule – for quality assurance, for statistics, for research, and for clinical genomics where we need to compare other patient's data to reach a diagnosis for the next patient. These activities are important for good quality healthcare and to ensure our patients get the best available treatment. All the exceptions to the rule make it a confusing and difficult landscape to operate in, allowing for different interpretations and an uncertainty that causes paralysis.

We believe we need to change our current perspective on data use and look to other examples, such as the ways the Danish Government uses data, for inspiration and guidance. We believe the rules should be written so that health data can be used to deliver up to date healthcare to patients in a safe way that balances and protects our privacy.

Implementation of precision medicine will move forward as the sum of many smaller decisions. A shared vision and understanding between all stakeholders is important when choosing the solutions of today. In 2012, the UK started their 100.000 genomes project with the following: "Over the next 10 years our ambition is to create the most advanced genomic healthcare system in the world, underpinned by the latest scientific advances, to deliver better health outcomes at lower cost."¹⁰⁸

We believe that a similarly clear message from the government and shared vision for Norwegian healthcare would support and accelerate the development of precision medicine.

8.4 AN ORGANISATION DESIGNED FOR CHANGE

Hospitals would benefit from development of a data strategy to facilitate the use of patient data for healthcare benefits. Considerations will include: what data do we need to start storing and in what formats? And what roles and responsibilities do we need for this?

In 2017, a BigMed stakeholder group identified legal issues as the biggest barrier to implementation of precision medicine. At the closing seminar of the project three years later, "organisation" was voted the biggest challenge when evaluating the same barriers. Many of the legal issues have been clarified, and there is a clear road ahead. The alignment of interpretation between different

organisational entities and different roles is still to be done.

The experience from our project shows a hospital system resistant to change and not constructed for continuous improvements. Incentives for change are often lacking at the decision makers level, as the unit benefiting from a change will often not be the same as the unit having to do the work or pay. "No change" will undeniably equal "no increased risk" for many parts of the system.

The ability of healthcare organisations to adjust and facilitate iterative development processes, both on a practitioner level and across the key-decision-maker landscape, are necessary for implementation of new technologies. The traditional hierarchical decision structure in public healthcare is posing a challenge to this. Roles and mandates can seem unclear for practitioners when observed from within the complex system. We saw an example of this when failure to agree on ownership of the IP address to a fibre optic cable inhibited attempts to move from manual data transfers to a cable between the hospital system and the research cloud at the University.

We need to clarify ownership of decisions and facilitate freedom of decision making at the right and competent level to avoid important processes getting lost in the maze that is the decision structure today.

8.5 BUILDING ON THE BIGMED LEGACY

The tools, solutions, and infrastructure created by BigMed, coupled with documented knowledge from the project, are fundamental to building new initiatives. Key competence and tools for NGS were brought into new precision medicine initiatives like InPreD at OUH, an infrastructure for clinical trials in molecular diagnostics in cancer.

Permanent groups and networks have been established to continue the best practices established in BigMed. The legal network has developed into Nordic Permed Law, and will continue in the role, building on a Nordic group of legal experts.

The balance between structured data and free text has been an important discussion throughout the project. Structured data are easily reused, yet clinical descriptions can rarely fit in a standard box. We see the value of both approaches and believe they need to co-exist moving forward, with each addressing different needs. Continued research on

In which area do you see the biggest barriers to the implementation of precision medicine?

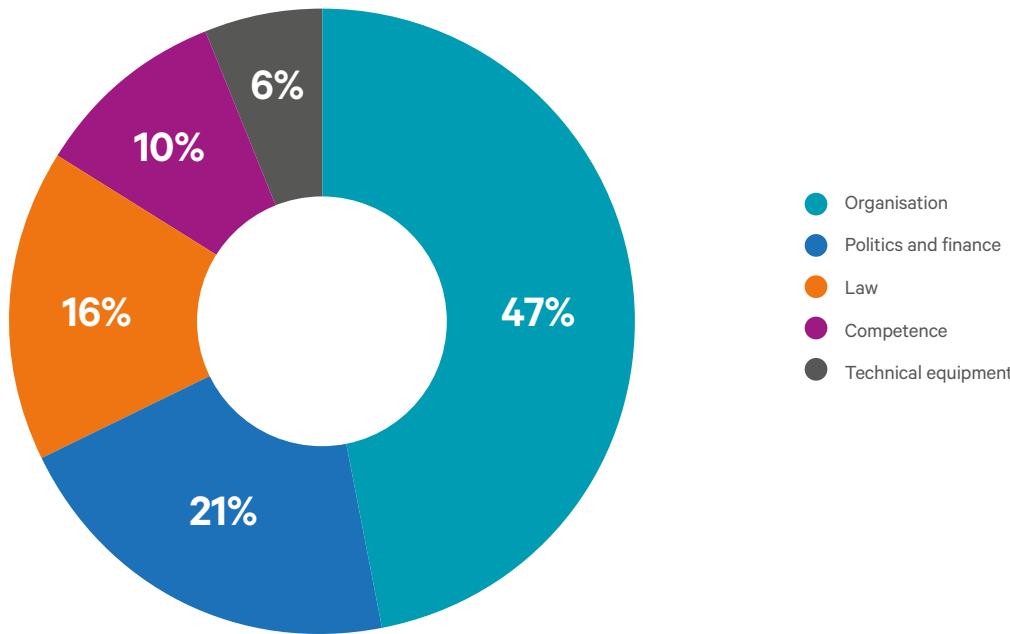


Figure 13. Biggest barriers to precision medicine in 2020. Source: BigMed 2020 conference (October 2020).

NLP and the creation of Norwegian clinical word vectors to harvest the rich context found in free text has gathered an NLP research network of clinicians, academics and industry partners from within the BigMed project and beyond.

Peer-to-peer discussion has been very valuable. Through BigMed, the ICT service provider of the region built a platform for innovation and used the process to map out and understand the future clinical needs for ICT. This kind of discussion forum between the clinical IT needs and the infrastructure provider may continue through the network of connections already made. BigMed has also been part of establishing a national competence sharing network for AI in healthcare (KIN) to ensure peer to peer discussions and for sharing the learnings from future initiatives.

With a vision of laying the foundation for clinical implementation of precision medicine, BigMed started by identifying barriers to implementation and set out to address these. We have not produced a complete recipe for how to overcome these barriers. Instead, the project has created specific solutions within the three model clinical areas, paired with an even more detailed understanding of the barriers.

Moving forward, more specialised initiatives that follow from BigMed will each carry on developing solutions and bringing important discussion topics to the stakeholders and the public. The right setup for allowing incremental changes in our system will allow us to continue working towards our common goal of implementing precision medicine in the healthcare system for the benefit of our patients.



Appendix: Deliverables overview

The list below is an attempt of mapping deliverables from the project that are confirmed per February 2021 and is not exhaustive to all the work and the knowledge generated through the cooperation between partners.

Deliverables	Result	Partners
Bioinformatics pipeline for somatic cancer: RNA sequencing and Whole exome DNA sequencing		
Somatic variant calling analysis pipelines: Gene panel DNA sequencing	Tool	OUH
List of requirements for a functional pipeline	Knowledge	OUH
Somatic genomic feature report WES	Tool	OUH
Quality control report – cancer	Tool	OUH
True negatives: CCallable CAncer IOci (CACAO)	Tool	OUH
Somatic RNA-Seq pipeline	Automated process	OUH
Bioinformatics pipeline for somatic cancer: Germ-line whole genome sequencing		
Dragen pipeline for "high speed pipeline" analysis of whole genome seq trios	Automated process	OUH
Quality control scheme (multiQC ++)	Knowledge/Process	OUH
Report on filtering of technical (false) variants	Tool	OUH
Genomics based reports and clinical reports		
Cancer Predisposition Sequencing Report (CPSR)	Tool	OUH
Cancer Predisposition Sequencing Reporter (CPSR): a flexible variant report engine for high-throughput germline screening in cancer, Sigve Nakken, Vladislav Saveliev, Oliver Hofmann, Pål Møller, Ola Myklebost, and Eivind Hovig, 2020	Paper	OUH
Personal Cancer Genome Report (PCGR)	Tool	OUH
Personal Cancer Genome Reporter: variant interpretation report for precision oncology, Sigve Nakken, Ghislain FournOUH, Daniel Vodák, Lars Birger Aasheim, Ola Myklebost and Eivind Hovig, Bioinformatics, 2018, 10: p. 1778	Paper	OUH
Clinical reporting of NGS data: A systematic Nordic collaborative, peer-reviewed benchmarking, DNV, 2018	Report	DNV
Molecular reporting and decision support in Dashboard	Demo	OUH

Deliverables	Result	Partners
Standards development and best practices for NGS		
Accuracy and efficiency of germline variant calling pipelines for human genome data, Sen Zhao, Oleg Agafonov, Abdulrahman Azab, Tomasz Stokowy and Eivind Hovig, 2020, Scientific Reports. 10	Paper	OUH, DNV
Implementing NGS-based diagnostics in cancer care: Technical and organisational factors in the Nordics, DNV, 2021		
Regulatory frameworks and quality assurance for NGS-based diagnostics	Report	DNV
Mapping of molecular diagnostics for cancer in the Nordic countries	Report	DNV
Clinical decision Support Software; Regulatory landscape in Europe from May 26th 2020, DNV, 2020.	Report	DNV
Clinical sequencing: Regulatory frameworks and quality assurance for NGS-based diagnostics, DNV, 2018	Report	DNV
Organisational maturity assessment for NGS labs	Knowledge - Recommended practice	DNV OUH
Consent for clinical genetic testing in Norway Considerations to the development of process and content, DNV, 2020	Report	DNV, OUH
High speed pipeline for NGS		
Drivers in rapid genetic diagnostics for rare diseases in infants, DNV, 2019	Report	DNV
Suggesting Reasonable Phenotypes to Clinicians, Laura Slaughter and Dag Hovland, 2019	Paper	UiO, OUH, DIPS
E-requisition with structured phenotypes	Product	DIPS, OUH, UiO
EHR models for semantic support on structured phenotyping	Demo	UiO, OUH
Mapping of ontologies for structured phenotyping	Knowledge	UiO
NICU high speed NGS pipeline (case)	Product	OUH DMG
Use of structured phenotyping in ELLA (variant classification)	Product	OUH DMG
Product	OUH DMG	
Genomic reference database version 1 and 2		
Design documents - functional and technical requirements for genomic decision support databases (TVX demo)	Knowledge	DNV

Deliverables	Result	Partners
Germline genomic medicine: A BigMed needs analysis, DNV, 2020	Report	DNV
Norvariome 1 - first version reference database (VCF collection with diagnostic samples consented for research and pseudonymized)	Demo	OUH DMG
Demonstrator and proof of concept for secure sharing of genomic data across European borders		
Variant Exchange - Sharing of interpreted genomic variants	Tool & service	DNV, OUH, Scilife Lab /Karolinska
Beacon - online sharing of single variants (GA4GH Beacon service)	Implemented & demonstrated solution	OUH, Scilife Lab /Karolinska
Matchmaker exchange - secure sharing/ querying of variants in context and with HPO phenotype (GA4GH)	Implemented & demonstrated solution	OUH, Scilife Lab /Karolinska
Risk assessment of federated sharing / matchmaker exchange	Knowledge	DNV, OUH
Variant interpretation decision support software		
Decision support for copy number variant interpretation (Ella core functionality)	Product	OUH DMG & RAD
Report on verification and validation approaches for variant interpretation decision support software (incl. ML/AI, MDR/IVDR) ("Clinical decision support software - Regulatory landscape in Europe from May 26th 2020")	Report	DNV
Infrastructure and data provisioning		
Mapping of needs for innovation zone	Knowledge	Sykehuspartner, OUH, IBM, DNV, DIPS, PubGene, Kunnskapsforlaget, UiO, Ahus
Data provisioning SP - pipeline for text extraction for EHR	Process	OUH ICT, DIPS, UiO, SP
Data provisioning Ahus - pipeline for text extraction for EHR	Process	OUH ICT, DIPS, UiO, SP
Project sandbox in TSD	Product	UiO ICT
Architecture for BIGMED Integrations (BIGMED-zone RIF)	Report (internal)	SP, OUH ICT
BigMed zone innovation platform with pipelines for data extraction (on RIF)	Product	Sykehuspartner, OUH ICT, IBM, DNV, DIPS
Risk analysis (BIGMED-zone in RIF)	Report (internal)	OUH ICT, DNV, IBM, DIPS, UiO, SP
MaxManus- automatic anonymisation for free text	Process and tool	OUH ICT, DIPS, UiO, SP
Seminar series between partners	Knowledge sharing	OUH, Sykehuspartner
Digital consent solution in TSD	Product	OUH ICT, UiO, SP

Deliverables	Result	Partners
Generic deliverables legal group		
Legal network meetings	Knowledge sharing	UiO, OUH
Establishing Nordic Permed Law	Network	UiO, OUH, DNV
BigMed legal conference 1 (juni 2018)	Conference	UiO, OUH
BigMed legal conference 2 (nov 2019)	Conference	UiO, OUH, DNV
Dialogue and discussions with regulators	Knowledge	UiO, OUH
Regulation of NGS		
Report: BigMed-konferansen 2019: Rettslige reguleringer av persontilpasset medisin, BigMed, 2019	Report	OUH, UiO
Persontilpasset medisin - rettslige perspektiver, Anne Kjersti Befring, 2019	Book	UiO
Artikkel om genetiske varianters rettslige stilling. Enkeltstående fortolkede genetiske varianter er anonyme.	Paper	UiO, OUH
AI for clinical use and secondary use of data		
Master thesis by Gjertrud Bøhn Magelli: Regulation of AI in Healthcare	Report	UiO
Kunstig intelligens og big data i helsesektoren, Anne Kjersti Befring and Inger-Johanne Sand, 2020	Book	UiO
Blog on regulating PM in Dagens medisin	Blog	UiO
Risk calculator for sudden cardiac death (SCD)	Tool	OUH, DIPS
ML: Automatic echo measurements for input to calculator (machine learning)	Demo of method	Inmeta /OUH
Mapping of clinical needs based on design thinking	Report (internal)	IBM, OUH
Product strategy based on AI	Knowledge	DIPS
Natural Language Processing (NLP)		
NLP and early risk identification		
NLP: Identification of patients at risk for SCD from electronic medical journal; identification of "syncope" to populate risk calculator.	Demo of method	Ahus & IFI LTG
Building a Norwegian Lexical Resource for Medical Entity Recognition, Ildikó Pilán, Pål H. Brekke, Lilja Øvrelid, 2020	Paper	UiO, OUH

Deliverables	Result	Partners
NLP: Pedigree tool- Extraction from free text - family relations relevant for medical condition	Demo & product	Ahus & IFI LTG, OUH
Dahl, F.A., Rama, T., Hurlen, P. et al. Neural classification of Norwegian radiology reports: using NLP to detect findings in CT-scans of children. BMC Med Inform Decis Mak 21, 84 (2021).	Paper	Ahus & IFI LTG
Iterative development of family history annotation guidelines using a synthetic corpus of clinical text, Taraka Rama, Pål H. Brekke, Øystein Nytrø and Lilja Øvrelid, ACL Anthology: Proceedings of the Ninth International Workshop on Health Text Mining and Information Analysis; 2018. 18: p. 111	Paper	UiO, NTNU, OUH
Synthetic data set of clinical texts for NLP	Available online	OUH
Word vectors	Demo & applications	Ahus
Data access and data flow for colorectal cancer		
Dashboard with timeline in the Electronic Health Register (EHR), DIPS arena	Tool	DIPS, OUH
Data extraction from registries to populate patient timeline – previOUH disease incidents.	Knowledge	CR, DIPS, OUH
Open EHR Archetypes definition for structured report to cancer registry, saving valuable clinician time and improving valuable data gathering	Knowledge	OUH ICT, OUH, DIPS
Automatic reporting tool from EHR to the cancer registry	Tool	DIPS, OUH, Cancer registry
Research data capture directly in EHR – (ProtheCT)	Clinical test	DIPS, OUH
Text mining from electronic health record for automatic population of Dashboard - NLP testcase for WatsonExplorer	Demo	IBM, OUH
DPIA for Access to data for Machine learning	Example and template	DNV, OUH
Patient similarity		
Patient similarity classifier predictor	Tool	OUH OCBE
Patient similarity networks for precision medicine, OUH, 2021	Report	OUH OCBE
Boolean search for research articles	Tool	PubGene, OUH
Indexed English dictionary for research articles	Tool	PubGene, OUH
Personalised statistics tool based on cancer registry data	Application	Cancer registry, OUH

Deliverables	Result	Partners
Dissemination		
Big data management for the precise treatment of three patient groups, BigMed, 2018	Report	DNV, all partners
Podcast series: Clinical implementation of precision medicine (Norwegian)	Podcast	All partners
Oslo Health hackathon 2019 for cancer	Event	DNV, Acando, Norwegian Cancer society, OUH
Establishing KIN - AI in Norwegian healthcare (Kunstig Intelligens i Norsk helsetjeneste)	Network	DNV, OUH, UiO, Health regions.
Digital webinar: BigMed at EHiN 2020, demonstrating practical solutions	Recording	Sykehuspartner, DNV, OUH, DIPS, Cancer registry
Digital webinar: Precision Medicine: A Health Economics Perspective	Recording	UiO, OUH
Digital webinar: NLP in Health – What is Possible, Useful and Allowed?	Recording	Ahus, UiO
Digital webinar: BigMed-konferansen 2020: Veien til presisjonsmedisin	Recording	All partners
Digital webinar: Federated Analytics of Health Data	Recording	DNV, Cancer registry
Digital webinar: Real-World Data, digitalisation and decentralisation of future clinical trials	Recording	DNV, OUH

Appendix: Consortium partners

Akershus University Hospital

The analysis department provides data extraction services for all purposes in the hospital. In NLP projects, we provide both infrastructure, preprocessing of data and analyzes. Works closely with research environments at the hospital with, for example, large clinical text corpus.

Lars Åge Møgster,
Head of Analytics
ahus.no

DIPS

DIPS aims to enable data collected in the DIPS record, including genomics data, to be processed and analysed to support precision medicine under the BigMed project.

Liv Bollvåg, Head of Research
dips.com/no

DNV

DNV is the independent expert in risk management and assurance, operating in more than 100 countries. Through its broad experience and deep expertise DNV advances safety and sustainable performance, sets industry benchmarks, and inspires and invents solutions. DNV is a founding partner of the Nordic Alliance for Clinical Genomics.

Stephen McAdam,
Director of Digital Health
dnv.com

Norwegian Armed Forces

Joint medical services. Supporting BigMed activities by learning from current initiatives in Norwegian military medicine in using big data methodologies for medical challenges in cold weather and arctic operations.

Hjelle, Brigader
forsvaret.no

The Cancer Registry of Norway

Providing decision support tools based on data in the clinical cancer registries

Jan F. Nygård
krefregisteret.no

IBM

IBM's strategy is to be an essential partner to organisations that want to digitally transform their business model using AI technologies based on a hybrid cloud platform. Using data as the driver for transformation, IBM delivers its Watson AI services and technologies to extract value from any type of data.

Loek Vredenberg, CTO IBM Norge
ibm.com/no-en

Karolinska Institutet / SciLifeLab

The clinical genomics facility provides a dedicated research infrastructure for projects utilising massively parallel / next generation sequencing technologies. The facility serves as a competence centre assisting the translation of genomics-based tools to routine clinical care.

Valtteri Wirta, Director of
Clinical Genomics Facility
scilifelab.se

Kunnskapsforlaget,

Gyldendahl Norsk Forlag

As part of Gyldendahl, Kunnskapsforlaget provides tools for extracting knowledge from unstructured and structured texts, and to resolve ambiguity and prepare texts for language technology analysis and output.

kunnskapsforlaget.no

The Norwegian Heart and Lung Patient Organisation (LHL)

LHL closely follows up patients with heart and lung disease and their relatives before, during and after treatment through research, political influence, public awareness, and professional treatment.

Are Helseth, Medical Director
lhl.no

Norwegian Cancer Society (Kreftforeningen)

The Norwegian Cancer Society (NCS) is one of the largest organisations in Norway representing the voices of those affected by cancer.

NCS works continuously to improve society's attitude to the prevention and treatment of cancer. We fight cancer locally, nationally and globally through research and preventive measures, information, support, advice and lobbying.

Ingrid Stenstavold Ross,
Secretary General
kreftforeningen.no

Norway Health Tech

An organisation to support and accelerate the development of new medical technology and e-health products, services and innovative solutions for the Norwegian and global health markets.

Kathrine Myhre, CEO
norwayhealthtech.com

Norwegian University of Science and Technology (NTNU)

Faculty of Information and Technology and Electrical Engineering, Department of Computer Science

Conducting research in the fields such as artificial intelligence, big data, computer architecture, computer graphics, computer security, databases, human computer interaction, information systems, operating systems, and software engineering.

Øystein Nytrø, Associate Professor
ntnu.edu/idi

Oslo University Hospital, Department of Medical Genetics (DMG)

DMG is the largest medical genetic department in Norway studying hereditary diseases and performing research on genetic causes of disease.

Dag Erik Undlien,
Professor and Head of DMG
med.uio.no/klinmed/om/organisasjon/klinikker/laboratoriemedisin/medisinsk-genetikk/index.html

Oslo University Hospital, Institute for Cancer Research, Department of Tumour Biology

Engaged in basic and translational cancer research all the way from experimental research on model organisms and human materials to clinical trials for advanced medical research across sciences, including

genomics and bioinformatics, for individualised diagnostics and treatments for cancer patients.

Eivind Hovig, Professor
ous-research.no/tumorbiology

Oslo University Hospital, the Intervention Centre

Multidisciplinary centre to develop new treatment methods for patients, for example, new algorithms for processing and understanding complex, large amount of data for high precision diagnosis, treatment, and follow-up.

Erik Fosse, Head of the Intervention Centre
ous-research.no/interventionalcentre

Oslo University Hospital, Legal Department to Oslo University Hospital CEO

OUS legal department dedicates resources to engage in research to identify and address the legal issues that need changing to meet the future of healthcare.

Randi Borgen, Legal director

Oslo University Hospital, ICT

Working on a strategic level to ensure optimal environment in regard to technical, software and workflow issues, suitable to support innovation and research in advanced medical treatment.

Sissel Jor, Section Manager

Oslo University Hospital, OCBE

Develops and applies statistical and machine learning methodology and algorithms to (i) extract understanding from clinical and genomic data and (ii) make predictions of future events / conditions. Biomarker discovery.

Patient safety monitoring based on electronic health records.

Arnoldo Frigessi, Director of OCBE
med.uio.no/imb/english/research/centres/ocbe

Pubgene

Patented biomedical research text mining (Coremine) for mining evidence for better diagnoses and possible treatments for every single patient.

Odd Arild Lehne, CEO
pubgene.com

Sykehuspartner

Shared IT services provider for hospitals in the South-Eastern health region of Norway.

Alia Zaka, Head of Development and Innovation, Customer services
sykehuspartner.no

The Norwegian Association for Children with Congenital Heart Disease (Foreningen for Hjertesyke Barn)

The Foundation aims to improve the overall quality of life of children with heart defects by supporting research addressing physical, psychological, social and spiritual aspects towards an integrated approach.

Marte Jystad, Special Adviser
ffhb.no

University of Oslo, The Faculty of Law

The faculty of law has established courses on precision medicine and dedicates resources to engage in research to identify and address the legal issues that need changing to meet the future of healthcare.

Anne Kjersti Befring,
Assistant Professor
jus.uio.no

**University of Oslo, Institute
of Health and Society**

Estimating cost-effectiveness
of precision medicine, which
a specific focus on small
non-randomised control trials.

Eline Aas, Associate Professor
med.uio.no/helsam

**University of Oslo, Services for
Sensitive Data (TSD)**

TSD is an e-infrastructure which
meets the strict requirements
of the law for the treatment and
storage of sensitive biomedical (and
other sensitive) research data.

Gard Thomassen, Assistant Director,
Department for Research Computing
uios.no/tjenester/it/forskning/sensitiv

**University of Oslo, Department
of Informatics, Language
Technology Group (LTG)**

Language Technology comprises
theoretical and applied informatics
that seeks to enable computers to
'make sense' of human language.
LTG performs data-driven
linguistic analysis of text using
machine learning and HPC.

Lilja Øvreliid, Associate Professor
mn.uio.no/ifi/forskning/grupper/ltg

**University of Oslo, Department
of Informatics, Logic and
Intelligent Data (LogID)**

The work in LogID is based on
well-established methods from logic,
which they extend and enhance to
tackle tomorrow's challenges in fields
like Semantic Web and Big Data.

Arild Waaler, Professor
mn.uio.no/ifi/english

Definitions

Artificial Intelligence

Artificial intelligence (AI) makes it possible for machines run by computers to learn from data and interaction with an environment, to adapt to new inputs and perform human-like tasks. Computers run machine learning algorithms to make decisions. Statistics delivers the understanding of risk and uncertainty. Robotics allows to translate decisions into physical actions. There is research ongoing that aims to produce algorithms that learn what to learn, and in this way express more autonomy.

Big Data

Big data describes data sets which are larger than usual for the domain of action. Their physical dimension is therefore variable. In Genomics, for example, big data might be of some GB; in sensor data we easily speak about TB and more. Big data are both designed and collected for a specific purpose or repurposed. The noise level of big data is varying, but can be very high, with elements of bias and informative missingness.

Clinical decision support software

A clinical decision support system (CDSS) is a health information technology system that is designed to provide physicians and other health professionals with clinical decision support (CDS), that is, assistance with clinical decision-making tasks.

Digital ecosystem

A digital ecosystem is a group of interconnected information technology resources that can function as a unit. Economic definition: «Two or n-sided markets that grow by network effects, not on traditional economies of scale». (Parker, van Alstyne, Choudary 2014). Technical definition: «A digital ecosystem consists of the collection of platform and the apps specific to it». (Tiwana, 2013).

Federated data sharing

Also known as: Decentralised data sharing, as opposed to centralised sharing. Traditionally, data resources have been stored, managed and processed within a centralised server. A decentralised approach utilises a distributed architecture of multiple independent machines that cooperate on storage, management and processing of data.

Federated network

A federated network is based on multiple networks or locations operating under agreed protocols with shared resources.

Genomic medicine

Use of genomic information generated by exome/genome sequencing as part of clinical care for diagnostic or therapeutic decision-making

Machine learning

Machine learning is a method of data analysis that automates model building for the purpose of prediction, classification, estimation and decision making. ML produces algorithms which can learn from data, identify patterns in data that allow to perform prediction, classification, estimation tasks and decision making. Statistics is very closely related to machine learning, but focuses more on uncertainty quantification and explainable models. ML is one of the pillars of AI.

MDR and IVDR

The EU mandated medical device and in-vitro device regulation will come into effect in May 2021 and May 2022. The MDR is the ruling regulation, with the IVDR aligning on MDR principles for the regulation of in-vitro diagnostics devices (including software). The MDR includes a wider inclusion definition of medical devices than in the earlier MDD: to include all devices which provide information used in medical prediction or prognosis of human beings (article 2(1)), specific caveats addressing technologies created within health institutions 'in-house' (article 5(5)), and re-classification of software as higher risk (according to rule 11 – Annex VIII).

OpenEHR

OpenEHR is a community driven non-profit IT platform for e-health, consisting of specifications, clinical models and software that can be used to create standards for building and storing information in healthcare and interoperability solutions for healthcare.

Precision medicine

Precision medicine is a medical model that proposes the customisation of healthcare, with medical decisions, practices, or products being tailored to the individual characteristics of a patient. Customisation of medicine to individuals means taking into account all relevant sources of information, from biomarkers like a person's genes, to social attributes.

RWD/RWE

Real world data can be defined as routinely collected data relating to a patient's health status or the delivery of health care from a variety of sources other than traditional clinical trials. Real world evidence is the clinical evidence that can be derived from analysis of this data.

Abbreviations

Ahus	Akershus University Hospital	LSTM	Long-short term memory
AI	Artificial Intelligence	MDCG	Medical Device Coordination Group
AI HLEG	The High-Level Expert Group on AI	MDD	Medical Device Directive
AIMD	Active implantable medical devices	MDR	Medical device regulations
AI-MDSW	Artificial Intelligence- Medical Device Software	MDSW	Medical device software
ALTAI	Assessment List for Trustworthy Artificial Intelligence	ML	Machine learning
API	Application programming interface	MRI	Magnetic resonance imaging
CDS	Clinical decision support	NACG	Nordic Alliance for Clinical Genetics
CEA	Cost-effectiveness analysis	NEM	National ethics committee
CNN	Convolutional neural network	NGS	Next generation sequencing
CPSR	Cancer Predisposition Sequencing Reporter	NICU	Neonatal Intensive Care Unit
CRC	Colorectal Cancer	NLP	Natural Language Processing
CT	Computed tomography	NOMA	Norwegian Medicines Agency
CUA	Cost-utility analysis	NPV	Net present value
DMG	Department of Medical Genetics at Oslo University Hospital	NTRK	Neurotrophic Tyrosine Receptor Kinase
DNA	Deoxyribonucleic acid	OCBE	Oslo Centre for Biostatistics & Epidemiology
DRUP	Drug Rediscovery Protocol	OUH	Oslo University Hospital
ECG	Electrocardiogram	PCAWG	Pan-Cancer Analysis of Whole Genomes
EHR	Electronic Health Records	PM	Precision Medicine
EMA	European Medicines Agency	PROMS	Patient Reported Outcome Measures
FHIR	Fast Healthcare Interoperability Resources	PREMS	Patient Reported Experience Measures
GA4GH	Global Alliance for Genomics and Health	PSA	Probabilistic Sensitivity Analysis
GDPR	General Data Protection Regulations	PSN	Patient Similarity Networks
HCLS	Health Care and Life Sciences	PVO	Privacy ombudsman
HCM	Hypertrophic Cardiomyopathy	QALYs	Quality-adjusted life-years
HCN	Health Compute Norway	QC	Quality control
HCPs	Health care providers	RCN	Research Council Norway
HL7	Health Level 7	RCTs	Randomised clinical trials
HOD	Health and Care Services	REC	Regional ethics committee
HPC	High Performance Computing	RNA	Ribonucleic acid
HPO	Human Phenotype Ontology	PVO	Real World Data
HTA	Health technology assessments	RWE	Real World Evidence
HTS	High Throughput Sequencing or screening	rWGS	rapid whole genome sequencing
IAM	Identity and Access Management	SCA	Sudden Cardiac Arrest
ICD-10	International Classification of Diseases	SCD	Sudden Cardiac Death
ICER	Incremental cost-effectiveness ratio	SEN	South-Eastern Norway health region
ICGC	International Cancer Genome Consortium	SIEM	Security information
ICT	Information and communications technology	SNP	and event management
InPreD	Infrastructure for precision diagnostics	SNOMED CT	Systematically organized computer processable collection of clinical terms
IVDD	In Vitro Device Directive	SVM	Single Nucleotide polymorphism
IVDR	In vitro device regulations	TSD	Support Vector Machines
KIN	National competence sharing network for AI in healthcare	UIO	Services for Sensitive Data
LDT	Lab developed test	WES	University of Oslo
LIMS	Laboratory Information Management System	WGS	Whole exome sequencing
		SW	Whole genome sequencing
		TAT	Software
			Turnaround time

References

- 1 Oslo University Hospital, DNV et al, *Big data management for the precise treatment of three patient groups*. 2018.
- 2 European Council, *Council conclusions on personalised medicine for patients*. 2015.
- 3 National Research Council (US), *Towards Precision Medicine – Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease*, 2011. [Cited 2021; available from: <https://pubmed.ncbi.nlm.nih.gov/22536618/>]
- 4 Helse- og omsorgsdepartementet: Høring [cited 2020, available from: <https://www.regjeringen.no/no/dokumenter/bruk-av-helseopplysninger-for-a-lette-samarbeid-laring-og-bruk-av-kunstig-intelligens/id2740599/>]
- 5 Perrault, R., Shoham, Y., Brynjolfsson, E., et al, AI Index Steering Committee, *Human-Centered AI Institute*, Stanford University, *The AI Index 2019 Annual Report*. 2019.
- 6 Nilssen, Y., et al., *Lung cancer treatment is influenced by income, education, age and place of residence in a country with universal health coverage*. *Cancer Epidemiol*35. 2015. 138: p. 1
- 7 European Commission, *High-Level Expert Group on Artificial Intelligence, Ethics guidelines for trustworthy AI*. 2019.
- 8 Beacon network: A global search engine for genetic mutations [cited 2021, available from www.beacon-network.org]
- 9 Matchmaker Exchange: *Genomic discovery* [cited 2020, available from www.matchmakerexchange.org]
- 10 Ewels, Phil, *MegaQC* [cited 2021, available from <https://megaqc.info>]
- 11 Haarbrandt B, Tute E, Marschollek M. *Automated population of an i2b2 clinical data warehouse from an openEHR-based data repository*. *J Biomed Inform*. 2016.
- 12 van der Wouden, C., et al, *Implementing Pharmacogenomics in Europe: Design and Implementation Strategy of the Ubiquitous Pharmacogenomics Consortium*. *Clinical Pharmacology and Therapeutics*. 2017. 101: p. 341
- 13 The Norwegian Directorate of health, *National strategy for personalised medicine in healthcare 2017-2021*. 2016.
- 14 Oslo University Hospital, DNV et al, *Big data management for the precise treatment of three patient groups*. 2018.
- 15 Oslo University Hospital, DNV et al, *Big data management for the precise treatment of three patient groups*. 2018.
- 16 BigMed, *Implementing NGS-based diagnostics in cancer care: Technical and organisational factors in the Nordics*. 2020.
- 17 Denning, S., *The Age of Agile: How Smart Companies Are Transforming the Way Work Gets Done*. Amaryllis Business. 2018.
- 18 Porter, Michael, *Redefining Health Care Creating Value-Based Competition on Results*, 2006.
- 19 U.S. Food and Drug Administration Guidance document, *Master Protocols: Efficient Clinical Trial Design Strategies to Expedite Development of Oncology Drugs and Biologics*. 2018.
- 20 Park, J., et al., *Systematic review of basket trials, umbrella trials, and platform trials: a landscape analysis of master protocols*. *Trials*. 2019. 20: p. 572
- 21 European Medicines Agency, *Vitrakvi*, [cited 2019; available from: <https://www.ema.europa.eu/en/medicines/human/EPAR/vitrakvi>]
- 22 European Medicines Agency, *Rozlytrek*, [cited 2020; available from: <https://www.ema.europa.eu/en/medicines/hum2an/EPAR/rozlytrek>]
- 23 van der Velden, D., et al, *The drug rediscovery protocol facilitates the expanded use of existing anticancer drugs*. *Nature*. 2019. 574: p. 127
- 24 van der Velden, D., et al, *The drug rediscovery protocol facilitates the expanded use of existing anticancer drugs*. *Nature*. 2019. 574: p. 127
- 25 Dagens medisin, *Vil starte nasjonal studie på kreftpasienter*. [cited 2020; available from: <https://www.dagensmedisin.no/artikler/2020/02/12/presisjonsmedisin-vil-starte-nasjonal-studie-pa-kreftpasienter/>]
- 26 U.S. Food and Drug Administration Guidance document, *Master Protocols: Efficient Clinical Trial Design Strategies to Expedite Development of Oncology Drugs and Biologics*. 2018.
- 27 Inan, O., et al, *Digitizing clinical trials*. *NPJ Digital Medicine*. 2020. 3: 101
- 28 Ali, Z., et al, *Virtual Clinical Trials: Perspectives in Dermatology*. *Dermatology*. 2020. 236: p. 375
- 29 U.S. Food and Drug Administration Guidance document, *Master Protocols: Efficient Clinical Trial Design Strategies to Expedite Development of Oncology Drugs and Biologics*. 2018.
- 30 European Medicines Agency, *Update on Real World Evidence Data Collection*. [cited 2016; available from: https://ec.europa.eu/health/sites/health/files/files/committee/stamp/2016_03_stamp4/4_real_world_evidence_ema_presentation.pdf]
- 31 European Medicines Agency, *EMA Regulatory Science to 2025*. 2020.
- 32 European Medicines Agency, *Guideline on registry-based studies – launch of public consultation*. [cited 2020; available from: <https://www.ema.europa.eu/en/news/guideline-registry-based-studies-launch-public-consultation>]
- 33 Menon economics. [cited 2017, available from: <https://www.menon.no/publication/verdien-kliniske-studier-norge/>]
- 34 Sekretariatet for Nye Metoder, *Revurdering og utfasing*, 2020, [cited 2021, available from: <https://hyemetoder.no/om-systemet/revurdering-og-utfasing/>]
- 35 Ministry of Social Affairs, *Retningslinjer for prioriteringer innen norsk helsetjeneste*, NOU1987: 23
- 35 Ministry of Care and Health Services, *Prioritering på ny – gjennomgang av retningslinjer for prioriteringer innen norsk helsetjeneste*, NOU 1997: 18
- 36 Ministry of Care and Health Services, *Åpent og rettfærdig – prioriteringer i helsetjenesten*, NOU 2014: 12.
- 37 The Norwegian Parliament, *Stortingsmelding 34 (2015–2016)*. Verdier i pasientenes helsetjeneste – melding om prioritering. 2016.
- 38 Ministry of Care and Health Services, Magnussen, J., et al, *På ramme alvor – alvorlighet og prioritering*. 2015.
- 39 The Norwegian Parliament, *Lovvedtak. Forskrift om stønad til deknинг av utgifter til viktige legemidler mv. (Blåreseptforskriften)*. 2007.
- 40 The Norwegian Directorate of Health, *Nasjonalt råd for prioritering*. 2008.
- 41 The Norwegian Directorate of Health, *The National System for Managed Introduction of New Health Technologies within the Specialist Health Service in Norway*. 2013.
- 42 The Norwegian Directorate of Health, *Nasjonalt råd for prioritering*. 2008.
- 43 Drummond, M., et al, *Methods for Economic Evaluation in Health Care Programmes*. Oxford University Press. 2005. 3rd Edition.
- 44 Neumann, P., et al, *Cost-effectiveness in health and medicine*. Oxford University Press. 2016.
- 45 Befring, Anne Kjersti, *Personalised medicine – legal perspectives*, Oslo: Gyldendal. 2019.
- 46 Befring and Sand: *Kunstig intelligens og big data i helsesektoren* (2020) <https://www.gyldendal.no/faglitteratur/jus/juridiske-fag/kunstig-intelligens-og-big-data-i-helsesektoren/p-825763-no/>
- 47 Befring, Anne Kjersti, *Personalised medicine – legal perspectives*, Oslo: Gyldendal. 2019.
- 48 Befring, Anne Kjersti, «Samtykkets faktiske og rettslige begrensninger» (Eng.: «The actual and legal limitations of consent») in *Kunstig intelligens og big data i helsesektoren – rettslige perspektiver* (Eng.: «Artificial Intelligence and big data in the health sector – legal perspectives»), Anne Kjersti Befring and Inger Johanne Sand (ed.), Oslo: Gyldendal, 2020, chapter 14.
- 49 Befring and Sand (eds.) (2020), Ch. 2 and 14 in *Kunstig intelligens og big data i helsesektoren – rettslige perspektiver* (Eng.: Artificial Intelligence and big data in the health sector – legal perspectives), Anne Kjersti Befring and Inger Johanne Sand (ed.), Oslo: Gyldendal. 2020.
- 50 Befring, Anne Kjersti, *Personalised medicine – legal perspectives*, Oslo: Gyldendal. 2019.
- 51 Befring, Anne Kjersti, *Personalised medicine – legal perspectives*, Oslo: Gyldendal. 2019.
- 52 Befring, Anne Kjersti, *Personalised medicine – legal perspectives*, Oslo: Gyldendal, (2019), ch. 1.1-1.2 and ch. 3
- 53 Befring, Anne Kjersti and Oda Bakken, *Overgangen fra personopplysninger til genetisk kunnskap* (Eng.: From personal data to genetic knowledge) in *Kunstig intelligens og big data i helsesektoren – rettslige perspektiver* (Eng.: Artificial Intelligence and big data in the health sector – legal perspectives), Anne Kjersti Befring and Inger Johanne Sand (ed.), Oslo: Gyldendal, 2020, chapter 8.2.
- 54 The human genome was first mapped in the early 2000s in the project: *The Human Genome Project*, see e.g. Befring (2019) Ch. 3. For further information, visit National Institute of Health, *The Human Genome Project*, [cited 2019; available from: <https://www.genome.gov/human-genome-project>].
- 55 For further information about the reference genome, see *The Genome Reference Consortium*, [cited 2019; available from: <https://www.ncbi.nlm.nih.gov/grc>] and *The National Human Genome Research Institute* [available at: <https://www.genome.gov/Funded-Programs-Projects/Human-Genome-Reference-Program>].
- 56 NIH, *The human genome project*, cited 2021, available from <https://www.genome.gov/human-genome-project>
- 57 One of the cases where patients have received improper health care is discussed in Befring, Anne Kjersti, *Personalised medicine – legal perspectives*, Oslo: Gyldendal, 2019. Ch. 8.3.
- 58 Cf. art. 4 of UNESCO's *International Declaration on Human Genetic Data* of 16 October 2003.
- 59 Raze, S., and Hall, A., *Genomic medicine and data sharing*. *British Medical Bulletin*, 2017. 123: p. 35
- 60 Wright, Caroline F. et al., “Genomic variant sharing: a position statement” (2019) *Welcome Open Research*, Vol. 4 (22), page 35-45.
- 61 Prop. 56 LS (2017-2018) Ch. 321.
- 62 Health Personnel Act of 2 July 1999 no. 64, Ch. 5. Befring (2019) Ch. 12.
- 63 Personal Data Act of 15 June 2018 no. 38.

- 64 Cf. Act of 20 June 2014 no. 42 relating to processing of health data for the purpose of providing healthcare [the Patient Record Act (No.: pasientjournalloven)] § 2, letter b, the Act of 20 June 2014 no. 43 relating to health registries and processing of health data [The Health Register Act (No.: helseregisterloven)] § 2, letter a, and the Act of 20 June 2008 no. 44 relating to medical and health research [the Health Research Act (No.: helseforskningsloven)] § 4, letter d.
- 65 The Health Personnel Act the Act of 2 July 1999 no. 63 relating to patient and user rights [the Patient Rights Act (No.: pasient- og brukerrettighetsloven)], the Patient Record Act, the Act of 2 July 1999 no. 61 relating to specialist health services and more [the Specialist Health Services Act (No.: spesialisthelsetjenesteloven), the Health Research Act and the Health Register Act.
- 66 See e.g. the Patient Rights Act §§ 1-1 and 4-1.
- 67 See e.g. the Health Personnel Act §§ 4, 16 and 21 and the Health Research Act § 5. This is also discussed in Befring (2019) Ch. 8.2 and 8.3, among other, based on a decision from the County Governor of Oslo and Viken (previously the County Governor of Oslo and Akershus) dated 22 May 2018.
- 68 Befring: Persontilpasset medisin (2019) <https://www.gyldendal.no/faglitteratur/jus/juridiske-fag/persontilpasset-medisin/p-545345-no/>
- 69 Human Rights Act of 21 May 1999 no. 30 § 3, cf. § 2.
- 70 Council of Europe, Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine: Convention on Human Rights and Biomedicine, Oviedo 1997.
- 71 The Health Personnel Act (No.: Helsepersonelloven) of 2 July 1999, the Patient Rights Act (No.: Pasient – og brukerrettighetsloven) of 2 July 1999 no. 63, the Patient Record Act (No.: Pasientjournalloven) of 20 June 2014 no. 42, the Specialist Health Services Act (No.: Spesialisthelsetjenesteloven) of 2 July 1999 no. 61, the Health Research Act (No.: Helseforskningsloven) of 20 June 2008 no. 44 and the Health Register Act (No.: Helseregisterloven) of 20 June 2014 no. 43.
- 72 See among other The National Research Ethics Committee for Medicine and Health Sciences (NEM), "Retningslinjer for bruk av genetiske undersøkelser av mennesker i medisinsk og helsefaglig forskning" (Eng.: "Guidelines for the use of genetic examinations of people in medical and health research"), 31 March 2016, Oslo, page 7-8, and Norwegian Directorate of Health, National strategy for personalised medicine in the health service 2017-2021 (IS-2446), June 2016, Oslo, page 61-62.
- 73 Shabani, Mahsa and Luca Marelli "Re-identifiability of genomic data and the GDPR: Assessing the re-identifiability of genomic data in light of the EU General Data Protection Regulation" (2019) Science & Society, EMBO Reports, Vol. 20 (6): e48316.
- 74 Relevant assessment factors can be derived from recital 26 of the GDPR and legal literature.
- 75 Ministry of Health and Care Services, "Informasjonshåndtering i spesialisthelsetjenesten" (Eng.: About information management in the specialist health service), (2019). Circular (No.: rundskriv) I-3/2019, page 14.
- 76 Wright et al., Shringapure, Suyash S. and Carlos D. Bustamante, "Privacy Risks from Genomic Data-Sharing Beacons" (2015) The American Journal of Human Genetics, Vol. 97, page 631-646, and Shabani and Marelli (2019).
- 77 The Norwegian Directorate of health, National strategy for personalised medicine in healthcare 2017-2021. 2016.
- 78 Befring, Anne Kjersti, Personalised medicine – legal perspectives, Oslo: Gyldendal, 2019. Ch. 13.
- 79 Befring, Anne Kjersti, Personalised medicine – legal perspectives, Oslo: Gyldendal, 2019. Ch. 12.5.2.
- 80 Befring, Anne Kjersti, Personalised medicine – legal perspectives, Oslo: Gyldendal, 2019. Ch. 8.3.
- 81 Quintero, D., and Lee, F., IBM reference architecture for high performance data and AI in healthcare and life sciences. IBM Redbooks, 2019.
- 82 ISO, Information technology – Artificial intelligence – Overview of trustworthiness in artificial intelligence. [cited 2020; available from: <https://www.iso.org/standard/77608.html>]
- 83 European Commission, High-level expert group on AI, Assessment List for Trustworthy Artificial Intelligence. 2020.
- 84 BigMed, Clinical decision Support Software; Regulatory landscape in Europe from May 26th 2020. 2020.
- 85 European Commission, High-level expert group on AI, Ethics Guidelines for Trustworthy AI. 2019
- 86 European Commission, High-level expert group on AI, Assessment List for Trustworthy Artificial Intelligence. 2020.
- 87 European Commission, White Paper on Artificial Intelligence: a European approach to excellence and trust. 2020.
- 88 de Vreede-Swagemakers J., et al, Out-of-Hospital Cardiac Arrest in the 1990s: A Population-Based Study in the Maastricht Area on Incidence, Characteristics and Survival. *J. Am. Coll. Cardiol.*, 1997, 30: p. 1500
- 89 La Gerche, A., et al., Cardiac imaging and stress testing asymptomatic athletes to identify those at risk of sudden cardiac death. *JACC Cardiovasc Imaging*, 2013, 6: p. 993
- 90 Rama, T., Brekke, PH., et al., Iterative development of family history annotation guidelines using a synthetic corpus of clinical text. *ACL Anthology: Proceedings of the Ninth International Workshop on Health Text Mining and Information Analysis*; 2018. 18: p. 111
- 91 Rama, T., Brekke, PH., et al., Iterative development of family history annotation guidelines using a synthetic corpus of clinical text. *ACL Anthology: Proceedings of the Ninth International Workshop on Health Text Mining and Information Analysis*; 2018. 18: p. 111
- 92 Pilan, I., Brekke, PH., et al., Classification of Syncope Cases in Norwegian Medical Records. *ACL Anthology: Proceedings of the 3rd Clinical Natural Language Processing Workshop*; 2020. p. 79
- 93 Ruwald, M., et al., Accuracy of the ICD-10 discharge diagnosis for syncope. *Europace: Eu. Pacing Arrhythm Card Electrophysiol: J Work Groups Card Pacing Arrhythm Card Cell Electrophysiol Eur Soc Cardiol.* 2013, 15: p. 595.
- 94 Haugaa, K., et al., Transmural differences in myocardial contraction in long-QT syndrome: mechanical consequences of ion channel dysfunction. *Circulation*, 2010, 122: p. 1355.
- 95 Klaeboe, L., Brekke, PH., et al., Impact of transcatheter aortic valve implantation on mechanical dispersion. *Open Heart*, 2020, 7: e001199.
- 96 Haugaa, KH., et al., Strain echocardiography improves risk prediction of ventricular arrhythmias after myocardial infarction. *JACC Cardiovasc Imaging*, 2013, 6: p. 841.
- 97 Aalen, J., et al., Imaging predictors of response to cardiac resynchronisation therapy: left ventricular work asymmetry by echocardiography and septal viability by cardiac magnetic resonance. *Eur Heart J*, 2020, 41: p. 3813
- 98 University of Copenhagen, PM Heart [cited 2019; available from: <https://healthsciences.ku.dk/news/faculty-news/2019/03/big-step-towards-more-personalised-medicine-for-heart-patients/>]
- 99 Zylyftari, N., et al., Contacts to the healthcare system prior to out-of-hospital cardiac arrests. *Eur Heart J*, 2020, 41: ehaa946.0766
- 100 Nasjonal Kompetansetjeneste for Sjeldne Diagnosir (NKSD) [cited 2017; Available from: <https://helsenorge.no/sjeldne-diagnosir>]
- 101 Department of Health and Social Care, The UK strategy for rare diseases: 2020 update to the implementation plan for England. 2020.
- 102 BigMed: Slaughter, H., and Hovland D., Suggesting Reasonable Phenotypes to Clinicians, [Cited 2021, Available at https://bigmed.no/assets/Reports/Suggesting_reasonable_phenotypes_Slaughter_Hovland.pdf]
- 103 HPO Annotations. [cited 2019, available from: <https://hpo.jax.org/app/download/annotation>]
- 104 BigMed, Drivers in rapid genetic diagnostics for rare diseases in infants. 2019.
- 105 BigMed, Consent for clinical genetic testing in Norway: Considerations to the development of process and content. 2020.
- 106 Pai S, Hui S, Isserlin R, Shah MA, Kaka H, Bader GD. netDx: interpretable patient classification using integrated patient similarity networks. *Mol Syst Biol*. 2019 Mar 14;15(3):e8497
- 107 Fretland et al, Laparoscopic Versus Open Resection for Colorectal Liver Metastases: The OSLO-COMET Randomized Controlled Trial. *Ann Surg*. 2018 Feb;267(2):199-207.
- 108 HM Government, 2020, GENOME UK – The future of healthcare.
- 109 European Commission, High-level expert group on AI, Ethics Guidelines for Trustworthy AI. 2019.

