## PERMIT – PERsonalised Medicine Trials
### Methods for stratification and validation cohorts

**Work Package 2 - Deliverable 2.1**

<table>
<thead>
<tr>
<th>Deliverable no</th>
<th>2.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deliverable Title</td>
<td>Methods in stratification /validation cohorts</td>
</tr>
<tr>
<td>Contractual delivery month</td>
<td>October 2020 (M10)</td>
</tr>
<tr>
<td>Responsible Partner</td>
<td>Parc Sanitari San Joan de Déu (PSSJD)</td>
</tr>
<tr>
<td>Author(s)</td>
<td>Teresa Torres Moral, Albert Sánchez-Niubó, Josep Maria Haro (PSSJD), Rita Banzí, Chiara Gerardi (Istituto di Ricerche Farmacologiche Mario Negri), Vanna Pistotti (CAeB) Paula Garcia, Jacques Demotes-Mainard (ECRIN)</td>
</tr>
<tr>
<td>Other contributors</td>
<td>Anna Monistrol (PSSJD)</td>
</tr>
</tbody>
</table>
Executive summary

Background
Personalised medicine is going to bring many changes that will not only influence patients’ diagnosis and treatment, but also the way pharmacological treatments are discovered and developed and even the allocation of resources in the health care system so that patients’ benefits are maximized.

Despite the diverse definitions and understandings of personalised medicine, three main positions can be identified; (a) personalised medicine is not a new concept, as medicine has always been individualized; (b) personalised medicine is holistic health care, centred around the needs of the individual patient; (c) personalised medicine is treatment targeted at stratified subgroups.

Regardless of differences in concept, definition or application, any approach to personalised medicine should undergo two different phases: discovery and validation, and definition of usefulness from a clinical perspective. Robust methodological approaches are needed to deal with the complexity and heterogeneity of the process, as well as the range of possible applications of stratification using multidimensional data, or “molecular profiling”, among other terms.

The partners of the PERMIT project agreed on a common operational definition of personalised medicine research. It is a set of comprehensive methods to be applied in the different phases of the development of a personalised approach to treatment, diagnosis, prognosis, or risk prediction. Ideally, robust and reproducible methods should cover all the steps between the generation of the hypothesis, its validation and pre-clinical development, up to the definition of its value in a clinical setting.

There are some common steps in any research programme in personalised medicine: first, stratification algorithms are run in a cohort most often an existing cohort, with extensive multimodal data and biosamples. This is the ‘stratification cohort’. Second, the reproducibility, robustness and validity of the clustering are assessed in another sufficiently large patient sample, which is called ‘validation cohort’. Third, a translational step to test validity and security is often necessary.

In some cases, the use of pre-clinical models (cellular, in-silico, organoid) might be useful to give confidence in the allocation of patients to specific treatment arms as identified through clustering. Alternatively, the multi-omics profiles from clinical samples can lead to the identification of new disease categories, prediction of disease prognosis, exploration of drug sensitivity and dose selection. Treatment options should be tested in the subgroups of patients in the context of clinical studies, ideally randomised clinical trials, to generate evidence informing regulatory, clinical and coverage decisions.

In this scoping review, we focus on the design of stratification and validation cohorts. Three case models have been explored: oncology, Alzheimer’s disease and stroke. These were chosen for their big impact on society and individual health, because they are in three different phases of personalised medicine, and because they cover different kinds of data to stratify patients.

Research Questions
The main research questions addressed by the scoping review are the following:

1. What are the approaches to define the optimal size of stratification/validation cohorts?
2. What are the differences, pros and cons of the prospective and retrospective nature of stratification and validation cohorts?
3. What are the prerequisites and methods used for the integration of multiple retrospective cohorts?
4. Which validation designs exist for the stratification (or clustering) in personalised medicine? Which methods and tools are used to build the cohorts of validation (external/sub-cohort)? What are their gaps?
5. What are the methods for the evaluation of the risk of bias?
6. How are the (-omics, imaging, exposome, lifestyle etc.) data generated?
7. What are the tools used for data management and multimodal data analysis used in personalised medicine (for instance, Galaxy)? What are their gaps?
8. What quality of data of cohorts is needed to obtain a biomarker or multimodal data profiling? Are there requirements to monitor the collection of associated clinical data?
9. What is the outlook of data generation seen as (CE-labelled) in-vitro diagnostics?

Methods
We conducted a scoping review following the methodological framework suggested by the Joanna Briggs Institute[1]. A study protocol reporting all methodological details was uploaded in the Zenodo repository before conducting the review and we used the PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) checklist to report our results.

Regarding study identification, we searched PubMed, EMBASE and the Cochrane Library for (systematic) reviews in the fields of cancer, stroke, and Alzheimer's disease to first identify the most common methodological approaches.

We focused on the description of the methods applied to cohorts that were used for patient stratification 'stratification cohorts' or for validation of patient clustering 'validation cohorts'. We limited the searches to reports in English, French, German, Italian and Spanish published from 2005 to April 2020.

Screening process was conducted by two reviewers (PSSJD) that screened independently all titles and abstracts of the retrieved citations to have a first selection of papers. Later, one of the reviewers (PSSJD) examined all full text of papers. The result of this process was reported through a PRISMA flowchart.

Results
Overall, the database searches retrieved 2362 records. Nine additional records were retrieved through the included reviews and grey literature. After the screening process, 50 reviews were included: 19 focused on oncology, two on Alzheimer's disease, two on stroke. Twenty-seven reviews analysed methods in one of our case models (oncology, Alzheimer's or stroke) but also included other illnesses. Considering that the information provided by these papers was relevant to the aim of review, we decided to include them. These reviews are referred to as “multiple disease reviews”.

Regarding question one (approaches to define the optimal size of stratification/validation cohorts), a review on Alzheimer’s Disease shows that different approaches to define optimal sizes and different elements, as target population or effect size, must be taken into account. The authors highlight that depending on the statistical model, the power calculations tend to be different. A review on cancer suggests that recommended sample sizes for training and testing samples in retrospective studies include about 110 subjects without cancer and 70 subjects with cancer in the test sample, and likely at least the same size for the training sample. In prospective studies, the sample size must be larger. Two multiple disease reviews point out that for each study of a candidate predictor, at least 10 events are required for categorical outcomes. For the prospective studies of time-to-event outcomes, at least several hundred outcome events are needed. For continuous outcomes, the effective sample size is determined by the number of participants included in the linear regression analysis.

Regarding question two (differences, pros and cons of the prospective and retrospective cohorts), the advantages reported of the prospective design are that it enables optimal measurement of predictors and outcome, the risk of recall bias is reduced or that causality is easier to establish. It is the preferred design for prognosis studies on oncology, and the results can be used as platforms for biomarker research. On the other hand, the advantages shown of retrospective design are that there is no
requirement for follow-up of participants, the costs are lower, and it enhances the ability to investigate rare tumours. For these reasons, some reviews recommend retrospective cohort design in some cases, if data are available. More information about more specific designs has been collected in this scoping review.

About question three (prerequisites and methods used for the integration of multiple retrospective cohorts), only one review was found. It focused on oncology and showed approaches concerning data aggregation (putting together different microarray experiments to form a single dataset) and meta-analysis (analysing each individual microarray experiment and then aggregating the statistical results of all individual experiments).

Addressing question four (validation designs for stratification; methods and tools used to build the cohorts of validation and their gaps) and regarding stratification cohorts, a review about Alzheimer’s disease concludes that a comprehensive multimodal definition will be useful for building stratification cohorts in combination with current advances in methods and power for computational analytics. A multiple disease review focuses on the classification of methods of stratification, and it found 5 types: clustering, dimensionality reduction, similarity, software tools and combination of clustering or similarity metrics and supervised approaches. A review focusing on methods for the staging of cancer concludes that these methods are suboptimal for treatment recommendation because they are not based on a multidisciplinary approach.

Regarding validation cohort building, a review about stroke concludes that propensity scoring and one-to-one matching are good methods, but propensity score requires a larger pool of enrolled controls and one-to-one matching requires computer programming support. Focusing on oncology, several validation cohort methods were classified into two groups: a) external validation (Conducted on different data used to generate the results) that is rarely possible, and b) internal validation (Conducted on the same data used to generate the results) that needs complex cross-validation procedures to assure the stability of the biomarker. One of the multiple disease reviews focuses on three designs in pharmacogenetic studies: targeted (or enriched), stratification or adaptive design. The other reviews focus on the advantages and disadvantages of the different methods found and classify them in external and internal validation methods.

Related to question five (methods for the evaluation of the risk of bias), and regarding oncology, two methods for addressing bias due to lack of data have been analysed: inverse probability weighting and multiple imputation. The conclusion is that the risk of bias is unavoidable, but these approaches can mitigate it. Multiple disease reviews found that to avoid bias it is very important to take into account its source, and analyse each source (Within-Subject Correlation, Multiplicity, Multiple Clinical Endpoints, Selection bias and Publication bias) and the possible solutions. In addition, the specific tool for the evaluation of bias ‘Quality in Prognosis Studies (QUIPS) tool’ was analysed and the analysis showed that three sources of risk of bias are usually studied: outcome measurement, study confounding and statistical analysis and reporting.

Regarding question six (how data are generated), reviews found focus on specific kinds of data. The papers were classified by the type of data that they analyse and show that, when compared to all other types of data, image and environmental data on different biological layers are hard to measure quantitatively, and that it is important to consider the clinical context of molecular signatures.

Related with question seven (tools used for data management and multimodal data analysis), a large diversity of methods was identified, and the strengths and limitations of the different methods were collected in order to facilitate the choice of the most appropriate one for each study. As a general conclusion, oncology reviews found that an important limitation is that integrative data sets often do not have a standard format for research use. Other reviews point out that current research needs to develop and improve multi-layer data integrative methods for multi-omics derived data to provide tailored therapies.
About question eight (quality of data of cohorts needed to obtain a biomarker or multimodal data profiling, requirements to monitor the collection of associated clinical data), we have not found specific information, although this question was partially answered in other questions of this review. We expected to find standards or some level of regulatory guidance in this area.

Regarding question nine (the outlook of data generation and its consideration as ‘in-vitro diagnostics’), information found points to the new regulation for in vitro diagnostics to be applied to the European Union before May 2022. Although it is difficult to anticipate the effects of this new regulation, it is likely that in vitro diagnostics will go through an exponential increase, implying that some diagnostic tools will probably be commercialized soon and will become easy to find.

**Gaps identified**

When the review was being conducted, we detected some areas with a scarcity of information or lack of standards that we would like to focus on, during the Gap Analysis Workshop. These areas include: sample size calculation, the prerequisites and methods used for integration of multiple retrospective cohorts, the quality of data of cohorts needed to obtain a biomarker or multimodal data profiling and the requirements to monitor the collection of associated clinical data in biomarkers studies.
DISCLAIMER

This document contains information which is the proprietary to the PERMIT Consortium. Neither this document nor the information contained herein shall be used, duplicated or communicated by any means to any third party, in whole or in parts, except with prior written consent of the PERMIT Coordinator.

The information in this document is provided "as is", and no guarantee or warranty is given that the information is fit for any particular purpose. The document reflects only the authors’ view. The PERMIT Consortium is not responsible for any use that may be made of the information it contains. The user uses the information at its solo risk and liability.

Document log

<table>
<thead>
<tr>
<th>Issue</th>
<th>Date (yyyy-mm-dd)</th>
<th>Comment</th>
<th>Author/partner</th>
</tr>
</thead>
<tbody>
<tr>
<td>Version 1.0</td>
<td>2020-11-30</td>
<td></td>
<td>PSSJD</td>
</tr>
</tbody>
</table>
### Table of Contents

Executive summary ............................................................................................................. 1  
Background .......................................................................................................................... 7  
Approaches (Methods) ......................................................................................................... 11  
Results .................................................................................................................................. 14  
Summary of findings and next steps..................................................................................... 36  
Acronyms ............................................................................................................................... 39  
References ............................................................................................................................. 40  
Appendix I: Searches ............................................................................................................. 45  
Appendix II: Outline of methods and tools for data-analysis and for data integration 59  
Appendix III – Full scoping review protocol......................................................................... 63
Background

The concept of personalised medicine is going to impact how pharmacological treatments are discovered and developed, how patients are diagnosed and treated, and how health care systems allocate their resources to maximize patient benefits.

Personalised medicine may be considered an extension of traditional approaches to understanding and treating disease. Ideally, it could serve to take clinical decisions based on a patient's profile (often molecular, but the concept is broader) to minimise harmful side effects, ensure a more successful outcome, and possibly help contain costs compared with a “trial-and-error” approach to disease treatment [2].

Personalised medicine stems on the broad concept that managing a patient's health should be based on the individual patient's specific characteristics, including age, gender, height/weight, diet, environment, etc. Different understandings of personalised medicine exist, in which three main positions can be identified [3]:
(a) personalised medicine is not a new concept as medicine has always been individualized;
(b) personalised medicine is holistic health care, centred around the needs of the individual patient;
(c) personalised medicine is treatment targeted at stratified subgroups (e.g. pharmacogenetics).

Even when the focus is restricted to the third position, there is not a unique definition of personalised medicine, nor a straightforward terminology to define this concept. While “personalised” emphasizes the notion of individualized— “this is exclusively designed for you”, other more scientifically rigorous terms such as stratified medicine refer to the identification of groups or strata of patients with specific molecular characteristics or other determining factors which predict susceptibility to disease, disease prognosis, and/or response to therapy. Some authors suggested that rather than considering personalised medicine as a precise scientific concept, it should be understood as an open and negotiable ideal that accounts for a plurality of visions, depending on people, reasons and interests behind these alternative conceptions [4].

Regarding the terminology, in the European context, the term personalised medicine is preferred, as this term best reflects the ultimate goal of effectively tailoring treatment based on an individual’s ‘personal profile’, as determined by the individual’s genetical and phenotypical characteristics. Other terms are widespread, for instance stratified medicine, mainly used in the UK, or precision medicine mostly used in US and broadly referred to 4 P (preventive, predictive, personalised and participatory) medicine. While there may be small nuances in the literal meanings of these terms, they usually refer to the same concept when applied in practice [5].
A recent review reported that the literature about personalised medicine usually refers to two different semantic approaches. Firstly, patients’ stratification, that is grouping individual patients in subpopulation according to their probability to have a therapeutic benefit from a drug or regimen. Secondly, treatment tailoring, that is the individual status of a patient (i.e., disease characteristics or subject’s genotype/phenotype) is the rationale basis for drug choice [6]

A broad community of stakeholders, including funders and people involved in medical research and care, are increasingly concerned with ensuring that the right patient receives the right therapy, at the right dose and at the right time. The identification of markers of mechanistic pathways or multiple variables characterising clusters of subjects that might inform meaningful disease stratification may have different clinical applications in the context of personalised medicine. Broadly, stratification may be applied at the diagnosis level (e.g., to identify a particular pathophysiological/clinical stratum within a heterogeneous patient population for diagnostic purposes), to predict disease course (prognostic value), the development of a disease (predictive value), or the response to therapy (theragnostic value).

Regardless of the application, any approach to personalised medicine should undergo different phases: discovery, validation and definition of usefulness from a clinical perspective. Robust methodological approaches are needed to deal with the complexity and heterogeneity of the process, as well as the range of possible applications to stratification using multidimensional data (what is meant among other by “molecular profiling”).

**Personalised medicine research**

This series of scoping reviews has mapped the general concept of methods for personalised medicine, to set the basis for the discussion on robustness and reproducibility of personalised medicine development programmes. The final goal is the identification of standards and needs in terms of methodology of data generation, management, analysis, and interpretation to improve clinical studies in personalised medicine.

The group of authors agreed on a common operational definition of **personalised medicine research**: a set of comprehensive methods, (methodological, statistical, validation or technologies) to be applied in the different phases of the development of a personalised approach to treatment, diagnosis, prognosis, or risk prediction. Ideally, robust and reproducible methods should cover all the steps between the generation of the hypothesis (e.g., a given stratum of patients could better respond to a treatment), its validation and pre-clinical development, and up to the definition of its value in a clinical setting.
The process leading from the hypothesis to the clinic is complex and not always linear. The Medical Research Council in UK recently developed a framework for the development, design and analysis of stratified medicine [7] that is structured in six themes:

**Theme 1:** Framing the Question/Defining the Population

**Theme 2:** Designing Stratum Discovery Studies; selecting variables, defining response and powering

**Theme 3:** Assay Design; managing complexity and variability

**Theme 4:** Defining Strata; data integration, linkage to existing knowledge, linkage to outcome

**Theme 5:** Stratum Verification

**Theme 6:** Progression Towards Clinical Utility

Any attempt for classifying the phases of personalised medicine may appear as an oversimplification. However, a typical research programme in personalised medicine would include: first a stratification cohort (in many cases a retrospective study reusing data and biosamples from existing cohorts) with extensive multimodal data on which stratification algorithms are run, then a validation cohort, normally prospective, that assesses the reproducibility, robustness and validity of the clustering in another sufficiently large patient sample. Thirdly, a translational step is often necessary. In some cases, the use of pre-clinical models (cellular, in-silico, organoid) might be useful to give confidence in the allocation of patients to specific treatment arms as identified through clustering. Alternatively, the multi-omics profiles from clinical samples can lead to the identification of new disease categories, prediction of disease prognosis, exploration of drug sensitivity and dose selection. Finally, treatment options should be tested in the subgroups of patients in the context of clinical studies, ideally randomised clinical trials, to generate evidence informing regulatory, clinical and coverage decisions.

However, many alternative pathways can be proposed. In some cases, the stratification provides detailed information on the mechanism of disease and strong indications on the treatments to be tested in each patient cluster. This is for instance the case where identification of driver somatic mutations in cancer cells suggests the targeted treatment to be tested. In other cases, the stratification cohort includes data on response to an established treatment, making the translational step less necessary. Research programmes may be limited to the stratification step, in particular when no treatment is available – this is the case for instance for taxonomy studies in neurodegenerative disorders, aiming at identifying homogeneous clusters of patients. In any case, personalised medicine research is a complex programme, with multiple steps and lasting many years.
We considered out of the scope of this review the methods used for the clinical implementation of personalised medicine, the manufacturing and use of individualized treatments, and the pragmatic approach to individual patient care, such as n-of-1 trials.

Considering this framework outlined by Figure 1, the scoping reviews have approached **personalised medicine research** focusing on four main phases:

1. Methods for stratification and validation cohorts
2. Methods for machine learning applied to stratification
3. Pre-clinical methods for translational development of stratified therapies and treatments selection
4. Methods for clinical trials in personalised medicine

**Figure 1: Main steps in personalised medicine research programmes**

**Scoping review on methods for stratification and validation cohorts**

In personalised medicine, cohorts have a very important role: First, stratification algorithms are run in a ‘stratification cohort’. This is a cohort with extensive multimodal data that normally contains data and biosamples from other existing cohorts. Second, the reproducibility, robustness and validity of the clustering are assessed in another cohort with a sufficiently large patient sample. This is called ‘validation cohort’. This scoping review focuses on:

The characteristics of cohorts that have been used for patient stratification or validation of patient clustering obtained through stratification cohorts. Stratification cohorts of patients are used to create the clustering, and validation cohorts of patients are used to assess the reliability (robustness, reproducibility, etc.) of patient clustering. The different methods and tools used in design and management of stratification and validation cohorts (especially complex in multimodal approaches) to understand their limitations.
General papers that describe methods and tools in the design and management of stratification and validation cohorts were assessed irrespective of the diseases field. Case examples of biomarkers or multimodal data profiling in different medical fields and coming from different sources (omics, neuroimaging, genetics…) were also analysed to explore the actual application of these methods and tools. Cancer, stroke and Alzheimer’s disease were the three areas where informative examples were collected. These three fields were chosen for their considerable impact on society and individual health. They are in three different development phases of personalised medicine, which allows us to know different methods and strategies in different levels of development, and they also use different kind of data to stratify patients. Oncology is the field where personalised medicine was firstly applied and where targeted therapies and diagnostics have been focused. Moreover, several applications of biomarkers for the successful stratification of patients with a given type of cancer exist, most of them based on molecular data, specially genomics. Alzheimer’s disease research in personalised therapies and diagnostics is nowadays giving its firsts results, based on imaging, cognitive and molecular data. Stroke is currently opening to personalised medicine, with some approaches and studies in more initial phases. Most of the data for patient stratification are imaging and molecular data. The main research questions planned to be covered by the scoping review are:

1. What are the approaches to define the optimal size of stratification/validation cohorts?
2. What are the differences, pros and cons of the prospective and retrospective nature of stratification and validation cohorts?
3. What are the prerequisites and methods used for integration of multiple retrospective cohorts?
4. Which validation designs exist for the stratification (or clustering) in personalised medicine? Which methods and tools are used to build the cohorts of validation (external/sub-cohort)? What are their gaps?
5. What are the methods for the evaluation of the risk of bias?
6. How are the (-omics, imaging, exposome, lifestyle etc.) data generated?
7. What are the tools used for data management and multimodal data analysis used in personalised medicine (for instance, Galaxy)? What are their gaps?
8. What quality of data of cohorts is needed to obtain a biomarker or multimodal data profiling? Are there requirements to monitor the collection of associated clinical data?
9. What is the outlook of data generation seen as (CE-labelled) in-vitro diagnostics?

Approaches (Methods)

We conducted a scoping review following the methodological framework suggested by the Joanna Briggs Institute [8,9]. The framework consists of six stages: 1) identifying the research questions, 2) identifying relevant studies, 3) study selection, 4) charting
the data, 5) collating, summarising and reporting results and 6) consultation. We will perform the last step through a workshop with partners of the PERMIT project planned on December 1-2, 2020.

A study protocol reporting all methodological details was uploaded in the Zenodo repository before conducting the present scoping review [10] (see Appendix III for full protocol). We used the PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) checklist [11] to report our results.

**Study identification**

We searched PubMed, EMBASE, the Cochrane Library, Web of Science, and PsycInfo (search dates: March-June 2020) for (systematic) reviews in the fields of cancer, stroke and Alzheimer’s disease to first identify the most common methodological approaches. The methods team (Istituto di Ricerche Farmacologiche Mario Negri) defined the search strategies with the support of the review team (PSSJD) for identifying relevant keywords.

We limited our search from 2005 to March-June 2020. We restricted inclusion to English, French, Spanish, Italian and German languages. Search strategies are presented in Appendix I. In addition, a review of grey literature was also conducted to obtain further information (Appendix I).

**Eligibility criteria**

For this report, we focused on reviews and other reports describing the methods applied to cohorts that have been used for patient stratification or validation of patient clustering obtained through stratification cohorts.

We also included reviews on methods to define the optimal size of cohorts, to design these cohorts, to integrate multiple retrospective cohorts, to evaluate risk of bias, and to manage data and analysis in personalised medicine. We were interested in the quality of data and monitoring of associated clinical data requirements and in the legal framework of data generated in personalised medicine.

The scoping review covered a broad range of multimodal data profiling studies and biomarkers based on all kinds of data: genetic, metabolomic, genetic expression, genomic, or radiomic. As already mentioned, we focused on three case models: oncology, Alzheimer’s disease and stroke.

**Study selection**

We exported the references retrieved from the searches into the Rayyan online tool [12]. Duplicates were previously removed by the methods team using the reference manager Endnote X9 (Clarivate Analytics, Philadelphia, United States).
All records were manually verified and duplicates that had not been previously detected were removed. Two reviewers (PSSJD) screened independently all titles and abstracts of the retrieved citations. All papers having a wrong design, focusing on other illnesses or clearly out of our aim were discarded. In case of disagreement, consensus was determined by discussion. The full-text screening was performed by one of the reviewers who examined all full-text copies. We reported the result of this process through a PRISMA flowchart [13].

Charting the data
We designed a data extraction form using an excel file (not shown). Study characteristics extracted were as follows: Title, author(s), year of publication, country of origin, aims/purpose, disease analyzed, disease risk/prognosis of disease or response to treatment, abstract, principal outcome, principal outcome related to the aim of this scoping review, number of publications included, type of data analyzed, conclusions about data generated, sample size, meta-analysis or review, quality of data of meta-analysis, validation method used in meta-analysis, tools for data-analysis used, methods to define size of cohort analysed, conclusions about size of cohorts, articles included about sample size, information regarding prospective and retrospective cohorts for studies of personalised medicine, conclusions about prospective and retrospective nature of cohorts in studies of personalised medicine, articles included about sample size, information regarding prospective and retrospective nature of cohorts, stratification cohort methods, stratification cohort tools, conclusions about stratification cohort methods and tools, articles included about stratification cohort methods and tools, methods for data-analysis used in personalised medicine, methods of different kind of data integration, tools for data-analysis in personalised medicine, conclusions about data-analysis in personalised medicine, articles included about data-analysis in personalised medicine, methods of validation cohort construction, tools used for validation cohort construction, conclusions about validation cohort construction, articles included about validation cohort construction, methods for cohort integration analysed, tools used for cohort integration analysed, conclusions about cohort integration in personalised medicine, articles included about cohort integration in personalised medicine, methods of risk of evaluation of bias, tools used for analysis of risk of bias, conclusions about risk of bias in personalised medicine, articles about risk of bias, information in the article about quality of data needed in personalised medicine and about the need of monitoring of associated clinical data in personalised medicine, conclusions about requirements in personalised medicine, articles included about requirements in personalised medicine, information about data generation seen as in-vitro diagnostics, conclusions about data generation in personalised medicine seen as in-vitro diagnostics, articles about data generation seen as in-vitro diagnostics included and conclusion of the study.

Collating, summarising, and reporting results
All relevant information of each review included was organized by questions in a Word document. After that, all the information of each question was put together, and then summarized and rewritten.

**Results**

Results of the study selection and general characteristics of reports
The screening process is summarized in a flow diagram (Figure 2). Out of the 2362 records retrieved, we finally included 50 reviews. We conducted the searches of the three illnesses: oncology, Alzheimer's disease and stroke, but we observed that some reviews analysed methods and they also included other illnesses, together with oncology, Alzheimer's or stroke. Considering that the information provided by them was relevant to the aim of review, we decided to include them. These reviews are referred to as “multiple disease reviews”. In the PRISMA flow chart we have added one more box with this information.
Records identified through database searching n = 2362
Oncology n = 1889
Stroke n = 227
Alzheimer’s disease n = 246

Records identified through reviews included n = 4
Records identified through grey literature n = 5

Records after duplicates removed
Oncology n = 1724
Stroke n = 223
Alzheimer’s disease n = 240

Records screened n = 2187
Oncology n = 1724
Stroke n = 223
Alzheimer’s disease n = 240

Records excluded n = 2034
Oncology n = 1602
Stroke n = 216
Alzheimer’s disease n = 216

Full-text articles assessed for eligibility n = 153
Oncology n = 122
Stroke n = 7
Alzheimer’s disease n = 24

Full-text articles excluded n = 103
Oncology n = 77
Stroke n = 5
Alzheimer’s disease n = 21

Total reviews included n = 50
Oncology n = 19
Stroke n = 2
Alzheimer’s disease n = 2
Multiple disease n = 27

Figure 2. PRISMA* Flowchart of the scoping review.

A summary of results found for each research question is shown in Table 1.

<table>
<thead>
<tr>
<th>Question</th>
<th>Reviews regarding oncology</th>
<th>Reviews regarding Alzheimer’s disease</th>
<th>Reviews regarding stroke</th>
<th>Multiple disease reviews</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. What are the approaches to define the optimal size of stratification/validation cohorts?</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>2. What are the differences and pros and cons of the prospective and retrospective nature of stratification and validation cohorts?</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>3. What are the prerequisites and methods used for integration of multiple retrospective cohorts?</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4. Which validation designs exist for the stratification (or clustering) in personalised medicine? Which methods and tools are used to build the cohorts of validation (external/sub-cohort)? What are their gaps?</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>5. What are the methods for the evaluation of the risk of bias?</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>6. How are the (-omics, imaging, exposome, lifestyle etc.) data generated?</td>
<td>10</td>
<td>2</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>7. What are the tools used for data management and multimodal data analysis used in personalised medicine, i.e. Galaxy? What are their gaps?</td>
<td>8</td>
<td>0</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>8. What quality of data of cohorts is needed to obtain a biomarker or multi-modal data profiling? Are there requirements to monitor the collection of associated clinical data?*</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>9. What is the outlook of data generation seen as (CE-labelled) in-vitro diagnostics?</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 1. Summary of results of reviews found

*Regarding question 8 we found that the information was already given in the answers to questions 1, 4 and 7
1. What are the approaches to define the optimal size of stratification/validation cohorts?

Information about the optimal size of stratification/validation cohorts was found in four reviews: one about Alzheimer [14], one about oncology [15] and two including multiple diseases [16,17].

The review about Alzheimer’s disease (AD) [14] explains that there are different approaches to define optimal sizes and that different elements must be taken into account, not only statistical calculation. These elements include:

- **Effect size**: When defining effect size as a percentage reduction in mean rate of decline, a smaller assumed mean rate of decline under the null hypothesis translates to smaller effect sizes powered for larger required sample size.

- **Target population**: The more restrictive criteria in the target population, the more statistical power obtained (less sample size needed).

- **Additional considerations**: Differences in the imaging processing methods and additional considerations as pilot data (e.g., ADNI data). Pilot data can be used to test initial implicit assumptions or to adjust sample size calculations and accommodate study subject dropout or loss to follow.

The authors also highlight that depending on the statistical model, the power calculations tend to be different. On one side there are two-stage “summary measures” analyses which require only the assumption that summary measures (i.e., change scores or least-squares slopes) are independent, identically distributed asymptotically normal random variables. On the other hand, parameterized longitudinal models and analysis plans can be used as linear mixed models either assuming longitudinal trajectories of decline are linear within-subject or assuming subjects have random intercepts but identical rates of decline within the arm. This latter results in general with smaller sample sizes. Moreover, when trials allocate unequal sample size in each arm, they are slightly less efficient and require a modest adjustment in the calculation of sample size, leading to a larger sample size required.

The review focused on cancer [15] suggests that recommended sample sizes for training and testing samples in retrospective studies include about 110 subjects without cancer and 70 subjects with cancer in the test sample, and likely at least the same size for the training sample. In prospective studies, the sample size must be larger.

The two multiple disease reviews [16,17] point out that, ideally, for each study of a candidate predictor, at least 10 events are required, although a recent study showed that this number could be lower in certain circumstances. This calculation is since, when the number of predictors is much larger than the number of outcome events, there is a risk of overestimating the predictive performance of the model. For the prospective studies of time-to-event outcomes, the effective sample size is also high (at least several hundred outcome events) and it is related to the number of participants.
who experience the event. For continuous outcomes, the effective sample size is determined by the number of participants included in the linear regression analysis.

2. What are the differences and pros and cons of the prospective and retrospective nature of stratification and validation cohorts?

Fourteen reviews related with this question have been found. Five of them involve multiple diseases [16–20], and nine focus on oncology [21–29]. We collect relevant information about pros and cons of these study designs in supplementary tables (S1 and S2)

### Oncology

<table>
<thead>
<tr>
<th>Study design</th>
<th>Sampling of specimens</th>
<th>Auxiliary data</th>
<th>Cons</th>
<th>Pros</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population-based cohort in the general population.</td>
<td>Sampling at baseline for a variety of purposes. Repeated sampling is frequently not possible, e.g., due to budgetary constraints. Biobanking has become an integral part of these studies.</td>
<td>Large set of data about subject and sample. Laboratory parameters not specifically selected for marker research.</td>
<td>Not conducted in a pre-defined at-risk population. Not aimed at marker research. No serial investigations. Insufficient cases with pre-diagnostic samples collected within 12 months prior to diagnosis.</td>
<td>Well-conducted according to good epidemiological practice (GEP). Supervision of the recruitment of subjects and samples. Providing reference samples for case-control comparisons.</td>
</tr>
<tr>
<td>Screening cohort in the target population at risk for the outcome to be detected.</td>
<td>Serial collection of pre-diagnostic samples. Biobanking not an integral part of screening in clinical settings.</td>
<td>Limited information about the subject and sample. Laboratory parameters frequently not according to common SOPs.</td>
<td>Efforts needed to implement GEP into practice. Lack of comparison of the marker(s) with usual care. Voluntary screens with limited compliance to attend the screens regularly.</td>
<td>Serial screens in the target population. Suitable for add-on studies on new markers. Well-described case group if diagnostic workup is performed in the center that conducts the screening.</td>
</tr>
</tbody>
</table>
### Specific prospective design in oncology

<table>
<thead>
<tr>
<th>Study design</th>
<th>Sampling of specimens</th>
<th>Auxiliary data</th>
<th>Cons</th>
<th>Pros</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Prospective, retrospective” studies.</td>
<td>Use of archived specimens collected from patients who have participated in other prospective studies.</td>
<td>Proposed solution to cons: Recently an international registry has been established so that investigators can prospectively document their intention to perform a tumor biomarker study, in a manner similar to “clinicaltrials.gov”. Thus, regardless of which strategy is pursued, an investigator can document that the study protocol, methods and analytical techniques were prospectively considered.</td>
<td>Specimen availability (it is almost always far less than 100%) Pre-analytical concerns and assay failure, and under-powering, since the parent trial is almost always powered for the main therapeutic effect, not the tumor biomarker sub-analysis. They need to be prospectively validated</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2.** Definition and pros and cons of different prospective cohort designs

The reviews that focus only on oncology describe different advantages of prospective stratification and validation cohorts. More detailed subtypes of designs and some of their specific pros and cons have been collected in Table 2. First, the source population that has given rise to tumours is well-defined, which diminishes issues due to selection bias. Risk of recall bias is also reduced. Due to better control of the biases and a better classification or preparation of the samples to be studied, causality is easier to establish. As an example, when examining tumour subtypes classified by the microbial profile in a prospective cohort design, we can examine the association of exposure to a specific factor (which can be microbes) and tumour subtypes. If the microbial data are available before cancer diagnosis, we can link a microbe with the incidence of tumour subtypes classified by molecular pathological signatures of the tumour [22]. Another advantage is that the prospective cohort is the preferred design for prognosis studies on oncology [24]. Finally, it is found that many institutions have highly recommended upgrading prospective designs to serve as platforms for biomarker research [27]. On the other hand, prospective designs are limited because they require large sample sizes, and, if the endpoint is cancer on biopsy, there may be bias due to overdiagnosis [15].

Regarding retrospective approaches, the inherent limitations have been mentioned.
Firstly, biomarker analyses in retrospective patient series are yet to be validated prospectively in broader patient series provided by large clinical trials [23]. Moreover, predictive analytic models based on retrospective data may reinforce existing biases in clinical care. One important bias is the lack of data from unrepresented populations, that could compromise the generalizability of the predictive model [29]. Also, the prognostic estimates cannot be revised. They can be measured only once on the original available samples, ignoring the clinical reality of changes in the level of the marker over time, e.g. as a consequence of treatment [24]. In addition, other limitations of retrospective studies have been found: the attribution of success to alterations being identified simply as actionable or leading to a change in therapy (not causal relationship), the heterogeneity of histologic tumour types, and the inflation of the value of broad-based NGS (Next-generation sequencing) profiling in the setting of including patients with well-characterized alterations (in reported studies) [26].

On the other hand, retrospective studies also have advantages. For example, if stored specimens from cohort studies with clinical cancer endpoints are available, retrospective studies can provide a quick and valid way to evaluate the performance of the markers or changes in the markers before the onset of clinical symptoms [21]. Moreover, there is no requirement for follow-up of participants, lower costs compared with prospective cohort studies, and the ability to investigate rare tumours [22]. Some reviews recommend retrospective cohort design if data are available because it enables cheaply and easily optimal measurement of predictors and outcome [25].

Finally, some studies have been defined as “prospective retrospective” studies. A summary of the features of this design is shown in Table 2. They are retrospective because they are based on specimens collected from patients who have participated in other prospective studies. However, these “prospective retrospective” studies are prospectively planned, preferably within a written protocol precisely stating how patients to be included will be selected, the exact analytical methodology for the tumour biomarker assay, and what cut points will be used to distinguish “positive” from “negative”. Such protocol should include a detailed statistical section about stating power and the statistical modelling and testing to be incorporated in the analysis. Even though performed “prospectively”, such studies contain biases not inherent in true prospective studies: availability (it is almost always far less than 100%), pre-analytical concerns and assay failure, and underpowering (since the parent trial is almost always powered for the main therapeutic effect, not the tumour biomarker sub-analysis). Thus, findings from such a trial should be validated using specimens from a second data set, preferably from another prospective trial designed and conducted similarly [28].

Multiple disease reviews
Some reviews found show that the use of prospective design for stratification and validation cohorts have different advantages, such as control of experiment features [18] and a general preference of this design from regulatory institutions [20]. Moreover,
some reviews mention that this design enables optimal measurement of predictors and outcome. Furthermore, it is interesting to point out that some strategies, like exploratory pharmacogenetics strategies, should go beyond the collection of samples by including prospective planning of design and analysis to be optimized [20].

In addition to this, there are inherent limitations of cross-sectional data collected retrospectively in case-control and retrospective approaches. As an example, it is difficult to assume causality [19]. Moreover, the defined strata in these approaches may not be homogeneous concerning other potential predictors of clinical outcome. This limits the ability to determine the prognostic and predictive power of individual biomarkers. Using a multivariate model to address this issue is common [19].

On the other hand, retrospective studies have great potential to validate these biomarkers, mainly due to the wealth of data that exist. Moreover, the time required to conduct a time-to-event study with an endpoint such as progression-free or overall survival is greatly reduced with prospective designs [19].

Furthermore, data collection needed in prospective designs is expensive, and given the large sample sizes necessary for successful molecular signature discovery, using existing datasets may be a more feasible approach [18].

3. What are the prerequisites and methods used for integration of multiple retrospective cohorts?

Only one review has been found addressing this question [30]. Focused on oncology, it shows two approaches concerning data aggregation and meta-analysis:

1. Putting together different microarray experiments to form a single dataset.
   - Clustering or intersection operations can then be easily performed.

2. Analysing each individual microarray experiment and then aggregating the statistical results of all individual experiments as e.g. in the rank aggregation approach.
   - Among the methods that can be used to explore the relationship between the study characteristics are the classical statistics, such as the Mantel–Haenszel method, meta-regression or the more advanced ones specifically developed for this purpose, such as the latent variable approach.

A more general conclusion that can be drawn from the reviews is that the integration of PACS (patient archiving systems for imaging data), genomic and pharmacogenomic databases, as well as other laboratory and patient-relevant data represent significant challenges for designers and administrators of information management systems. The lack of international standards for patient care and management and different accounting systems will require the development and installation of country-specific (or even regional-specific) systems. Security issues arising from the sensitivity of certain
types of information need to be addressed and resolved properly. This requires institutional commitment. However, the necessary resources should not be underestimated. While institutional solutions are being developed, researchers should be able to fully exploit the potential for integrating biomolecular and clinical data. Meanwhile, a more pragmatic approach is to establish a medium-scale solution at the departmental level. Such a focused infrastructure requires only a fraction of the cost and time compared to an institutional one. It should be noted that, due to the available technology, the computational infrastructure can be dissociated from the actual site where the data are generated.

4. Which validation designs exist for the stratification (or clustering) in personalised medicine? Which methods and tools are used to build the cohorts of validation (external/sub-cohort)? What are their gaps?

Thirteen reviews related with this question have been found, seven are multiple disease reviews [17,20,31–35], four deal with oncology [23,25,36,37], one with stroke [38], and one with Alzheimer’s disease (AD) [39].

Building of stratification cohorts

- Designs and methods

**Alzheimer’s Disease**

Regarding AD, a list of clustering methods has been mentioned: Ward's clustering; unsupervised graph-theory-based clustering approach; non-negative matrix factorization, unsupervised random-based clustering; assignment of patients to subtypes in a probabilistic fashion methods [39].

More attention should be given to pathology-specific PET-imaging data in data-driven work due to the distinct advantages it offers for understanding the regional distribution of molecular disease processes. More systematic cross-modal subtyping comparisons and the direct combination of high-dimensional information from multiple modalities in future clustering approaches may achieve a more comprehensive definition of distinct disease subtypes in AD and related dementias. A comprehensive multimodal definition will strongly benefit from the increasing availability of large-scale standardized datasets from well-phenotyped patient cohorts in combination with current advances in methods and power for computational analytics [39].

**Multiple disease reviews**

A review organized the methods to stratify cohorts in four types, as summarized in Table 3 [35]
<table>
<thead>
<tr>
<th>Methods and approaches</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clustering</td>
<td>The developed methodology has the main goal of creating groups of patients with similar disease evolution.</td>
</tr>
<tr>
<td>Dimensionality reduction</td>
<td>The main novelty is the selection of representative features to characterize specific groups of patients. Reviews in this group can either present a single feature selection approach, or a feature selection approach which is then coupled to clustering to obtain groups of similar patients.</td>
</tr>
<tr>
<td>Similarity</td>
<td>The main goal is to define a novel measure of similarity among patients.</td>
</tr>
<tr>
<td>Software tools</td>
<td>The main contribution is the availability of a software solution implementing the proposed approach.</td>
</tr>
<tr>
<td>Combination of clustering or similarity metrics and supervised approaches</td>
<td>The features that characterize a group of patients are then used to solve a classification task.</td>
</tr>
</tbody>
</table>

*Table 3. Classification of methods for stratifying cohorts*

**Oncology**

A review collects different staging systems for hepatocellular carcinoma approved and widely used in clinical systems. These are the Barcelona Clinic Center Liver Cancer (BCLC), the Hong Kong Liver Cancer Staging System, the Japanese Okuda, the Tumor Node Metastasis (TNM) staging system for primary liver cancer, and the Vauthey’s simplified staging of the American Joint Committee on Cancer (AJCC)/International Union Against Cancer (UICC) TNM systems [37].

In summary, these methods of staging should be relevant for treatment recommendations. Nevertheless, these methods are suboptimal for this aim. An effective prevention and accurate patient stratification both demand a well-concerted multidisciplinary approach. The novel methods present various degrees of portability, ranging from general to very specific algorithms, which depend on the context and design of the disease [37].

**Building of validation cohorts**

- **Designs and methods**

**Stroke**

The review found deals with methods based on propensity scores and one to one matching. Both methods aim to match a control/patient group as a method to build validation cohorts.

Statistically, p-value and standardized difference to gauge balance between patients and control cohorts was suggested [38].

As a conclusion, propensity score matching and one-to-one matching are both good options for enrolling matched control groups into stroke biomarker studies. Propensity
scoring requires a larger pool of enrolled controls but maybe more practical for investigators who lack a large EHR or who already have many control samples in hand. One-to-one matching is a more targeted approach but requires computer programming support [38].

**Oncology**

Multivariate prognostic analysis of several biomarkers including all clinical, pathological prognostic variables and stratification factors of the clinical trial needs statistical corrections for multiple analyses to avoid false-positive result reporting. At best, external validation in a comparable series of patients is suitable but rarely possible. Alternatively, internal cross-validation complex procedures could assure the stability of the biomarker [23].

Below, the methods defined in the reviews that were found has been schematized in Table 4 [23,25,36,37]

<table>
<thead>
<tr>
<th><strong>External validation</strong></th>
<th>Conducted on different data that was used to generate the results.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Internal validation.</strong></td>
<td>Conducted on the same data that was used to generate the results.</td>
</tr>
</tbody>
</table>

A training set of patients, randomly selected from the whole series of patients, allows us to obtain a multivariable prediction model including the biomarkers of interest and the clinical predictors. Then, this model is tested without any change in the rest of the same patients’ series, used as a validation set. Such a strategy needs a large initial sample size, to get precise estimates.

*Leave-one-out* cross-validation approach.

**Bootstrap method.** Bootstrapping consists of drawing samples (with replacement) from the original data set to generate a large number of training sets (several hundred) of the same size as the original sample. A prediction model is developed on each training set and tested on the original data. Validation results are then reported as the average performance over the whole process.

**Interactive Bayesian model and artificial neural networks (ANN)** that used random split technique between training and validation data sets.

Table 4. Methods for validation defined in the oncology reviews found

Random-split technique to build validation cohorts can result in over and underfitting of the model, particularly as details of cross-validation were not given. More sophisticated techniques of data splitting that exploit the structure of the data exist and provide more confident results, but at a higher computational cost [25].

**Multiple diseases**
One of the reviews focuses on designs in pharmacogenetic studies [20].

· **Targeted or enriched design**
  It involves a pre-screening step whereby patients are selected for the study based on genotype. Patients carrying the negative genotype are excluded from the study whereas positive patients are then randomised to one of the treatment groups.
  This design has advantages as it can result in smaller studies when the effect of treatment is greater in the positive group. However, this design does not provide information on treatment effect in the excluded population and can only be used when there is substantive prior knowledge about the impact of a single genetic marker.

· **Stratification design**
  It involves a pre-screening step where all subjects are allocated to groups based on genotype, then randomised to treatment.
  This design is useful when there is an interest in collecting information about the treatment effect in the marker negative group. In addition, this enables the operating characteristics, such as sensitivity and specificity, of a genetic-based test to be estimated. However, this design also requires considerable prior knowledge about the genetic marker.

· **Adaptive design.**
  Enrolled patients are randomised to treatment groups and the treatments are compared as part of a primary objective of the study. If the study fails to meet the primary objective (i.e. there is no difference in the treatments), then patients are subdivided into groups based on genotype and a comparison of treatments is performed within these groups.

  Because multiple statistical tests are performed in this approach, the overall false-positive error rate is greater and needs to be controlled. This design can result in larger studies, but it is more flexible than the targeted or stratified design because it enables the testing of multiple objectives and can be modified to include the evaluation of several genetic markers.

Another review focuses on propensity score methods, useful methods for matching, and one-to-one matching. Both choices are good for enrolling matched control groups in biomarker studies [33].

Some reviews focus on the replication methods used, and their advantages and disadvantages are defined in the reviews found [17,34]:

· **Internal validation**
  Conducted on the same data that was used to generate the results:
- Analysis of random (cross-validated) subsamples of the development dataset.
- Analysis of using resampling techniques (like bootstrapping).

In theory, internal cross-validation or bootstrapping methods may achieve unbiased estimates of classification accuracy, but in practice, it has been shown that such procedures often give exaggerated estimates of biomarker performance, probably because of diverse biases, including violations of independence in training and testing, feature selection bias, population selection bias, and optimization biases.

· External validation

Conducted on different data that was used to generate the results:
It is the most rigorous form of model validity assessment, although biases may even creep also into this kind of validation.

Some reviews suggest that empirical evidence from other technologies imply that classification performance decreases when more rigorous validation practices are adopted in large and better-conducted studies [25,34]. External validation is performed only in a minority of studies.

It is also interesting that external independent validation can have different levels of stringency:
- Different samples but the same team of investigators. The analysis may also be performed by the same team.
- Different samples but validation performance conducted by different centres and different teams analysing the data.
- Samples come from different centres and the analysis is performed by the laboratory involved in the discovery phase.

Ideally, the validation samples should come from completely different teams and the analysis performed also by different investigators than those involved in the original analysis. This further safeguards the independence of the process and offers stronger support for the reproducibility of the results.

Some reviews show that most candidate biomarkers are assessed only once without any replication plan [31,36]. Replication is not a readily funded step for most funding programs. Most biomarker studies never have an independent replication. When they do, it is more frequent for investigators of the original team to be involved than not. Systematic reviews of biomarkers, as prognostic or diagnostic tests, are becoming very popular. However, they are usually more useful in assessing the problems with the evidence rather than the evidence itself. Larger studies typically tend to show less promising performance than smaller studies which may nevertheless be highly cited and drive interest in the field owing to preferential citation bias for the most exciting results. Depending on the
methodological challenges of each measurement platform, additional special problems may arise in the validation phase. These problems may result in incomplete, suboptimal validation analyses that yield inflated estimates of performance, as has been described in the case of markers based on microarray signatures and may also affect proteomics. Given these subtle biases, most, if not all, published biomarker studies may end up having statistically significant results. Instead of being a sign of perfect validation, this excess significance may be an indirect indicator of bias in the field and the need to perform some carefully done, large, preregistered studies. Options to improve the quality and trustworthiness of the validation evidence include collaborative analyses in large consortia and multiple teams of investigators [31].

5. What are the methods for the evaluation of the risk of bias?

Eight reviews have been found on this topic. Four of them deal with more than one disease [18,19,40,41], and four reviews deal with oncology [22,24,25,27].

Methods and tools to avoid bias

**Oncology**

Two methods for addressing bias due to lack of data are found in the reviews [22]:

- Inverse probability weighting.
- Multiple imputation.

The reviews conclude that there are unavoidable sources of bias, but the above approaches can reduce the risk of bias. For example, they can be used to mitigate selection bias due to the availability of tissue specimens. Certain general scientific standards across many fields should be established for overall scientific rigour and reproducibility, while there is also a great need for standardisation of statistical tools and methodologies in specific fields.[22,27]

**Multiple diseases**

Different methods have been reported to try to mitigate different kinds of bias [18,19,40,41]. These methods and tools are summarized in Table 5.

<table>
<thead>
<tr>
<th>Methods to reduce originated bias depending on the kind.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Methods and tools</strong></td>
</tr>
<tr>
<td><strong>Within-Subject Correlation</strong></td>
</tr>
<tr>
<td>· To quantify intraclass correlation: modifications of Pearson’s product-moment correlation coefficient.</td>
</tr>
<tr>
<td>· Comparisons based on the generalized estimating equations generated by mixed-effect models.</td>
</tr>
</tbody>
</table>
### Methods to reduce originated bias depending on the kind.

<table>
<thead>
<tr>
<th>Multiplicity</th>
<th>Multiple Clinical Endpoints</th>
<th>Selection bias</th>
<th>Publication bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>· Controlling the family-wise error rate (Tukey, Bonferroni, Scheffe, and other).</td>
<td>· The selection of a single primary endpoint for formal statistical inference, considering that the endpoints are possibly biologically related and positively correlated.</td>
<td>· To adjust for age, stage, treatment, and so forth.</td>
<td>· To publish positive and negative results.</td>
</tr>
<tr>
<td>· Approach to controlling the false discovery rate used in biomarker studies: Benjamini and Hochber.</td>
<td>· Creating a univariate outcome by combining multiple clinical endpoints (weighted measures taking into account the relevance of each endpoint).</td>
<td>· Matched samples.</td>
<td>· Encourage the objective assessment of molecular signatures by reporting both positive and negative outcomes.</td>
</tr>
<tr>
<td>· Analysis of data using a methodology that controls the family-wise error rate.</td>
<td>· To compare the two samples based on the endpoint of highest priority first, and, if no winner can be determined, would one move to the endpoint of the next highest priority.</td>
<td>· Analysis of data using a multivariate model to simultaneously adjust for confounders</td>
<td>· Make data publicly available after publication.</td>
</tr>
<tr>
<td>· Analysis of data by prioritizing the relevant endpoints or by using a composite endpoint.</td>
<td></td>
<td>· Obtention of matched samples.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>· Propensity score weighted.</td>
<td></td>
</tr>
</tbody>
</table>

Table 5. A summary of methods, tools and strategies proposed to avoid different types of bias.

Some reviews also suggest that the main problem in personalised medicine is that there are errors in the experimental phase, in fundamental data, that cannot be avoided. Long cohort studies are a tool to improve the results. Also, maybe an interesting approach could be to try to separate—rather than combine—bench researchers (or technologists) and clinical epidemiologists to cover both kinds of research. If the fundamental data that will be analysed are not strong, then better analytic methods have limited use in improving the situation. Recognition of this situation is an important first step toward improving the quality of clinical research on markers aiming to detect the illness of interest.[40]

Another bias with non-statistical origin is the ‘publication bias’. It can occur because research groups tend to report only the best results among many attempted approaches and only positive results are published. To publish positive and negative outcomes is a need to reduce this kind of bias [18].
Moreover, in order to try to reduce bias, it is very important to take into account its source [19].

Methods and tools to assess risk of bias

**Oncology**

There are tools to analyze the risk of bias, such as the Quality in Prognosis Studies (QUIPS) tool. This tool assesses the risk of bias in six domains: study participation, study attrition, prognostic factor measurement, outcome measurement, study confounding and statistical analysis and reporting. In addition to these tools, reviews examine the main source of bias analysed in the papers. Information provided shows that there is a black box in the analysis of the risk of bias. However, three sources of risk of bias are usually analysed: outcome measurement, study confounding and statistical analysis and reporting [24,25].

6. How are the (-omics, imaging, exposome, lifestyle etc.) data generated?

Twenty-five reviews clearly are focused on specific kind of data. Ten of these reviews fall into the oncology field [15,22,26,30,37,42–46], and two deal with Alzheimer disease [14,39] and two with stroke [38,47]. The other eleven are multiple disease reviews [16–18,20,34,40,48–52].

We have summarized the number of results in Table 6

<table>
<thead>
<tr>
<th>Alzheimer’s disease</th>
<th>Nº reviews</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multimodal</td>
<td>1</td>
</tr>
<tr>
<td>Imaging (MRI/hippocampal volume)</td>
<td>2</td>
</tr>
<tr>
<td>Cognition</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Multiple diseases</th>
<th>Nº reviews</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transcriptomic</td>
<td>1</td>
</tr>
<tr>
<td>Genetic</td>
<td>3</td>
</tr>
<tr>
<td>Metabolomic</td>
<td>1</td>
</tr>
<tr>
<td>Multimodal</td>
<td>5</td>
</tr>
<tr>
<td>Proteomics</td>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Oncology</th>
<th>Nº reviews</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multimodal</td>
<td>4</td>
</tr>
<tr>
<td>Radiomics</td>
<td>1</td>
</tr>
</tbody>
</table>
Microbiome | 3
---|---
Genomic/Genetic | 3
Transcriptomics | 3
Metilation (epigenetics) | 2
Proteomics | 1
Metabolomics | 1

Type of data | Nº reviews
---|---
Genomic | 1
Multimodal | 1

<table>
<thead>
<tr>
<th>Stroke</th>
<th>Nº reviews</th>
</tr>
</thead>
</table>

Table 6. Summary of the quantity of information found by type of data.

1. Genetic variation: It is an unbiased source of genetic basis of disease and allows the direct inference of causality.
2. Epigenetics: Useful to know the functional impact and the typically easy to infer causality. It is not applicable for all phenotypes.
3. Gene expression: It is an inexpensive assay for an intermediate step towards the phenotype.
4. Proteomics and metabolomics: They are likely to be very close to the phenotype. Although, they are expensive and difficult to scale (specially proteomics).
5. Microbiome: It is likely to be very close to the phenotype and measures a combination of genetic and environmental influences. Nevertheless, combination of genetic and environmental influences makes it difficult to infer the direction of causality.

As shown in Table 6, most of the results are focused on oncology. The reviews show that compared to the other kind of data, the image and environmental data on different biological layers are hard to be quantitatively measured [42]. Moreover, they show that one of the biggest advantages of radiogenomic studies is the opportunity to assess the relationships between genotype features (such as genomic variants), intermediate phenotype features (such as transcriptomics and epigenetic variables), radiomic features (image phenotype) and phenotypic clinical outcomes [44]. Furthermore, the reviews highlight that the future research projects should include diverse populations to circumscribe and associate the functional properties of the microbiome with other features of these populations. There is a need to develop tools and algorithms that can integrate 16S sequence information (information obtained by sequence of 16S rRNA) and metagenomics with metatranscriptomics and metaepigenomics [43]. As a more general conclusion, it can be said that it is important to consider the clinical context of the molecular signature [18].

7. What are the tools used for data management and multimodal data analysis used in personalised medicine, i.e. Galaxy? What are their gaps?

Twenty-four reviews have been found on this topic. Fifteen are focused on more than one disease [17,18,35,41,48–58], eight on cancer [22,30,42,44–46,59,60] and one on stroke [47].

Due to the great diversity of the methods and tools necessary, an outline has been
made. This information is collected in Appendix II, and divided into two sections: methods and tools for data-analysis and methods for different kind of data integration. The most relevant information and conclusions are described below.

**Oncology**

Due to the diversity of methods, the strengths and limitations of each method must be clear in order to choose the most appropriate one for the study in question.

Regarding Joint Modelling approaches (in a supervised - applied with prior knowledge of the system- or unsupervised manner), the reviews point out that these approaches are recommended to integrate both large-scale omics and non-omics data because they account for the correlation structure between the two data types and capture a larger complexity than the conditional or independent modelling. The decision of which modelling strategy (multi-staged or meta-dimensional) to follow should be done following the main objective of the analysis: association testing or risk prediction. Multi-staged analysis that models the relationship between the different layers of information will probably be preferable when the interest is to increase our biological knowledge of the disease mechanisms. On the other hand, the metadimensional approach will be more suitable when the goal is to improve prediction or prognosis for personalised medicine and modelling the mechanisms is not so relevant. Concatenation-based integration combines the different data types into a joint data matrix and performs a variable selection or dimension reduction to the whole data set. The concatenation approach cannot ignore that the different data types are expected to have different relevance to the outcome and the joint analysis should take this into account [60].

Another point highlighted in the reviews is network analysis. The reviews point out that inferring network structure from multi-omics data is challenging because there are large numbers of highly correlated variables. If the study aims to find a multi-omics signature comprised of a small set of biomarkers to discriminate a biological outcome of interest, statistical methods such as Multiblock Partial least squares (or projection to latent structures, MBPLS) can be employed. These methods seek to maximize covariance between summary vectors derived from each omics data block and a biological phenotype or response. The identified molecular features derived from multi-omics layers can be thought of as jointly contributing to a biological trait. Software for performing these analyses is available in Matlab and R [59]. Biological networks are powerful for measuring the potential relationship among different molecular elements [42].

Concerning additional tool-based conclusions, some reviews show that there are R packages that manage several sets of biological experiments and facilitate the coordination of different types of operations, such as data visualization, data manipulation, subsetting, data integration and reshaping. Moreover, these data
containers enable subsetting of data by different items, such as clinical or pathologic variables, genes, genomic ranges and assays. Additionally, some data warehouses enable users to dynamically interrogate clinico-pathologic data in a multidimensional manner are developed in this context, such as Data Warehouse for Translational Research (DW4TR) [44].

As a general conclusion, an important limitation is that integrative data sets often do not have a standard format for research use, let alone for use in a structured clinical system. The infrastructure to house and manage these data will be required, which introduces financial and administrative burdens. In particular, health informaticians will be tasked with building a robust infrastructure for storing genetic and transcriptomic data in the Electronic Health Record (EHR). Moreover, determining which information will be reported back to a patient and incorporated into an EHR will require concerted efforts from clinicians and researchers [61].

Multiple diseases

An integrative approach using multi-omics data is a powerful strategy to decipher the mechanistic details of the information flow in a cell. Currently, there is a wide array of tools and methods available in the public domain to integrate multi-omics data sets to derive meaningful insights. Multi-omics data repositories, visualization portals, and challenges in the integration of data sets are studied [58].

Classification of the approaches presented in the literature as multi-omics methods is a non-trivial task for at least three reasons. First, most of the computational approaches developed so far are pipelines of analysis that apply several methods to carry out a sequence of tasks; therefore, different pipelines share some methods: for example, partial least squares regression is included in both Integromics and sMBPLS. Second, pipelines presented for addressing a particular problem can also be used, with minor modifications, to solve another problem, possibly with other types of omics. Third, several tools can be used in a supervised or unsupervised setting, according to the formulation of the problem [57].

As the tools and methods are largely isolated, there is a need to have a uniform framework that can effectively process and analyse multi-omics data in an end-to-end manner along with easy and biologist-friendly visualization and interpretation. The improvement in this uniform framework may lead to a reduction in spurious signal caused by technical limitations of individual platforms, and an increased ability to identify molecular signatures associated with the underlying mechanistic roles in disease pathogenesis [18]. Nevertheless, for greater improvement there are some remaining issues and challenges include strategies for combining data from multiple time points and multiple tissues; data normalization and scaling issues when considering multiple data types; consideration of complex biological processes,
including feedback loops and compensatory mechanisms, which will require nonlinearity models; and the possibility of combining different data types from different individuals or samples. The difficulty when developing new strategies is the inability to know what the true models should include and therefore what the most effective modelling strategies will be [49].

Constraint-based modelling methods have two main advantages: first, they do not need dynamic or kinetic data as they are based on mass balance across the metabolic network; second, they are suitable for integration of different omics layers at genome-scale to improve their predictive performance. In particular, multi-omics “vertical” integration methods have been proposed to include omics layers (mainly transcriptomics and proteomics) [62,63]. Conversely, “horizontal” integration methods have focused on modelling different environments, cancers or growth conditions starting from the same model [64,65]. Such multi-omics integration in genome-scale models has provided a mechanistic link between the genotype and their phenotypic observables [50].

Moreover, conclusions focusing on tools have been drawn, mainly about machine learning. Within machine learning approaches, deep learning-based approaches are highlighted. These approaches allow the computational flexibility to effectively model and integrate almost any type [35,51].

The reviews also point out that deep learning model architectures and training techniques share many similarities with biological message-passing systems. Deep learning models contain a minimum of three layers: input, hidden, and output. This could mimic representation of relationships between gene transcription, protein expression, and metabolite concentrations, but can also extend to other omics layers. Interesting parallels between computational and biological optimizations such as backward propagation in deep learning and signal inhibition in omics have also emerged. Personalised medicine requires complex data encoding and integration tasks, which are well suited for deep learning. Currently, identification of causality in complex phenotypes requires custom analyses and domain expert interpretation; however, one might envision a future of medical data accessibility, quality, and scale, which could enable near automated deep learning-based detection of many clinically relevant events [35,51].

One review has also analysed methods taking into account the analytical features which these methods help to improve [53]:

- Dealing with velocity, one of the possible future directions is the utilization of so-called “anytime algorithms” that can learn from streaming data and that still return a valuable result if their execution is interrupted at any time. With the increasing number of measured features and the increasing time span of the measurements, a key challenge will be to find a data integration model that will directly mine time series measurements for which the time spans and frequencies of measurements
Variety (i.e. heterogeneity) has been addressed by many methods. Matrix factor-based methods are promising for mining heterogeneous datasets. Although Graph-regularized non-negative matrix tri-factorization (GNMTF) is a versatile data integration framework its computational complexity increases with the number of data types to be integrated. Thus, integrating large numbers of heterogeneous data types within the Matrix factor-based framework needs novel algorithmic improvements. Extracting the complementary information conveyed in data of different formats and types is another challenge that is partially addressed by the presented integrative methods. Integrating genetic interaction network with Protein-protein interaction network and other molecular networks is beneficial in many biological problems. Moreover, many data types including exposomic and metagenomic data are yet to be analysed and their integration with other data will be a focus of future studies. EHR data are increasingly becoming available for academic research purposes and they present numerous computational challenges that are yet to be addressed. Two major computational challenges include developing algorithms for: (i) individual phenotyping (i.e. annotating patient records with disease conditions) and (ii) integration of EHR data with omics data for better understanding of disease mechanisms and treatments. The biggest obstacles of the first challenge are nosiness and incompleteness of the EHR data that need to be properly taken into account. On the other hand, the biggest obstacles of the second challenge are heterogeneity and different format types of EHR and genomic data.

Regarding criterion for the election of predictor, papers show that the most reported criterion in multivariable models was a p-value of 0.05. Other criteria, such as Akaike’s Information Criterion or coefficient of determination R², were used much less frequently.

Finally, as a conclusion, all reviews point out that the existing statistical methods mainly addressed data analysis and management. Current research needs to develop and improve multi-layer data integrative methods for multi-omics derived data. Although multi-omics research is still challenging, it will allow early diagnosis and will help clinicians and families to forecast and make informed decisions about the prognosis and, prospectively, will provide tailored therapies [52].

8. What quality of data of cohorts is needed to obtain a biomarker or multimodal data profiling? Are there requirements to monitor the collection of associated clinical data?

Regarding the quality of data, we have not found specific information through the scoping review. In the cohort building step within the global process of personalised medicine, we have only found guidelines related to sample size (answered in question
one), or regarding the significance of the analysis conducted (in answers to questions four and seven). We expected to also find standards of minimum experimental data quality in order to conduct a biomarker or multimodal data profiling.

In addition to the lack of information about the quality of data, we have also noticed an absence of information regarding the monitorization of associated clinical data. In this sense, we would expect to find some clear regulation.

As a conclusion, we think there is an important lack of information regarding these two topics, making it interesting that experts deepen on them with the objective of designing a useful guideline.

9. What is the outlook of data generation seen as (CE-labelled) in-vitro diagnostics?

Five documents have been found in grey literature regarding this question [66–70]. “In vitro diagnostics” are tests done on samples such as blood or tissue that have been taken from the human body. These in vitro diagnostics can detect diseases or other adverse conditions and can also be used to monitor a person’s overall health in order to help cure, offer a treatment, or prevent diseases. They can be used in a laboratory or other professional health settings and some other tests are for consumers to use at home. In vitro diagnostics can include next-generation sequencing. In the context of personalised medicine, they can be used to identify patients who are likely to benefit from specific treatments or therapies [68].

Regarding regulation, information found shows that there is a new regulation for in vitro diagnostics to be applied to the European Union before May 2022 [66,67].

One of the effects derived from the new regulation implies that devices will improve in quality and effectiveness. This means that there will be the need for small companies to form alliances with big Pharma companies since it will become harder for a small company to develop and improve such devices. In the field of personalised medicine, the collaboration between companies may translate in advances in drug development and diagnostic tests, which will foster precision medicine. In 2017, 35% of all new drugs approved by the FDA were personalised, and it is assumed that this number will continue growing [69,70].

However, within such a wide field, it is difficult to anticipate the effects of this new regulation or make predictions about the consequences it will have specifically on personalised medicine and in vitro diagnostics.

It is likely that in vitro diagnostics will go through an exponential increase, implying that some of them will probably be commercialized soon and easy to find [71–74]. At the moment, some techniques are already advancing particularly fast, as CRISPR Technology [75], liquid biopsy [76], and epigenetics [77].
Identified gaps

When reviewing the information collected, we have detected a lack of standards and information in some specific areas:
- Sample size calculation
- Prerequisites and methods used for integration of multiple retrospective cohorts
- Quality of data of cohorts needed to obtain a biomarker or multimodal data profiling
- Requirements to monitor the collection of associated clinical data in biomarkers studies.

Summary of findings and next steps

Summary of findings

Regarding question one (approaches to define the optimal size of stratification/validation cohorts), the information found was not uniform among the different disease areas. There are differences between diseases in terms of the factors that must be taken into account in the calculation of sample size and in the statistical methods recommended. Therefore, there is a need for standardisation of sample size calculation.

Question two (differences, pros and cons of the prospective and retrospective cohorts) addressed pros and cons of the prospective and retrospective designs and the final conclusion is that the use (or choice) of one type of cohort or another depends on the objective of the study.

About question three (prerequisites and methods used for the integration of multiple retrospective cohorts), only one review was found. The review found focused on oncology and showed approaches concerning data aggregation (putting together different microarray experiments to form a single dataset) and meta-analysis (analysing each individual microarray experiment and then aggregating the statistical results of all individual experiments). The scarcity of reviews on this topic evidence a lack of information, suggesting that there is a need of standardization of prerequisites for integration of multiple retrospective cohorts and that there is a need for research of new tools to this integration.

Addressing question four (validation designs for stratification; methods and tools used to build the cohorts of validation and their gaps) reviews found useful information of different designs and methods for building stratification and validation cohorts, and for dealing with different kinds of validation. Methods were analysed and classified in groups by their characteristics.
Regarding question five (methods for the evaluation of the risk of bias), most of the information in the reviews refers to different kinds of bias and methods and tools to avoid it. There is also some information about the assessment of risk of bias.

Papers related to question six (how data are generated), classified the information according to the type of data analysed: Multimodal, imaging (MRI/hippocampal volume), cognition, radiomics, microbiome, genomics, transcriptomics, methylation (epigenetics), proteomics or metabolomics. Reviews show that, compared to the other kind of data, image and environmental data on different biological layers are hard to measure quantitatively, and that it is important to consider the clinical context of molecular signatures.

Concerning question seven (tools used for data management and multimodal data analysis), a huge diversity of methods was identified, and the strengths and limitations of the different methods were collected in order to facilitate the choice of the most appropriate one for each study. As a general conclusion, oncology reviews found that an important limitation is that integrative data sets often do not have a standard format for research use. Other reviews point out that current research needs to develop and improve multi-layer data integrative methods for multi-omics derived data to provide tailored therapies.

About question eight (quality of data of cohorts needed to obtain a biomarker or multimodal data profiling, requirements to monitor the collection of associated clinical data), no specific information was retrieved, although this topic was partially responded in other questions of the present review. We expected to find standards of minimum experimental data quality to conduct a biomarker or multimodal data profiling together with some clear regulation about monitoring of associated clinical data.

For question nine (outlook of data generation seen as in-vitro diagnostics) five documents were found in grey literature. According to these documents, a new regulation for in vitro diagnostics will be applied to the European Union before May 2022. The effects of this new regulation cannot be anticipated, and it is difficult to predict which consequences this new regulation will have on personalised medicine. However, because at present in vitro diagnostics shows an unlimited growth capacity, it seems likely that it will continue increasing, pointing to the possibility of commercialization of some diagnostics in the near future.

**Next steps**

Bearing in mind the above findings, we wish to focus our discussions at the Gap Analysis Workshop on questions about:
- The lack of uniform sample size calculation methods to ensure the quality and credibility of the clustering.
- Prerequisites and methods used for integration of multiple retrospective cohorts to ensure reliability when combining individual data from different cohorts within the same clusters.
- Quality of data of cohorts needed to obtain a biomarker or multimodal data profiling
- Requirements to monitor the collection of associated clinical data in biomarkers studies.

Once deep gap analysis and Gap Analysis Workshop is completed, the next step is to conduct a first draft of recommendations. We will aim to select experts for those areas where questions of interest are generated. They may be identified through searches on web pages of important research centres or universities or relevant projects and initiatives. Experts will be given specific questions related to the areas of interest, together with the first draft of recommendations and a workshop will be organized, with focused discussions per areas. Bearing in mind the comments of the experts, the first draft of recommendations will be improved and consolidated.
### Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alzheimer’s disease</td>
<td>AD</td>
</tr>
<tr>
<td>Bayesian consensus clustering</td>
<td>BCC</td>
</tr>
<tr>
<td>Convolutional neural network</td>
<td>CNN</td>
</tr>
<tr>
<td>Consensus PCA</td>
<td>CPCA</td>
</tr>
<tr>
<td>Deep neural network, Mean decrease accuracy</td>
<td>DNN-MDA</td>
</tr>
<tr>
<td>Data Warehouse for Translational Research</td>
<td>DW4TR</td>
</tr>
<tr>
<td>Electronic Health Record</td>
<td>EHR</td>
</tr>
<tr>
<td>Electronic Health Record</td>
<td>EHR</td>
</tr>
<tr>
<td>Elastic Net</td>
<td>ENET</td>
</tr>
<tr>
<td>Factor analysis - linear discriminant analysis</td>
<td>FALDA</td>
</tr>
<tr>
<td>Feature extraction</td>
<td>FE</td>
</tr>
<tr>
<td>Feature selection</td>
<td>FS</td>
</tr>
<tr>
<td>Feature selection multiple kernel learning</td>
<td>FSMKL</td>
</tr>
<tr>
<td>Graph-regularized non-negative matrix tri-factorization</td>
<td>GNMTF</td>
</tr>
<tr>
<td>Joint and Individual Variation Explained</td>
<td>JIVE</td>
</tr>
<tr>
<td>Least Absolute Shrinkage and Selection Operator</td>
<td>LASSO</td>
</tr>
<tr>
<td>Multiple-block PCA</td>
<td>MBPCA</td>
</tr>
<tr>
<td>Multiple co-inertia analysis</td>
<td>MCIA</td>
</tr>
<tr>
<td>Multiple dataset integration</td>
<td>MDI</td>
</tr>
<tr>
<td>Multiple Factor Analysis</td>
<td>MFA</td>
</tr>
<tr>
<td>Multiple factor analysis</td>
<td>MFA</td>
</tr>
<tr>
<td>Multi-omics factor analysis</td>
<td>MOFA</td>
</tr>
<tr>
<td>Molecular Regularized Consensus Patient Stratification</td>
<td>MRCPS</td>
</tr>
<tr>
<td>Multiple similarities collaborative matrix factorization</td>
<td>MSCMF</td>
</tr>
<tr>
<td>Network-based</td>
<td>NBS</td>
</tr>
<tr>
<td>Next-generation sequencing</td>
<td>NGS</td>
</tr>
<tr>
<td>Non-negative matrix factorization</td>
<td>NMF</td>
</tr>
<tr>
<td>Non-negative matrix factorization</td>
<td>NMF</td>
</tr>
<tr>
<td>Patient archiving systems for imaging data</td>
<td>PACS</td>
</tr>
<tr>
<td>Pathway Recognition Algorithm using Data Integration on Genomic Models</td>
<td>PARA-DIGM</td>
</tr>
<tr>
<td>Principal Component Analysis</td>
<td>PCA</td>
</tr>
<tr>
<td>Penalized multivariate analysis</td>
<td>PMA</td>
</tr>
<tr>
<td>Protein–protein interaction</td>
<td>PPI</td>
</tr>
<tr>
<td>Patient-specific data fusion</td>
<td>PSDF</td>
</tr>
<tr>
<td>Random forest</td>
<td>RF</td>
</tr>
<tr>
<td>Reproducing Kernel Hilbert Spaces Regression models</td>
<td>RKHS</td>
</tr>
<tr>
<td>Reproducing Kernel Hilbert Spaces Regression models</td>
<td>RKHS</td>
</tr>
<tr>
<td>Regularized multiple kernel learning- locality preserving projection</td>
<td>rMKL-LPP</td>
</tr>
<tr>
<td>Sparse generalized canonical correlation analysis</td>
<td>sGCCA</td>
</tr>
<tr>
<td>Sparse multi-block partial least squares</td>
<td>smBPLS</td>
</tr>
<tr>
<td>Similarity network fusion</td>
<td>SNF</td>
</tr>
<tr>
<td>Support vector machine</td>
<td>SVM</td>
</tr>
<tr>
<td>Single vector decomposition</td>
<td>t-SVD</td>
</tr>
</tbody>
</table>
References

2. Personalized medicine coalition http://www.personalizedmedicinecoalition.org. no date; .
5. Challenge AG for S. Health, Demographic Change and Well-being. no date; .
47. Swerdlow DI, Hingorani AD, Humphries SE. Genetic Risk Factors and Mendelian Randomization in Cardiovascular Disease. Curr Cardiol Rep 2015; 17:.
64. Eyassu F, Angione C. Modelling pyruvate dehydrogenase under hypoxia and its role in cancer metabolism. R Soc Open Sci 2017; 4:.
65. Yaneske E, Angione C. The poly-omics of ageing through individual-based metabolic modelling. BMC Bioinformatics 2018; 19:.
76. Palmirotta R, Lovero D, Cafforio P, Felici C, Mannavola F, Pellè E, Quaresmini
### Supplementary tables

<table>
<thead>
<tr>
<th></th>
<th>Oncology</th>
<th></th>
<th>Pros</th>
<th>Cons</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pros</strong></td>
<td></td>
<td></td>
<td><strong>Pros</strong></td>
<td><strong>Cons</strong></td>
<td><strong>Pros</strong></td>
</tr>
<tr>
<td>Better definition of included population that developed tumours (diminution of selection bias).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recall bias is reduced.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preferred design for prognosis studies.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Healthy-volunteer and spectrum bias (All human studies are voluntary where subjects with a healthy lifestyle, like non-smokers, are more likely to take part in health research. Spectrum bias is observed when subjects are healthier than the target population in which the marker will be later applied).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cons</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bias due to overdiagnosis (If the endpoint is cancer on biopsy).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Large sample size requirement.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lead-time bias (A marker detects a disease earlier than usual care and does not result in a lower disease-specific mortality).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length bias (Slowly growing tumours may be captured by diagnostic markers better than rapidly growing cancers, which more frequently appear between screening intervals).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Healthy-volunteer and spectrum bias (All human studies are voluntary where subjects with a healthy lifestyle, like non-smokers, are more likely to take part in health research. Spectrum bias is observed when subjects are healthier than the target population in which the marker will be later applied).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Pros</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quick and valid way.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No requirement for follow-up of participants.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The ability to investigate rare tumours.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enables optimal measurement of predictors and outcome.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Longer follow-up time easily available.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normally, poorer quality data.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cons</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Propensity to recall and selection biases (in single-cohort studies lacking a control arm).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Possible reinforcement of existing biases in clinical care.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Need to be prospectively validated.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Compromised generalizability of the predictive model (Due to lack of data from unrepresented populations).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heterogeneity of histologic tumour types.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table S1. Pros and cons about study designs in oncology found in reviews

<table>
<thead>
<tr>
<th>Multiple diseases</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Prospective</strong></td>
</tr>
<tr>
<td><strong>Pros</strong></td>
</tr>
<tr>
<td>High control of all aspects of the experiment.</td>
</tr>
<tr>
<td>Regulatory preference for prospectively designed studies.</td>
</tr>
<tr>
<td>Optimal measurement of predictors and outcome.</td>
</tr>
</tbody>
</table>

Table S2. Pros and cons of study designs found in multiple disease reviews

Appendix I: Searches

Alzheimer’s Disease Searches

Pubmed 15/5/2020

#14

Search: #9 OR #13 Sort by: Publication Date 56 10:26:47
#13

Search: #7 AND #12 Sort by: Publication Date 45 10:25:29
#12
<table>
<thead>
<tr>
<th>No.</th>
<th>Query</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>#16</td>
<td>#14 OR #15</td>
<td>206</td>
</tr>
<tr>
<td>#15</td>
<td>#13 AND [review]/lim</td>
<td>127</td>
</tr>
<tr>
<td>#14</td>
<td>#8 AND #12 AND ((cochrane review)/lim OR [systematic review]/lim OR [meta analysis]/lim)</td>
<td>122</td>
</tr>
<tr>
<td>#13</td>
<td>#8 AND #12</td>
<td>3250</td>
</tr>
<tr>
<td>#12</td>
<td>#9 OR #10 OR #11</td>
<td>378330</td>
</tr>
<tr>
<td>#11</td>
<td>dementia OR alzheimer</td>
<td>378330</td>
</tr>
<tr>
<td>#10</td>
<td>'alzheimer disease'/exp</td>
<td>197786</td>
</tr>
<tr>
<td>#9</td>
<td>'dementia'/de</td>
<td>116670</td>
</tr>
<tr>
<td>#8</td>
<td>#3 AND #6 AND #7</td>
<td>63646</td>
</tr>
<tr>
<td>#7</td>
<td>'validation study'/exp OR 'procedures'/exp OR validation:ti,ab OR method*:ti,ab</td>
<td>29820061</td>
</tr>
<tr>
<td>#6</td>
<td>#4 OR #5</td>
<td>1619793</td>
</tr>
<tr>
<td>#5</td>
<td>'cohort analysis'/exp OR 'cross-sectional study'/exp OR 'prospective study'/exp</td>
<td>1374626</td>
</tr>
<tr>
<td>#4</td>
<td>'cohort studies':ti,ab OR 'cohort study':ti,ab OR 'cohorts design':ti,ab OR 'prospective cohort':ti,ab OR 'retrospective cohort':ti,ab OR 'data integration':ti,ab OR bias:ti,ab OR 'cross study':ti,ab OR 'cross studies':ti,ab</td>
<td>518448</td>
</tr>
<tr>
<td>#3</td>
<td>#1 OR #2</td>
<td>543287</td>
</tr>
<tr>
<td>#2</td>
<td>'biological marker'/exp OR 'personalized medicine'/exp</td>
<td>338118</td>
</tr>
<tr>
<td>#1</td>
<td>'stratified medicine':ti,ab OR biomarker*:ti,ab OR 'precision medicine':ti,ab OR 'personalized medicine':ti,ab OR 'personalised medicine':ti,ab OR 'individualized medicine':ti,ab OR 'individualised medicine':ti,ab OR 'individualized therapy':ti,ab OR 'individualisation therapy':ti,ab OR 'patient stratification':ti,ab OR pharmaco:ti,ab OR 'patient specific modeling':ti,ab OR 'personalized clinical decision making':ti,ab OR 'personalised clinical decision making':ti,ab OR 'prediction of response':ti,ab OR 'prediction of responses':ti,ab</td>
<td>427868</td>
</tr>
</tbody>
</table>

**Web of science 15/5/2020**

<table>
<thead>
<tr>
<th># 10</th>
<th>67</th>
<th>#9 OR #7</th>
<th>Edit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Indexes=SCI-EXPANDED, SSCI, A&amp;HCI, CPCI-S, CPCI-SSH, ESCI</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Timespan=All years</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th># 9</th>
<th>40</th>
<th>#8 AND #6</th>
<th>Edit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Indexes=SCI-EXPANDED, SSCI, A&amp;HCI, CPCI-S, CPCI-SSH, ESCI</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Timespan=All years</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th># 8</th>
<th>280,095</th>
<th>TOPIC: (&quot;systematic review&quot; OR &quot;meta analysis&quot;)</th>
<th>Edit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Indexes=SCI-EXPANDED, SSCI, A&amp;HCI, CPCI-S, CPCI-SSH, ESCI</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Timespan=All years</td>
<td></td>
</tr>
</tbody>
</table>
PERMIT has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N. 874825

# 7 60  #5 AND #4
Refined by: DOCUMENT TYPES: (REVIEW)
Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI
Timespan=All years

# 6 351  #5 AND #4
Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI
Timespan=All years

# 5 265,230 TOPIC: (dementia OR alzheimer)
Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI
Timespan=All years

# 4 7,402  #3 AND #2 AND #1
Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI
Timespan=All years

# 3 11,248,557 TOPIC: ((validation OR method*))
Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI
Timespan=All years

# 2 764,078 TOPIC: ("cohort studies" OR "cohort study" OR "cohorts design"
OR "prospective cohort" OR "retrospective cohort" OR "data integration"
OR bias OR "cross study" OR "cross studies"))
Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI
Timespan=All years

# 1 375,740 TOPIC: ("stratified medicine" OR biomarker* OR "precision medicine"
OR "personalized medicine" OR "personalised medicine" OR "individualized Medicine" OR "individualised Medicine" OR "individualized therapy" OR "individualised therapy" OR "patient stratification"
OR pharmacogenetics OR "patient specific modeling" OR "personalized clinical decision making" OR "personalised clinical decision making"
OR "prediction of response" OR "prediction of responses"))
Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI
Timespan=All years

Psycinfo 15/5/2020
<table>
<thead>
<tr>
<th>Set</th>
<th>Search</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>S8</td>
<td>((TI,AB(&quot;stratified medicine&quot; OR biomarker* OR &quot;precision medicine&quot; OR &quot;personalized medicine&quot; OR &quot;personalised medicine&quot; OR &quot;individualized Medicine&quot; OR &quot;individualised Medicine&quot; OR &quot;individualized therapy&quot; OR &quot;individualised therapy&quot; OR &quot;patient stratification&quot; OR pharmacogenetics OR &quot;patient specific modeling&quot; OR &quot;personalized clinical decision making&quot; OR &quot;personalised clinical decision making&quot; OR &quot;prediction of response&quot;) AND TI,AB(&quot;cohort studies&quot; OR &quot;cohort study&quot; OR &quot;cohorts design&quot; OR &quot;prospective cohort&quot; OR &quot;retrospective cohort&quot; OR &quot;data integration&quot; OR bias OR &quot;cross study&quot; OR &quot;cross studies&quot;) AND TI,AB(validation OR method*)) AND TI,AB(dementia OR alzheimer)) AND TI,AB(review* OR systematic OR &quot;meta analysis&quot;)</td>
<td>21</td>
</tr>
<tr>
<td>S7</td>
<td>TI,AB(review* OR systematic OR &quot;meta analysis&quot;)</td>
<td>590,444</td>
</tr>
<tr>
<td>S6</td>
<td>TI,AB(&quot;stratified medicine&quot; OR biomarker* OR &quot;precision medicine&quot; OR &quot;personalized medicine&quot; OR &quot;personalised medicine&quot; OR &quot;individualized Medicine&quot; OR &quot;individualised Medicine&quot; OR &quot;individualized therapy&quot; OR &quot;individualised therapy&quot; OR &quot;patient stratification&quot; OR pharmacogenetics OR &quot;patient specific modeling&quot; OR &quot;personalized clinical decision making&quot; OR &quot;personalised clinical decision making&quot; OR &quot;prediction of response&quot;) AND TI,AB(&quot;cohort studies&quot; OR &quot;cohort study&quot; OR &quot;cohorts design&quot; OR &quot;prospective cohort&quot; OR &quot;retrospective cohort&quot; OR &quot;data integration&quot; OR bias OR &quot;cross study&quot; OR &quot;cross studies&quot;) AND TI,AB(validation OR method*)) AND TI,AB(dementia OR alzheimer)</td>
<td>114</td>
</tr>
<tr>
<td>S5</td>
<td>TI,AB(dementia OR alzheimer)</td>
<td>97,265</td>
</tr>
<tr>
<td>S4</td>
<td>TI,AB(&quot;stratified medicine&quot; OR biomarker* OR &quot;precision medicine&quot; OR &quot;personalized medicine&quot; OR &quot;personalised medicine&quot; OR &quot;individualized Medicine&quot; OR &quot;individualised Medicine&quot; OR &quot;individualized therapy&quot; OR &quot;individualised therapy&quot; OR &quot;patient stratification&quot; OR pharmacogenetics OR &quot;patient specific modeling&quot; OR &quot;personalized clinical decision making&quot; OR &quot;personalised clinical decision making&quot; OR &quot;prediction of response&quot;) AND TI,AB(&quot;cohort studies&quot; OR &quot;cohort study&quot; OR &quot;cohorts design&quot; OR &quot;prospective cohort&quot; OR &quot;retrospective cohort&quot; OR &quot;data integration&quot; OR bias OR &quot;cross study&quot; OR &quot;cross studies&quot;) AND TI,AB(validation OR method*))</td>
<td>385</td>
</tr>
<tr>
<td>S3</td>
<td>TI,AB(validation OR method*)</td>
<td>863,142</td>
</tr>
<tr>
<td>S2</td>
<td>TI,AB(&quot;cohort studies&quot; OR &quot;cohort study&quot; OR &quot;cohorts design&quot; OR &quot;prospective cohort&quot; OR &quot;retrospective cohort&quot; OR &quot;data integration&quot; OR bias OR &quot;cross study&quot; OR &quot;cross studies&quot;)</td>
<td>102,458</td>
</tr>
<tr>
<td>S1</td>
<td>TI,AB(&quot;stratified medicine&quot; OR biomarker* OR &quot;precision medicine&quot; OR &quot;personalized medicine&quot; OR &quot;personalised medicine&quot; OR &quot;individualized Medicine&quot; OR &quot;individualised Medicine&quot; OR &quot;individualized therapy&quot; OR &quot;individualised therapy&quot; OR &quot;patient stratification&quot; OR pharmacogenetics OR &quot;patient specific modeling&quot; OR &quot;personalized clinical decision making&quot; OR &quot;personalised clinical decision making&quot; OR &quot;prediction of response&quot;)</td>
<td>19,287</td>
</tr>
</tbody>
</table>
Oncology Searches

Pubmed 27/3/2020

<table>
<thead>
<tr>
<th>Search</th>
<th>Actions</th>
<th>Details</th>
<th>Query</th>
<th>Results</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>#26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Search: #16 OR #19 Filters: English, French, German, Italian, Spanish</td>
<td></td>
<td></td>
<td></td>
<td>1,047</td>
<td>11:05:09</td>
</tr>
<tr>
<td>#22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Search: #16 OR #19 Filters: English, from 2005 - 2020</td>
<td></td>
<td></td>
<td></td>
<td>1,047</td>
<td>11:03:19</td>
</tr>
<tr>
<td>#21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Search: #16 OR #19 Filters: from 2005 - 2020</td>
<td></td>
<td></td>
<td></td>
<td>1,123</td>
<td>11:03:08</td>
</tr>
<tr>
<td>#20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Search: #16 OR #19</td>
<td></td>
<td></td>
<td></td>
<td>1,345</td>
<td>11:02:53</td>
</tr>
<tr>
<td>#19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Search: #5 AND #18</td>
<td></td>
<td></td>
<td></td>
<td>515</td>
<td>11:01:27</td>
</tr>
<tr>
<td>#18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Search: &quot;systematic review&quot;</td>
<td></td>
<td></td>
<td></td>
<td>186,019</td>
<td>11:01:04</td>
</tr>
<tr>
<td>#16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Search: #1 AND #2 AND #3 AND #4 Filters: Meta-Analysis, Review</td>
<td></td>
<td></td>
<td></td>
<td>1,142</td>
<td>10:59:57</td>
</tr>
<tr>
<td>#5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Search: #1 AND #2 AND #3 AND #4</td>
<td></td>
<td></td>
<td></td>
<td>42,105</td>
<td>10:59:41</td>
</tr>
<tr>
<td>#4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Search: “cohort studies” OR “cohort study” OR “cohorts design” OR “prospective cohort” OR “retrospective cohort” OR “data integration” OR bias OR “cross study” OR “cross studies” OR cohort studies [Mesh] OR Cross-Sectional Studies [Mesh] OR Prospective Studies[Mesh]</td>
<td></td>
<td></td>
<td></td>
<td>2,487,093</td>
<td>10:52:12</td>
</tr>
<tr>
<td>#1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Search: “stratified medicine” OR biomarker* OR “precision medicine” OR “personalized medicine” OR “personalised medicine” OR “individualized Medicine“ OR “individualised Medicine“ OR “individualized therapy“ OR “individualised therapy“ OR “patient stratification” OR pharmacogenetics OR “patient specific modeling” OR “personalized clinical decision making” OR “personalised clinical decision making” OR “prediction of response”</td>
<td></td>
<td></td>
<td></td>
<td>983,810</td>
<td>10:51:41</td>
</tr>
</tbody>
</table>
**PERMIT**

**PERsonalised Medicine Trials**

**OR “prediction of responses” OR “Biomarkers” [Mesh] OR "Precision Medicine" [Mesh]**

**Embase 27/3/2020**

<table>
<thead>
<tr>
<th>Query</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>#24 #21 AND #22 AND ([english]/lim OR [french]/lim OR [german]/lim OR [italian]/lim OR [spanish]/lim)</td>
<td>429</td>
</tr>
<tr>
<td>#23 #21 AND #22</td>
<td>438</td>
</tr>
<tr>
<td>#22 [embase]/lim NOT [medline]/lim</td>
<td>9595932</td>
</tr>
<tr>
<td>#21 (#18 OR #19) AND [2005-2020]/py</td>
<td>1302</td>
</tr>
<tr>
<td>#20 #18 OR #19</td>
<td>1355</td>
</tr>
<tr>
<td>#19 #17 AND [review]/lim</td>
<td>779</td>
</tr>
<tr>
<td>#18 #4 AND #9 AND #13 AND #16 AND ([cochrane review]/lim OR [systematic review]/lim OR [meta analysis]/lim)</td>
<td>818</td>
</tr>
<tr>
<td>#17 #4 AND #9 AND #13 AND #16</td>
<td>20180</td>
</tr>
<tr>
<td>#16 #14 OR #15</td>
<td>5241925</td>
</tr>
<tr>
<td>#15 cancer*:ti,ab OR carcinoma*:ti,ab OR tumor*:ti,ab OR tumour*:ti,ab OR oncolo*:ti,ab OR leukemia*:ti,ab OR lymphoma*:ti,ab OR sarcoma*:ti,ab</td>
<td>4465029</td>
</tr>
<tr>
<td>#14 'neoplasm'/exp/mj</td>
<td>3703629</td>
</tr>
<tr>
<td>#13 #10 OR #11 OR #12</td>
<td>29597086</td>
</tr>
<tr>
<td>#12 validation:ti,ab OR method*:ti,ab</td>
<td>9431867</td>
</tr>
<tr>
<td>#11 'procedures'/exp</td>
<td>28915500</td>
</tr>
<tr>
<td>#10 'validation study'/exp</td>
<td>81960</td>
</tr>
<tr>
<td>#9 #5 OR #6 OR #7 OR #8</td>
<td>1592114</td>
</tr>
<tr>
<td>#8 'prospective study'/exp</td>
<td>588853</td>
</tr>
<tr>
<td>#7 'cross-sectional study'</td>
<td>339070</td>
</tr>
<tr>
<td>#6 'cohort analysis'/exp</td>
<td>560295</td>
</tr>
<tr>
<td>#5 'cohort studies':ti,ab OR 'cohort study':ti,ab OR 'cohorts design':ti,ab OR 'prospective cohort':ti,ab OR 'retrospective cohort':ti,ab OR 'data integration':ti,ab OR 'cross study':ti,ab OR 'cross studies':ti,ab</td>
<td>511111</td>
</tr>
<tr>
<td>#4 #1 OR #2 OR #3</td>
<td>535142</td>
</tr>
<tr>
<td>#3 'personalized medicine'/exp</td>
<td>41477</td>
</tr>
<tr>
<td>#2 'biological marker'/exp</td>
<td>296253</td>
</tr>
<tr>
<td>#1 'stratified medicine':ti,ab OR 'biomarker':ti,ab OR 'precision medicine':ti,ab OR 'personalized medicine':ti,ab OR 'personalised medicine':ti,ab OR 'individualized medicine':ti,ab OR 'individualised medicine':ti,ab OR 'individualized therapy':ti,ab OR 'individualised therapy':ti,ab OR 'patient stratification':ti,ab OR 'pharmacogenetics':ti,ab OR 'patient specific modeling':ti,ab OR 'personalized clinical decision making':ti,ab OR 'personalised clinical decision making':ti,ab OR 'prediction of response':ti,ab OR 'prediction of responses':ti,ab</td>
<td>420990</td>
</tr>
</tbody>
</table>

**Web of Science 27/3/2020**

<table>
<thead>
<tr>
<th>#</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>590</td>
</tr>
<tr>
<td>9</td>
<td>463</td>
</tr>
</tbody>
</table>

Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI
Timespan=All years

PERMIT has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N. 874825
# 8 274,177 TOPIC: ("systematic review" OR "meta analysis")

Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI
Timespan=All years

# 7 350 #4 AND #3 AND #2 AND #1


Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI
Timespan=All years

# 6 2,340 #4 AND #3 AND #2 AND #1


Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI
Timespan=All years

# 5 2,402 #4 AND #3 AND #2 AND #1

Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI
Timespan=All years

# 4 3,811,935 TOPIC: (neoplasm* OR Cancer* OR carcinoma* OR tumor* OR tumour*OR oncolo* OR leukemia* OR lymphoma* OR sarcoma*)

Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI
Timespan=All years

# 3 11,099,593 TOPIC: (validation OR method*)

Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI
Timespan=All years

# 2 753,244 TOPIC: ("cohort studies" OR "cohort study" OR "cohorts design" OR "prospective cohort" OR "retrospective cohort" OR "data integration" OR bias OR "cross study" OR "cross studies")

Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI
Timespan=All years

# 1 368,188 TOPIC: ("stratified medicine" OR biomarker* OR "precision medicine" OR "personalized medicine" OR "personalised medicine" OR "individualized Medicine" OR "individualised Medicine" OR "individualized therapy" OR "individualised therapy" OR "patient stratification" OR pharmacogenetics OR "patient specific modeling" OR "personalized clinical decision making" OR "personalised clinical decision making" OR "prediction of response" OR
"prediction of responses")

Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI
Timespan=All years

Cochrane Library 27/3/2020

#1 "stratified medicine":ti,ab 29
#2 biomarker*:ti,ab OR 'precision medicine':ti,ab OR 'personalized medicine':ti,ab OR 'individualised medicine':ti,ab OR 'individualized medicine':ti,ab OR 'individualized therapy':ti,ab OR 'individualised therapy':ti,ab OR 'patient stratification':ti,ab OR pharmacogenetics:ti,ab OR 'patient specific modeling':ti,ab OR 'personalized clinical decision making':ti,ab OR 'personalised clinical decision making':ti,ab OR 'prediction of response':ti,ab OR 'prediction of responses':ti,ab 34529
#3 #1 or #2 34542
#4 'cohort studies':ti,ab OR 'cohort study':ti,ab OR 'cohorts design':ti,ab OR 'prospective cohort':ti,ab OR 'retrospective cohort':ti,ab OR 'data integration':ti,ab OR bias:ti,ab OR 'cross study':ti,ab OR 'cross studies':ti,ab 125366
#5 validation:ti,ab OR method*:ti,ab validation:ti,ab OR method*:ti,ab 637228
#6 neoplasm*:ti,ab OR cancer*:ti,ab OR carcinoma*:ti,ab OR tumor*:ti,ab OR tumour*:ti,ab OR oncolo*:ti,ab OR leukemia*:ti,ab OR lymphoma*:ti,ab OR sarcoma*:ti,ab 189985
#7 #3 AND #4 AND #5 AND #6 with Publication Year from 2005 to 2020, in Trials 1047
#8 REVIEW* 1621582
#9 #7 AND #8 1047
#10 "accession number" near pubmed 659110
#11 "accession number" near EMBASE 533181
#12 #10 or #11 993434
#13 #9 not #12 67
#14 clinicaltrials.gov 207571
#15 #13 NOT #14 29
#16 www.who.int 135542
#17 #15 not #16 13

Stroke Searches

Pubmed 3/6/2020

<table>
<thead>
<tr>
<th>Search</th>
<th>Actions</th>
<th>Details</th>
<th>Query</th>
<th>Results</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>#34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Search: #4 AND #28 Filters: Meta-Analysis, Review, Systematic Reviews
#33

Search: #4 AND #28 Filters: Review, Systematic Reviews
#32

Search: #4 AND #28 Filters: Systematic Reviews, from 2005 - 2020
#31

Search: #4 AND #28 Filters: from 2005 - 2020
#30
PERMIT has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N. 874825
<table>
<thead>
<tr>
<th>No.</th>
<th>Query</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>#15</td>
<td>'brain ventricle hemorrhage'/exp</td>
<td>1027</td>
</tr>
<tr>
<td>#14</td>
<td>'lacunar stroke'/exp</td>
<td>3336</td>
</tr>
<tr>
<td>#13</td>
<td>'brain infarction'/exp</td>
<td>74495</td>
</tr>
<tr>
<td>#12</td>
<td>'brain ischemia'/exp</td>
<td>185994</td>
</tr>
<tr>
<td>#11</td>
<td>'transient ischemic attack'/exp</td>
<td>38968</td>
</tr>
<tr>
<td>#10</td>
<td>'cerebrovascular disease'/de</td>
<td>63275</td>
</tr>
<tr>
<td>#9</td>
<td>'cryptogenic stroke':ti,ab OR 'transient ischemic attack*':ti,ab OR 'brain infarction*':ti,ab OR 'lacunar stroke':ti,ab OR 'intraparenchymal hemorrhage':ti,ab OR 'intraventricular hemorrhage':ti,ab OR 'ischemic stroke':ti,ab OR 'ischemic':ti,ab OR 'ischaemic':ti,ab OR 'stroke*':ti,ab OR 'cerebrovascular accident*':ti,ab OR cva:ti,ab</td>
<td>650270</td>
</tr>
<tr>
<td>#8</td>
<td>#3 AND #6 AND #7</td>
<td>63599</td>
</tr>
<tr>
<td>#7</td>
<td>'validation study'/exp OR 'procedures'/exp OR validation:ti,ab OR method*':ti</td>
<td>2933023</td>
</tr>
<tr>
<td>#6</td>
<td>#4 OR #5</td>
<td>1636917</td>
</tr>
<tr>
<td>#5</td>
<td>'cohort analysis'/exp OR 'cross-sectional study'/exp OR 'prospective study'/exp</td>
<td>1390377</td>
</tr>
<tr>
<td>#4</td>
<td>'cohort studies':ti,ab OR 'cohort study':ti,ab OR 'cohorts design':ti,ab OR 'prospective cohort':ti,ab OR 'retrospective cohort':ti,ab OR 'data integration':ti,ab OR 'cross study':ti,ab OR 'cross studies':ti,ab</td>
<td>523306</td>
</tr>
<tr>
<td>#3</td>
<td>#1 OR #2</td>
<td>547946</td>
</tr>
<tr>
<td>#2</td>
<td>'biological marker'/exp OR 'personalized medicine'/exp</td>
<td>341290</td>
</tr>
<tr>
<td>#1</td>
<td>'stratified medicine':ti,ab OR 'biomarker*':ti,ab OR 'precision medicine':ti,ab OR 'personalized medicine':ti,ab OR 'personalised medicine':ti,ab OR 'individualized medicine':ti,ab OR 'individualised medicine':ti,ab OR 'individualized therapy':ti,ab OR 'industrialised therapy':ti,ab OR 'patient stratification':ti,ab OR pharmacogenetics:ti,ab OR 'patient specific modeling':ti,ab OR 'personalized clinical decision making':ti,ab OR 'personalised clinical decision making':ti,ab OR 'prediction of response':ti,ab OR 'prediction of responses':ti,ab</td>
<td>431979</td>
</tr>
</tbody>
</table>

**Web of Science 3/6/2020**

# 11 61 #10 OR #8

Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI
Timespan=All years

# 10 51 #9 AND #7

Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI
Timespan=All years

# 9 282,750 TOPIC: (((("systematic review" OR "meta analysis") )))

Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI
Timespan=All years

# 8 33 #5 AND #4

Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI
PERMIT has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N. 874825

Timespan=All years

# 7 395  #5 AND #4
Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI
Timespan=All years

# 6 396  #5 AND #4
Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI
Timespan=All years

# 5 515,470  TOPIC: ("Cryptogenic Stroke" OR "transient ischemic attack**" OR "TIA" OR "brain infarction**" OR "Lacunar stroke" OR "intraparenchymal hemorrhage" OR "intraventricular hemorrhage" OR "ischemic stroke" OR "TIA OR ischemic OR ischaemic OR stroke* OR "CEREBROVASCULAR ACCIDENT**" OR cva)
Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI
Timespan=All years

# 4 7,499  #3 AND #2 AND #1
Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI
Timespan=All years

# 3 11,303,677  TOPIC: (((validation OR method*)) ))
Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI
Timespan=All years

# 2 768,507  TOPIC: ((("cohort studies" OR "cohort study" OR "cohorts design" OR "prospective cohort" OR "retrospective cohort" OR "data integration" OR bias OR "cross study" OR "cross studies") ))
Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI
Timespan=All years

# 1 378,951  TOPIC: ((("stratified medicine" OR biomarker* OR "precision medicine" OR "personalized medicine" OR "personalised medicine" OR "individualized Medicine" OR "individualised Medicine" OR "individualized therapy" OR "individualised therapy" OR "patient stratification" OR pharmacogenetics OR "patient specific modeling" OR "personalized clinical decision making" OR "personalised clinical decision making" OR "prediction of response" OR "prediction of responses") )))
Indexes=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, ESCI
Timespan=All years
<table>
<thead>
<tr>
<th>Date of search</th>
<th>Source</th>
<th>Url</th>
<th>Relevant document identified (Y/N)</th>
<th>SEARCH STRATEGY</th>
</tr>
</thead>
<tbody>
<tr>
<td>May 11, 2020</td>
<td>Heads of Medicines Agencies</td>
<td><a href="https://www.hma.eu">https://www.hma.eu</a></td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>May 11, 2020</td>
<td>EUROPEAN ALLIANCE FOR PERSONALIZED MEDICINE</td>
<td><a href="https://www.euapm.eu">https://www.euapm.eu</a></td>
<td>N</td>
<td></td>
</tr>
<tr>
<td>May, 15</td>
<td>European commission</td>
<td><a href="https://ec.europa.eu/">https://ec.europa.eu/</a></td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>March 11, 2020</td>
<td>FDA</td>
<td><a href="https://www.fda.gov/">https://www.fda.gov/</a></td>
<td>Y</td>
<td>&quot;Personalised medicine&quot;</td>
</tr>
<tr>
<td>June 4, 2020</td>
<td>Explore the British Library</td>
<td><a href="http://explore.bl.uk">http://explore.bl.uk</a></td>
<td>Y</td>
<td>&quot;Personalised medicine&quot;</td>
</tr>
<tr>
<td>June 4, 2020</td>
<td>ESMO</td>
<td><a href="https://www.esmo.org/">https://www.esmo.org/</a></td>
<td>N</td>
<td>(&quot;personalised medicine&quot;+&quot;legal&quot;); (&quot;personalised medicine&quot; + &quot;regulation&quot;) filtered by &quot;ethics, legal...&quot;</td>
</tr>
<tr>
<td>June 5, 2020</td>
<td>OpenGrey</td>
<td><a href="http://www.opengrey.eu/">http://www.opengrey.eu/</a></td>
<td>Y</td>
<td>&quot;Personalised medicine&quot;; &quot;cohorts validation&quot;; &quot;cohorts stratification&quot; (&quot;legal&quot; + &quot;personalised medicine&quot;); (&quot;ethics&quot; + &quot;personalised medicine&quot;); (&quot;data management&quot;+&quot;personalised medicine&quot;)</td>
</tr>
<tr>
<td>June 5, 2020</td>
<td>National Technical Information Service (NTIS)</td>
<td><a href="http://www.ntis.gov/">www.ntis.gov/</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>June 5, 2020</td>
<td>European Medicines Agency (EMA)</td>
<td><a href="https://www.ema.europa.eu/">https://www.ema.europa.eu/</a></td>
<td>Y</td>
<td>Personalised medicine; (&quot;legal frame&quot; + &quot;personalised medicine&quot;); (&quot;ethics&quot; + &quot;personalised medicine&quot;)</td>
</tr>
<tr>
<td>June 5, 2020</td>
<td>Canadian Agency for Drugs and Technologies in Health (CADTH)</td>
<td><a href="https://www.cadth.ca/">https://www.cadth.ca/</a></td>
<td>N</td>
<td>personalised medicine</td>
</tr>
<tr>
<td>Year</td>
<td>National Institute for health and Care Excellence</td>
<td>June 5, 2020</td>
<td><a href="https://www.nice.org.uk/">https://www.nice.org.uk/</a></td>
<td>N</td>
</tr>
<tr>
<td>--------</td>
<td>--------------------------------------------------</td>
<td>-------------</td>
<td>--------------------------</td>
<td>---</td>
</tr>
</tbody>
</table>
Appendix II: Outline of methods and tools for data-analysis and for data integration

Methods and tools for data-analysis

-Oncology

· Strategies:
  Independent modelling, conditional modelling and joint modelling.
  
  · Methods and tools:
    
    A. Statistical method approaches.
    - Dimension reduction
      Principal Component Analysis (PCA) and Factor Analysis and variations. These variations include Multiple Factor Analysis (MFA), consensus PCA (CPCA), multiple-block PCA (MBPCA) and non-negative matrix factorization (NMF).
    - Data integration or data fusion.
      Multivariate, concatenation-based and transformation-based methods.
      Tools: R/Bioconductor packages MultiDataSet and MultiAssay experiment (between others), and some tools more specific: Caleydo StratomeX, CAS-viewer, cBio Cancer Genomic Portal, Genboree Workbench, MARIO, mixOmics, ModulOmics, Omics Integrator and XENA UCSC browser.
    
    - Causal inference
      Parallel regression, Bayesian networks, sparse generalized canonical correlation analysis (sGCCA), Multi-omics factor analysis (MOFA), Mendelian randomization, and Joint and Individual Variation Explained (JIVE), Inverse probability weighting and component-based method.
    
    -Criterion for predictor selection:
      · Method of predictor selection used within multivariable analysis
        Backward selection, forward selection, added value of a specific predictor to existing predictors or model, all predictors included regardless of statistical significance and similar predictors combined
      
      · Criterion for selection of predictors in multivariable analyses
        p-Value cut-off at <0.05 or lower, p-Value cut-off higher than 0.05, Akaike’s information criterion, Bayesian information criterion, explained variance (R2) and change in C-statistic.
    
    - Multiple imputation
B. Machine learning approach

- Support vector machine (SVM)
- Random forest (RF)
- Convolutional neural network (CNN)
- Deep learning

C. Network modelling approach

Types of networks: Protein–protein interaction (PPI) network, miRNA-mRNA regulatory network and gene co-expression network.

Analysis methods for networks:

- Network analyses
  Tools: Cytoscape, ARACNe, WGCNA, IPA, Bibliosphere
- Enrichment analysis (Gene Set Enrichment Analysis)
  Tools: DAVID, g:Profiler, AmiGO, Onto-Tools, ClueGO Golorize, FatiGO, GoStat
- Analysis of protein interactions
  Tools: HPRDR, Bioconductor, Matlab, GenePattern, BIND, DIP, BioGRID, STRING, Annotator, Pfam, PROSITE, InterPro, ProDom, SMART, BLOCKS, UniProt
- Pathway analysis
  Tools: Reactome, Cancer Cell Map, PathwayExplorer, GenMAPP, INOHGEO, ArrayExpress, SMD, Oncomine, PANTHER, Science signaling map
- Multistage or meta-dimensional fashion (as an example, for drawing inferences from these data, involves pairwise analyses of data sets).

- **Multiple diseases**

A. Statistical method approaches

- Dimension reduction
  Principal Component Analysis (PCA) and Factor Analysis and variations. These variations include Multiple Factor Analysis (MFA), non-negative matrix factorization (NMF) of different kinds: iNMF, Joint NMF and regularized NMF.

- Data integration or data fusion.

Multivariate analysis:
Multiple co-inertia analysis (MCIA), Joint and Individual Variation Explained (JIVE), Joint NMF, Single vector decomposition (t-SVD), Sparse multi-block partial least squares (smBPLS), Penalized multivariate analysis (PMA), Feature selection
multiple kernel learning (FSMKL), Non-negative matrix factorization (NMF), Graph-regularized non-negative matrix tri-factorization (GNMTF), Multiple factor analysis (MFA), MoCluster and Regularized multiple kernel learning- locality preserving projection (rMKL-LPP), Joint kernel matrices, Multiple similarities collaborative matrix factorization (MSCMF), CNAmet

Multivariate, concatenation-based and transformation-based methods.

**Tools:** R/Bioconductor packages mixOmics, Integromics, iPAC, Camelot

- Causal inference
- Bayesian inference

Joint and Individual Variation Explained (JIVE), multi-omics factor analysis (MOFA), Bayesian consensus clustering (BCC), patient-specific data fusion (PSDF), Pathway Recognition Algorithm using Data Integration on Genomic Models (PARADIGM), iCluster, iClusterPlus, LRAcluster, Joint Bayesian factor, multi-omics factor analysis (MOFA), Multiple dataset integration (MDI), Molecular Regularized Consensus Patient Stratification (MRCPS), factor analysis - linear discriminant analysis (FALDA)

**B. Machine learning approach**
Classification taking into account five specific computational challenges associated with integrative analysis:
- Curse of dimensionality
Feature extraction (FE) or feature selection (FS)

- Data heterogeneity
Tree-based learning, penalized linear models, Multiple Kernel Learning, Graphs and Networks, Latent Sub-space Clustering and Deep Learning)
- Missing data
Single Imputation, Maximum likelihood, Multiple imputation, Matrix Factorization, Deep Learning/Autoencoder, Integrative imputation
- Class imbalance
Data Sampling, Cost-Sensitive Learning and Evaluation Measure-based
- Scalability issues
Efficient Algorithms for Big Data, Online Machine Learning, Distributed ML implementations and Cloud computing solutions

- Specific methods of deep learning
Autoencoder, convolutional neural network, recurrent neural network, DNN-MDA (DNN, deep neural network; MDA, mean decrease accuracy) and DeepNovo.

**C. Network modelling approach**
Pathway Recognition Algorithm using Data Integration on Genomic Models (PARADIGM), similarity network fusion (SNF), network-based (NBS), Coupled network propagation, Network completion, and NetLCS.

- **Stroke**

  A. Statistical method approaches

  - Causal inference
  
  Mendelian randomisation: Analytical tool to dissect the roles of risk factors, to separate causal markers from only markers (those without causal relation).

**Methods for different kinds of data integration**

- **Multiple disease**
  
  - Genetic risk scores, regression methods and Bayesian networks.

  - Constraint-based modelling

  - Methods that allow to overcome multicollinearity problem while modelling huge number of variables:

  - Regularized regression methods, such as ridge regression, the Least Absolute Shrinkage and Selection Operator (LASSO), a combination of the last two, the Elastic Net (ENET), the Bayesian LASSO, or other Bayesian regularized regression methods such as Reproducing Kernel Hilbert Spaces Regression models (RKHS)

- **Oncology**
  
  Similarity Network Fusion and Multiblock Partial least squares (or projection to latent structures, MBPLS), Naïve Bayesian integration model.
Appendix III – Full scoping review protocol
Methodological approaches for personalised medicine: a series of scoping reviews
Protocol V.2– 29 April 2020

Authors
PERMIT WP 2 group
Methods team
Rita Banzi, Chiara Gerardi, Maddalena Fratelli
Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Milan, Italy

Review teams:
Teresa Torres, Josep Maria Haro Abad, Albert Sanchez Niubo
Parc Sanitari San Joan de Déu, San Boi del Llobregat, Barcelona, Spain
Enrico Glaab
Luxembourg Centre for Systems Biomedicine, Université du Luxembourg, Luxembourg
Emanuela Oldoni, Florence Bietrix, Vibeke Fosse, Emmet McCormack
European infrastructure for Translational Medicine(EATRIS), Amsterdam, The Netherlands and University of Bergen, Bergen, Norway
Raphael Porcher, Cecilia Superchi
Centre de Recherche Épidémiologie et Statistiques Université de Paris, Université Paris Descartes, Paris, France

PERMIT coordination
Paula Garcia, Jacques Demotes-Mainard
European Clinical Research Infrastructure Network (ECRIN), Paris, France

Protocol reviewed by WP2 contributors:
Lorena San Miguel, Frank Hulstaert
Belgium Health Care Knowledge Centre, Bruxelles, Belgium
Alexander Grundmann, Rainer Girgenrath
Deutsche Zentrum für Luft- und Raumfahrt, Köln, Germany
Luis M. Sánchez-Gómez, Mar Polo-de Santos, Iñaki Imaz-Iglesia, Setefilla Luengo
Instituto de Salud Carlos III, Madrid, Spain
Table of Contents
Introduction ......................................................................................................................... 3
Personalised medicine research ......................................................................................... 4
Review methods .................................................................................................................. 6
  Identification of the research questions ........................................................................ 7
Scope & Research Questions ............................................................................................... 7
  1. Methods for stratification and validation cohorts ....................................................... 7
  2. Methods for machine learning applied to stratification ............................................ 8
  3. Pre-clinical methods for translational development of stratified therapies and treatments selection .......................................................................................................................... 9
  4. Methods for clinical trials in personalised medicine .................................................. 9
Study identification ............................................................................................................ 10
Eligibility Criteria ................................................................................................................ 10
  1. Methods for stratification and validation cohorts ....................................................... 10
  2. Methods for machine learning applied to stratification ............................................ 11
  3. Pre-clinical methods for translational development of stratified therapies and treatments selection .......................................................................................................................... 11
  4. Methods for clinical trials in personalised medicine .................................................. 12
Study selection .................................................................................................................. 12
Data extraction ................................................................................................................... 12
Study quality ..................................................................................................................... 13
Plan for presenting the results ............................................................................................ 13
Consultation exercise .......................................................................................................... 13
Funding ............................................................................................................................... 14
References ......................................................................................................................... 14
Introduction

The concept of personalised medicine is going to impact how pharmacological treatments are discovered and developed, how patients are diagnosed and treated, and how health care systems allocate their resources to maximize patient benefits. Personalised medicine may be considered an extension of traditional approaches to understanding and treating disease. Ideally, it could serve to take clinical decisions based on a patient’s profile (often molecular, but the concept is broader) to minimise harmful side effects, ensure a more successful outcome, and possibly help contain costs compared with a “trial-and-error” approach to disease treatment (1).

Personalised medicine stems on the broad concept that managing a patient's health should be based on the individual patient's specific characteristics, including age, gender, height/weight, diet, environment, etc. Different understandings of personalised medicine exist, in which three main positions can be identified (2):
(a) personalised medicine is not a new concept as medicine has always been individualized;
(b) personalised medicine is holistic health care, centred around the needs of the individual patient;
(c) personalised medicine is treatment targeted at stratified subgroups (e.g. pharmacogenetics).

Even when the focus is restricted to the third position, there is not a unique definition of personalised medicine, nor a straightforward terminology to define this concept. While "personalised" emphasizes the notion of individualized— “this is exclusively designed for you”, other more scientifically rigorous terms such as stratified medicine refer to the identification of groups or strata of patients with specific molecular characteristics or other determining factors which predict susceptibility to disease, disease prognosis, and/or response to therapy. Some authors suggested that rather than considering personalised medicine as a precise scientific concept, it should be understood as an open and negotiable ideal that accounts for a plurality of visions, depending on people, reasons and interests behind these alternative conceptions (3).

Regarding the terminology, in the European context, the term personalised medicine is preferred, as this term best reflects the ultimate goal of effectively tailoring treatment based on an individual’s ‘personal profile’, as determined by the individual’s genetic and phenotypical characteristics. Other terms are widespread, for instance stratified medicine, mainly used in the UK, or precision medicine mostly used in US and broadly referred to the 4 P (preventive, predictive, personalised and participatory) medicine. While there may be small nuances in the literal meanings of these terms, they usually refer to the same concept when applied in practice (4).

A recent review reported that the literature about personalised medicine usually refers to two different semantic approaches. Firstly, patients’ stratification, that is grouping individual patients in subpopulation according to their probability to have a therapeutic benefit from a drug or regimen. Secondly, treatment tailoring, that is the individual status of a patient (i.e., disease characteristics or subject’s genotype/phenotype) is the rationale basis for drug choice (5).

Box 1 reports a collection of definitions, along with their references.

A broad community of stakeholders, including funders and professionals involved in medical research and care, are increasingly concerned with ensuring that the right patient receives the right therapy, at the right dose and at the right time. The identification of markers of mechanistic pathways or multiple variables characterising clusters of subjects that might inform meaningful disease stratification may have different clinical applications in the context of personalised medicine. Broadly, stratification may be applied at the diagnosis level (e.g., to identify a particular pathophysiological/clinical stratum within a heterogeneous patient population for diagnostic purposes), to predict disease course (prognostic value), the development of a disease (predictive value), or the response to therapy (theragnostic value).
Regardless of the application, any approach to personalised medicine should undergo different phases: discovery, validation and definition of usefulness from a clinical perspective. Robust methodological approaches are needed to deal with the complexity and heterogeneity of the process, as well as the range of possible applications to stratification using multidimensional data (what is meant by “molecular profiling” among other terms).

**Personalised medicine research**

This series of scoping reviews will map the general concept of methods for personalised medicine, to set the basis for the discussion on robustness and reproducibility of personalised medicine development programmes. The final goal is the identification of standards and needs in terms of methodology of data generation, management, analysis and interpretation to improve clinical studies in personalised medicine.

The group of authors agreed on a common operational definition of **personalised medicine research**: a set of comprehensive methods, (methodological, statistical, validation or technologies) to be applied in the different phases of the development of a personalised approach to treatment, diagnosis, prognosis, or risk prediction. Ideally, robust and reproducible methods should cover all the steps between the generation of the hypothesis (e.g., a given stratum of patients could better respond to a treatment), its validation and pre-clinical development, and up to the definition of its value in a clinical setting.

The process leading from the hypothesis to the clinic is complex and not always linear. The Medical Research Council in UK recently developed a framework for the development, design and analysis of stratified medicine (6) that is structured in six themes:

- **Theme 1: Framing the Question/Defining the Population**
- **Theme 2: Designing Stratum Discovery Studies; selecting variables, defining response and powering**
- **Theme 3: Assay Design; managing complexity and variability**
- **Theme 4: Defining Strata; data integration, linkage to existing knowledge, linkage to outcome**
- **Theme 5: Stratum Verification**
- **Theme 6: Progression Towards Clinical Utility**

Any attempt for classifying the phases of personalised medicine may appear as an oversimplification. However, a typical research programme in personalised medicine would include: first a stratification cohort (in many cases a retrospective study reusing data and biosamples from existing cohorts) with extensive multimodal data on which stratification algorithms are run, then a validation cohort, normally prospective, that assesses the reproducibility, robustness and validity of the clustering in another sufficiently large patient sample. Thirdly, a translational step is often necessary. In some cases, the use of pre-clinical models (cellular, in-silico, organoid) might be useful to give confidence in the allocation of patients to specific treatment arms as identified through clustering. Alternatively, the multi-omics profiles from clinical samples can lead to the identification of new disease categories, prediction of disease prognosis, exploration of drug sensitivity and dose selection. Finally, treatment options should be tested in the subgroups of patients in the context of clinical studies, ideally randomised clinical trials, to generate evidence informing regulatory, clinical and coverage decisions.

However, many alternative pathways can be proposed. In some case, the stratification provides detailed information on the mechanism of disease and strong indications on the treatments to be tested in each patient cluster. This is for instance the case where identification of driver somatic mutations in cancer cells suggests the targeted treatment to be tested. In other cases, the stratification cohort includes data on response to an established treatment, making the translational step less necessary. Research programmes may be limited to the stratification step, in particular when no treatment is available – this is the case for instance for taxonomy studies in
neurodegenerative disorders, aiming at identifying homogeneous clusters of patients. In any case, personalised medicine research is a complex programme, with multiple steps and lasting many years.

We consider out of the scope of this review the methods used for the clinical implementation of personalised medicine, the manufacturing and use of individualized treatments, and the pragmatic approach to individual patient care, such as n-of-1 trials.

Considering this framework outlined by Figure 1, the scoping reviews will approach **personalised medicine research** focusing on four main phases:

- Methods for stratification and validation cohorts
- Methods for machine learning applied to stratification
- Pre-clinical methods for translational development of stratified therapies and treatments selection
- Methods for clinical trials in personalised medicine

---

**Figure 1: Main steps in personalised medicine research programmes**
Review methods

We aim to perform a set of scoping reviews investigating various aspects of the methodology applied in personalised medicine research programmes as outlined in the Scope & Research Questions section.

Scoping reviews are used to present a broad overview of the evidence pertaining to a topic; they are useful to examine areas that are emerging, to clarify key concepts and identify gaps. Scoping reviews have great utility for synthesizing research evidence and are often used to map existing literature in a given field in terms of its nature, features, and volume. They differ from standard systematic reviews that are usually aimed to answer a specific question or series of questions according to a rigid set of a priori eligibility criteria. Scoping reviews have a broader approach, generally with the aim of mapping literature and addressing broader research questions. Due to the iterative nature of scoping reviews, deviations from the protocol are expected, differently from what happens in systematic reviews. Anyway, the discrepancies from the protocol will be clearly detailed and justified in the ‘Methods’ section of the scoping review report, if and when they occur.

To ensure the transparency and reproducibility of the review process, we will follow the methodological guidance for the conduct of scoping reviews suggested by the Joanna Briggs Institute (7, 8). The main steps of the process are summarised in Figure 2.

Figure 2: Main steps in the preparation of scoping review (in grey optional steps).

This overall process will be applied to the four themes outlined below by a dedicated review team supported by a methods team. Each step may require small adaptations given the nature of the research questions and scope defined and the type of literature/data that will be retrieved.

The four reviews will be part of a unique report covering the different aspects of methodology to inform the gap analysis and the subsequent phases of the PERMIT project (https://permit-eu.org/).
Identification of the research questions

The first step of any scoping review is to define the objective and research questions of interest. For the purpose of these reviews, four main aspects of the general concept of personalised medicine have been identified and will be the focus of this analysis:

- Methods for stratification and validation cohorts
- Methods for machine learning applied to stratification
- Pre-clinical methods for translational development of treatment options and treatments selection
- Methods for clinical trials in personalised medicine

Through several rounds of joint discussion and one face-to-face meeting (Paris, Jan 24, 2020), the four review teams had clarified the scope and defined research questions to the purpose of the scoping reviews. As the four topics are connected, this step also served to harmonise the four parts and avoid possible overlaps.

The outcome of this exercise is reported in the following section Scope & Research Questions.

Scope & Research Questions

1. Methods for stratification and validation cohorts

The scoping review will focus on:

- The characteristics of cohorts that have been used for patient stratification or validation of patient clustering obtained through stratification cohorts. Stratification cohorts of patients are used to create the clustering, and validation cohorts of patients are used to assess the reliability (robustness, reproducibility, etc.) of patient clustering.
- The different methods and tools used in design and management of stratification and validation cohorts (especially complex in multimodal approaches) to understand their limitations.

The review will not be restricted to a given type of data for stratification, i.e., genetic, metabolomics, gene expression, genomic, neuroimaging, etc.

General papers that describe methods and tools in the design and management of stratification and validation cohorts will be assessed irrespective of the diseases field. Case examples of biomarkers or multimodal data profiling in different medical fields and coming from different sources (omics, neuroimaging, genetics...) will be also analysed to explore the actual application of this methods and tools. Cancer, stroke and Alzheimer’s disease will be the three areas where informative examples will be collected, as they are complex conditions (many biological and environmental factors involved) and are representative of different approaches and degree of success in personalised medicine.

The main research questions addressed by the scoping review will be:

- What are the approaches to define the optimal size of stratification/validation cohorts?
- What are the differences, pros and cons of the prospective and retrospective nature of stratification and validation cohorts?
- What are the prerequisites and methods used for integration of multiple retrospective cohorts?
• Which validation designs exist for the stratification (or clustering) in personalised medicine? Which methods and tools are used to build the cohorts of validation (external/sub-cohort)? What are their gaps?
• What are the methods for the evaluation of the risk of bias?
• How are the (-omics, imaging, exposome, lifestyle etc.) data generated?
• What are the tools used for data management and multimodal data analysis used in personalised medicine (for instance, Galaxy)? What are their gaps?
• What quality of data of cohorts is needed to obtain a biomarker or multimodal data profiling? Are there requirements to monitor the collection of associated clinical data?
• What is the outlook of data generation seen as (CE-labelled) in-vitro diagnostics?

2. Methods for machine learning applied to stratification

The scoping review will focus on:

• Supervised and unsupervised machine learning methods for biomedical stratification using omics data. Few examples for other data modalities, e.g. imaging data, digital pathology and mobile sensor data will also be explored but not as the major focus.
• Cover both disease-based stratification (patient omics clustering, major focus) and drug-based stratification (clustering of drug-induced changes in patient-derived cells, minor focus)
• Methodologies that have been successfully validated and applied in clinical practice. New emerging approaches, which have not yet been sufficiently validated will also be explored but not as the major focus.
• Pros/cons, opportunities/limitations of different stratification methodologies and the associated validation approaches.
• Examples of successful applications.

Methodologies that have led to clinically validated biomarker signatures will be prioritised, as well as methodologies that have been cross-validated and externally tested on large sample sizes (preferably across multiple patient cohorts). Methods that lack statistical validation and a demonstrated biomedical application will be excluded.

The main research questions addressed by the scoping review will be:

• What are the main types of supervised and unsupervised machine learning methods for omics-based stratification in biomedicine (structured categorization)?
• What are the used and recommended workflows for supervised and unsupervised omics-based stratification (pre-processing, quality control, model building, model validation, model interpretation)?
• What are the specific strengths/weaknesses and opportunities/limitations of different types of omics-based stratification methods?
• Which validation methods exist for omics-based stratification models (assess accuracy, confirm biomedical relevance, test robustness) and what are their pros and cons?
• Which practical utility has been demonstrated for omics-based stratification and validation methods in real-world biomedical applications in the past (representative examples for previous success and/or failure stories, lessons learned)?
• What are the current gaps in standardization and methodological guidelines, and what is the outlook for the future of the field of omics-based machine learning stratification (new emerging approaches, new initiatives for data sharing, quality improvement, FAIRification)?
3. Pre-clinical methods for translational development of stratified therapies and treatments selection

The scoping review will focus on two aspects:

**3.1. Personalised clinical decision-making based on pre-clinical models, aimed to explore drug sensitivity screening step (cellular based assay, organoids, PDX model) to predict therapy response and allocation of patients to different treatment arm, dose ranges and other aspects relevant for initiation of clinical trials. Suitable use cases will be selected in fields other than oncology, where clinical trials have been performed using pre-clinical models for personalised clinical decision-making.**

The main research questions addressed will be:

- What are the fields of medicine other than oncology where pre-clinical models for personalised clinical decision-making have been applied?
- What are the pre-clinical models preferentially used in this context?
- How many drugs have been developed/are currently under development based on multi-omics profiling programs? What is the estimated success rate of the trial using this approach?
- What are the current gaps for broad implementation of pre-clinical testing for treatment selection?
- What information was collected at the pre-clinical stage to inform the clinical study design?

**3.2. Stratified medicine development, to show which pre-clinical models (cellular, animal, organoid, in silico) are currently used as validation methods prior to personalised medicine clinical trials, both in academia and in industry. The example use case will be oncology.**

As prospects, the review will discuss how to adapt the existing pre-clinical model systems to personalised medicine, and emerging models (such as in silico) which can replace the traditional animal models (3Rs). We will also perform a categorisation based on relevance and interpretation of models in the context of personalised medicine.

The main research questions addressed will be:

- Which pre-clinical models are currently used to provide validity data prior to therapeutic clinical trials of personalised medicine in oncology?
- What are the pros and cons of the various pre-clinical methods?
- Are the current pre-clinical models predictive for personalised medicine trials in oncology?

4. Methods for clinical trials in personalised medicine

The scoping review will focus on:

- Clinical trials, especially randomised trial designs, for personalised medicine.
- Trials evaluating a treatment in a subgroup of patients defined e.g. by a biomarker, in several clusters or subgroups of patients (e.g., basket or umbrella trials), trials comparing a personalised medicine strategy to a non-personalised strategy, or trials aiming at defining a subgroup of patients with enhanced response to treatment (e.g., adaptive enrichment design, adaptive signature design).
• Elements of clinical trial design applied to personalised medicine improving their appropriateness for HTA decision (e.g., external validity, choice of comparator, use of clinically meaningful outcome measure).
• Methodological reports (e.g., a scientific piece of work aiming at describing and evaluating the operational characteristics of a particular design) and guidance documents issued by regulatory or agencies for health technology assessment.
• Examples of published or ongoing trials in personalised medicine.

The review will not be restricted to a given medical field, although several examples in oncology are expected.

The main research questions addressed by the scoping review will be:

• What are the available designs for clinical trials applied to personalised medicine?
• What are the examples of current applications of these approaches?
• What are the pros and cons of the different approaches?
• What are the gaps in the current research on personalised medicine clinical trials?
• How is a personalised medicine strategy vs. non-personalised strategy evaluated?

Study identification
Relevant studies and documents will be identified balancing feasibility with breadth and comprehensiveness of searches.

Formal literature searches will be conducted on relevant databases (i.e., Medline, Embase, Cochrane Library) by the methods team. The keywords for the search strategy will be defined with the support of the review teams. Additional rounds of literature searches may be needed to refine specific aspects. The reference list of all identified reports and articles will be searched for additional studies.

To identify reports not published as scientific journal papers and unpublished (grey literature) information each review team will hand searching of relevant literature and websites (including conference meetings). Review teams may also contact relevant stakeholders to retrieve additional studies.

Documents published between 2005-2020 written in English, French, Spanish, Italian, German will be sought. Other specific time window, if deemed necessary by each review team, will be applied. Appropriate and clear justification for choices will be provided.

Appendix 1 reports examples of the search strategies planned for the four parts of this scoping review.

Eligibility Criteria
Each review team defined broad eligibility criteria based on the four “Scope & Research Questions’ sections.

1. Methods for stratification and validation cohorts
We will include articles and other reports describing the methods applied to cohorts that have been used for patient stratification or validation of patient clustering obtained through stratification cohorts.

We will also include reports on methods to define the optimal size of cohorts, to design these cohorts, to integrate multiple retrospective cohorts, to evaluate risk of bias, and to manage data and analysis in personalized medicine. We are also interested in the quality of data and monitoring
of associated clinical data requirements and in the legal framework of data generated in personalized medicine.

Three case models will be explored: oncology, Alzheimer’s disease and stroke.

These three fields were chosen for their big impact on society and individual health, because they are in three different phases of personalized medicine, which allows us to know different methods and strategies in different levels of development, and because they cover different kind of data to stratify patients. Oncology is the field where personalised medicine was firstly applied and where targeted therapies and diagnostics have been focused. Moreover, several applications of biomarkers for the successful stratification of patients with a given type of cancer exist, most of them based on molecular data, specially genomics. Alzheimer’s disease research in personalized therapies and diagnostics is nowadays giving its firsts results, based in imaging, cognitive and also molecular data. Stroke is currently opening up to personalized medicine, with some approaches and studies in more initial steps. Most of the data for patient’s stratification are imaging and molecular data. The review will cover a broad range of multimodal data profiling studies and biomarkers based on all kinds of data: genetic, metabolomic, genetic expression, genomic, or radiomic.

As general approach, we will search for (systematic) reviews to first identify the most common methodological approaches. Subsequent rounds of more specific searches will be conducted according to the results obtained from the scan of the reviews and to cover detailed aspects.

2. Methods for machine learning applied to stratification

We will include articles and other reports with methodology descriptions or reviews/opinion articles on supervised and unsupervised machine learning approaches and associated validation methods for omics-based stratification that have been tested on real-world biomedical data.

We will prioritize reports describing methodologies that have led to clinically validated biomarker signatures and those describing methodologies that have been cross-validated and externally tested on large sample sizes (preferably across multiple patient cohorts)

Articles reporting on methods that lack appropriate validation statistics and a demonstrated biomedical application will be excluded

There will be no restrictions in terms of types of publication or medical areas.

3. Pre-clinical methods for translational development of stratified therapies and treatments selection

3.1. Pre-clinical models for personalised clinical decision-making.

We will include articles and other reports describing methods (i.e. cellular based assay, organoids, animal models) used to assign treatment options to patient clusters. The case model will be mental disorders disease, chosen as non-oncology medical field. Indeed, this therapeutic area is included in the FDA Table of Pharmacogenomic Biomarkers in Drug Labelling as one of the most represented after oncology (9). Biomarkers in the table include but are not limited to germline or somatic gene variants (polymorphisms, mutations), functional deficiencies with a genetic etiology, gene expression differences, and chromosomal abnormalities; selected protein biomarkers that are used to select treatments for patients are also included.

3.2. Stratified medicine development in oncology

We will include articles and other reports describing translational medical approach, specifically pre-clinical validation methods applied prior to personalised medicine clinical trials. The case
model will be oncology, chosen as the field where personalised medicine was firstly applied and where targeted therapies and diagnostics have, for the most part, been focused.

The review will have a broad focus on the preclinical methodologies used for personalised medicine i.e. animal (mainly PDX), organoid, cellular models and in silico/computerised models, assessing the validity, reliability and predictive value of the various models. As general approach, we will include papers which describe the concept of the methods and exclude those which only deal with models applied to a specific type of cancer and original biomarker research.

Subsequent rounds of more specific searches will be conducted if needed, according to the results obtained from the scan of the first set of articles to cover detailed aspects.

There will be no restrictions in terms of types of publications included.

4. Methods for clinical trials in personalised medicine
We will include methodological and statistical articles and reviews describing or evaluating designs and validation of randomised controlled trials for personalised medicine, assessing both pharmaceutical and non-pharmaceutical interventions. We will also include articles reporting on personalised medicine trials and trial protocols, either published or available on trial registries. Finally, guidance documents issued by regulatory or health technology assessment agencies will be assessed.

There will be no restrictions in terms of types of publication or medical areas.

Study selection
The title and abstracts of records identified by the literature search will be screened by two independent reviewers. The full text publication of relevant articles will be retrieved and checked for confirming eligibility. Discrepancies will be solved by discussion among the review team and the method group if needed. An iterative approach to study selection is expected: each major change from what is reported in this protocol will be recorded and justified.

The screening process will be summarised in flow diagrams as suggested by the PRISMA guidelines for reporting scoping review (10).

Data extraction
The main feature of each report considered eligible, as providing information of a given aspect covered by one or more research questions, will be summarised in tables by one reviewer and checked by a second to ensure data quality. As we expect the reviews to include a variety of scientific articles and other documents, we will not develop a common pre-defined extraction form. However, the following information will be sought and summarised for each included report. This list will be adapted according to the needs of the different review teams.

- Author(s)/reference/title
- Year of publication
- Source origin/country of origin
- Type of publication (e.g. article, editorial, report, poster, etc.)
- Concept/Aims/purpose
- Study population and sample size (if applicable)
- Methodology/Study design
- Intervention type and comparator (if applicable)
- Duration of the intervention/time horizon (if applicable)
- Outcome measures (if applicable)
Main results/findings
Key findings that relate to the review question

This list will be adapted according to the needs of the different review teams.

Study quality
As general approach, we will not perform a formal assessment of methodological quality of the included studies as it is generally not performed in scoping reviews. However, the evaluation of risk of bias of clinical studies included as case examples may be considered.

Plan for presenting the results
The collected evidence will be assembled, summarized and reported to address the research questions defined in the Scope & Research Questions section. The format will be refined toward the end of the process when we will have the increased awareness of the contents of their included studies. Results will be discussed considering the gaps in methodology and the implications for policy, practice and research to inform the consultation exercise.

Consultation exercise
The activities of the review teams (WP3-WP6 in the PERMIT project, permit-eu.org/) will cover this aspect, through dedicated consultations and workshops with field experts. The discussion will involve PERMIT participants and associated partners, and the PERMIT project Scientific Advisory Board.
Funding
This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 874825.

References
8. Tricco AC. et al. A scoping review on the conduct and reporting of scoping reviews. BMC Medical Research Methodology 2016;16:15.
# Box 1: Main definitions of personalised medicine

<table>
<thead>
<tr>
<th>Proponent</th>
<th>Definition</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Horizon 2020 Advisory Group and Strategic Research and Innovation Agenda (SRIA) of PerMed</strong></td>
<td>Personalised medicine is ‘a medical model using characterisation of individuals’ phenotypes and genotypes (e.g. molecular profiling, medical imaging, lifestyle data) for tailoring the right therapeutic strategy for the right person at the right time, and/or to determine the predisposition to disease and/or to deliver timely and targeted prevention.</td>
<td><a href="https://ec.europa.eu/info/research-and-innovation/research-area/health-research-and-innovation/personalised-medicine_en">https://ec.europa.eu/info/research-and-innovation/research-area/health-research-and-innovation/personalised-medicine_en</a> <a href="http://www.permed2020.eu">http://www.permed2020.eu</a></td>
</tr>
<tr>
<td><strong>European Council conclusions on personalised medicine for patients (2015/C 421/03)</strong></td>
<td>Medical model using characterisation of individuals’ phenotypes and genotypes, or tailoring the right therapeutic strategy for the right person at the right time, and to determine the predisposition to disease and/or deliver timely and targeted prevention.</td>
<td><a href="https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52015XG1217(01)&amp;from=EN">https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52015XG1217(01)&amp;from=EN</a></td>
</tr>
</tbody>
</table>
| **UK Medical Research Council** | Stratified medicine is the identification of key sub-groups of patients within a heterogeneous disease population; these being distinguishable groups with differing mechanisms, risk or course of disease, or particular responses to treatments. Stratification can be used to:  
• Improve mechanistic understanding of disease processes and enable the identification of new targets for treatments  
• Develop biomarkers for disease risk, diagnosis, prognosis and response to treatment  
• Allow treatments to be developed, tested and applied in the most appropriate patient groups | [https://mrc.ukri.org/research/initiatives/precision-medicine/stratified-medicine-methodology-framework/](https://mrc.ukri.org/research/initiatives/precision-medicine/stratified-medicine-methodology-framework/) |
| **Personalized Medicine Coalition (PMC)** | Personalized medicine is an evolving field in which physicians use diagnostic tests to determine which medical treatments will work best for each patient. By combining the data from those tests with an individual’s medical history, circumstances and values, health care providers can develop targeted treatment and prevention plans. Personalized medicine is the tailoring of medical treatment to the individual characteristics of each patient. The approach relies on scientific breakthroughs in our understanding of how a person’s unique molecular and genetic profile makes them susceptible to certain diseases. This same research is increasing our ability to predict which medical treatments will be safe and effective for each patient, and which ones will not be. Personalized medicine may be considered an extension of traditional approaches to understanding and treating disease. Equipped with tools that are more precise, physicians can select a therapy or treatment protocol based on a patient’s molecular profile that may not only minimize | [http://www.personalizedmedicinecoalition.org/](http://www.personalizedmedicinecoalition.org/) [http://www.personalizedmedicinecoalition.org/Userfiles/PMC-Corporate/file/pmc_age_ofPMC_factsheet.pdf](http://www.personalizedmedicinecoalition.org/Userfiles/PMC-Corporate/file/pmc_age_ofPMC_factsheet.pdf) |
harmful side effects and ensure a more successful outcome, but can also help contain costs compared with a “trial-and-error” approach to disease treatment.

| **Precision Medicine Initiative (US NIH)** | Precision medicine is "an emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle for each person." This approach will allow doctors and researchers to predict more accurately, which treatment and prevention strategies for a particular disease will work in which groups of people. It is in contrast to a one-size-fits-all approach, in which disease treatment and prevention strategies are developed for the average person, with less consideration for the differences between individuals. | https://ghr.nlm.nih.gov/primer/precisionmedicine-definition |
| **Food and Drug Administration** | Precision medicine, sometimes known as "personalized medicine" is an innovative approach to tailoring disease prevention and treatment that takes into account differences in people's genes, environments, and lifestyles. The goal of precision medicine is to target the right treatments to the right patients at the right time. | https://www.fda.gov/medical-devices/vitro-diagnostics/precision-medicine |
| **Schleidgen et al.** | Personalized medicine seeks to improve stratification and timing of health care by utilizing biological information and biomarkers on the level of molecular disease pathways, genetics, proteomics as well as metabolomics. | BMC Medical Ethics 2013, 14:55 |
| **Sadée and Dai** | Pharmacogenomics is a harbinger of personalised medicine, a paradigm shift from the mindset of 'one-drug-fits-all' to 'the right drug for the right patient at the right dose and time.' This does not mean that each patient will be treated differently from every other patient, an economically untenable proposition. Rather, patients are divided into groups by genetic and other markers that predict disease progression and treatment outcome. | Human Molecular Genetics 2005;14(suppl_2):R207–R214 |
Appendix 1: Examples of search strategies

### 1. Methods for stratification and validation cohorts

<table>
<thead>
<tr>
<th>Query</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>#24 #21 AND #22 AND [(english)/lim OR [french]/lim OR [german]/lim OR [italian]/lim OR [spanish]/lim)</td>
<td>429</td>
</tr>
<tr>
<td>#23 #21 AND #22</td>
<td>438</td>
</tr>
<tr>
<td>#22 [embase]/lim NOT [medline]/lim</td>
<td>9595932</td>
</tr>
<tr>
<td>#21 (#18 OR #19) AND [2005-2020]/py</td>
<td>1302</td>
</tr>
<tr>
<td>#20 #18 OR #19</td>
<td>1355</td>
</tr>
<tr>
<td>#19 #17 AND [review]/lim</td>
<td>779</td>
</tr>
<tr>
<td>#18 #4 AND #9 AND #13 AND #16 AND [(cochrane review]/lim OR [systematic review]/lim OR [meta analysis]/lim)</td>
<td>818</td>
</tr>
<tr>
<td>#17 #4 AND #9 AND #13 AND #16</td>
<td>20180</td>
</tr>
<tr>
<td>#16 #14 OR #15</td>
<td>5241925</td>
</tr>
<tr>
<td>#15 cancer*:ti,ab OR carcinoma*:ti,ab OR tumor*:ti,ab OR tumour*:ti,ab OR oncolo*:ti,ab OR leukemia*:ti,ab OR lymphoma*:ti,ab OR sarcoma*:ti,ab</td>
<td>4465029</td>
</tr>
<tr>
<td>#14 'neoplasm'/exp/mj</td>
<td>3703629</td>
</tr>
<tr>
<td>#13 #10 OR #11 OR #12</td>
<td>29597086</td>
</tr>
<tr>
<td>#12 validation:ti,ab OR method*:ti,ab</td>
<td>9431867</td>
</tr>
<tr>
<td>#11 'procedures'/exp</td>
<td>28915500</td>
</tr>
<tr>
<td>#10 'validation study'/exp</td>
<td>81960</td>
</tr>
<tr>
<td>#9 #5 OR #6 OR #7 OR #8</td>
<td>1592114</td>
</tr>
<tr>
<td>#8 'prospective study'/exp</td>
<td>588853</td>
</tr>
<tr>
<td>#7 'cross-sectional study'/exp</td>
<td>339070</td>
</tr>
<tr>
<td>#6 'cohort analysis'/exp</td>
<td>560295</td>
</tr>
<tr>
<td>#5 'cohort studies':ti,ab OR 'cohort study':ti,ab OR 'cohorts design':ti,ab OR 'prospective cohort':ti,ab OR 'retrospective cohort':ti,ab OR 'data integration':ti,ab OR bias:ti,ab OR 'cross study':ti,ab OR 'cross studies':ti,ab</td>
<td>511111</td>
</tr>
<tr>
<td>#4 #1 OR #2 OR #3</td>
<td>535142</td>
</tr>
<tr>
<td>#3 'personalized medicine'/exp</td>
<td>41477</td>
</tr>
<tr>
<td>#2 'biological marker'/exp</td>
<td>296253</td>
</tr>
<tr>
<td>#1 'stratified medicine':ti,ab OR biomarker*:ti,ab OR 'precision medicine':ti,ab OR 'personalized medicine':ti,ab OR 'personalised medicine':ti,ab OR 'individualized medicine':ti,ab OR 'individualised medicine':ti,ab OR 'individualized therapy':ti,ab OR 'individualised therapy':ti,ab OR 'patient stratification':ti,ab OR pharmacogenetics:ti,ab OR 'patient specific modeling':ti,ab OR 'personalized clinical decision making':ti,ab OR 'personalised clinical decision making':ti,ab OR 'prediction of response':ti,ab OR 'prediction of responses':ti,ab</td>
<td>420990</td>
</tr>
</tbody>
</table>
2. Methods for machine learning applied to stratification

Query

<table>
<thead>
<tr>
<th>Query</th>
<th>#25 #23 AND #24</th>
<th>688</th>
</tr>
</thead>
<tbody>
<tr>
<td>#24</td>
<td>[embase]/lim NOT [medline]/lim</td>
<td>9568801</td>
</tr>
<tr>
<td>#23</td>
<td>#20 AND #21 AND ((english)/lim OR [french]/lim OR [italian]/lim OR [spanish]/lim)</td>
<td>1423</td>
</tr>
<tr>
<td>#22</td>
<td>#20 AND #21</td>
<td>1433</td>
</tr>
<tr>
<td>#21</td>
<td>omic*:ti,ab OR 'machine learning':ti,ab OR 'personalized medicine':ti,ab OR 'personalised medicine':ti,ab</td>
<td>59092</td>
</tr>
<tr>
<td>#20</td>
<td>#4 AND #10 AND #16 AND #19</td>
<td>4830</td>
</tr>
<tr>
<td>#19</td>
<td>#17 OR #18</td>
<td>6287177</td>
</tr>
<tr>
<td>#18</td>
<td>validation:ti,ab OR validity:ti,ab OR validated:ti,ab OR 'cross validation':ti,ab OR 'cross validated':ti,ab OR 'clinical utility':ti,ab OR accuracy:ti,ab OR robustness:ti,ab OR reliability:ti,ab OR sensitivity:ti,ab OR specificity:ti,ab OR reliability:ti,ab OR 'cross study':ti,ab OR 'cross studies':ti,ab</td>
<td>6150811</td>
</tr>
<tr>
<td>#17</td>
<td>'validation study'/exp OR 'reliability'/exp OR 'sensitivity and specificity'/exp OR 'benchmarking'/exp</td>
<td>580344</td>
</tr>
<tr>
<td>#16</td>
<td>#14 OR #15</td>
<td>1174400</td>
</tr>
<tr>
<td>#15</td>
<td>omic*:ti,ab OR 'omic based':ti,ab OR 'multi omic*':ti,ab OR genomic*:ti,ab OR transcriptomic*:ti,ab OR proteomic*:ti,ab OR metabolomic*:ti,ab OR lipidomic*:ti,ab OR epigenomic*:ti,ab OR microarray:ti,ab OR 'rna seq':ti,ab OR 'mass spectrometry':ti,ab</td>
<td>852339</td>
</tr>
<tr>
<td>#14</td>
<td>#11 OR #12 OR #13</td>
<td>758269</td>
</tr>
<tr>
<td>#13</td>
<td>'mass spectrometry'/exp</td>
<td>455591</td>
</tr>
<tr>
<td>#12</td>
<td>'microarray analysis'/exp</td>
<td>68369</td>
</tr>
<tr>
<td>#11</td>
<td>'omics'/exp OR 'genomics'/exp OR 'epigenetics'/exp</td>
<td>299009</td>
</tr>
<tr>
<td>#10</td>
<td>#5 OR #6 OR #7 OR #8 OR #9</td>
<td>5000844</td>
</tr>
<tr>
<td>#9</td>
<td>'individualized medicine':ti,ab OR 'individualised medicine':ti,ab OR 'individualized therapy':ti,ab OR 'individualised therapy':ti,ab</td>
<td>3459</td>
</tr>
<tr>
<td>#8</td>
<td>'personalised medicine':ti,ab</td>
<td>1713</td>
</tr>
<tr>
<td>#7</td>
<td>'personalized medicine':ti,ab</td>
<td>13669</td>
</tr>
<tr>
<td>#6</td>
<td>'stratified medicine':ti,ab OR cluster*:ti,ab OR 'sub group*':ti,ab OR subgroup*:ti,ab OR biomarker*:ti,ab OR diagnos*:ti,ab OR prognosis*:ti,ab OR 'precision medicine':ti,ab</td>
<td>4904827</td>
</tr>
<tr>
<td>#5</td>
<td>'biological marker'/exp OR 'personalized medicine'/exp</td>
<td>330768</td>
</tr>
<tr>
<td>#4</td>
<td>#1 OR #2 OR #3</td>
<td>200079</td>
</tr>
<tr>
<td>#3</td>
<td>'machine learning'/exp</td>
<td>193633</td>
</tr>
<tr>
<td>#2</td>
<td>'statistical learning'/exp</td>
<td>46</td>
</tr>
<tr>
<td>#1</td>
<td>'machine learning':ti,ab OR 'statistical learning':ti,ab OR 'supervised learning':ti,ab OR 'unsupervised learning':ti,ab</td>
<td>34557</td>
</tr>
</tbody>
</table>
3. Pre-clinical methods for translational development of stratified therapies and treatments selection

#15 #13 AND #14 680
#14 [embase]/lim NOT [medline]/lim 9583919
#13 #3 AND #6 AND #9 AND #12 1133
#12 #10 OR #11 2126059
#11 'disorders of higher cerebral function'/exp/mj OR 'mental disease'/exp/mj OR 'psychosis'/exp/mj 1448175
   'psychiatric disease*':ti,ab OR 'mental disorder*':ti,ab OR 'psychiatric disorder*':ti,ab OR 'mental illness*':ti,ab OR depression:ti,ab OR 'bipolar disorder*':ti,ab OR bipolarism:ti,ab OR anxiety:ti,ab OR 'personality disorder*':ti,ab OR 'psychotic disorder*':ti,ab OR schizophrenia:ti,ab OR 'eating disorder*':ti,ab OR 'trauma related disorder*':ti,ab OR 'post traumatic stress disorder*':ti,ab OR 'substance abuse disorder*':ti,ab OR 'asperger syndrome':ti,ab OR autism:ti,ab OR 'delirium tremens':ti,ab OR epilepsy:ti,ab OR 'hallucinogen related disorder*':ti,ab OR hysteria:ti,ab OR 'minor depressive disorder*':ti,ab OR 'major depressive disorder*':ti,ab OR 'obsessive compulsive disorder*':ti,ab OR 'obsessive compulsive personality disorder*':ti,ab OR 'schizoid personality disorder*':ti,ab OR alzheimer:ti,ab OR dementia:ti,ab
#10 #7 OR #8 1372445
#9 #7 OR #8 6083613
#8 'drug therapy'/mj 243267
   'therapy selection':ti,ab OR 'therapeutic selection':ti,ab OR 'treatment':ti,ab OR 'patient allocation':ti,ab OR 'drug therapy':ti,ab OR 'trial success rate':ti,ab OR 'therapy selected':ti,ab OR 'therapeutic selected':ti,ab
#7 #4 OR #5 5925767
#6 #4 OR #5 433005
#5 'biological marker'/exp/mj OR 'personalized medicine'/exp/mj 93445
   'stratified medicine':ti,ab OR 'biomarker*':ti,ab OR 'precision medicine':ti,ab OR 'personalized medicine':ti,ab OR 'personalised medicine':ti,ab OR 'individualized medicine':ti,ab OR 'individualised medicine':ti,ab OR 'individualized therapy':ti,ab OR 'individualised therapy':ti,ab OR 'patient stratification':ti,ab OR 'patient specific modeling':ti,ab OR 'personalised clinical decision making':ti,ab OR 'personalised clinical decision making':ti,ab OR 'prediction of response':ti,ab
#4 #1 OR #2 412990
#3 #1 OR #2 680024
#2 'animal model'/exp/mj OR 'organoid'/exp/mj OR 'in vitro study'/exp/mj OR 'translational research'/exp/mj OR 'disease model'/exp/mj OR 'cell culture'/exp/mj OR 'preclinical study'/exp/mj 236722
   'cellular model*':ti,ab OR 'drug development*':ti,ab OR 'drug response*':ti,ab OR 'drug evaluation*':ti,ab OR 'drug evaluated':ti,ab OR 'patient specific modeling':ti,ab OR organoid*:ti,ab OR 'in silico':ti,ab OR 'drug response assay':ti,ab OR 'drug sensitivity screening':ti,ab OR 'pdx models':ti,ab OR 'patient derived xenografts':ti,ab OR 'preclinical pdx':ti,ab OR 'humanised mouse model':ti,ab OR 'preclinical model*':ti,ab OR 'preclinical stage' OR 'pre clinical testing':ti,ab OR 'translational medical research':ti,ab OR 'disease model*':ti,ab OR 'translational model*':ti,ab OR 'animal model*':ti,ab OR xenograft*:ti,ab
#1 474356
### 4. Methods for clinical trials in personalised medicine

<table>
<thead>
<tr>
<th>No.</th>
<th>Query</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>#14</td>
<td>#11 AND #12 AND ([english]/lim OR [french]/lim OR [german]/lim OR [italian]/lim OR [spanish]/lim)</td>
<td>927</td>
</tr>
<tr>
<td>#13</td>
<td>#11 AND #12</td>
<td>929</td>
</tr>
<tr>
<td>#12</td>
<td>[embase]/lim NOT [medline]/lim</td>
<td>9610086</td>
</tr>
<tr>
<td>#11</td>
<td>#7 OR #10</td>
<td>1221</td>
</tr>
<tr>
<td>#10</td>
<td>#4 AND #5 AND #8 AND [2020-2020]/py</td>
<td>202</td>
</tr>
<tr>
<td>#9</td>
<td>#4 AND #5 AND #8</td>
<td>7669</td>
</tr>
<tr>
<td>#8</td>
<td>'clinical trial'**:ti,ab</td>
<td>514125</td>
</tr>
<tr>
<td>#7</td>
<td>#3 AND #4 AND #5 AND [2005-2020]/py</td>
<td>1026</td>
</tr>
<tr>
<td>#6</td>
<td>#3 AND #4 AND #5</td>
<td>1033</td>
</tr>
<tr>
<td>#5</td>
<td>design**:ti,ab OR methods**:ti OR method**:ti,ab</td>
<td>4793126</td>
</tr>
<tr>
<td>#4</td>
<td>'biological marker'/exp/mj OR 'personalized medicine'/exp/mj OR 'stratified medicine':ti,ab OR biomarker':ti,ab OR 'precision medicine':ti,ab OR 'personalized medicine':ti,ab OR 'individualized medicine':ti,ab OR 'individualised medicine':ti,ab OR 'individualised therapy':ti,ab OR 'individualised therapy':ti,ab</td>
<td>431819</td>
</tr>
<tr>
<td>#3</td>
<td>#1 OR #2</td>
<td>52941</td>
</tr>
<tr>
<td>#2</td>
<td>'clinical trial'/exp/mj</td>
<td>50652</td>
</tr>
<tr>
<td>#1</td>
<td>'basket trial'<strong>:ti,ab OR 'basket stud'</strong>:ti,ab OR 'multi arm'<strong>:ti,ab OR 'master protocol'</strong>:ti,ab OR 'platform stud'<strong>:ti,ab OR 'platform trial'</strong>:ti,ab OR 'umbrella trial'<strong>:ti,ab OR 'adaptive study'</strong>:ti,ab OR 'adaptive trial'<strong>:ti,ab OR 'umbrella stud'</strong>:ti,ab</td>
<td>2402</td>
</tr>
</tbody>
</table>