TALLINN UNIVERSITY OF TECHNOLOGY
School of Information Technologies

Ivelisse Rodriguez Aguirre 184710YVEM

FEASIBILITY STUDY OF SCLEROSIS MULTIPLE REGISTRY BASED ON ESTONIAN HEALTH INFORMATION SYSTEM

Master's Thesis

Supervisor: Janek Metsallik
MSc

Katrin Gross-Paju
MD, PhD

Tallinn 2020

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond

Ivelisse Rodriguez Aguirre 184710YVEM

EESTI TERVISE INFOSÜSTEEMIL PÕHINEVA SCLEROSIS MULTIPLEX REGISTRI TEOSTUVUSUURING

Magistritöö

Juhendaja: Janek Metsallik
MSc

Katrin Gross-Paju
MD, PhD

Tallinn 2020

Author’s declaration of originality

I hereby certify that I am the sole author of this thesis. All the used materials, references to the literature and the work of others have been referred to. This thesis has not been presented for examination anywhere else.

Author: Ivelisse Rodriguez Aguirre

20.05.2020

Abstract

Estonia has one of the most comprehensive health information systems in the world, but the country still lags behind other regions in terms of epidemiological registers. On the basis of the health information system, Estonia should be able to create new epidemiological registers more efficiently than countries that do not have such a national data exchange layer.

This paper investigates the extent to which the data structure of the Estonian Health Information System corresponds to the data structure necessary for the multiple sclerosis registry.

Two different methods were used in the research. First, the assessments of specialists in the treatment of multiple sclerosis on the quality of the data collected today and the composition of the data expected from the registry were studied. Second, the correspondence between the International Multiple Sclerosis Registry (MSBase) data dictionary and the Estonian Health Information System data models was mapped.

A comparison of the data set of the health information system with the data set of MSBase showed that the health information system is already able to collect a large part of the necessary data. After some additions, it is possible to perform a multiple sclerosis registry based on data moving through the Estonian Health Information System.

This thesis is written in English and is 48 pages long, including 7 chapters, 8 figures and 3 tables.

Annotatsioon

EESTI TERVISE INFOSÜSTEEMIL PÕHINEVA SCLEROSIS MULTIPLEX REGISTRI TEOSTUVUSUURING

Eestis on üks terviklikumaid terviseinfosüsteeme maailmas, kuid siiski jääb riik epidemioloogiliste registrite osas teistest piirkondadest maha. Terviseinfosüsteemi baasil peaks Eesti suutma luua uusi epidemioloogilisi registreid tõhusamalt kui riigid, kus selline üleriigiline andmevahetuskiht puudub.

Käesolevas töös uuritakse, mil määral vastab Eesti Terviseinfosüsteemi andmestruktuur sclerosis multiplex registry jaoks vajalike andmete kogumiseks..

Uurimistöös kasutati kahte erinevat meetodit. Esiteks uuriti sclerosis multiplex raviga tegelevate spetsialistide hinnaguid täna kogutavate andmete kvaliteedi ja registrilt oodatavate andmete koosseisu kohta. Teiseks kaardistati vastavus etaloniks valitud rahvusvahelise sclerosis multiplex registry (MSBase) andmesõnastiku ja Eesti Terviseinfosüsteemi andmemudelite vahel.

Terviseinfosüsteemi andmekoosseisu võrdlemine MSBase andmekoosseisuga näitas, et terviseinfosüsteem on juba täna võimeline koguma suure osa vajalikest andmetest. Mõningate täienduste järel on võimalik sclerosis multiplex registri teostamine terviseinfosüsteemi kaudu liikuvate andmete põhjal.

Lõputöö on kirjutatud inglise keeles ning sisaldab teksti 48 leheküljel, 7 peatükki, 8 joonist, 3 tabelit.

List of abbreviations and terms

MS	Multiple Sclerosis
TUT	Tallinn University of Technology
HIS	Health Information System
TAI	Tervise Arengu Instituut
TEHIK	Health and Welfare Information Systems Centre
GDPR	General Data Protection Regulation
WHO	World Health Organization
MSIF	Multiple Sclerosis International Federation
RRMS	Relapsing-remitting multiple sclerosis
PPMS	Primary progressive multiple sclerosis
SPMS	Secondary-progressive multiple sclerosis
CSF	Cerebral Spinal Fluid
MRI	Magnetic Resonance Imaging

Table of contents

Author's declaration of originality 3
Abstract 4
Annotatsioon EESTI TERVISE INFOSÜSTEEMIL PÕHINEVA SCLEROSIS MULTIPLEX REGISTRI TEOSTUVUSUURING 5
List of abbreviations and terms 6
Table of contents 7
List of figures 9
List of tables 10
1 Introduction 11
2 Background 12
  2.1 Definition of multiple sclerosis 12
    2.1.1 Multiple sclerosis in Estonia 12
  2.2 Definition of patient registries 12
    2.2.1 Patients registries in Estonia 13
  2.3 Estonian health information system 13
    2.3.1 Quality measures of Estonian health information system 13
  2.4 Data governance and ethics 14
  2.5 Introduction to multiple sclerosis patient registries collaboration initiatives 14
    2.5.1 MSBase 14
    2.5.2 Big MS Data (BMSD) 15
    2.5.3 Atlas of MS 15
  2.6 Quality measures of patient registries 15
    2.6.1 Quality assurance of data 16
    2.6.2 Quality assurance of registry procedures 16
    2.6.3 Quality assurance of computerized systems 16
  2.7 Benefits of patients' registries 16
3 Research methodology 18
  3.1 Objective of the research 18
  3.2 Hypothesis 18
  3.3 Research method 18
    3.3.1 Survey to healthcare professionals 19

3.3.2 Compilation of the data requirements for the registry 19
3.3.3 Identification of the data sources about HIS 20
3.3.4 Analysis of data fields 20
3.3.5 Analysis of the feasibility of the MS Registry 22
4 Results 23
4.1 Specialists' survey 23
4.1.1 Demographics 23
4.1.2 Results of survey 23
4.2 Analysis of fields 28
4.2.1 Fields present 29
4.2.2 Free text fields 30
4.2.3 Mandatory fields 32
5 Discussion 34
6 Conclusions 40
7 References 41
Appendix 1 – Survey to healthcare professionals 43
Appendix 2 – MSBase data dictionary categories 46

List of figures

Figure 1. View of MSBase Data Dictionary 19
Figure 2. View of Excel file used to compare MSBase and Estonia HIS 20
Figure 3. View of Excel file filled comparing MSBase and Estonia HIS 21
Figure 4. Challenges to set a registry 24
Figure 5. Lack of patient registries 25
Figure 6. Data to guide treatment decisions 26
Figure 7. Retrospective data quality 27
Figure 8. Retrospective data quality for prescription medications 28

List of tables

Table 1. Fields Present 30
Table 2. Free text fields 31
Table 3. Mandatory fields 33

1 Introduction

As information technology capabilities grow, new ways to help develop research and treatment of diseases have been established. One of the multiples uses that information technologies have brought into the healthcare sector consist of patients' registries which besides containing more data are easier to manage and analyze.

Currently, most of the research about multiple sclerosis is done through observational studies, but this data can include potential source of bias, which in turn can led to this information not being accepted by the regulatory agencies or the rest of the medical community. Moreover, the results from randomized controlled clinical trials might not apply to the general clinical population due to constrains of the trial, for example age or comorbidities of the participants, disease type, or treatment environment. [1]

While the use of pooled data in patients registries for the research of multiple sclerosis have proved to be an effective way to gain knowledge about the evolution of the disease with the goal to reduce future complications, still there is a large need for big data that currently country-level or even international registries can't provide due to their limitations in size and amount of information collected. [2]

Besides this, there is a need for data quality and harmonization, the sharing of data across different jurisdictions, each one operating with different legal and ethical regulations, and the sustainability of the data sources. [1]

Estonia, being a small country with limited resources and a reduced pool of people suffering from multiple sclerosis, can benefit greatly collaborating in initiatives aimed to collect and share data about multiple sclerosis patients.

2 Background

2.1 Definition of multiple sclerosis

Multiple sclerosis can be defined as an immune mediated disease in which the myelin in the central nervous system (composed by the brain, spinal cord and optical nerves) is attacked by the body's immune system, causing inflammation which damages nerves fibers, myelin and the cells which produce myelin. [3]

The disease is categorized in three different courses, also known as types or phenotypes: relapsing-remitting multiple sclerosis (RRMS), primary progressive multiple sclerosis (PPMS), and secondary progressive multiple sclerosis (SPMS). [4]

Currently there isn't a cure for multiple sclerosis, instead disease-modifying therapies are used, together with proper management of relapses, treating the symptoms and overall good management of the health and wellness of the people suffering from the disease. [5]

While the exact cause of multiple sclerosis is unknown, it is believed to be caused by a combination of genetic susceptibility, abnormalities of the immune system and environmental factors. [3]

2.1.1 Multiple sclerosis in Estonia

In Estonia, is estimated that there is around 1,300 to 1,500 persons living with multiple sclerosis, with a prevalence of approximately 100 to 110 cases per 100,000 habitants. [6]

2.2 Definition of patient registries

According to the United States' Agency for Healthcare Research and Quality, a patient registry can be defined as an "organized system that uses observational study methods to collect uniform data (clinical and other) to evaluate specified outcomes for a population defined by a particular disease, condition, or exposure, and that serves one or more predetermined scientific, clinical, or policy purposes". [7]

2.2.1 Patients registries in Estonia

According to Tervise Arengu Instituut (TAI, in English the National Institute for Health Development), Estonia has six medical registries: Estonian Medical Birth Registry and Estonian Abortion Registry; Estonian Causes of Death Registry; Estonian Cancer Registry; Estonian Tuberculosis Registry; Estonian Drug Treatment Database and the Estonian Cancer Screening Registry. [8]

2.3 Estonian health information system

Under the Health Services Organization Act of the Republic of Estonia, the Health Information System (HIS) is a database which belongs to the State Information Systems and the controller of it is the Ministry of Social Affairs. The management and development of the HIS is done by the Tervise ja Heaolu Infosüsteemide Keskus (TEHIK, in English, the Health and Welfare Information Systems Centre). [9] [10]

The goals of the HIS is to use data related to healthcare for ensuring quality of services, guaranteeing the rights of patients, the protection of public health (including maintaining registers, organizing health statistics and for the management of healthcare), and for the performance of contracts related to the provision of healthcare-related services. [9]

2.3.1 Quality measures of Estonian health information system

Tervise Arengu Instituut (TAI) is the Estonian institution in charge of the evaluation of the quality of the data included in the HIS. [10]

Currently, healthcare providers are mandated to submit the documentation of treatment procedures to different systems causing that this information is duplicated; also, since information is submitted on a treatment case basis, the documentation may need to be submitted multiple times per each treatment case, that’s why one of the priorities of TAI is the reduction of the workload related to documentation done by the healthcare providers which in turn will help to create more detailed statistics related to healthcare and of higher quality. [10]

2.4 Data governance and ethics

Governance is a key component of patient registries because ensures that the processes and procedures needed for the proper collection of the data needed in the registry are set in place and helps with transparency as part of the decision making, the operation and the reporting of results. [11]

While patients' registries are an important tool for research, the right to privacy and to the protection of data are rights included in the European Charter of Fundamental Rights. Therefore, one of the most important tasks while establishing and managing patient registries is to ensure the privacy of the patients included in the registry. A key aspect of this, is that participants must give their consent for the use of the information in the registry. [7] [11]

2.5 Introduction to multiple sclerosis patient registries collaboration initiatives

As the need for data and collaboration has arose, multiple initiatives have been established around to the world aimed to work together to improve the data quality related to multiple sclerosis.

2.5.1 MSBase

Since 2004, MSBase has acted as an online international platform where MS clinics associated to the project can contribute patient's information. [12]

MSBase acts as a real-time secure database where the participating clinics share, track and analyze data. [13]

Membership to MSBase is open to all neurologists and their supporting teams around their world, they just need to register to the platform. [14]

The projects' database fills primarily two purposes:

- Helps to simplify the collection of epidemiological data thanks to the use of an online-based database.

- Allows to answer epidemiological question which can only be addressed analyzing data from thousands of participants. [13]

MSBase makes available on their website a reference file known as the MSBase Data-entry Software Data Dictionary, which contains all the data fields which need to be captured and stored as part of a MS patient registry. [15]

2.5.2 Big MS Data (BMSD)

Was established after MSBase, France’s OPSEP and the national MS registries of Denmark, Italy and Sweden merged certain data for specific projects. [2]

2.5.3 Atlas of MS

Published in 2008 as a partnership project of the Multiple Sclerosis International Federation (MSIF) and the World Health Organization (WHO) with the aim to analyze the epidemiology and the availability and accessibility of resources for people suffering of MS around the world. [16]

The second edition was published in 2013, containing data from 104 countries, and while the 2008 edition included information from 112 countries, it’s estimated that between both editions the data covers roughly 87% of the world population, hence, Atlas of MS is truly one of the most global compilations in information about multiple sclerosis. [17]

A third edition of Atlas of MS is expected to be published in 2020. [16]

2.6 Quality measures of patient registries

Due to patient registries having different purposes: some can be used for policy making while others are just to track the evolution of a disease is obvious that the quality measures needed varies depending of the aforementioned purpose.

Primarily, quality assurance is divided in three categories: quality assurance of data, quality assurance of registry procedures and quality assurance of computerized systems. [7]

2.6.1 Quality assurance of data

The assurance of quality of data refers to prevent errors including in interpretation or coding of the data; in data entry, transfer or in the accuracy during transformation; errors of intention, for example “cherry-picking” some data. [7]

Some of the measures to ensure the quality of the data include: training, provide feedback related to data missing or with inconsistencies, compare data across sources and over time, do audits of samples of data and carry for-cause audits to identify possible data sources which are suspicious of inaccuracies or intentional errors while reporting data. [7]

2.6.2 Quality assurance of registry procedures

To ensure that the process of the registry adhere to the required quality standards several procedures can be followed including: use of external auditors; verify that the registry’s systems operate with integrity and security; ensure that software used for the patient registry adheres to the basic principles of software development, for example that a software development life-cycle model is followed. [7]

2.6.3 Quality assurance of computerized systems

Since patient registries collect health information, is important that first they adhere to local data protection regulations, for example the General Data Protection Regulation (GDPR) in the European Union.

Besides this, is important that they have a system security plan, which include documentation of policies and standard operating procedures; performance of security assessments; education and training of people handling data; correct assignment of access rights and the implementation of access controls to the data; use of data enclaves and electronic signatures. [7]

2.7 Benefits of patients’ registries

Having patient registries can offer a wide range of benefits not only for improving the quality of life and health outcomes of the person suffering from the condition but also as a tool to gain scientific knowledge and from the epidemiological and societal point of view.

The benefits of having a patient registry can be divided in four main categories:

- To make reporting of data easier: which for example, when aggregated with other data can be useful for advocating in favor of the patients.
- For retrospective and prospective research: for example, using the patient registry database for exploratory research.
- For professional development: for example, to publish findings, connect with a community of healthcare professionals working in the same topic and to learn about “best practices” for data.
- For service improvement: for example, can be used as a tool to save money helping to improve purchasing process and to deliver in an efficient way patient care of good quality. [18] [19] [20]

3 Research methodology

3.1 Objective of the research

The objective of this study is to explore the current dataset stored in Estonia's Health Information System (HIS) in relationship to data needed for the creation of a multiple sclerosis patient registry. The research questions are:

- Is the data available in the HIS could be used to establish a patient registry for multiple sclerosis?

- Does this data available is well structured that the need for manual inputs of extra information would be minimal?

3.2 Hypothesis

Research shows that the use of data from HIS for the creation and maintenance of patients' registries has already been done with successful results. However, despite the availability of the information, and that the Estonian law under which the HIS operates allow the use of this data for the creation and maintenance of patient registries, this option hasn't been widely used: patient registries obtain information from other sources, instead of extracting it from the HIS. Therefore, this study is designed to show that this could be possible to a greater extend, to obtain a large volume of information which would allow Estonia to have a patient registry for multiple sclerosis.

3.3 Research method

To validate and to prove the hypothesis two different researches were carried out:

- A survey to healthcare professionals working with multiple sclerosis was carried as a way to raise qualitative data.
- For the quantitative method, an analysis based in MSBase Data Dictionary of which fields of data needed to create the patient registry where already present in the HIS.

3.3.1 Survey to healthcare professionals

A survey was applied to specialists working with MS patients using the snowball method. The first person to whom it was distributed was to Dr. Katrin Gross-Paju, who then proceeded to share the survey with other specialists.

This survey consisted of fifteen questions to gain their opinion about the needs of a patient registry, how they consider is the current state of the data gathered and what information they consider is useful for their work.

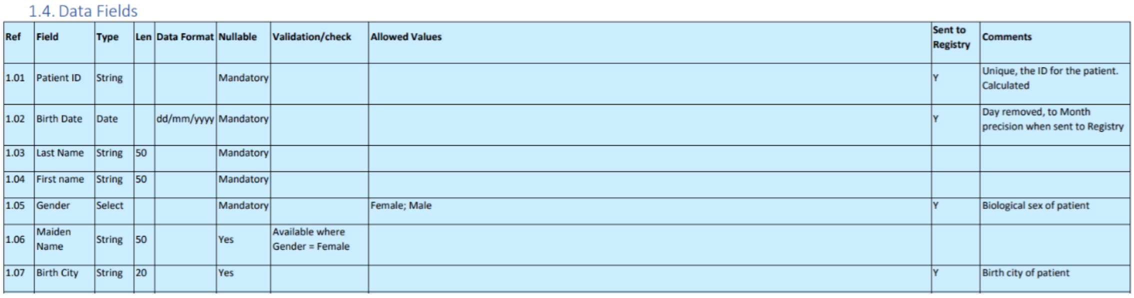
In total, five surveys were filled, and the analysis of the information gathered is presented in the results sections of this research.

The survey is included in the present research in the Annex section.

3.3.2 Compilation of the data requirements for the registry

MSBase makes available on their website a reference file known as “MSBase Data-entry Software (MDS) Data Dictionary”. This file contains a list of all the data fields needed to be captured as part of patient registry for MS, the data fields are grouped in several categories, like patient profile, medical history, MS diagnosis, among others. [15]

For each data field, certain attributes are included in the data dictionary: a reference number, field name, type of field, length, date format, if is nullable, a validation/check field, allowed values, a “sent to registry” yes/no field plus an area for comments.



1.4. Data Fields

Ref	Field	Type	Len	Data Format	Nullable	Validation/check	Allowed Values	Sent to Registry	Comments
1.01	Patient ID	String			Mandatory			Y	Unique, the ID for the patient. Calculated
1.02	Birth Date	Date		dd/mm/yyyy	Mandatory			Y	Day removed, to Month precision when sent to Registry
1.03	Last Name	String	50		Mandatory				
1.04	First name	String	50		Mandatory				
1.05	Gender	Select			Mandatory		Female; Male	Y	Biological sex of patient
1.06	Maiden Name	String	50		Yes	Available where Gender = Female			
1.07	Birth City	String	20		Yes			Y	Birth city of patient

Figure 1. View of MSBase Data Dictionary

For this research, all categories included in the MSBase Data Dictionary were used, excluding those known as Flexifields, for a total of 625 fields divided across 27 categories.

3.3.3 Identification of the data sources about HIS

For the data matching, TEHIK provided access to several of the operating manuals used in the HIS. Those manuals are published as an aiding tool to help filling the data needed in the HIS. [21]

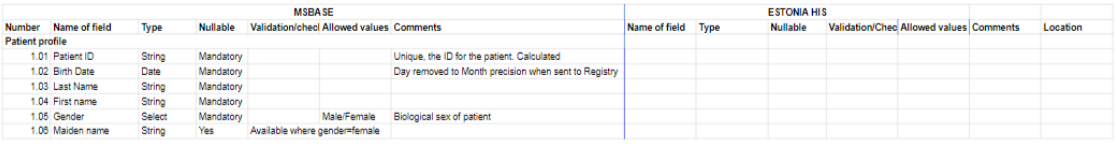
Since the manuals cover several different scenarios of who and when the data is introduced, Kerli Linna, the Head of Data Management Division at TEHIK, suggested that the manuals which could be more appropriate for the data matching of this research where *ambulatoorse epikriisi täitmise juhend* (guide to filling out an outpatient epicrisis), *kiirabikaardi täitmise juhend* (instructions for filling the ambulance report), *saatekirja vastuse täitmise juhend* (instructions for completing the referral letter reply) and *statsionaarse epikriisi täitmise juhend* (instructions for filling in a stationary epicrisis).

The manuals are published fully in Estonian language, with no English version available, hence, the files were translated using the Google Translate web tool.

3.3.4 Analysis of data fields

The analysis of data fields was done comparing the data fields used in patients' registries which are part of MSBase project against those data fields already present in the HIS.

Those 625 fields referenced by MSBase were exported to a Microsoft Excel file, where a table including only certain attributes (the reference number, field name, type of field, if is nullable, a validation/check field, allowed values and comments) were added. Since the main objectives of this research was based in trying to find the number of matches of data fields present, there was no need to match all the attributes contained in the MSBase Data Dictionary.



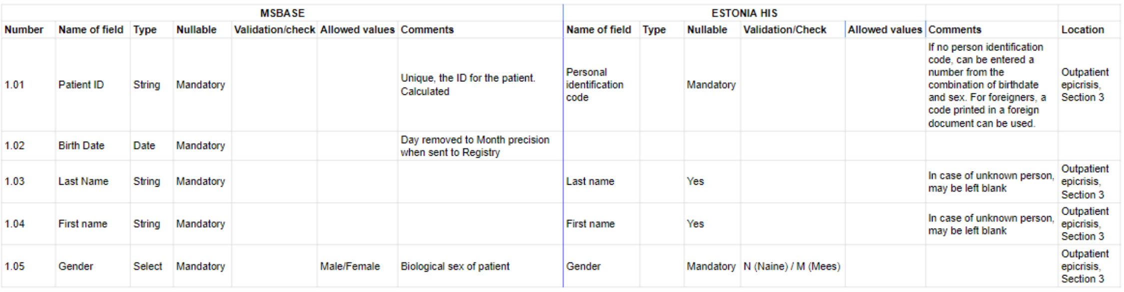
MSBASE							ESTONIA HIS						
Number	Name of field	Type	Nullable	Validation/checl	Allowed values	Comments	Name of field	Type	Nullable	Validation/Chec	Allowed values	Comments	Location
Patient profile													
1.01	Patient ID	String	Mandatory			Unique, the ID for the patient. Calculated							
1.02	Birth Date	Date	Mandatory			Day removed to Month precision when sent to Registry							
1.03	Last Name	String	Mandatory										
1.04	First name	String	Mandatory										
1.05	Gender	Select	Mandatory		Male/Female	Biological sex of patient							
1.06	Maiden name	String	Yes	Available where gender=female									

Figure 2. View of Excel file used to compare MSBase and Estonia HIS

After the manuals provided by TEHIK were translated, a manual process of matching the corresponding field in MSBase data dictionary with the equivalent field in the HIS was done: this was an extremely labor intensive task, since included reading all the manuals translations and trying to find the closest matching field.

Besides this, several sections of the MSBase data dictionary covers info pertaining to laboratory tests, which while they are included in the HIS and mentioned in the manuals, this is done based in their LOINC code. This added another layer to complexity to the matching, since all the laboratory tests included in the MSBase data dictionary had to be searched in TEHIK’s laboratory tests database to confirm those tests are performed in Estonia. [22]

When a field was found in the manuals, this was added in the columns to the right side of the Microsoft Excel file next to the corresponding field included in the MSBase Dictionary. Included was the name of the field according to the manual, the type of field (numeric, string, free text field, among others), if was nullable or mandatory, any validation/check information, the allowed values in the field (for example, male or female), any comments related to the field and the location where this field was found, including the name of the manual and in which section of it was found.



MSBASE							ESTONIA HIS						
Number	Name of field	Type	Nullable	Validation/check	Allowed values	Comments	Name of field	Type	Nullable	Validation/Check	Allowed values	Comments	Location
1.01	Patient ID	String	Mandatory			Unique, the ID for the patient. Calculated	Personal identification code		Mandatory			If no person identification code, can be entered a number from the combination of birthdate and sex. For foreigners, a code printed in a foreign document can be used.	Outpatient epicrisis, Section 3
1.02	Birth Date	Date	Mandatory			Day removed to Month precision when sent to Registry							
1.03	Last Name	String	Mandatory				Last name		Yes			In case of unknown person, may be left blank	Outpatient epicrisis, Section 3
1.04	First name	String	Mandatory				First name		Yes			In case of unknown person, may be left blank	Outpatient epicrisis, Section 3
1.05	Gender	Select	Mandatory		Male/Female	Biological sex of patient	Gender		Mandatory	N (Naine) / M (Mees)			Outpatient epicrisis, Section 3

Figure 3. View of Excel file filled comparing MSBase and Estonia HIS

After completing the matching of fields, a numerical analysis was performed, including calculation of percentages of data matching and elaboration of several graphics and tables. All this information is included in the results chapter.

The reason for choosing this method was because it was a straightforward way to calculate the percentage of data fields already present and estimate how much data is missing and would have to be captured manually if the data of the HIS is transferred to the MS patient registry.

3.3.5 Analysis of the feasibility of the MS Registry

The analysis of the feasibility was based in three main parameters: number of data fields present in the HIS compared with those included in the MSBase data dictionary, number of fields which are marked as free text fields and the number of fields which are marked as mandatory. The survey results were used as supporting evidence to confirm the need of a patient registry and to understand the current situation of the data.

4 Results

The results of this research are divided in two categories:

1. The information gathered as part of the survey to specialists to gain understanding of the current situation and needs for a patient registry for MS.
2. Analysis of the number of fields present compared against to those required by the MSBase project.

4.1 Specialists' survey

The survey consisted of 15 different questions, which can be grouped into demographics questions and those related to the patient registry data.

Of the 15 questions, five of them were open answer questions, one was yes/no option, one was a numeric value as answer and in eight questions the participants were asked to rate their opinion from 1 to 5, while 1 was assigned the value of "strongly agree" and 5 was assigned to "strongly disagree"

4.1.1 Demographics

The survey was applied to five neurologists who work at West Tallinn Central Hospital and at Estonia's MS Centre.

The neurologists had from 10 to 32 years of experience in their medical careers, with an average of 17.6 years practicing.

4.1.2 Results of survey

When asked *"Which do you consider is the main use of the data collected in Estonian registries?"* the answers ranged from "collection, exchange and comparison of data/comparative information with other countries", "scientific research", "analysis of treatments" and "epidemiological purposes".

For the question of *"In your opinion, which are the main challenges while setting a registry from a data perspective point of view?"* participants were asked to rate their opinion as "strongly agree", "agree", "neutral", "disagree" and "strongly disagree" for three different categories: data update, data quality and data coverage (how many patients are included) . The results for this question were as following:

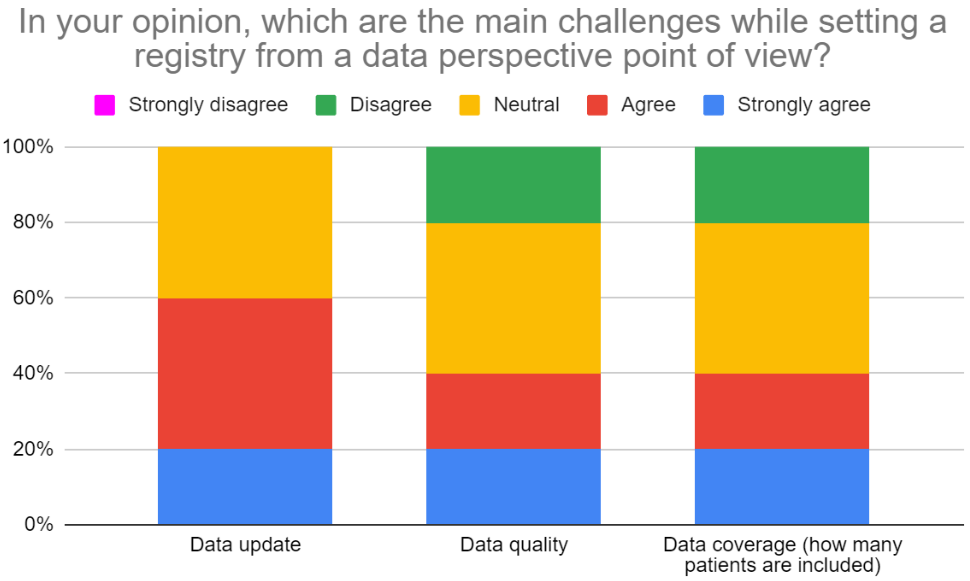


Figure 4. Challenges to set a registry

Related to the number of patient registries, the participants were asked *"According to Tervise Arengu Instituut (TAI), there are only six patient registries in the country. In your opinion, why is this number so small?"* for four different categories: no need, legal/ethical concerns, lack of resources and lack of interest, ranking for each category their opinion as "strongly agree", "agree", "neutral", "disagree" and "strongly disagree". The results for this question were as following:

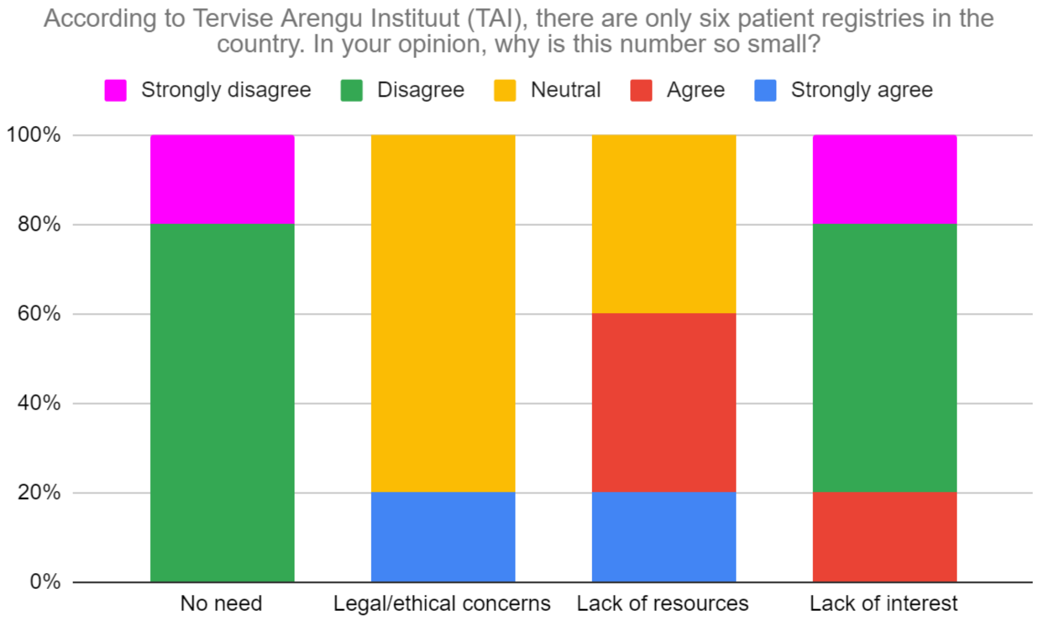


Figure 5. Lack of patient registries

To the question of *"Is it important to have a patient registry for MS in Estonia?"*, a 100% of the respondents answered that yes.

When asked about what in their opinion would be the main aim of the MS patient registry, the most popular opinion was related to "have a better overview of the MS population in Estonia", other comments included "collect information for scientific research" and "to obtain information about treatment regimens and efficacy".

On the question of *"Currently which data do you use to guide your treatment decisions for patients with MS?"* they were asked to rate their opinion as "strongly agree", "agree", "neutral", "disagree" and "strongly disagree" for three different categories: number of relapses, MRI activity and progression. The results for this question were as following:

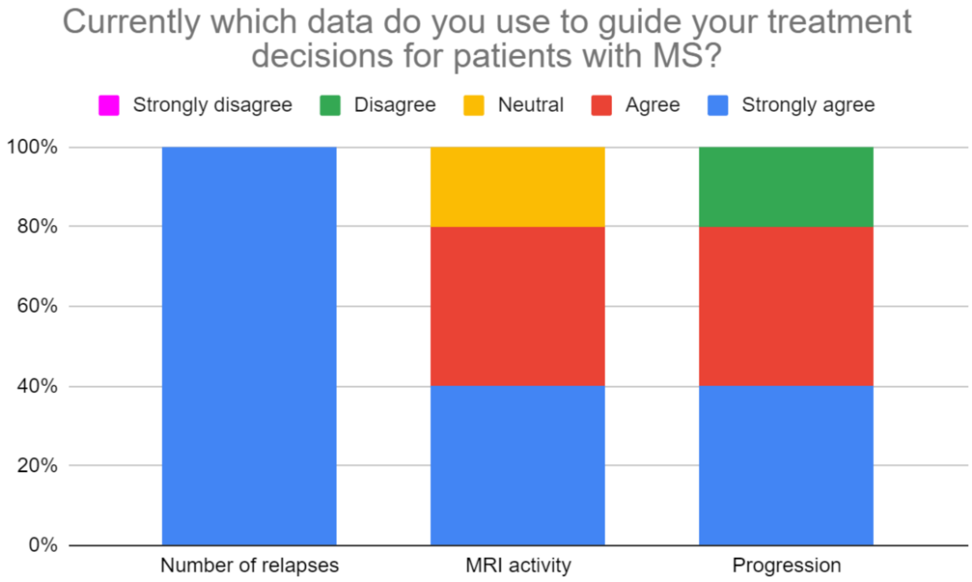


Figure 6. Data to guide treatment decisions

Referring to the question of *"For the patient registry for MS, which do you think the data would be primarily used?"* the most popular answer was "study of MS epidemiology" but others mentioned included "scientific research", "analysis of treatment availability", "better overview of individual disease courses and patient cohort", "can be used for reimbursement issues", "treatment regimens comparisons", "long-term comparison", "quality" and "management".

For the question of *"Who should be included in the patient registry?"*, 100% of the respondents agreed that should be patients with relapsing-remitting multiple sclerosis (RRMS), patients with primary progressive multiple sclerosis (PPMS) and patients with secondary-progressive multiple sclerosis (SPMS).

On the question of *"How would you rate the quality of the data in the process of retrospective data collection?"* the specialists were asked to rate the quality of data on the following rating: "highest quality", "good quality", "average" "bad quality", "lowest quality" for five different categories: MS first symptoms, MS diagnosis, relapse history, relapse treatment history and EDSS in time points. The results for this question were as following:

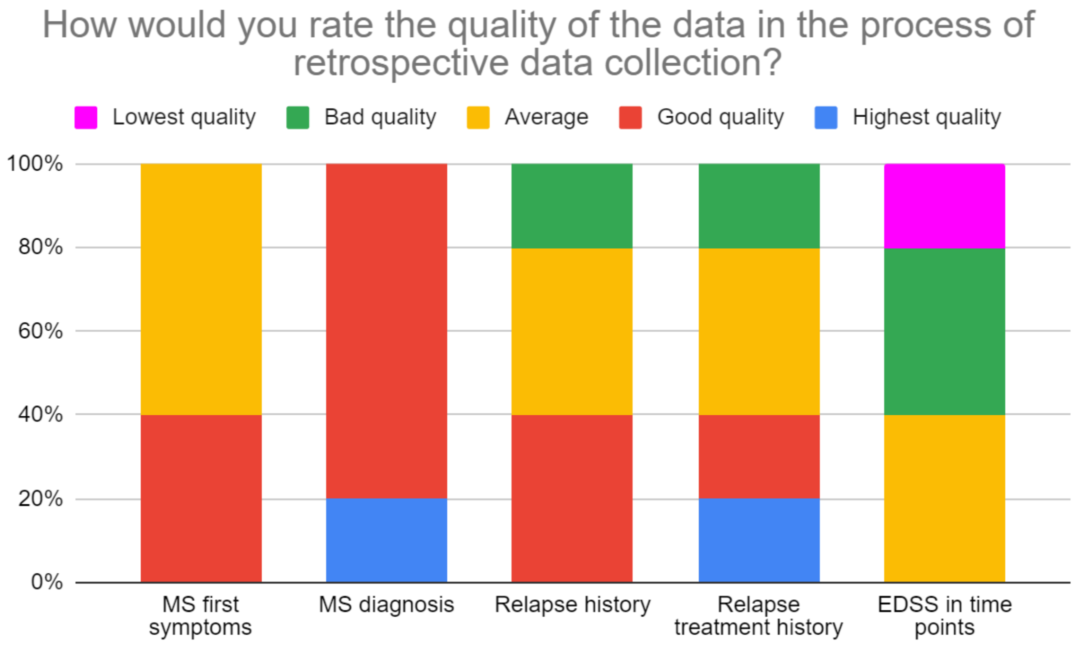


Figure 7. Retrospective data quality

On the question of *“How would you rate the quality of the data in the process of retrospective data collection on MS prescription medications?”* the specialists were asked to rate the quality of data on the following rating: “highest quality”, “good quality”, “average” “bad quality”, “lowest quality” for four different categories: interferons, GA, teriflunomide and dimethylfumarate. The results for this question were as following:

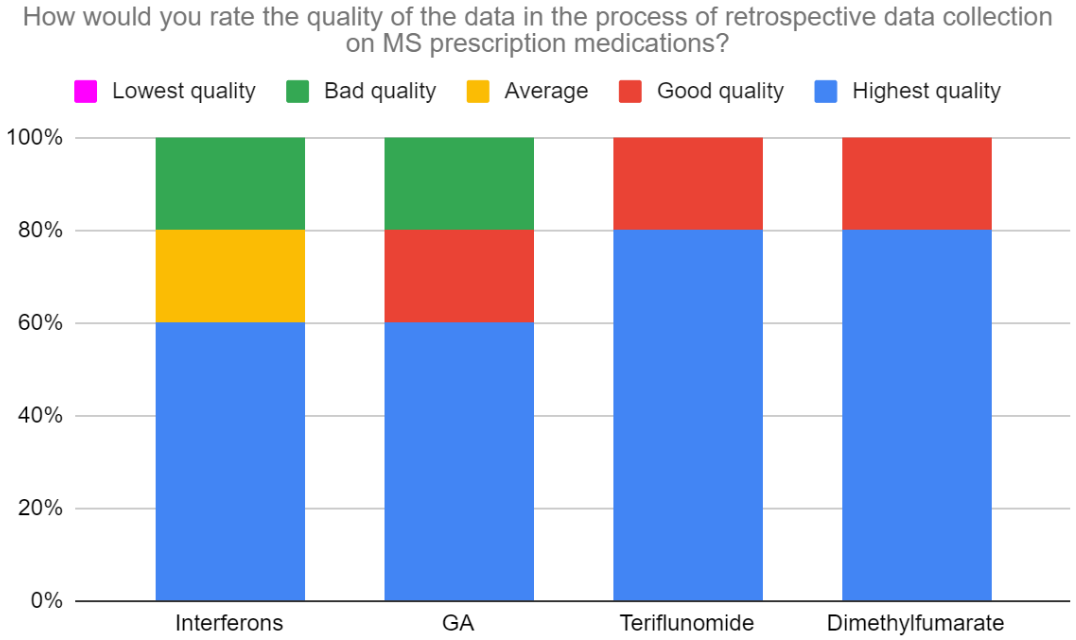


Figure 8. Retrospective data quality for prescription medications

For the question of *"How would you rate the quality of the data in the process of retrospective data collection on MS hospital budget medications?"* 100% of the participants answered that the retrospective data collection was of the highest quality for the four categories they were asked: natalizumab, ocrelizumab, alemtuzumab and fingolimod.

Regarding the question of "How would you rate the quality of the data in the process of prospective data collection on MS medications?" for the categories of interferons, GA, teriflunomide, dimethylfumarate, ocrelizumab, alemtuzumab and fingolimod 100% of the participants answered that the retrospective data collection was of the highest quality for all the categories.

4.2 Analysis of fields

The MSBase Data Dictionary comprises a total of 625 data fields divided across 27 categories.

The analysis of the fields between those included in the MSBase Data Dictionary and those present in Estonia's HIS is presented based in three categories:

- Fields present: Refers to those fields that are required in the MSBase Data Dictionary and an equivalent was found in the HIS.

- Fields which are free text field: While some fields can be present, sometimes their formatting in the HIS is of a free text field, which means that doesn't have a necessary structure for entering the data, hence, the information might be or might be not entered into the field.

- Mandatory fields: For all the fields present, how many of them are mandatory to fill and can't be left empty.

4.2.1 Fields present

For the field's present analysis, the objective was to find for each field mentioned in the MSBase Data Dictionary an equivalent in the HIS.

The 27 categories used are the same that MSBase Data Dictionary groups their fields of data for a total of 625 fields.

Of the 625 fields analyzed, it was found out that 473 of them were present, which equals to a 75.68% of the total of fields.

But this distribution wasn't equal among all categories: while some categories like EDSS, Safety, Relapses, Immuno-suppression associated, Other events, MRI, Evoked Potentials and Non-pharmacological where found to be present in a 100% rate in the HIS for all the category's fields, other categories like Family history, MusiQoL and Pregnancy had a 0% matching rate.

A higher percentage of matching fields is considered better since means more information present in the HIS can be extracted and used for the implementation of the MS patient registry.

All of the fields found to be present, were contained in the *ambulatoorse epikriisi täitmise juhend* (guide to filling out an outpatient epicrisis) manual provided by TEHIK.

Table 1. Fields Present

Category	Total fields	Present	Missing	Present as %
Patient profile	28	22	6	78.57%
Medical history	10	5	5	50.00%
Family history	10	0	10	0.00%
MS diagnosis	24	21	3	87.50%
Basic details	40	36	4	90.00%
Symptoms	39	38	1	97.44%
EDSS	11	11	0	100.00%
MusiQoL	33	0	33	0.00%
Safety	6	6	0	100.00%
Relapses	18	18	0	100.00%
Malignancy	15	13	2	86.67%
NMSC	15	13	2	86.67%
Herpes zoster	6	4	2	66.67%
Immuno-suppression associated	8	8	1	100.00%
Pregnancy	17	0	17	0.00%
Other events	16	16	0	100.00%
MRI	20	20	0	100.00%
CSF	20	18	2	90.00%
Evoked potentials	20	20	0	100.00%
Laboratory tests (Haematology)	58	44	14	75.86%
Laboratory tests (Blood chemistry)	52	39	13	75.00%
Laboratory tests (thyroid function)	20	15	5	75.00%
Laboratory tests (Serological tests)	45	34	11	75.56%
Laboratory tests (Auto-antibody tests)	72	54	18	75.00%
MS Specific	9	7	2	77.78%
Symptomatic	9	7	2	77.78%
Non-pharmacological	4	4	0	100.00%
TOTAL	**625**	**473**	**153**	**75.68%**

4.2.2 Free text fields

The free text fields is an important category of the data matching because, while there can be a matching field between MSBase data dictionary and Estonia’s HIS, the fact that the field in the HIS is a free text field means that maybe not in all cases the data is present, since the main characteristic of the free text field is the fact that is only an open text box area where healthcare professionals can enter information regarding the corresponding field but they aren’t mandated to do it and this information doesn’t need to follow any structure.

Of the 625 fields across 27 categories analyzed, it was found that 167 fields were free text fields, for a 26.72% of all the total fields analyzed.

When compared the percentage of free text field against the number of fields present in the HIS (473 fields), it was found that the 167 free text fields are equal to a 35.51% of all the present fields.

The distribution of free text fields wasn’t equal among all categories: while some categories like Patient profile and all of those related to laboratory tests had a 0% of free text fields, some others like EDSS, Relapses and Evoked potentials had a 100% rate of free text fields for all the fields present.

A lower percentage of free text fields in each category is considered better since means the data contained in the HIS is more structured and can be extracted in a more straightforward way and used for the implementation of the MS patient registry without the need of costly data analysis, machine learning or artificial intelligence tools to analyse the field and discover if the text contained is the one needed for that specific field.

Table 2. Free text fields

Category	**Total fields**	**Present**	**Free text field**	**Free text field as % of total fields**	**Free text field as % of present fields**
Patient profile	28	22	0	0.00%	0.00%
Medical history	10	5	1	10.00%	20.00%
Family history	10	0	0	-	-
MS diagnosis	24	21	18	75.00%	85.71%
Basic details	40	36	18	45.00%	50.00%
Symptoms	39	38	33	84.62%	86.84%
EDSS	11	11	11	100.00%	100.00%
MusiQoL	33	0	0	-	-
Safety	6	6	0	0.00%	0.00%
Relapses	18	18	18	100.00%	100.00%
Malignancy	15	13	0	0.00%	0.00%
NMSC	15	13	1	6.67%	7.69%
Herpes zoster	6	4	3	50.00%	75.00%
Immuno-suppression associated	8	8	5	62.50%	62.50%
Pregnancy	17	0	0	-	-
Other events	16	16	15	93.75%	93.75%

MRI	20	20	20	100.00%	100.00%
CSF	20	18	0	0.00%	0.00%
Evoked potentials	20	20	20	100.00%	100.00%
Laboratory tests (Haematology)	58	44	0	0.00%	0.00%
Laboratory tests (Blood chemistry)	52	39	0	0.00%	0.00%
Laboratory tests (thyroid function)	20	15	0	0.00%	0.00%
Laboratory tests (Serological tests)	45	34	0	0.00%	0.00%
Laboratory tests (Auto-antibody tests)	72	54	0	0.00%	0.00%
MS Specific	9	7	0	0.00%	0.00%
Symptomatic	9	7	0	0.00%	0.00%
Non-pharmacological	4	4	4	100.00%	100.00%
TOTAL	**625**	**473**	**167**	**26.72%**	**35.31%**

4.2.3 Mandatory fields

The mandatory fields are an important part of the analysis of data fields because they are impossible to ignore: to proceed and continue saving the data, a value is necessary to enter in the field.

Of the 625 fields across 27 categories analyzed, it was found that 146 fields were mandatory, which equals to a 23.36% of all the data fields analyzed.

When the 146 fields deemed as mandatory were compared against the 473 fields present in the matching between the HIS and the MSBase Data Dictionary, it was found out they equal to 30.87% of those aforementioned present fields.

But this distribution wasn’t equal among all categories: for certain categories like Other events and MRI a 100% of the fields were considered as mandatory, but for categories like EDSS, Safety and Relapses, a 0% of the fields were deemed as mandatory.

It’s important to notice that a higher percentage of fields established as mandatory is better for the establishment of a patient registry, since this allows the data to just be extracted from the HIS and inserted into the patient registry without need of much extra data treatment.

Table 3. Mandatory fields

Category	**Total fields**	**Present**	**Mandatory**	**Mandatory as % of total fields**	**Mandatory as % of present fields**
Patient profile	28	22	4	14.29%	18.18%
Medical history	10	5	4	40.00%	80.00%
Family history	10	0	0	-	-
MS diagnosis	24	21	6	25.00%	28.57%
Basic details	40	36	5	12.50%	13.89%
Symptoms	39	38	5	12.82%	13.16%
EDSS	11	11	0	0.00%	0.00%
MusiQoL	33	0	0	-	-
Safety	6	6	0	0.00%	0.00%
Relapses	18	18	0	0.00%	0.00%
Malignancy	15	13	0	0.00%	0.00%
NMSC	15	13	2	13.33%	15.38%
Herpes zoster	6	4	3	50.00%	75.00%
Immuno-suppression associated	8	8	0	0.00%	0.00%
Pregnancy	17	0	0	-	-
Other events	16	16	16	100.00%	100.00%
MRI	20	20	20	100.00%	100.00%
CSF	20	18	17	85.00%	94.44%
Evoked potentials	20	20	0	0.00%	0.00%
Laboratory tests (Haematology)	58	44	15	25.86%	34.09%
Laboratory tests (Blood chemistry)	52	39	13	25.00%	33.33%
Laboratory tests (thyroid function)	20	15	5	25.00%	33.33%
Laboratory tests (Serological tests)	45	34	11	24.44%	32.35%
Laboratory tests (Auto-antibody tests)	72	54	18	25.00%	33.33%
MS Specific	9	7	1	11.11%	14.29%
Symptomatic	9	7	1	11.11%	14.29%
Non-pharmacological	4	4	0	0.00%	0.00%
TOTAL	**625**	**473**	**146**	**23.36%**	**30.87%**

5 Discussion

The analysis both of the opinion of the doctors and the matching of fields between MSBase and Estonia’s HIS demonstrated that while there is a need and a will for the MS patient registry and there is a good amount of information already contained in the HIS, the implementation of the patient registry will not be a straightforward work of just pulling data from the HIS and export to a new database: while for some areas this can be done, some others areas will need data treatment and specially, will be necessary to make that healthcare workers follow certain standards or good practices while entering the data in the HIS.

My hypothesis was confirmed: to a great extent, the data included in Estonia’s HIS can be used to set up for a patient registry for MS. The analysis showed that of the 625 fields that MSBase Data Dictionary requests to be part of the patient registry, 473 fields are present in the HIS, which equals to a 75.68%.

And while 167 fields (26.72% of total) are marked as free text fields and only 146 (23.36% of total) are mandatory, those figures shouldn’t be used to discourage the creation of patient registries, this should be seen as a challenge and an opportunity to innovate and create better information systems: while isn’t ideal to have free text fields or non-mandatory ones because can impact the quality of the data, their mere presence can be seen as a way to bring updates and innovations to the HIS, for example to promote the use of data analysis algorithms and machine learning.

Regarding the main challenges to implement a patient registry from the data perspective point of view, the participants evaluated this from three different angles: data update, data quality and data coverage, and it’s interesting to notice that a 60% of participants agree or strongly agree that data update is a main challenge while implementing a patient registry.

From this can be inferred that updating data can be se seen as a monotonous or repetitive task, hence makes sense to follow the US’ Agency for Healthcare Research and Quality recommendation of integrating data collection procedures in the day-to-day practice of the healthcare practitioners, which would allow them to enter and update data with minimal disruption to their regular work processes. [7]

For the data quality and data coverage that percentage is reduced to 40% and a further 20% for those two categories disagree that they can be a challenge.

Another curious fact from the opinion gathered via the survey to healthcare professionals, is that when they were asked about the small number of patient registries in Estonia, a 100% of them disagreed or strongly disagreed that that was because the country didn't need them and an 80% of them also disagreed or strongly disagreed that was because of a lack of interest.

Following with this line of thought, 100% of the doctors surveyed answered "Yes" to the question of *"Is it important to have a patient registry for MS in Estonia?"*, it's important to notice two factors which could had influence on them to give this answer: all of them are neurologists and they work with MS patients. Those two factors could make them more aware of the need of a patient registry since they realize they lack a tool to have aggregated information about their patients and benchmark them against other patients and makes difficult the collaboration with international initiatives and research for MS.

To further validate the need of a patient registry, this is inferred from their answers to the question of "what's the main aim of a MS patient registry?", in which the most popular opinion was "have a better overview of the MS population in Estonia", which also confirms that the healthcare professionals need and want the patient registry as a tool to better serve their patients.

Two other popular answers to this question were "collect information for scientific research" and "to obtain information about treatment regimens and efficacy", which closely aligns to the benefits of patient registries mentioned by Jeffrey P. Trotter in his article Patient Registries: A New Gold Standard for "Real World" Research. [20]

Those topics were further referred when the participants were asked their opinion about the primary use of the data of an MS patient registry, where the most common answer to this was "study of MS epidemiology", but also "scientific research", "analysis of treatment availability", "better overview of individual disease courses and patient cohort", "can be used for reimbursement issues", "treatment regimens comparisons", "long-term comparison", "quality" and "management" were mentioned. This leads the author of this research to believe that the healthcare professionals besides using the patient registry as a tool to provide better treatment for their patients and study about MS they are also interested in other aspects, like healthcare quality and financing.

While an 80% of the survey participants had a neutral opinion regarding the fact that legal/ethical concerns could be a reason why Estonia has such a small number of patient registries, a 60% of the participants agreed or strongly agreed that lack of resources was to blame for the lack of patient registries (with the further 40% of responses having a neutral opinion about this). While sadly, isn’t possible to know in this instance if the lack of resources refers to human, material, or financial, this should be kept in mind for further studies regarding patient registries in Estonia.

To prove the strong link between the use of data for medical decisions and how having that data present in HIS can help healthcare professionals to take more informed decisions, lies on the fact that 100% of the respondents strongly agreed that number of relapses is an important piece of data to guide them on decisions for their patients for MS and on the analysis of the fields needed to implement a patient registry in Estonia, for the category of relapses, of the 18 fields recommended by MSBase for this particular category, a 100% of them was found to already exist in the HIS. Unfortunately, 100% of them were free text fields and 0% of them were mandatory, which makes the author believe that the information contained on them lacks structure and data quality which would make it more difficult to extract and analyze if is transferred to a patient registry.

For the category of MRI activity, an 80% agreed or strongly agreed that this data helped them guide their treatment decisions for their MS sclerosis patients, this also successfully transferred to the data fields analysis, since a 100% of the data fields for the MRI category where found to be present in the HIS. While 100% of those fields were also mandatory, all of them are stored as free text fields.

Regarding the progression category, an 80% agreed or strongly agreed that this data helped them guide their treatment decisions for their MS sclerosis patients, but MSBase in his data dictionary doesn’t include a category for progression, instead, fields related to this area are spread across several categories in this file. [15]

As MS is a disease classified in three different courses (relapsing-remitting multiple sclerosis, primary progressive multiple sclerosis, and secondary progressive multiple sclerosis), all the healthcare professionals agreed that patients in the three courses should be included in the patient registry. [4]

Regarding retrospective data quality, the healthcare professionals rated it across five different categories, and their opinions were measured ranging from highest quality to lowest quality:

For MS first symptoms category, a 60% rated the data quality as average and a further 40% mentioned it was of good quality, this could directly correlate to the fact that for the symptoms category in the analysis of the fields, of the 39 fields mentioned in the MSBase data dictionary, 38 of them are already present in the HIS, equal to a 97.44% of them, but of the 39 fields, 33 of them (84.62%) are free text fields and only 5 fields (12.82%) are marked as mandatory, which makes believe the author of this research that while a large percentage of the information can be entered, due to the nature of the free text fields and the lack of enforcement product of not many of them being mandatory, the data quality isn’t as good as could be.

For the MS diagnosis category, an 80% of them rated the data quality as good and a further 20% rated it as of the highest quality. This good opinion about the data quality is also reflected in the fact that in the data fields analysis, of the 24 fields in the MS diagnosis category which are part of the MSBase data dictionary, 21 of them (87.50%) already exits in the HIS. Of this amount, 18 (75% of the total) fields are considered as free text field and only 6 (25% of the total) are marked as mandatory.

For the relapse history category, 20% rated it as of bad quality, a 40% as average and another 40% rated it as of good quality. Those mixed opinion in the middle ranges, can be understood based on the fact that while 100% of the 18 fields recommended by MSBase are present in the HIS, all of them are free text fields and none is a mandatory one.

Regarding the relapse treatment history, this category had really mixed opinions: a 20% regarded it as of bad quality, 40% rated as average quality, 20% as good quality and a further 20% of highest quality. Those results could be assumed based on the fact that the MS specific category of the data fields analysis, besides being quite small, only nine fields recommended by MSBase only seven of them (77.78%) are present in the HIS.

And while is optimistic that none of them is a free text field, only one is marked as mandatory.

For the EDSS in time points, the opinion of the healthcare professionals ranged on the low end of the scale: a 20% rated as lowest quality, 40% rated it as of bad quality and a further 40% as average. In the analysis of the data fields, while the EDSS category has all 11 fields present in the HIS for a matching rate of 100%, all of those fields are free text field, and none is marked as mandatory. This high amount of free text fields for sure has an impact in the data quality and explains the low opinion the healthcare professionals have of this category.

Regarding the opinions of the healthcare professionals for data quality of MS medications like interferons, GA, teriflunomide and dimethyl fumarate and of the data quality of MS hospital budget medications (natalizumab, ocrelizumab, alemtuzumab and fingolimod), all those categories received really high mark, where the lowest percentage of highest quality score was 60% of respondents and for some others this percentage was raised up to 100%. This could be understood from the fact that the MS specific category, which covers MS treatments, while only a 77.78% of the fields are present (seven out of nine), none of them is a free text fields, which makes believe the author of this research that the quality of the data is good.

As can be seen from the previous discussion, is obvious that the answers and opinions given by the healthcare professionals closely matches the overall state of the fields suggested in MSBase data dictionary with those present in the HIS: those categories which were better rated by the healthcare professionals, have a higher percentage of fields present in the HIS.

This has implications for further developments of patient registries: healthcare professionals are trying to introduce data of good quality (even if at the moment when they do it they don’t directly associate good quality of data to patient registries), which can help the future projects of data migration from HIS to a patient registry database much easier.

Also, is important to notice that while designing or updating the HIS further consideration should be put in making the data more structured: reduce the number of free text fields, make more of them mandatory.

A prime example of this, is shown in the results of the data analysis for all the fields of laboratory tests, which besides having good percentage of data present and of mandatory fields, none of them includes free text fields. Obviously, this can be explained due to the nature of laboratory tests, which always must include a value (the result of the test), the unit used to measure the result and if the result was normal or abnormal, but can be also used as proof of the importance of having required and well-structured data.

Situation like this would lead to greatly simplify data extraction not only for patient registries, but also for data analysis of epidemiological surveillance, analysis related to financial spending, allocation and use of resources (human or material) and in overall all areas which could benefit from data science.

6 Conclusions

As information technologies grow and the ability to analyze larger volumes of information grows, it’s important that the healthcare sector pays attention to the power that information technologies can bring to the field.

One area with unmatched potential is patient registries: as can be proven by this research, a large amount of the data needed to establish a patient registry for MS patients in Estonia is already contained in the HIS, just needs to be organized and structured to be interpreted in a meaningful way.

Besides this, is important that healthcare and information technology professionals sit together and listen to each other about how they can collaborate together: while healthcare professionals deal with the human body and the technology ones do it with computers and machines, their aims are always the same: make life better for people.

7 References

[1] Cohen J, Trojano M, Mowry E et al., “Leveraging real-world data to investigate multiple sclerosis disease behavior, prognosis, and treatment.,” *Multiple Sclerosis,* vol. 26, no. 1, pp. 23-37, 2020.

[2] Iaffaldano P, Butzkueven H, Hillert J et al., “Big Multiple Sclerosis Data network: data sharing among five large MS registries”.

[3] National Multiple Sclerosis Society, “Definition of MS : National Multiple Sclerosis Society,” [Online]. Available: https://www.nationalmssociety.org/What-is-MS/Definition-of-MS. [Accessed 10 March 2020].

[4] National Multiple Sclerosis Society, “Types of MS : National Multiple Sclerosis Society,” [Online]. Available: https://www.nationalmssociety.org/What-is-MS/Types-of-MS. [Accessed 10 March 2020].

[5] National Multiple Sclerosis Society, “Brochure-The-MS-Disease-Modifying-Medications,” 2019. [Online]. Available: https://www.nationalmssociety.org/NationalMSSociety/media/MSNationalFiles/Brochures/Brochure-The-MS-Disease-Modifying-Medications.pdf. [Accessed 10 March 2020].

[6] Eesti Haigekassa, “Eesti Haigekassa Tervishoiuteenuste Loetelu Muutmise Taotlus Koos Täitmisjuhistega,” 2019. [Online]. Available: https://www.haigekassa.ee/sites/default/files/TTL/2019/1295_taotlus.pdf. [Accessed 10 March 2020].

[7] R. Gliklich, N. Dreyer and M. Leavy, Registries for Evaluating Patient Outcomes: A User's Guide, Rockville, MD: Agency for Healthcare Research and Quality, 2014.

[8] Tervise Arengu Instituut, “Registers,” 1 December 2015. [Online]. Available: tai.ee/en/r-and-d/registers. [Accessed 11 March 2020].

[9] Riigikogu, “Health Services Organisation Act,” 2002.

[10] K. V., “Imputation of injury data of uninsured patients using Estonian Health Insurance Fund,” National Institute for Health Development, Tallinn, 2019.

[11] M. Zaletel and M. Kralj, “Methodological guidelines and recommendations for efficient and rational governance of patient registries,” National Institute of Public Health, Slovenia, Ljubljana, 2015.

[12] T. D. Spelman, Multiple Sclerosis MSBase registry: using real-world data to define MS outcomes and optimise treatment strategies, Melbourne: The University of Melbourne, 2016.

[13] H. Butzkueven, J. Chapman, E. Cristiano, F. Grand'Maison, M. Hoffmann, G. Izquierdo, D. Jolley, L. Kappos, T. Leist, D. Pöhlau, V. Rivera, M. Trojano, F. Verheul and J.-P. Malkowski, “MSBase: an international, online registry and platform for collaborative outcomes research in multiple sclerosis,” *Multiple Sclerosis,* pp. 769-774, 2006.

[14] MSBase, “MSBase Registry Quick Start Guide,” [Online]. Available: msbase.org/media/1179/msbase-registry-quick-start-guide.pdf. [Accessed May 2020].

[15] MSBase, “User Guide,” [Online]. Available:

https://www.msbase.org/media/1197/mds-data-dictionary.pdf. [Accessed May 2020].

[16] Multiple Sclerosis International Federation, “The Next Atlas of MS is underway,” 24 April 2019. [Online]. Available: https://www.msif.org/news/2019/04/24/the-next-atlas-of-ms-is-underway/. [Accessed May 2020].

[17] P. Browne, D. Chandraratna, C. Angood, H. Tremlett, C. Baker, B. V. Taylor and A. J. Thompson, “Atlas of Multiple Sclerosis 2013: A growing global problem with widespread inequity,” *Neurology,* p. 1022–1024, 2014.

[18] E. C. Nelson, M. Dixon-Woods, P. B. Batalden, K. Homa, A. D. Van Citters, T. S. Morgan, E. Eftimovska, E. S. Fisher, J. Ovretveit, W. Harrison, C. Lind and S. Lindblad, “Patient focused registries can improve health, care, and science,” *BMJ,* 2016.

[19] World Federation of Hemophilia, “Fact Sheet,” [Online]. Available: http://www1.wfh.org/publication/files/pdf-1381.pdf. [Accessed May 2020].

[20] J. P. Trotter, “Patient Registries: A New Gold Standard for “Real World” Research,” *Ochsner J.,* vol. 4(4), p. 211-214, 2002.

[21] Tervise ja Heaolu Infosüsteemide Keskus, “Standardite ja klassifikaatorite publitseerimiskeskus,” 2020. [Online]. Available: http://pub.e-tervis.ee/manuals. [Accessed May 2020].

[22] Tervise ja Heaolu Infosüsteemide Keskus, “ELHR,” [Online]. Available: https://elhr.digilugu.ee/data/algandmedList.html. [Accessed May 2020].

Appendix 1 – Survey to healthcare professionals

1) Which is your current position?

2) In which institution you primarily practice?

3) How many years of professional experience do you have?

4) In your opinion, is it important to have a patient registry for MS in Estonia?

Yes / No

5) Please comment shortly what would be the main aim of MS registry:

6) Currently which data do you use to guide your treatment decisions for patients with MS?

	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
Number of relapses					
MRI activity					
Progression					

7) In your opinion, which are the main challenges while setting a registry from a data perspective point of view?

	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
Data update					
Data quality					
Data coverage (how many patients are included)					

8) According to Tervise Arengu Instituut (TAI), there are only six patient registries in the country. In your opinion, why is this number so small?

	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
No need					
Legal/Ethical concerns					
Lack of resources					
Lack of interest					

9) Which do you consider is the main use of the data collected in Estonian registries?

10) For the patient registry for MS, which do you think the data would be primarily used?

11) Who should be included in the patient registry?

	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
Patients with RRMS					
Patients with PPMS					
Patients with SPMS					

12) How would you rate the quality of the data in the process of retrospective data collection?

	Highest quality	Good quality	Average	Bad quality	Lowest quality
MS first symptoms					
MS diagnosis					
Relapse history					
Relapse treatment history					
EDSS in time points					

13) How would you rate the quality of the data in the process of retrospective data collection on MS prescription medications?

	Highest quality	Good quality	Average	Bad quality	Lowest quality
Interferons					
GA					
Teriflunomide					
Dimethyl fumarate					

14) How would you rate the quality of the data in the process of retrospective data collection on MS hospital budget medications?

	Highest quality	Good quality	Average	Bad quality	Lowest quality
Natalizumab					
Ocrelizumab					
Alemtuzumab					
Fingolimod					

15) How would you rate the quality of the data in the process of prospective data collection on MS medications?

	Highest quality	Good quality	Average	Bad quality	Lowest quality
Interferons					
GA					
Teriflunomide					
Dimethyl fumarate					
Ocrelizumab					
Alemtuzumab					
Fingolimod					

Appendix 2 – MSBase data dictionary categories

MSBase data dictionary is grouped into 27 different categories, each one explained in detail here.

Patient Profile: This category captures basic demographic details, for example age, sex and other demographic details of the patient.

Medical History: Captures information about the patient’s medical history.

Family History: This category is an important one since allows to capture familial medical conditions.

MS Diagnosis: The diagnosis form allows the user to enter information about the MS diagnosis of the patient. It is in this section that the MS Course of the patient is calculated (also allows a RIS or NMO diagnosis).

Basic details: Provides basic information about the patient, like marital status, weight, height.

Symptoms: This category allows the user to enter any symptoms being experienced by the patient for that visit.

Expanded Disability Status Scale (EDSS): The Expanded Disability Status Scale (EDSS) is a method of quantifying disability in MS. This category allows the neurologist to enter the scores for each functional system.

Multiple Sclerosis International Quality of life questionnaire (MusiQoL): A self-administered and multidimensional questionnaire. There are 31 questions over 9 dimensions, with a final score being an average of these 9 dimensions.

Safety: The category allows the user to report changes in the medical conditions of the patient, such as malignancies, adverse reactions or event.

Relapses: This category captures the date and details of MS relapse events.

Malignancy: The malignancy category allows to record occurrences and details of malignancy.

NMSC: This category allows the user to record occurrences and details of NMSC.

Herpes Zoster: This category allows the user to record occurrences and details of herpes zoster.

Immuno-suppression associated: Category which allows the user to record occurrences and details of infection.

Pregnancy: The pregnancy category allows the user to record occurrences and details of pregnancy.

Other Events: This category captures other medical events.

MRI: The MRI category allows user to record Magnetic Resonance Imaging (MRI) scan details.

CSF: This category provides details on the collection of Cerebral Spinal Fluid (CSF).

Evoked Potentials: This category records details about the evoked potentials exam, measuring rate of conduction and amplitude of the nerve impulses.

Laboratory Tests (Hematology): The hematology category, part of the broader laboratory tests, enables the user to enter data from exams and comment where the test value is abnormal.

Laboratory Tests (Blood chemistry): The blood chemistry category, part of the broader laboratory tests, enables the user to enter data from exams and comment where the test value is abnormal.

Laboratory Tests (Thyroid function): The thyroid function category, part of the broader laboratory tests, enables the user to enter data from exams and comment where the test value is abnormal.

Laboratory Tests (Serological tests): The serological tests category, part of the broader Laboratory Tests, enables the user to enter data from exams and comment where the test value is abnormal.

Laboratory Tests (Auto-antibody Tests): The auto-antibody tests category, part of the broader Laboratory Tests, enables the user to enter data from exams and comment where the test value is abnormal.

MS Specific: This category identifies MS specific treatments.

Symptomatic: This category identifies symptomatic treatments.

Non-pharmacological: This category identifies non-pharmacological treatments.