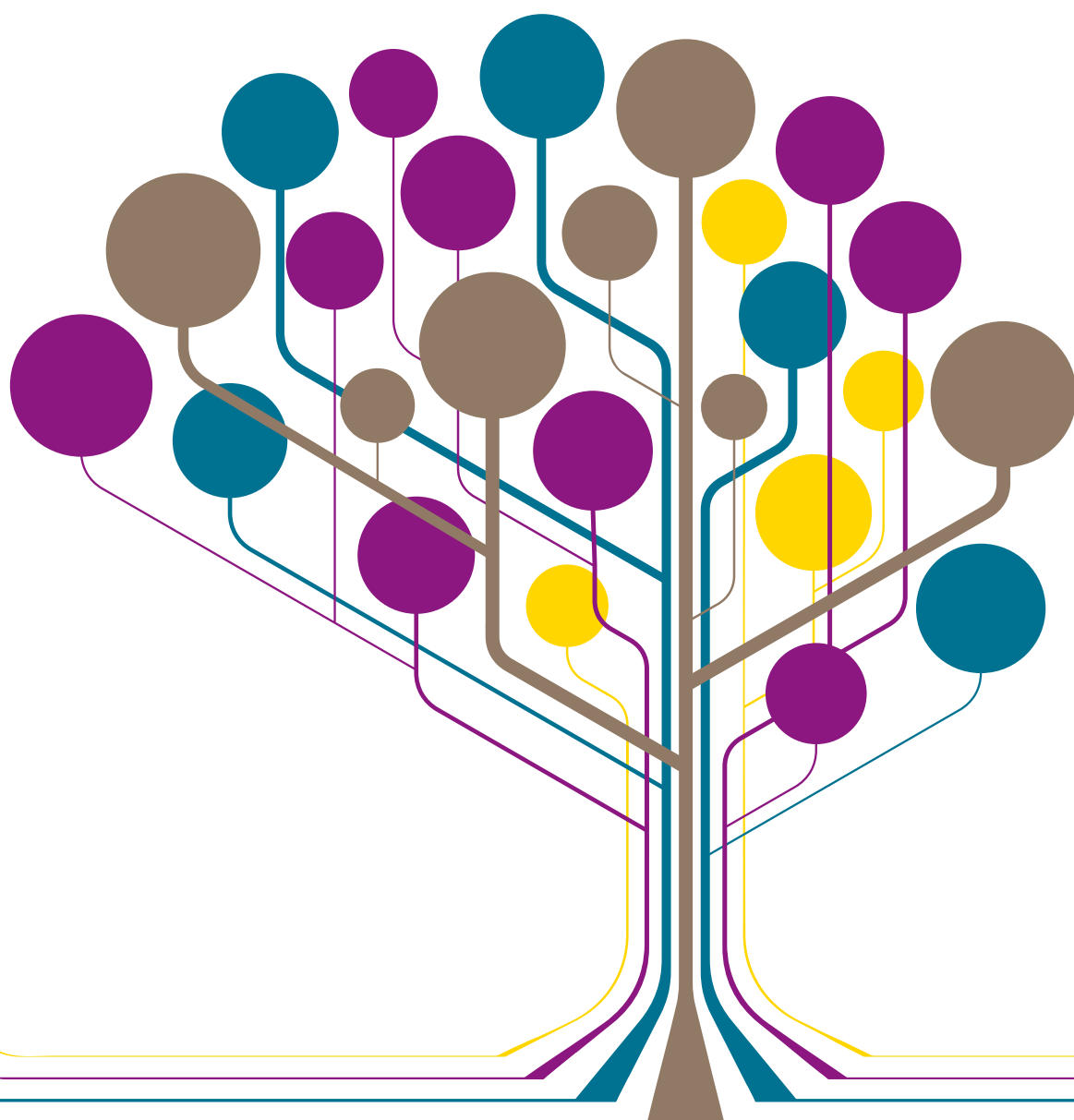


# The MRC Framework for the Development, Design and Analysis of Stratified Medicine Research





ISBN 978-0-903730-28-0

© Medical Research Council, a part of UK Research and Innovation.

The MRC Framework for the Development, Design and Analysis of Stratified Medicine Research by Medical Research Council is licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>).

When providing a citation for this document we recommend the following:

Medical Research Council (2018) The MRC Framework for the Development, Design and Analysis of Stratified Medicine Research. Publisher: Medical Research Council, Swindon, UK. ISBN: 978-0-903730-28-0  
<https://mrc.ukri.org/research/initiatives/stratified-medicine/stratified-medicine-methodology-framework/>

# The MRC Framework for the Development, Design and Analysis of Stratified Medicine Research

## Enabling Stratified, Precision and Personalised Medicine

**Authors:** David Crosby<sup>1</sup>, Patrick Bossuyt<sup>2</sup>, Peter Brocklehurst<sup>3</sup>, Chris Chamberlain<sup>4</sup>, Caroline Dive<sup>5</sup>, Chris Holmes<sup>6</sup>, John Isaacs<sup>7</sup>, Richard Kennedy<sup>8</sup>, Fiona Matthews<sup>7,9</sup>, Mahesh Parmar<sup>3,10</sup>, Jonathan Pearce<sup>1</sup>, David Westhead<sup>11</sup>, John Whittaker<sup>12,13</sup>, Stephen Holgate<sup>14</sup>.

<sup>1</sup>Medical Research Council Head Office; <sup>2</sup>Academisch Medisch Centrum, Amsterdam; <sup>3</sup>University College London; <sup>4</sup>UCB; <sup>5</sup>University of Manchester; <sup>6</sup>University of Oxford; <sup>7</sup>Newcastle University; <sup>8</sup>Queens University Belfast; <sup>9</sup>MRC Biostatistics Unit; <sup>10</sup>MRC Clinical Trials Unit; <sup>11</sup>University of Leeds; <sup>12</sup>Glaxosmithkline; <sup>13</sup>The London School of Hygiene and Tropical Medicine, <sup>14</sup>University of Southampton

This Framework was developed by the authors, informed by the presentations and discussions taking place at the Medical Research Council (MRC) Methodology for Stratified Medicine (M4SM) workshop, 9th-10th July 2015 at Sunninghill, Ascot, UK, which was funded by the MRC. The authors of the Framework are the M4SM Steering Group, chaired by Professor Stephen Holgate. All workshop attendees were consulted on the development of the Framework, and are listed in [Annex 2](#). Other experts provided helpful comments during the drafting process, as listed in [Annex 3](#).

# 0. Summary

Whether described as stratified, precision or personalised medicine, all funders and deliverers of medical research and care are increasingly concerned with ensuring that the right patient receives the right therapy at the right time. The complexity of this evolving discipline requires the application of robust methodologies to disentangle the myriad potential markers and mechanistic pathways that might inform meaningful disease stratification.

This document, applying across the related terms of *stratified*, *precision* and *personalised medicine*, attempts to aid investigators in addressing this complexity by providing a methodological framework for the design, conduct, analysis and interpretation of research attempting to divide complex patient groups into sub-classes (or strata), as defined by differences in mechanism, disease course, risk of developing disease or response to therapy.

Stratification of complex, heterogeneous patient groups is possible through measurement of traits assessed through any number of modalities – genetic, biochemical, imaging, clinical scores, behavioural/psychological assessment etc. Strata are increasingly able to be defined by multiple variables, measured through multiple modalities. It is the intent of this Framework to consider stratification by any means, and is not limited to the traditional genetic or biochemical variables. Despite the strategic importance of the field, there remains a lack of clarity in the stratified medicine research community around best practice in study design, analysis and reporting.

This Framework will focus on particular issues and methodological challenges in:

- The discovery and verification of stratifying biomarkers
- The definition of strata by integration of marker information
- The methodologies for design and development of diagnostics to facilitate this
- Methods for deriving mechanistic insight from these markers/strata
- Provision of the requisite level of evidence to inform development of diagnostics and trials able to test new stratifying hypotheses

While there are numerous methodological challenges in the design of stratified trials and in the analysis of the health economics of stratification, these issues are not explored in depth in this Framework, but are considered as key downstream drivers.

This Framework:

- Presents a pathway for stratified medicine research that starts with the end in mind and covers stratum discovery, verification and early clinical assessment
- Highlights to stratified medicine Investigators the critical questions that they should ask themselves when designing studies, and illustrates potential design pitfalls and sources of bias
- Describes the suite of methodological approaches and resources available to inform study design and analysis, drawing on appropriate guidance and illustrative case studies
- Enables an investigator working at one phase of the stratified medicine pipeline to understand the outputs and requirements of, and iterative interrelations between, the up- and down-stream phases, such that they can tailor their own work to these requirements

The Framework is structured into 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

Ultimately, this Framework aims to illustrate to investigators the various phases of stratified medicine research, the challenges to be faced, the sources of bias and design flaws to be avoided and examples of good practice to be considered. The Framework is directed to enable optimisation of stratified medicine research design in order to increase the likelihood of translation to impact; to ensure that the right patient will receive the right therapy at the right time.

# 1. Background

Stratified Medicine is a key strategic priority for biomedical and health research. Indeed, whether described as stratified, precision or personalised medicine, all funders of medical research and deliverers of healthcare are increasingly concerned with ensuring that, with the increasing recognition of disease complexity and heterogeneity, the right patient receives the right therapy at the right time. The medical/healthcare industry is equally invested in stratification, as it promises to deliver more targeted and more successful therapy which is less likely to fail during development. While some may take e.g. stratified, precision and personalised medicine to have different implications, it should be recognised that for the purposes of this document they are part of the same continuum, and that this document applies broadly across these terms.

The complexity of this evolving discipline requires the application of robust methodologies and will require the development of innovative new methodologies to disentangle the myriad potential markers and mechanistic pathways that might inform meaningful stratification.

Methodological challenges and uncertainty exist at every stage of the stratified medicine research continuum, from initial biomarker discovery and verification, through integration of data to inform characterisation of strata, to studies of clinical and health economic effectiveness. Examples of methodological challenges include methods for stratification of complex diseases by multiple, overlapping and interrelated phenotypic traits and biomarkers (which may develop and change rapidly), and modelling of the relationship between complex, marker/information-rich endotypes and response to intervention.

Excellent guidance publications have been produced from an array of sources concerning the various individual steps along the stratified medicine pathway. However, there has not previously been an attempt to bring this field together in order to take a holistic view of all phases of research intended to support stratification. It is this joined-up approach which this Framework seeks to achieve; to provide a framework about which Investigators can structure the methodology (for the design, conduct, analysis and interpretation) of stratified medicine research projects, encouraging consideration of design pitfalls, sources of bias and examples of current best practice. Furthermore, it is the aim of this Framework to encourage investigators working at any stage of the pathway to consider the wider context, including clinical need, mechanistic plausibility, and the eventual path towards clinical utility; that is, to start with the end in mind and to be mindful of the reciprocal interactions between each phase of the stratified medicine research spectrum.

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

We refer throughout this Framework to 'patients' but, for the purposes of this document, this term should be taken to also encompass people who might not be conventionally classified as patients, e.g. apparently healthy individuals undergoing screening and stratification by risk.

In this context, it is important to note that stratification of complex, heterogeneous patient groups is possible through measurement of traits assessed through any number of modalities – genomic, epigenomic, metabolomics, proteomic, microbiomic, histological, imaging, clinical scores, behavioural/psychological assessment, demographic etc. Strata are increasingly able to be defined by multiple variables, measured through multiple modalities. It is the intent of this Framework to consider stratification by any means, and is not limited to the traditional genetic or biochemical variables.

Despite the strategic importance of the field for the MRC and UK biomedical science, there remains a lack of clarity in the stratified medicine research community around best practice in study design, across the stratified medicine spectrum from marker discovery, through integration of marker information to define strata, to verifying findings and informing the design of stratified trials.

The ability to more accurately and differentially diagnose, to monitor/predict disease course and response to intervention, and to better inform mechanistic understanding underpins the stratified medicine agenda, but as yet, methods development to model and interpret such data has not kept pace with the capacity to phenotype more deeply.

To address these challenges, the MRC convened a Steering Group (the authors of this Framework) and held a workshop to inform development of this Framework. The Workshop involved the [MRC stratified medicine consortia](#)<sup>A</sup> and members of the wider stratified medicine community including industry, plus methodology experts from within and without stratified medicine research, such that ideas and tools could be drawn in from across the fields of biostatistics, informatics, computational/systems biology, epidemiology and beyond. The Workshop also involved representatives from regulatory agencies such as the UK's Medicines and Healthcare products Regulatory Agency (MHRA)<sup>B</sup> and National Institute for health and Care Excellence (NICE)<sup>C</sup> to help ensure that stratified medicine discovery outputs meet the needs of these critical stakeholders. Workshop attendees are listed in [Annex 2](#).

The intended target audience for this Framework are stratified medicine Investigators, to help them to understand the wider context in which their work sits so positioning their studies along the stratified medicine development pathway, and to help them to identify and address the methodological challenges they will need to manage.

## 1.1. Glossary

For the purposes of this document, the following definitions will be used:

- **Stratified Medicine** - the identification of key sub-groups of patients within a heterogeneous population; these being distinguishable groups with differing mechanisms 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; the development of biomarkers for disease risk, diagnosis, progression and response to treatment; and treatments to be tested and applied in the most appropriate patient groups.
- **Stratum** - a subgroup within a heterogeneous disease classification/patient group, which can be defined by a distinct functional or pathobiological mechanism, and/or by differences in risk, clinical course or response to therapy.
- **Endotype** - a subtype of a condition, which must be characterised by a distinct functional or pathophysiological mechanism.
- **Biomarker/Marker** – for the purposes of this Framework, a biomarker is any measurable trait which could potentially be associated with, and used to define, a stratum. Examples of types of measurement by which markers may be identified include (but are not limited to) biochemical assays, genotyping, histology, clinical assessment, imaging, psychological/behavioural assessment.
- **Assay** – for the purposes of this Framework, an assay is a means of measuring a biomarker, as defined above. The definition therefore extends beyond the traditional usage of a lab-based/biochemical measurement, to any pheno- or geno-typic measurement.
- **(Stratum/Marker) Discovery** - identification of a patient sub-group(s), with characteristic biomarker(s), associated with a differential and clinically relevant outcome(s).

<sup>A</sup><http://www.mrc.ac.uk/research/initiatives/stratified-medicine/research/>

<sup>B</sup><https://www.gov.uk/government/organisations/medicines-and-healthcare-products-regulatory-agency>

<sup>C</sup><https://www.nice.org.uk/>

- **(Stratum/Marker) Verification** - replication of the discovered association, between biomarker, biomarker(s)-defined stratum and outcome, using samples/data drawn from a new population and using a fit-for-purpose assay (the latter potentially providing a technological bridge between discovery and clinical application).
- **(Stratum/Marker) Validation** – the process of assessing the biomarker measurement performance characteristics, determining the range of conditions under which the biomarker/stratum will give reproducible and accurate data, and the process of demonstrating a linkage between the biomarker/stratum and clinical end points. It entails the ability of an assay to conform to predefined quality specifications, and the ability of a biomarker/stratum to conform to predefined clinical specifications in detecting patients with a particular clinical condition and affecting patient treatment/outcome.
- **Diagnostic marker** – a marker which identifies a patient having a particular disease/condition.
- **Prognostic marker** – a marker which predicts disease course and outcome.
- **Theragnostic marker** – a marker which predicts response to therapy.
- **Mechagnostic marker** – a marker which identifies a distinct mechanistic stratum but may not, in the immediate term, predict e.g. response to therapy.



## 2. Aims

This Framework aims to help the stratified medicine community to understand what study design/conduct/analysis/interpretation issues and challenges should be considered in planning and undertaking stratified medicine studies, and what methodological tools and examples of guidance and best practice already exist that could help them address these challenges. It is intended that the Framework will focus on particular issues and methodological challenges in:

- The discovery and verification of stratifying biomarkers
- The definition of strata/subgroups/endotypes by integration of marker information
- The methodologies for design and development of diagnostics to facilitate this
- The methods for deriving mechanistic insight from these markers/strata
- The provision of the requisite level of evidence to inform development of diagnostics and trials able to test new stratifying hypotheses

While there are numerous methodological challenges in developing assays to a commercialisable/clinically implementable level, in the design of stratified trials and in the analysis of the health economics of stratification, these issues are not considered in depth in this Framework. Rather, these issues were considered as key downstream elements of the process; the upstream components of stratified medicine research which are covered by the Framework must understand the needs of these downstream elements, and take these into account and be informed by them in their design, and deliver against them.

The Framework will not seek to prescribe which methodologies to employ, given that there is unlikely to be a single correct design/method; rather there may be several appropriate or inappropriate methods (dependent on context), to which the Framework will seek to alert the reader. Furthermore, methodologies will evolve as research progresses, such that prescriptive methodological guidance may be unhelpful, restricting progress and rapidly becoming obsolete. However, the Framework will seek to summarise the state of the art, and to describe fundamental principles of stratified medicine methodology, by listing the questions that investigators should ask themselves when designing their studies, highlighting design pitfalls and sources of bias and, where appropriate, suggesting examples of current best methodological practice.

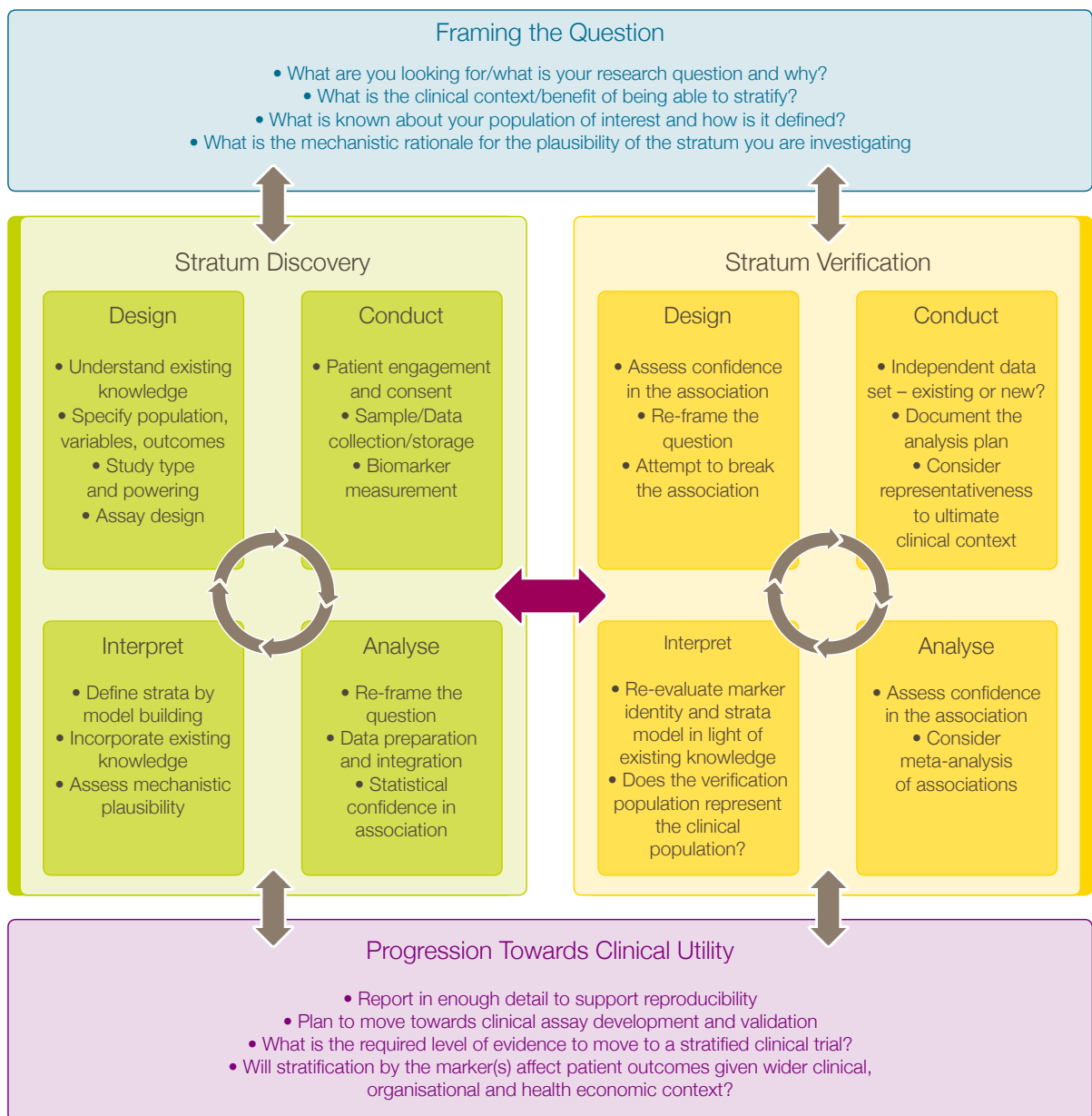
### 2.1. Framework Mission

The mission of this MRC Framework is:

- To present a pathway for stratified medicine that starts with the end in mind and covers stratum discovery, verification and early clinical assessment, to support stratified medicine study design, execution and analysis.
- To highlight to stratified medicine Investigators the critical questions that they should ask themselves when designing studies, and to illustrate potential design pitfalls and sources of bias.
- To describe the suite of methodological approaches and resources available to inform study design and analysis, drawing on appropriate guidance and illustrative case studies.
- To enable an investigator working at one phase of the stratified medicine pipeline to understand the outputs and requirements of the up- and down-stream phases, such that they can tailor their own work to these criteria.

### 3. Pathway Diagram for Stratified Medicine Research

The below diagram is an illustration of the Pathway of stratified medicine research; a simplified representation of the stages of study design, conduct, analysis and interpretation intended to produce stratified treatment of patients. The Pathway of stratified medicine research should not be considered a linear process from marker discovery to clinical impact. Each phase should interact iteratively and reciprocally with the others, and be informed by them in their design and interpretation, e.g. initial framing of the research question should consider clinical need and eventual impact; verification of markers may inform mechanistic understanding and further exploratory research etc. New markers may be identified while existing ones are being validated; new technologies may refine and change the fundamental characteristics of existing biomarkers to improve their performance; new therapies may be developed which change the relationship between the biomarker and response to treatment, etc. Investigators need to plan to be responsive to these changes as the research is designed and conducted.



## 4. Framework Themes

The following Themes, about which this Framework is structured, represent phases of the process of stratified medicine research:

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

Within each Theme, challenges and hypotheses are identified concerning the methodological complexity of the stratified medicine research process and questions have been listed which investigators should ask themselves when designing stratified medicine studies.

The Framework seeks to alert investigators to the potential design pitfalls and sources of bias which might otherwise undermine their work. The Framework also seeks to identify for investigators existing guidance and published examples of best methodological practice, for each of the Themes, such that these can inform study design.

The Framework features a series of case studies of lessons learned and good methodological practice (see [Annex 1](#)), derived from Workshop participants' and other experts' experience, which illustrate the principles described in the Themes below.

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

#### **Mechanistic/Diagnostic/Prognostic/Theragnostic Stratification**

There are different methodological pathways of discovery/verification/validation of biomarkers and strata depending on their intended utility e.g. discovery-type biomarkers which are more mechanistic in nature or intended only to give a signal regarding the involvement of a pathway and are not necessarily therapeutically actionable at the current time ("mechanistic" markers), versus full-blown diagnostic/prognostic markers or those intended to predict response to therapy ("theragnostic" markers) or those which are subject to regulatory approval alongside development/approval of a therapy (companion diagnostics). Each will require different methods and studies, and levels of evidence, and indeed within each of these themes the type of investigation (biological biomarker, clinical biomarker, behavioural biomarker etc.) will also impact on these issues.

There is a case to stratify clinically and a case to stratify mechanistically – neither is inherently right or wrong; investigators should ask themselves whether they have fully considered both paradigms in framing their question, and then direct the design of the study to delivering against the primary aim. When considering the initial research question, it is important to consider the needs and drivers of downstream phases, and how these may reciprocally inform each other. Given the usual time taken from discovery to application, this is likely to be a changing landscape; however having a view of the potential pathway is important. Plans should be in place from the outset for checking and verifying any associations that may be found. During discovery phases, there will likely be false positives; particularly if hypothesis-free screens are employed, there may be a large number of (potentially spurious) associations – how will these be verified (see Theme 5)? Numerous methods exist to test and increase confidence in discovered associations<sup>1</sup>. A special edition of Nature reviewed broad issues in research reproducibility<sup>2</sup>.

#### **Team Composition**

Study team composition is crucial, from the very earliest stages of conceptualisation of the planned study. All relevant expertise should be engaged **from the outset**, to properly understand the context and logistics of the planned project, and to properly inform its design, conduct, analysis and interpretation.

<sup>1</sup><http://www.nature.com/news/reproducibility-1.17552>

It is inappropriate to engage e.g. statistical or informatics support only after the study has already been planned in the assumption that these are “downstream” activities.

Proper team composition will lead to a deeper understanding of important study design issues, and of what data and resource already exist which may impact on the project; for example having an informatician involved at the conception and design stage may help point out to the Investigator appropriate study design aspects or additional data sources of importance, as well as ensuring effective planning for data storage, management and analyses (and dissemination as appropriate). Beyond engaging statisticians and informaticians from the outset, inclusion of project management and delivery plus other technical expertise, as appropriate, in the study design team (e.g. in analytic or diagnostic technologies) should also be considered.

Industrial partnership and patient/public involvement should always be considered at an early stage, such that stratified diagnosis and treatment are ultimately clinically feasible and acceptable. The importance of multidisciplinary team composition in a complex project is illustrated in [Annex 1; Case Study 4](#).

### Early Consideration of Intellectual Property (IP) Issues

Consideration of IP issues should happen as early in the process as possible; discussions involving University technology transfer offices or industrial partners may be protracted; expectations across partners will need to be made explicit. Contractual and IP issues should not impede the science, so early consideration will circumvent future delays – see [Annex 1; Case Study 2](#). With regards specifically to stratified medicine research, it is important to acknowledge that while it is increasingly difficult to patent biomarkers, particularly in the USA where they may be considered natural phenomena, the methodology/technology developed to measure markers or to define strata may be more amenable to IP protection.

### Reviewing Existing Evidence

Reviewing the literature systematically is likely to be a highly beneficial first step, in order to properly frame the question, to ensure that the spotlight used to look for markers is not too narrow (see theme 2), to ensure that the questions asked are not redundant and to identify other relevant studies or external information sources, including the availability of meta-data. While some efforts in stratification have based their assessment of which variables to measure (when looking for markers) on expert opinion, this will not necessarily illuminate the full range of potential candidates. Comprehensive systematic reviews would take into account not only the published literature (peer-reviewed and grey), but also other relevant data sources e.g. cohorts (including clinical trial cohorts), biobanks, ‘omic databases, e-health records, autopsy data.

### Defining the Study Population

When thinking about the study population, there is a need to understand the balance between hypothesis-driven and hypothesis-free investigations: do you start by stating that there are e.g. four strata (perhaps based on clinical observation) and looking for markers which define these, or do you attempt to stratify openly, assuming that the data will tell you the number of strata? The outcome measures selected will define the strata found – an initially hypothesis-generating, high-throughput screen means that the net will be cast more widely, but this will be more expensive and may reduce clarity and lead to too many small (and potentially clinically irrelevant) strata. To start with a defined set of “types” may end up missing important currently unknown factors and causing false negative results.

**Hypothesis-free research:** a discovery approach attempting to characterise a system, usually focusing on finding significant correlations in very large data sets such as sequencing or expression data, without a defined supposition or proposed explanation to be tested.

**Hypothesis-driven research:** The investigator begins with an a priori supposition or proposed explanation made on the basis of the currently available evidence and theoretical deduction, and then designs a study that will test (and thereby refute or support) the hypothesis.

Within one programme of stratified medicine research, the hypothesis-free and hypothesis-driven approaches are not mutually exclusive, and both approaches may be suitable depending on the circumstances.

When defining the study population, it is important to consider and explicitly describe its relation and relevance to the population of interest, in order to prospectively address issues of representativeness, and of utility of the ultimate stratification method.

It is also important to consider whether your chosen population would be engaged/recruitable (if recruitment of a new cohort is required). For example if you wish to stratify pregnant women at risk of pre-eclampsia by means of an intervention, what is the evidence to suggest that this population will be willing to participate in such a study? Will the intervention be acceptable to patients in practice?

It is important to consider phenotype complexity; many studies and disease areas employ relatively crude phenotype/stratum descriptions in defining the population of study and/or the clinical population. It is important to reflect uncertainty in defining phenotypes/strata and associated marker groups in order to explore whether real strata exist. Investigators should be able to fully quantitatively describe the phenotype/s of interest. In some areas, there may be a problem in application of pattern recognition at a clinical level to define complex conditions such as asthma, rheumatoid arthritis, depression etc. It is important that investigators start to question clinical dogma, and question whether deeper insights might be gained by considering the search for strata in a clinical symptom-agnostic manner. Investigators might look mechanistically for biomarkers, across what we currently define as different/separate diseases. **See [Annex 1; Case Study 7](#)**, for example. The concept of a mechanistically-driven rationale for selecting the study population, and selection of outcomes/potential markers to measure beyond those classically used in current trials based on a mechanistic rationale are illustrated in **[Annex1; Case Study 10](#)**.

Industry are starting to take steps in this direction, looking at drugs which target underlying mechanisms of pathology (for example auto-immunity), which may operate across diseases as currently defined (such as rheumatoid arthritis, lupus, Sjögren's syndrome, asthma etc.). Evidence has started to emerge which supports such a paradigm, e.g. targeting the underlying eosinophilia across inflammatory respiratory diseases with a T lymphocyte helper type 2 molecular signature<sup>2</sup>. The primary driver should not be dividing an existing group into smaller and smaller sub-classes; rather the investigator should consider whether the overarching disease group, as currently defined, is real/relevant.

It is also important to consider whether the phenotype is stable. What do we already know about variation and the natural history of the condition? How big is the variation within a phenotype?

### Defining Outcome(s)

It is crucial, from the outset, to consider how well-defined the outcome/s of interest is/are, depending on the mode of stratification under investigation. If you are looking at response to therapy, do you know how to define response? If you are considering prognosis, do you know how to define disease progression/outcome? If you are stratifying by risk, do you know how to define disease? If you are looking at a behavioural factor do you know the individual variability, across time etc.? Outcome is also dependent on the biomarker under investigation and the relationship expected.

Studies should clearly lay out and document the primary outcome variables of interest against which the stratification is seeking to identify sub-groups, as well as other secondary outcomes to be explored should the primary outcome show no evidence of stratification (or fails to be verified). This is important for the open reporting of findings.

Response to therapy has to take into account both efficacy and safety. Efficacy responses may be widely variable; in their simplest form, it may be possible to dichotomise this (no response/response, for example when all-cause mortality is the end-point). However, in most cases, an efficacious drug response is not binary; it may be continuous and the location of the cut-off point between response/non-response needs to be carefully defined and justified. Indeed, a 2005 review of priorities and standards in pharmacogenetic research<sup>3</sup> noted that "In pharmacogenetics, the first step is usually the hardest: careful thought must go into choosing the most appropriate way to define response, and this should precede genetic analysis.

It seems too rarely appreciated that the appropriate definition of response (in terms of safety and efficacy) is often not obvious....It can be surprisingly difficult to represent even “simple” phenotypes like dose-response”.

In selecting and defining outcomes, investigators should not always (in the discovery phase) default to only using classical (and regulatory-required) measures of outcome, which may not be sensitive to important stratifying effects. For example, in asthma the standard measures frequently include indices of lung function (e.g. FEV1/FVC, PEF). These are reliable outcome measures for bronchodilator therapies but may not be responsive in newer-generation therapeutic modalities such as anti-IL5. However when a more mechanistic endpoint is used (e.g. eosinophil levels), signals then become apparent which may correlate with endpoints of importance to patients. As such, anti-eosinophilic therapies in asthma have lesser effects on lung function but great effect on frequency of exacerbations<sup>4</sup>. However, with further refinement of the Th2-type eosinophilic phenotype, efficacy is now revealed on other asthma end points including lung function<sup>5,6</sup>. See [Annex 1; Case Study 7](#).

For safety end-points, the characterisation of whether a patient is suffering from an adverse reaction can be complicated by phenotypic heterogeneity between different patients. This is important to consider as different manifestations may have different predisposing factors. Thus, pre-specified standardised phenotypes should be considered from the outset of the study<sup>7</sup>. Where the safety end-point is a continuous variable, for example liver function tests, again there is a need to carefully consider the boundary between normal and abnormal values.

### **Mechanistic Plausibility**

For the purposes of this paper, mechanistic plausibility is the consideration of the evidence (empirical or theoretical) for the likelihood of the proposed marker/stratum being involved/existing in the disease/group to be stratified. In various fields this construct has parallels in terms such as biological plausibility, mechanistic plausibility or theoretical plausibility/rationale – for the purposes of this paper, mechanistic plausibility will be used to convey all of these constructs.

It is appropriate to emphasise that mechanistic plausibility of markers associated with strata may increase confidence in the validity of the association. Conversely, mechanistic implausibility may suggest that putative markers should be more closely scrutinised. However there is a danger in being beholden to this; if little is known about disease mechanism then anything one discovers would be unexpected and therefore, potentially at the outset, implausible. A balance should be struck in maintaining a broad enough view that can probe new areas and discover unexpected associations, while maintaining a sense-check on putative associations by considering their mechanistic plausibility, insofar as the state of the art knowledge of mechanism allows.

An example from the behavioural science field of stratified interventions being developed and evaluated based on mechanistic plausibility/hypothesis is described in [Annex 1; Case Study 8](#). Another example, from neurology, is in [Annex 1; Case Study 10](#).

### **Consideration of Variable Response**

A Clear definition of what constitutes response to intervention is essential and is discussed in Defining Outcome(s) above. Investigators should consider whether that response to therapy may be influenced by factors of variability between individuals but are not part of the disease process itself and so may be inadvertently overlooked when framing the question.

For example, with biologic therapies it is increasingly recognised that the therapy can be immunogenic, which may affect the response – a high proportion of patients may become unresponsive to therapy because of anti-drug antibodies (ADAs), e.g. to Adalimumab<sup>8</sup>, rather than mechanistic insensitivity to the therapy. There may be genetic reasons why patients mount ADA, but unless this possibility is taken into account, the ‘outcome’ (lack of response to drug) will be heterogeneous and include different types of non-responsiveness.



Variability in response may also be due to inter-individual differences in the pharmacokinetics of a drug therapy, which may lead to marked differences in drug exposure despite the same administered dose. For example, with warfarin, patients with the genotype CYP2C9\*3/\*3 have a 90% reduction in clearance when compared with wild-type (CYP2C9\*1/\*1) patients, which can lead to prolonged international normalised ratio (INR) and increase the risk of bleeding<sup>9</sup>. Variability can also occur because of pharmacodynamic factors; here, dose and exposure may be similar between patients, but a drug target is variable, predisposing to either lack of efficacy or toxicity. For example, hypersensitivity to the anti-HIV drug abacavir is most likely to occur in patients who are positive for the HLA allele HLA-B\*57:01; this is evidenced by the reduction in the incidence of hypersensitivity from 7% to <1% through the introduction of pre-prescription genotyping<sup>10</sup>.

It is also important to explore and understand whether patients stay in the same group over time - do they give reproducible measures? Do they move between strata? The possibility that individual patients may not respond consistently to therapy (usually in chronic conditions where therapy may be repeated) is one that investigators should consider.

When considering a stratified medicine research project, investigators should consider the cycle of four phases; **Design, Conduct, Analyse, and Interpret**.

## Questions Investigators Should Consider when Framing the Question/Defining the Population:

### Design

- Starting with the end in mind - What are you looking for/what is your research question? What kind of stratification are you looking to establish - patients can be stratified by e.g. risk, prognosis, early detection, diagnosis, response to therapy, surrogate markers – there are many different types of marker, and each may potentially be differently assessed.
  - Are you looking for a stratum that:
    - Identifies a particular pathophysiological/clinical stratum within a heterogeneous patient population (**diagnostic**)?
    - Predicts disease course (**prognostic**)?
    - Predicts response to therapy (“**theragnostic**”)?
    - Tells you about mechanism (“**mechagnostic**”)?
    - Predicts risk of developing disease (“**predictive**”)?
  - What is the clinical context/benefit of being able to stratify?
  - Have you considered the ultimate requirements of the clinical end users and the regulators of the potential stratification tool that your work may give rise to?
- What are the known characteristics and distribution/variability of your population of interest?
  - Known mechanisms?
  - Known differential response to therapy?
  - Known variations in clinical course?
  - Range of severity?
  - Variation within and between individuals?
  - Known correlated co-morbidities and their treatments?
- Why have you selected the population you intend to study?
  - Do you need to recruit a new population or will existing populations be acceptable e.g. existing cohorts or trial populations?
  - What characteristics and metrics will you use to define the population?
  - Have you sufficiently captured the complexity of your chosen population? Is your spotlight wide enough? Are there other variables you should consider, beyond traditional genomic or biochemical readouts, e.g. environmental factors, behavioural analysis, mental health status, the microbiome etc.?

- How representative of the population of interest is your study population? If it differs significantly, why, and what impact is that likely to have? Is the population you will study appropriate for ultimate intended clinical use?
- What level of heterogeneity is it appropriate to include, i.e. not so heterogeneous that a meaningful signal is undetectable, but not so homogenous and controlled that clinically useful strata cannot be distinguished (and may have been excluded)?
- Have you considered defining your population by underlying mechanism, rather than by traditional clinical disease categories?
- For newly recruited populations, will they be engageable or recruitable into the proposed study? Is there evidence to reassure as to feasibility of recruitment?
  - Have you appropriately engaged with patient groups to explore this?
  - Are there sufficient numbers available who are not already engaged in research endeavours?
  - Are there appropriate and feasible methods for data collection in this population?
- Have you thought about the pathway to verification e.g. independent sample availability?
- Have you thought through study feasibility (both in terms of appropriate numbers and time-frames) with regards to available population/endotype size/frequency, and market feasibility of the ultimate stratification tool/therapy that may be produced?
- What is the argument for the mechanistic plausibility of measuring the range of potential markers you have selected?
- What are the requisite capabilities and types of expertise necessary to deliver your planned programme and to address your identified research question?
  - Have you thought through the composition of your study team and included all relevant individuals?
  - Have you **fully** embedded statistical, informatics, project management and other technical expertise **from the outset** to inform design of your study and to plan for necessary infrastructure?
  - Have you explored potential contractual and intellectual property issues with all relevant partners at an early stage?

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

As a general pathway for biomarker discovery and verification, Cancer Research UK have published a helpful series of Roadmaps for developing biomarkers in diagnostic<sup>E</sup>, prognostic<sup>F</sup>, screening<sup>G</sup>, pharmacological<sup>H</sup> and imaging<sup>I</sup> studies.

### Marker Discovery – Selecting Variables to Measure

When selecting variables to measure in order to attempt to discover and define strata, it is important to widen one's spotlight and consider non-standard measures; investigators should think more broadly than the standard, orthodox markers (while balancing this against the risk that selective inference and the testing of multiple markers can reduce statistical power and potentially inflate the evidence of association unless properly adjusted for). For example, in stroke patients with rheumatoid arthritis (RA), the arthritis often does not affect the side that is affected by the stroke<sup>11,12</sup> – it would not necessarily be intuitive to measure neurological factors in an RA study. Likewise other factors beyond the classical autoimmune mediators have been implicated in RA which, a priori, might not have been obvious candidates by which to stratify, such as diet or vaccine history<sup>13</sup>. Conversely, markers which would appear to be logical predictors of response to therapy may not, in practice, be so e.g. the presence of autoantibodies would intuitively seem as though it should predict response to B-cell depletion therapy in RA; it does, to a certain extent, but not completely<sup>14</sup>.

In conditions with a variable or relapsing/remitting course, looking for and understanding markers of remission is just as important as markers of disease progression; such markers may provide windows into mechanism, as well as being useful prognostically and theragnostically.



The less mechanistically well-understood the disease, the wider and more inclusive the choice of variables to measure should be. Data can always be refined at a later stage. If possible, the choice of variables should maximise integration with data from other sources and facilitate later meta-analysis.

**Annex 1; Case Study 10** describes a complex phenotyping project in Alzheimer's disease which selected outcomes/potential markers to measure beyond those classically used in current trials, based on a mechanistic rationale taking into account the clinical limitations and objectives.

### **Defining Effect Size and Calculating Sample Size**

When calculating the sample size for a study, the expected effect size is a key factor. This may be effect size in terms of a response to therapy, or in terms of the expected level of variability of a marker. It is crucial for the means of effect size derivation to be clear, reportable and evidence-based. Investigators should think through, and be able to justify, why a particular effect size has been proposed.

Focussed pilot studies can help to try to understand variability and potential effect size. This may help not only in informing power/sample size calculations for the main study, but also to help inform assay development and to understand non-mechanistic variation in the assay measurements (see Theme 3 below). Investigators should consider how they will integrate pilot data with existing knowledge (possibly derived from a systematic review) to help to select variables to measure, to gauge effect size and again to inform assay specification – a pilot analytical component within a discovery phase. It is important for investigators to acknowledge uncertainty around effect and sample size. Investigators must exercise caution in that pilot studies, by their nature, are likely to be underpowered. Investigators must give careful thought to pilot design, such that they can assess what they are designed to assess. In marker discovery work, pilot studies can be useful to get an initial conception of the number of candidate markers to be investigated in a full study.

Currently sample size calculations are generally only conducted based on a single variate. Increasingly it is desirable to use multiple variates and it is statistically challenging to account for multiple variates and prevent overfitting<sup>15</sup>. Several papers have reviewed these challenges and in some cases proposed workaround solutions<sup>16</sup>. Similar challenges also apply to developing a multivariate model for the relationship between variates and outcomes which can be used to predict outcome e.g. survival<sup>17</sup> – see Theme 4.

### **Use of Retrospective Data and Secondary Analysis**

Use of retrospective data and secondary analysis are legitimate strategies in biomarker and stratum discovery studies, however care must be taken in their employment since the original study may have been designed for a quite different purpose. For example, if using trial data which was powered to detect a treatment effect, it may not be powered to detect a biomarker/outcome interaction, but it might give an initial signal to pursue. When looking at studies retrospectively in a meta-analysis, Isaacs et al showed how a biomarker may behave differently in different populations<sup>18</sup>.

The utility/reliability of this signal depends on whether investigators are working in discovery mode (when one can accept false positives that one will follow up on) or clinical validation mode.

A 2012 review of the reliability of subgroup analyses<sup>19</sup> helpfully provides criteria to assess the credibility of claimed subgroup effects, which may inform future study design.

<sup>15</sup><http://www.cancerresearchuk.org/sites/default/files/diagnostic.pdf>

<sup>16</sup>[http://www.cancerresearchuk.org/sites/default/files/prognostic\\_and\\_predictive.pdf](http://www.cancerresearchuk.org/sites/default/files/prognostic_and_predictive.pdf)

<sup>17</sup><http://www.cancerresearchuk.org/sites/default/files/screening.pdf>

<sup>18</sup><http://www.cancerresearchuk.org/sites/default/files/pharmacological.pdf>

<sup>19</sup><http://www.cancerresearchuk.org/sites/default/files/imaging.pdf>

## Anticipating Adherence Issues

When designing stratification studies involving an intervention, the likelihood of patient adherence to the therapeutic regimen is an issue which investigators should carefully consider in the design stage, since there is substantial variation in adherence between patients: for example, Osterberg and Blaschke

suggested adherence rates of 43 to 78% in trials of chronic conditions<sup>20</sup>. This may have substantial effects on study outcomes<sup>21</sup>. There is a large literature on both adherence rates in trials and ‘real world’ settings, and strategies to improve adherence<sup>22,23,24</sup>. Ensuring, as far as possible, adherence (and planning analysis and interpretation of results accounting for the lack of it) may be particularly challenging in therapies with a significant adverse effect profile, or in behavioural or other therapies requiring patient engagement or self-administration.

Investigators should consider how to differentiate non-response from non-adherence to therapy. This is particularly important where adherence may differ systematically between treatment groups, e.g. because adverse event profiles differ. The importance of adherence is discussed in [Annex 1; Case Study 5](#).

## Planning for Verification

It is important for investigators to consider, at the time of initial discovery study design, how any putative associations will subsequently be verified. Verification of markers and strata is explored further in Theme 5, below. An example of planning for verification hand in hand with discovery study design is the (in progress) Europe-wide Biomarker study, which will look for markers predictive of cardiovascular risk (see [Annex 1; Case Study 1](#)).

Investigators need to be mindful of the volume and complexity of data to be generated and how this will affect verification - if you are going to potentially identify and then check hundreds of putative markers, the likelihood is that false positives will be generated. Investigators need to be explicit on false discovery rate and how this will be assessed going forward.

Patient/sample/data availability is a crucial issue – at the stage of initial discovery study design, Investigators should be mindful of the need for a subsequent independent sample/data set against which to verify their findings. Plans should be made from the outset for a feasible strategy for the verification stage.

Investigators should also take account of the technical challenges of verification when planning discovery studies, without letting the technical limitations of verification techniques constrain their discovery plan. For example, in protein biomarker discovery experiments, one may end up with a large number of high-confidence potential biomarkers discovered through proteomic approaches. The current gold standard for follow up is ELISA (which can be expensive and highly dependent on the quality of available antibodies), so the proteins which are followed up tend to be those previously identified as of interest in the literature, and therefore already have “good” reagents available. A consequence of this approach is the possibility of re-treading of old ground, such that many potential markers remain unexplored. Investigators should be aware of the temptation to default to well-worn markers due to ease of verification studies.

## Planning for Data Curation

Investigators should be mindful of not discarding data; plans should be made to collect, store and report all possible data (while also being mindful of relevant data retention policies). While the investigator may not find immediate use for them, it is quite possible that others will. Investigators should avoid collecting and reporting only crude aggregate data or mean values; there will be richness beneath these which may be informative not necessarily in addressing your primary goal or hypotheses, but for other questions. Think about data management and curation for optimising access and usability, and the necessary resources for ensuring this.

An illustration of the use of (and complexities around) data-sharing platforms is provided in [Annex 1; Case Study 4](#).

For maximum transparency, visibility and accessibility of data, investigators are encouraged to consider the use of well-supported data management platforms based on open source software (e.g. [tranSMART<sup>J</sup>](http://transmartfoundation.org/about-the-pta/), [i2b2<sup>K</sup>](https://www.i2b2.org/)) and well-studied architectures (e.g. [eLabs<sup>L</sup>](http://www.researchobject.org)). Such platforms may help standardisation of data curation, and hence interoperability with other data sets. A case study of a complex multi-modal data stratification project using such data infrastructure is presented in [Annex 1; Case Study 11.](#)

When making clinical data publicly available, investigators must carefully think through issues of privacy and anonymity. How will the data be de-identified? Investigators are directed to the MRC Policy and Guidance on Sharing of Research Data from Population and Patient Studies<sup>M</sup>.

### Considering Alternate Study Designs

While the majority of marker/stratum discovery studies may be built on an observational/epidemiological characterisation approach, or conducted as part of a randomised interventional trial/study, investigators should be mindful of, and open to, alternate designs, e.g. factorial designs, adaptive designs, n=1 studies etc.

### Questions Investigators Should Consider when Designing Stratum Discovery Studies:

#### Design

- Existing knowledge – have you made systematic efforts to understand the relevant existing knowledge and use it to inform your study design (including selection of candidate markers to measure)?
  - What is the optimal way for you to incorporate mechanistic knowledge in biomarker discovery?
- Is there value in a pilot study during the discovery phase to gauge variance in marker/s between potential strata, and to inform marker selection and powering of a larger discovery programme?
- Have you considered the array of potential study types and design features which may be employed to address your research question, e.g.
  - Case control, cohort, cross-sectional, trial etc.
  - Prospective/retrospective
  - Blinding of operator/analyst etc.
- Selection of variables – what will you measure and why? What is the mechanistic rationale for measuring the prospective marker/s?
  - Are you sufficiently covering the complexity of your patient population?
  - Are you selecting variables to measure because they are the variables that you routinely study, rather than being those which are most informative or appropriate? Have you thought about broadening your spotlight?
  - Have you considered whether there are “treatable traits” which might be identified and exploited, i.e. a marker of pathology by which patients can be diagnosed and stratified, but that also responds to therapy?
  - Is there an assay available that is sufficiently accurate to measure the chosen variable meaningfully?
- Defining response – if your study involves stratification by a therapy, how will you define response/non-response?
  - Outcome measure(s) – what type of outcomes do you plan on measuring? Why have these been selected? Are they clinically relevant? What readouts will you use to assess response?
  - Will thresholds/cutoffs be used to define response? If so, why, and how will these be set? See Theme 3.
  - Over what timescale will you measure response – acute vs. chronic effects?

<sup>J</sup><http://transmartfoundation.org/about-the-pta/>

<sup>K</sup><https://www.i2b2.org/>

<sup>L</sup>[www.researchobject.org](http://www.researchobject.org)

<sup>M</sup><http://www.mrc.ac.uk/research/research-policy-ethics/data-sharing/policy/>

- Have you appropriately considered the mechanism of action of any intervention you are using, e.g. for a drug, does knowledge of the pharmacology suggest stratifiers? What is known about the drug's pharmacokinetics/dynamics? Would further study of PK/PD be informative, e.g. does the intervention reach the intended site of action?
  - For biologic therapies have you considered potential immunogenicity and anti-drug antibody formation?
- How will you account for variability/noise in the data you collect (see Assay Design – sources of variability, below)?
- Have you considered baseline variability in disease activity? Some disease will be mild, some severe; this may affect levels and patterns of potential markers.
  - In measuring response to therapy, is it more appropriate to do so in absolute terms or in improvement from baseline?
  - Can you ensure that your recruitment strategy will capture an appropriate cross-section of disease activity?
- Powering – Have you designed a study with adequate power to detect the intended effect/association, properly adjusting for potential strata discovery? How have you derived the expected effect size? How have you calculated the sample size to be used, and what is the expected power?
  - Has your powering taken account of the intended use of the marker (e.g. discovery/mechanistic insight versus diagnosis/clinical stratification)?
  - Have you considered number of variables, sample heterogeneity, assay characteristics, required confidence etc.?
  - How much variability is there in your analyte set between potential strata?
- Sample/Data type – what kind of data/samples should you look to generate, and what information could be extracted from these? How will these be measured? Is there patient acceptability of the sampling method? Have you engaged with technical experts to ensure that you collect the most appropriate samples and process them according to the requirements of the intended use e.g. for proteomic studies?
- Sample/Data access and storage – how will you collect/process/store your samples/data? How can you best define and report your procedures, to maximise likelihood of reproducibility? Have you considered storage and reporting of all data collected (not just aggregates/means, or those data that you found useful)?
  - Are there standards or common platforms for the kinds of data you will generate, which could potentially allow the interoperability and pooling of your data with those from other sources?
  - How will you ensure appropriate metadata are captured to ensure that your data adhere to relevant standards?
- Feasibility – based on the above considerations, is your required sample size feasible in terms of recruitment, cost etc.?
- If the study will be multi-site, have efforts been made to standardise procedures across sites, and what form will these efforts take? Will there be appropriate distribution of e.g. cases and controls across sites?
- Have you considered (and taken steps to mitigate) the bias that may be introduced by:
  - Mixed cohorts, populations, trials
  - Data not designed for the application
  - Missing information
  - Missing confounders, confounding by indication
  - High false discovery rate – particularly with high volume omics data
  - Potential for exaggerated effect sizes
  - Operator bias in data mining studies – searching for subsets

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

The variability introduced by the multitude of types and sources of variation in assay signals make biomarker discovery, development, verification and delivery a daunting task. It is the understanding and, where applicable, control and minimisation of these sources of variability that will allow stratified medicine to achieve its potential.

Great care should be taken to avoid the confounding of key measures of interest with other sources of variability, such as batch, day or operator effects<sup>26</sup>. Appropriate randomisation at the assay stage is key; this is obvious, but often neglected, particularly in early application of new technologies where sources of variation may not be well understood. Examples of such confounding issues emerge in early microarray work<sup>27</sup>, and in the ENCODE project<sup>28</sup>.

If possible, the choice of assay platform should maximise integration with data from other sources and facilitate later meta-analysis. However, choice of platform is often constrained by practical limitations: e.g. how the available tissue has been processed/stored, the need for standardisation across multiple sites, cost, etc. For this reason there is often a change in platform as the assay moves from discovery to clinical use: for example, a gene expression-based assay may move from a genome wide platform such as RNA Seq to a simpler assay of a single or small number of transcripts. Careful work needs to be done to understand the comparability of assays as this transition takes place.

It is also important to report on the level of quality control under which the assay was developed. How strict were the Standard Operating Procedures (SOPs) to ensure quality control and reproducibility? Not every assay needs to be developed in a fully Good Clinical Practice – Quality Controlled hospital-standard laboratory, but it is important to report on the environment in which it was developed, such that readers can assign the relevant amount of confidence to it.

#### Pre-analytic Variation

There are many potential sources of variability which may arise from inconsistent or non-standardised procedures around pre-analytic sample collection and processing. Investigators should be fully aware of and document all such factors e.g. sample/data collection method, sample processing such as tissue fixation, potential for degradation in sample over time since collection etc. The major pre-analytic issues are summarised, and mitigation strategies suggested, in a review<sup>29</sup>.

#### Operator and Centre Variability in Assay Design and Development

In developing the assay, Investigators should be cognisant of the potential for operator variability – some people may be more adept at operating the assay than others (this may be particularly prominent in psychological/behavioural/other complex intervention/assay scenarios). If the same star technician is always engaged to develop the favoured assay technology/technique, bias is being introduced. Investigators should not only understand that operator variability can happen, but should measure it, so its effects are not confounded with biology. Assay development and design can account for this by randomising operators.

There may also be confounding effects between centres in multi-centre studies<sup>30</sup>. Different pre-assay SOPs across centres (e.g. in sample collection/processing/storage) may affect the signal. Standardisation of training and SOPs for assay operators is crucial.

#### Dissecting Mechanistic Variation from Measurement Variation; Assay Accuracy and Variability

For the purposes of this Framework, mechanistic variation should be taken to mean inherent variability arising from the real differences in levels of the variable/analyte/measure in question within individuals (rather than noise derived from inaccuracy/instability of the assay itself – **measurement variation**). This will be termed mechanistic variation here, regardless of the origin of signal or mode of assay e.g. genotyping, a six-minute walk test, behavioural questionnaire etc.

In order to understand the ability of the assay to measure mechanistic variation, it is important to also understand the sources of technical variation affecting assay results. This process is crucial in assay development as ultimately, as described by the “STrengthening the Reporting of OBservational studies in Epidemiology: Molecular Epidemiology” paper<sup>36</sup> (STROBE-ME, which provides a checklist to support reporting of the use of biomarkers in epidemiological studies), “Without information on measurement error,

intra-individual variation and inter-individual variation biomarker studies are uninterpretable.” Often this can be done by use of standard conditions; for example, ligand binding assays are typically tested on sets of reference samples of known content, as described in the US Food and Drug Administration’s Guidance for Industry on Bioanalytical Method Validation<sup>N</sup>. More generally, in the development of assays, statistical design of experiments is often of great value in both understanding the sources of technical variation and optimising assay conditions, for example in cytometry<sup>31</sup>.

Ideally, in developing and selecting an assay, it would be possible to compare the assay against a reference standard to check its accuracy, although it is recognised that in some instances there will be no standard. Investigators should attempt to identify whether there is a standard the assay may be calibrated against. Consideration should be given to repeated validation against a stable reference wherever possible to check for variation over time. If no standard is available at onset, this should be reviewed periodically as the position may change.

Potential sources of assay noise include instrument drift during bedding-in of a new assay, instrument degradation through aging, reagents varying across batches or labs, plate effects if using a high-throughput assay, time drift due to environmental factors including diurnal and seasonal effects, the impact of different operators and technicians etc.

An example of iteration of an assay to best identify subgroups can be found in behavioural or psychological assay development. Here, a set of questions could be asked that will be expected to cover both the breadth and depth of the condition to enable e.g. the ascertainment of high risk or low risk groups. Such investigation requires variability in response to enable the correct identification. Item response theory (or other similar methods) can then be used to reduce the question set down. Further development work can involve adaptive testing methods with computerised item banks<sup>32,33,34</sup> to augment the identified aspects. Subgroup refinement can be undertaken by using an adaptive assay that changes based on individuals responses, whilst still reflecting the variability across the whole population. A worked example of computerised adaptive testing can be found in a study of activities of daily living in cardiovascular patients<sup>35</sup>.

Investigators should think about how to standardise, and should report on, sample preparation and storage, as well as analysis procedures; standard operating procedures should be prepared and used throughout, wherever possible. Recommendations for study design considering varying impacts of marker assay variability due to e.g. batch, storage, freeze-thaw and change in technology are helpfully made in a paper on Design Options for Molecular Epidemiology Research within Cohort Studies<sup>36</sup>.

The STARD Statement for Reporting Studies of Diagnostic Accuracy<sup>37</sup> provides a helpful checklist and flowchart; it is recommended that these should be followed in reports of studies of diagnostic accuracy.

In terms of considering the performance of an assay, it may be helpful to distinguish between two related constructs – repeatability and reproducibility: Repeatability is closeness of agreement between the results of successive measurements carried out under the same conditions on the same samples. Reproducibility is closeness of agreement between the results of measurements performed under changed conditions (e.g. of time, operators, calibrators, reagent lots) on different samples.

Repeatability and reproducibility are of crucial importance; are you designing, developing and testing your assay (and reporting on this) in such a way that the tests of its performance will be repeatable and reproducible? Part of this is making sure the data are stored, curated and open. Doing early work on reproducibility of an assay can prevent a programme of work being undermined by a poor assay at a later date.

A publication from the Test Evaluation Working Group of the European Federation of Clinical Chemistry and Laboratory Medicine reviews the need for outcome-based specifications of analytic performance, the most relevant types of outcomes to be considered, and the challenges and limitations faced when setting outcome-based analytical performance specifications<sup>38</sup>.

<sup>N</sup>[www.fda.gov/downloads/Drugs/Guidances/ucm070107.pdf](http://www.fda.gov/downloads/Drugs/Guidances/ucm070107.pdf)



## Considering Sources of Mechanistic Variation

Markers within a single individual may not be consistent. For example, within a given tumour there may be significant heterogeneity in cell types which will appear as noise in an assay<sup>39</sup>. Investigators also need to account for the possibility that any treatment given may affect heterogeneity within the sample to be assayed e.g. a chemotherapeutic agent will kill off susceptible cells, such that those remaining are different. One cannot assume that the original diagnostic assay will always be representative of what is happening in the patient after treatment.

Within one individual there may be temporal/diurnal/seasonal variations in certain factors e.g. blood pressure; the timing of sampling should be considered and meta-data collected on the time at which sample was taken/assay conducted in order to look for potential circadian effects. Such sources of variation may not only be intra-individual but also inherent in the patient care pathway – for example, particular patient subtypes (e.g. mild versus severe) might be seen at certain times of day, dependent on how clinics are organised.

The potential should also be considered for introduction of inter-individual confounding environmental factors affecting variables to be assayed, e.g. diet, exercise, alcohol etc.

If banked samples/samples collected for purposes other than the project itself will be used for the project's assay, the original intended purpose for which the samples were obtained (diagnostic, resection, other research project) may also introduce variability and affect the properties of the sample. For example, measured markers may differ dependent on whether samples were from needle biopsy or tumour resection<sup>40</sup> (where samples may include more normal tissue).

If you are working with e.g. a psychological assay, there will likely be a large drift in assay results within individuals due to practice/learning effects. The same may also apply to functional imaging. This should be taken account of in interpretation of results.

It is also important to consider the potential responsiveness of the marker to therapy – the background level of a particular analyte might be high, such that the signal to (analytic) noise ratio is high and the consequences of assay technical noise are low, but after therapy the level of the analyte might be e.g. tenfold smaller, where technical assay noise will be much greater, in proportion to signal. In discovery studies, it is likely that the size of the signal will be unknown prior to conducting the study; this needs to be considered and potentially pilot work conducted in order to iteratively work up the assay.

## Meta-Data and Transparency on Data Cleaning/Processing

Investigators should think carefully about collection and reporting of meta-data such that one can subsequently look for drift and artefact. Such meta-data should include not only data on how the data were collected, but also how they were cleaned/processed/ transformed. All data processing should be reported and accessible in an open and standardised way, to support reproducibility. Samples and data should be made available in as raw a form as is useful. The level of resolution of data needs to be thought through and reported on e.g. for imaging it is not possible to store and make available all raw data, so for the data that are stored, used and reported, investigators must be very clear and transparent what has been done to clean and process the data. Meta-data on cleaning, processing and analysing data should be sufficiently granular for someone to be able to repeat the process. As a gold standard this might include the name and version of software packages used for each step, as well as method and parameters, and the exact version of any reference datasets used in comparisons. Data should be made available in standard open data formats where possible (rather than proprietary data formats) to encourage reuse.

## Longitudinal Effects

Assays will improve rapidly as technology changes, particularly in omics-based technologies (for example the replacement of microarrays by RNA-seq for gene expression measurement). In the context of stratified medicine, investigators might therefore develop a stratification tool using one technology only to find it is becoming obsolete before it reaches the clinic or trials. Investigators need to have a long view of technologies that are appearing 'on the horizon' and a clear strategy by which emergent methods may be compared to the current method, validated and ultimately adopted. Keeping old samples may be important in benchmarking new assay technologies against old, and also large-scale reference samples for regular quality control.

Furthermore, as analytic sensitivity in an assay tends to improve over time during the course of long-term data collection, this may give the appearance of trends which aren't actually there, especially if operators improve in their handling of the assay technology/technique as time passes. Investigators must be cognisant of the possibility of such effects, and take steps to mitigate the bias this may introduce.

It should also be noted that the demands on the assay will be different from a first early discovery study through to a clinically implementable assay technology; the former needs to be more agile, the latter needs to be more robust. For example in biomarker discovery studies in schizophrenia, PET scanning is used to look for signals; however PET scanning would be prohibitively expensive and impractical in routine clinical stratification.

### ***Questions Investigators Should Consider when Considering Assay Design:***

#### **Conduct**

- What consents and approvals will be required for the data you wish to collect?
- How will samples/data be collected? What resources and infrastructure will be required to deliver this?
  - What archived/clinical samples are available for discovery and/or verification and is suitable consent in place for this use?
- When designing the assay, investigators should consider:
  - What will the detection limit of the assay be? What technical constraints may affect this? What is the mechanistic rationale for the appropriateness of the intended detection limit?
  - What are the potential pre-analytic sources of variability?
- Sample collection type (biopsy, resection, liquid etc.)
- Sample site collection (e.g. primary versus metastatic tumour)
- Sample collection timing (e.g. pre/post treatment, diurnal variation)
- Sample handling (storage, shipping, stability etc.)
- Sample preservative (FF, FFPE, RNA later, EDTA etc.)
- Sample objective (Diagnostic, Biomarker, Tumour boundary etc.)
- Clinical Site effect (variation in methodology, +/- execution of SOPs; a composite of the above)
  - What are the sources of variability in the assay itself (i.e. "technical noise")? Will your design allow you to quantify these?
- What is the expected natural variation/noise for a given marker, such that the technical noise can be understood and quantified?
- How many measurements must be taken before variation is understood and a signal can be verified?
- Knowing the limitations of technology, how will you minimise noise?
- How will you design your assay to be robust to unavoidable noise?
- How will you ensure noise is controlled during assay delivery?
  - Is there a potential for circadian/seasonal variation in the marker to be assayed?
  - What is the likely extent of responsiveness of the marker to therapy? If high, will assay technical noise be a greater confounding factor and source of bias post-treatment, when levels of the marker are low?
  - What sample quantity will be required to make the proposed measurement? Is that quantity clinically feasible/acceptable to patients?
  - Which technology platform should be used?



- Is the candidate platform sensitive, specific and precise enough to reliably detect your candidate analyte(s)?
- Does it have sufficient breadth of coverage for de novo biomarker discovery?
- Will the discovery platform be the same platform for validation and commercial delivery?
- If not, what studies are needed to migrate and when should these be performed?
- Is the chosen platform likely to be long-lived enough to be relevant throughout the study period? If horizon-scanning has indicated that a newer method may be forthcoming, what strategy will be adopted to incorporate any necessary changes?
  - How will the assay be calibrated?
  - How will outliers be defined?
  - How will the assay be validated?
- If comparing the assay to a gold standard, have you considered that the gold standard itself will also have measurement error?
  - What is the expected target profile of the ultimate clinical stratification tool that your work will inform? How can your study design take account of this?
- What is the statistical justification of the proposed sample size for assay development?
- How will assay data be cleaned/processed and how will you document and make this transparent?
- When developing the assay technology/technique, has the potential for operator/machine/centre variability been taken into account?
  - How will you attempt to standardise data/sample collection and processing? Have you developed a Standard Operating Procedure? Do you have appropriate training programmes and quality control systems in place?
- Have you considered the possibility for improvement in assay technology and operator performance over time, and taken steps to mitigate potential introduction of bias?
- If working with psychological/behavioural assays, have you considered the possibility for learning/practice effects?

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

### Re-framing the Question

In most studies, prior knowledge in the field will be a strong influence on initial ideas for relevant strata. Researchers should be cognisant of the wider context of their work and look for opportunities to maximise the use of prior knowledge to inform interpretation of their own study and proposed strata and markers. However, it is always important to consider the possibility of new paradigms, particularly in cases when existing disease categories have been developed without detailed information on molecular mechanisms, and may be partly historical in nature. Equally, new external data and discoveries will emerge in the course of most studies, and investigators need to be prepared to revisit the initial epistemological schema regularly and consider re-framing the research question. This principle is illustrated in [Annex 1; Case Study 9](#).

Defining strata is almost always challenging, and there may be many different (potentially overlapping) strata that are relevant to different contexts. Investigators should ensure that their clinical population is sufficiently deeply characterised that there is the possibility of redefining strata based on the mechanistic/molecular findings of the study. At each stage, investigators should return to the principles outlined in Theme 1, consider what question they are trying to answer, and proceed with that end in mind.

Investigators should ask themselves whether the putative biomarker/s by which they will define strata is/are likely to be on the causal pathway, or rather an after-the-fact marker of disease; this will affect how it should be built into the stratum/model and how any genetic component should be integrated.

### **Strata Defined by Multiple Marker Profiles**

It is suggested that in modern medicine and in stratified medicine research programmes in particular, single biomarkers or single clinical symptoms are unlikely to be sufficient to fully capture patient

complexity and guide treatment. However there should be a balance, and an avoidance of unnecessary information which increases complexity without enabling further insight. In building useful stratification models, it is important to achieve a “dimensional reduction in phenotype space”; there is an abundance of phenotypic measures one can measure, but statistically it may be best to prune these down.

The importance of multiple markers measured through multiple modalities is illustrated by the fact that rheumatoid arthritis has no single diagnostic test – classification criteria across a number of modalities are used, and an algorithm is employed to give a score. This system was developed for research, but is now useful clinically in diagnosis<sup>41</sup>.

Adopting a single marker approach to stratification can often be too simple and reductive a strategy. If a single marker approach is employed, there is a potential to lose the complexity, interactions and associations between markers. There is a need for methods to integrate and co-analyse multiple variables, even within a single modality data set. However investigators should be clear on the proposed value of the data to be integrated, and consider if it is possible to determine whether the addition of more data is improving the overall model. This may well be context-dependent – e.g. if the intended purpose is to impact clinical therapy choice, will the data set have relevance to that end?

The work of the STELAR consortium<sup>42</sup> in identifying an asthma endotype with multiple allergic sensitisations through a complex model, which would not have been apparent a priori, provides an example of a multivariate stratification model – see [Annex 1; Case Study 3](#). Another example of the additional insight that can be derived from characterising and stratifying a population through use of multiple modalities of measurement is described in [Annex 1; Case Study 4](#). A third example considers prediction of response to pharmacologic therapy through multimodal data in a clinical cohort – see [Annex 1; Case Study 5](#). Two further examples from the field of neurology are described in [Annex 1; Case Study 9](#) and [Annex 1; Case Study 10](#).

An example of a trial assessing the use of multiple markers to inform stratification and affect clinical decisions and endpoints is found in a randomised controlled trial of point-of-care cardiac markers in the emergency department from 2010<sup>43</sup>.

High dimensional data, for instance from ‘omics studies or imaging, are becoming increasingly available in the clinical setting and the special challenges of these data are worthy of separate consideration. The issues outlined in other sections, including assessing potential false discovery rates, dealing with noise and potential bias, and the need for careful verification, all apply as they do for other data types but they are made more acute by the high-dimensionality of the data, which can be significantly larger than the size of the study population. Failures to verify markers effectively have been well publicised<sup>44</sup>, and should serve as a salutary lesson for investigators.

Applications in the clinic or trials are, at present, dominated by relatively simple biomarkers, but the opportunity to generate data from multiple imaging or ‘omics modalities raises the possibility of using more complex markers/strata based on data integration. The advantages of such an approach might include increased robustness in identifying a pathological mechanism gained by integrating independent sources of evidence, and the possibility of identifying broader patient strata with a shared disease mechanism, for example dysregulation of a common pathway which may be initiated by different mutations in different patients but evidenced by common effects in downstream data. The incorporation of existing knowledge in omics-based stratification studies is related to data integration, and the two can be considered together. Most commonly in this area, existing knowledge is expressed in the form of a molecular network, for example known protein interaction, signalling or regulatory networks. There are many useful databases of such networks which can be used to inform data integration.

For example, a technique known as network-based stratification (NBS)<sup>45</sup> has been developed to integrate somatic tumour genomes with gene networks, enabling stratification of cancer into subtypes by clustering together patients with mutations in similar network regions. Its utility was demonstrated in ovarian, uterine and lung cancer cohorts from [The Cancer Genome Atlas](http://cancergenome.nih.gov/)<sup>9</sup>, in which it identified subtypes that are predictive of clinical outcomes such as patient survival, response to therapy or tumour histology.

Biomarkers based on more complex data integration have so far made limited progress into the clinic, but remain an area of active research that is likely to grow and should be monitored where relevant in stratified

medicine studies. Here we give a list of relevant example approaches, but this is not intended to be exhaustive. A publication from Cell in 2014<sup>46</sup> demonstrates an example of the development of an algorithm that integrates genomic data from primary tumours with data from functional RNAi screens in order to identify genetic drivers of breast cancer (and which was verified by prediction of known drivers). Another publication<sup>47</sup> describes the use of integrative Bayesian approaches to develop a computational framework that integrates chromosomal copy number and gene expression data for detecting aberrations that promote cancer progression, and provides a good example of the integration of marker information enabling new mechanistic insight. Another example is OncoIMPACT<sup>48</sup> which aims to integrate driver mutations with 'omic profiles reflecting changes in cell state, using an existing gene interaction network, with the aim of identifying patient specific driver genes. Finally, a study of pancreatic cancer identified relevant biomarkers by integration of imaging modalities<sup>49</sup>.

### Setting Cut-offs

While actual levels of markers will generally be continuous variables, in order to define strata (and eventually to inform treatment decisions), it may be necessary to dichotomise results and set cut-off levels. How will you make sure that your chosen cut-off level is calibrated and verified? The choice of the best cut-off/threshold for discriminatory biomarker tests is dependent on the consequences of false positives and false negatives – a decision analytical approach is called for that accounts for the overall net benefit at different biomarker threshold values<sup>50</sup>.

The likelihood is that there may be multiple markers defining your strata, not just one, and it will be an algorithm that defines the stratum, but that algorithm is also subject to the same considerations regarding cut-offs, calibration, verification, validation and regulation.

### Interpret New Associations and Effects with Caution

A 2008 publication<sup>51</sup> discusses the reasons why many marker associations may be reported or interpreted in an over-inflated manner, and presents a series of caveats investigators should employ in their interpretation.

Investigators should be aware of statistical methods to implement parsimony and increase confidence in putative associations, for example the LASSO (Least Absolute Shrinkage and Selection Operator) method<sup>52</sup>. However, there are many methodological possibilities with advantages and disadvantages in different situations, and specific choices require a team with appropriate expertise.

It is possible that newly discovered putative markers might not correlate closely with historical measures or assays associated with a disease. At the time, this can call into question the validity of the association, but this may be because a new paradigm has been hinted at; the accepted paradigm might be flawed. The story of the development of mepolizumab illustrates this principle<sup>53</sup> - see [Annex 1; Case Study 7](#).

It is suggested that a helpful philosophy is to actively challenge each putative association found – to try to break the association, rather than being subject to confirmation bias and only looking for and acknowledging data which support the association. Investigators should search for factors potentially driving an association which may be due to something not initially considered in the model, e.g. the putative marker may vary diurnally, or with age (as has been shown to be the case with glycans<sup>54</sup>).

<sup>9</sup><http://cancergenome.nih.gov/>

### Clear and Adaptable Analysis Plans

The plan for analysis of data should be clear and well-specified from the outset of the study design, and investigators should know and be explicit about the limitations of the analytic tools to be used (see Theme 3). However it should be acknowledged that this is an evolving field and there needs to be flexibility and regular review – analytical tools and the requirements of the study will adapt and change and the most appropriate tools should be used in the study. It should not be considered mandatory to rigidly adhere to the original analysis plan should better tools have emerged during the project, as long as a clear and transparent rationale for their employment is articulated and sufficient metadata are kept to clearly show how each dataset was actually analysed. Adherence to data standards is also important, particularly in

ensuring that data can be integrated with other sets in the future. It should be acknowledged that data standards themselves may be emergent/evolving for some types of data and this should also be taken into consideration when used within a study.

### Questions Investigators Should Consider when Defining Strata:

#### Analyse

- How will a robust and statistically significant association between the marker and the stratum/response/phenotype of interest be demonstrated?
  - What statistical steps can be taken to increase confidence in association?
- Have you considered the likelihood that a single marker will not be sufficient to define a meaningful stratum?
- How can the putative association be tested? How will the discriminative nature of the marker be established?
- How can you establish whether the marker is genuinely linked to strata rather than just to generic severity of condition that is the same across mechanistic strata?
- What is the argument for the mechanistic plausibility of the putative markers/strata you have discovered?
- If the discovery phase produces competing candidate markers or alternative stratifications, what considerations/methodologies might help the decision of what to take forwards to prospective evaluation?
- How will data be integrated, particularly if using data from multiple modalities (e.g. proteomics, imaging, clinical evaluation) to describe strata? What kinds of data are you trying to integrate?
  - Integrating multimodal biochemical/“non-clinical” data sets – genomics with metabolomics, proteomics etc. or with imaging
  - Integrating the above data types with health records/clinical data
  - Integrating data from the same mode (e.g. genome) from different sources e.g. different exome arrays, or comparing your sequencing hit to a large population data set from another source e.g. Genomics England (GEL).
- Are sufficient metadata available to ensure that mechanistically relevant datasets are integrated?
- How will a (statistical/mathematical/computational) model be built to analyse the data?
  - Will you look for features in the data in absence of outcome data and then check whether these are predictive of outcome? Or will you use outcome as an input?
  - How will you incorporate prior knowledge, including network/pathway analyses?
  - What is normal variation or noise for a given marker or set of markers? Could this help to prioritise markers?
  - How might multi-modal analysis be used to add to and prioritise components?
  - Do you limit number/type of components to support translation to a clinically applicable test?

- How will you evaluate performance of the model? How will you attempt to test/disprove the putative association between marker and stratum? Could you make use of external data sources? If so, what needs to be considered in order to ensure appropriate and timely access?
- How will you present your model to the community? How will you provide sufficient information on your data and the derivation of the model to allow others to test and verify it?

**Interpret** – when considering the data and model/s from previous steps and attempting to confirm the existence of strata, investigators should consider:

- How will you assess whether your markers are causal?
- How will you incorporate existing knowledge to establish mechanistic plausibility of your proposed strata/markers?
- How will you build a mechanistic/systems model, incorporating existing knowledge and your work, which can inform stratification and increase understanding of the underlying mechanisms of disease?
- How might such a model unlock potential mechanistic insights from the associations you have discovered, which might be aligned with hypothesis-driven models to form comprehensive systems models of disease?

## Theme 5: Stratum Verification

### The Nature of Verification

A 2011 opinion paper in *Nature*<sup>55</sup> described the “dismal patchwork of fragmented research on disease-associated biomarkers”, and estimated that of 150,000 papers describing putative biomarkers, only ~100 are in routine clinical use. Investigators are encouraged to ask themselves, as most stratifiers are never used, why will mine be? Proper verification of markers/strata is a key process in the path to their widespread use.

Verification is likely to require a series of steps, with each step developing greater confidence in a fit-for-purpose stratification assay and the strength of the association of its outputs with outcomes.

While not solely pertinent to biomedical research, the International Organisation for Standardisation’s ISO9001 standard<sup>P</sup> may be a useful tool for improving process and likelihood of reproducibility and verification in stratification. This standard addresses various aspects of quality management and provides guidance and tools for companies and organisations wanting to ensure that their products and services consistently meet the customer’s requirements, and that quality is consistently improved. It gives a strong framework for quality management, and while it might seem onerous at the start, adherence to its recommendations is likely to circumvent later problems. This is particularly true if there is a requirement for regulatory approval.

### Discovery and Verification; Differences in Design

When considering verification of a putative marker/stratum, investigators should consider broadly the same four phases as proposed above for discovery studies (Design, Conduct, Analyse, Interpret), but should recognise that verification studies present unique challenges and are not simple repetitions of discovery studies. The majority of the design questions listed in Themes 1, 2 and 3 should be re-considered at the verification stage. It should not be assumed that the answers to these questions will remain the same between discovery and verification stages.

Data from the discovery set may be used to inform the power calculation for the verification set as an effect size will be known. Investigators should think at an early stage about whether the effect size from the discovery study is large enough to give confidence to proceed towards a definitive answer, given the strong likelihood that it will get smaller during verification (and ultimately clinical implementation). Investigators should look for examples in their field of attrition in effect size for markers between discovery and verification to inform planning and decision making. If examples in the same field are not available, then investigators should look for examples in related fields.

<sup>P</sup>[http://www.iso.org/iso/home/standards/management-standards/iso\\_9000.htm](http://www.iso.org/iso/home/standards/management-standards/iso_9000.htm)



Whether a single study looking at multiple associations or multiple studies looking at single associations, the estimate of the highest observed association is likely to be upwardly biased (Winners-curse), so it is a good idea to account for this in the powering of the verification study. This phenomenon has been shown to be a problem e.g. for genetic epidemiology studies and methods have been proposed to adjust for it in this context<sup>57,58,59</sup>, and in biomarker discovery and validation studies more generally<sup>60,61</sup>.

A useful way to think of some of the components of verification studies is to consider how they may be used, not to confirm the association, but to attempt to break it i.e. to look for reasons why the association may be an artefact/spurious. In this process, Investigators should consider all the reasons why the association might be spurious and design a verification study which can assess whether these explanations do not hold. For example if a new prognostic factor is found which appears to predict the outcome for patients with a particular disease, then one of the verification studies might include allowance for all other known prognostic factors to show that the new factor really provides new information and is not simply incorporating information which is already known from other known factors.

Many verification studies also include further opportunistic discovery work. Investigators need to be clear about what they are trying to primarily achieve and design their study and set their statistical power to deliver on that objective. It is possible to add on additional discovery questions as part of the verification strategy, but investigators should be clear and open about their intentions in the study plan, and ensure that additional work does not jeopardise the primary verification objective.

It is strongly recommended that investigators should formalise and document a firm protocol and analysis plan for their verification phase. Verification itself might consist of a number of 'facets' of the marker, for example for a new prognostic biomarker one may need to verify both the assay and also its prognostic value. Consequently the process of verification is likely to need more than one study. For assays with a continuous score where a cut-off score is used for classification, it is important that a specific cut-off value is pre-specified for verification. The use of cut-offs based on the verification dataset population as a whole, such as the median score or quartile values, is not appropriate as these could not be used prospectively in the clinic to stratify on a patient by patient basis. The cut-off levels describing the different strata should have been set at earlier stages and they should not be re-set at the verification stage.

### **Verifying a Stratum Intended for Clinical Use**

A helpful framework for biomarker evaluation studies of those intended for use in disease diagnosis, screening, or prognosis contexts is provided by Pepe et al<sup>62</sup>. This paper describes how best practice dictates that the investigator should be blinded to the clinical outcome data when assessing the biomarker to prevent bias in the interpretation.

If developing a diagnostic/prognostic/theragnostic stratum intended for clinical use, it is often useful to set out a verification protocol that is as close as possible to the ultimate clinical environment in which the stratification would be applied. It is likely that, with the intent of moving closer to the ultimate clinical context, the population for the verification study may be broader than for the more strictly defined discovery stage. The verification dataset(s) are also less likely to be convenience samples.

### **Verifying Markers/Strata using Retrospective Versus Prospective Data**

There are numerous good examples of theragnostic stratifiers in cancer, many of which have been discovered and verified on retrospective data sets (e.g. KRAS status and immunohistochemical markers in colorectal cancer). Several of these stratifiers have subsequently been clinically validated and approved by the regulators on these retrospective data, supporting this as an acceptable and efficient strategy. Some have also been validated prospectively; both approaches are legitimate. Whether the verification study is retrospective or prospective, the range of issues that need to be addressed at its design stage are the same and include how the study deals with major issues such as bias, confounding, appropriate sample size and generalisability. It should be noted that when a retrospective approach is used, more datasets might be required than for prospective verification to allow for the inherent scepticism in the use of retrospective data. This is often a very worthwhile 'price to pay' for retrospective validation as it still often remains more cost-effective than a prospective process, and together with the multiple verifications adds real strength to this approach. For the verification process, at least one additional data set is required. It is suggested that there could be a structured approach to thinking about data sets for verification – the first is retrospective data (if the data you require already exist, why generate more?). The second is retrospective data with added fields; can you add or interpret some extra information that

will make the retrospective data fit for your verification purposes? Then, if those two options are not possible, the third is prospective data collection for a new data set against which to verify.

### **Confounding of Verification by Unexpected Factors**

Investigators should reflect carefully on the context in which the data used for verification were obtained. The link between marker and treatment can be complex e.g. tumour markers are often specific to the site of origin of the tumour; the tumour marker Kras (wild type or mutant) is known to be predictive for the effect of the monoclonal antibodies cetuximab and panitumumab in colorectal cancer. However, this tumour marker is not necessarily predictive of the effect of these therapies in lung cancer. Response may also be affected by other concurrent treatments, for example KRas monoclonal antibody therapy may not work in patients also being treated with oxaliplatin.

### **Meta-Analysis of Discovered Associations**

A useful step in verifying putative associations between markers and outcome may be in performing a meta-analysis of multiple studies all of which have considered these same markers. Guidance for the conduct of such meta-analyses for genetic associations is found in the US Centers for Disease Control and Prevention's [HuGE Reviews guidance](#)<sup>63</sup>, and the 2014 update publication<sup>63</sup>. A good example of differential responses to Quality Improvement strategies, identified through systematic review and meta-analysis, was reported by Tricco et al in 2012<sup>64</sup>.

### **Questions Investigators Should Consider when Considering Stratum Verification:**

In designing studies intended to verify strata, investigators should re-consider the questions from Themes 1, 2, 3 and 4 above, and bear in mind that the answers may not be the same as for the discovery phase. Additionally, investigators should consider:

- Is the discovered association sufficiently strong to warrant continued investigation? Have you considered the likelihood that the effect size may decrease from the initial discovery work?
  - What is the level of confidence in the association to be verified?
- How will you use data from the discovery work to power your verification study?
- How will you use external/independent data sets to verify your proposed association/strata?
  - Are there retrospective data sets available which could/should be used?
  - Are the retrospective data sets fit for purpose? If not, can additional data be added/interpreted to make them so?
- How will you ensure your verification protocol and study population are as close as possible to the ultimate applied clinical context? If this is not the case then you need to set out the reason why, and consider the implications of doing this.
- Have you designed your verification study to actively look for reasons why the proposed markers/strata may be spurious? To confirm the association, attempt to break it.
- How will you look and account for misclassification? I.e. to identify those that have been classified as stratifier negative but are actually positive. Investigators should actively look for this and, if possible, quantify it, and try to reduce it during the verification process.
- How will you establish that your strata have a pre-defined and important level of predictive value in diagnosis/prognosis/theragnosis?
- What is the potential application of your stratifier (application type and commercially/lab developed?) and are there any regulatory implications of this?
  - If the stratifier is intended as a companion diagnostic, then it will be highly regulated.
  - If the marker/assay is for prognostic use, then the regulatory requirement is for a CE mark
  - If the marker is of mechanistic but not clinically predictive value, there are no regulatory requirements
- How will you formalise and document a firm protocol and analysis plan for your verification study?

<sup>63</sup><http://www.cdc.gov/genomics/hugenet/participate.htm>

## Theme 6: Progression Towards Clinical Utility

A publication from Pepe et al<sup>65</sup> reviews methods for, and provides advice on, how to evaluate the performance of markers intended to clinically stratify and guide treatment decisions. This justifies why the population distributional characteristics of the biomarker should be taken account of when trying to devise the best biomarker-guided strategy, for the overall benefit and its distribution among strata are dependent both on the size of the effect modification and the distribution of the marker in the population.

Another valuable perspective on ensuring putative markers and strata are developed thoroughly with effective guidance of stratification and treatment in mind is provided by a publication emanating from the pharmaceutical industry statisticians' group [PSI](http://www.psiweb.org/)<sup>R</sup>.

The MRC [Progress Strategy Partnership](http://progress-partnership.org/)<sup>S</sup> made a number of recommendations for the design and reporting of the development and testing of prognostic and predictive models to guide clinical practice<sup>67</sup>, including their use in stratified medicine<sup>68</sup>. Broadly these publications have outlined the main phases of model development, namely internal verification, external validation, and investigations of a model's impact on clinical practice including the cost effectiveness of any change in practice. Briefly (and in accordance with the principles outlined in this Framework), reliable models for clinical practice are more likely to be obtained when they are: developed using a large, high quality dataset; based on a pre-defined study protocol with a sound statistical analysis plan; and verified in independent datasets obtained from different locations<sup>45</sup>. Furthermore, the development of models for stratified medicine to guide the choice of therapy should lead to independent testing for treatment/biomarker interactions within randomised clinical trials<sup>46</sup>.

The National Health Service represents an excellent environment for the development and implementation of precision medicine through its ability to record accurate patient follow-up with electronic health records and its close ties to UK academia.

A publication from Aronson and Rehm<sup>69</sup> describes the "precision medicine ecosystem" from a number of viewpoints (researcher, patient, clinician etc.), and considers how, in order to accelerate advances in genomic profiling towards healthcare impact, fundamental changes are needed in the infrastructure and mechanisms for data collection, storage and sharing. The paper discusses the creation of a continuously learning healthcare system with seamless cycling between clinical care and research and the importance of patient education about the benefits of sharing data. The importance of considering eventual clinical utility and implementability when designing a stratum discovery study is illustrated in [Annex 1; Case Study 6](#). This is further illustrated in [Annex 1; Case Study 9](#), which describes a stratum discovery project which aims for mechanistic understanding while also pursuing near term, clinically implementable theragnostic stratification.

### Re-framing the question

Once strata have been identified and defined by means of associated markers, it is important to revisit the research question as originally framed, and to consider its relevance and appropriateness in the progression towards clinical utility and impact. It is important to remember that the aspiration is to treat patients, not to treat a biomarker; an example of this is the Prostate Specific Antigen (PSA) story, a non-specific biomarker of prostate cancer, but also of benign prostatic hypertrophy, where harm may be caused by inappropriate intervention. The use of PSA for population-based screening remains controversial<sup>70,71,72</sup> and underlines the need for contextualising biomarkers within more precise, multi-factorial models of the whole patient.

It should also be emphasised that although thinking around stratification tends to gravitate to theragnosis, there are also significant clinical opportunities in risk and prognostic stratification and prevention of disease. A recent publication<sup>73</sup> discusses the potential of genetic stratification in disease prevention. It should also be borne in mind that not all work in stratified medicine must immediately seek to directly develop a therapy or stratifying assay – there is significant value in mechanistic stratification that creates better understanding of the disease and which might suggest future therapeutic targets or markers.

<sup>R</sup><http://www.psiweb.org/>

<sup>S</sup><http://progress-partnership.org/>



In considering clinical utility, investigators should think through demographic variation (ethnicity, age, gender etc.) and consider whether the population in which the initial studies were conducted was representative of the wider clinical population.

### **Clinical Assay Development and Patient Involvement**

If the intended stratification will require a (currently unavailable) clinically implementable assay, thought should be given to its development at an early stage. Investigators should consider the cost and timeline of assay/diagnostic development, and the potential need for (industrial) partnership – clinical assay development/validation needs to sit alongside the timeline for stratified therapy development and progression to a clinical trial. It is very difficult to lever in a diagnostic development study at a late stage. The path towards clinical impact of stratification will intrinsically involve patient acceptability of the stratification. Investigators should explore at the outset whether there are patient groups/support organisations in existence with which the study team can engage in order to understand patient needs and the acceptability of stratification, or of the proposed assay/sample collection method. It will also be important to recognise that, as multiple assays become available to stratify for different treatment interventions, sample availability may become a limiting factor for some technologies. It will be important that the requirements to run a single assay will not preclude patients from being tested for potential benefit from other interventions. Ideally multiple assay results should be available from a single diagnostic sample using the same testing platform (e.g. exome sequencing, MRI or RNA-Seq) thereby reducing cost, turnaround time and providing multiple treatment options for the patient. Another approach to deal with limited diagnostic material may be the analysis of samples where additional samples can be easily acquired using an approach acceptable to the patient such as the use of blood, urine or saliva. An example of patient engagement to explore acceptability of sampling/marker measurement is described in [Annex 1; Case Study 10](#).

Investigators should also bear in mind that there may be value in developing a slightly inferior marker/assay if it is more easily clinically implementable or of greater patient-related utility, with regards to e.g. patient acceptability of sampling method, logistics and expense<sup>74</sup>. A report from the International Society for Pharmacoeconomics and Outcomes Research's risk-benefit management working group provides guidelines on the research methods for how risk/benefit trade-offs from patients should be elicited<sup>75</sup>.

Equally, investigators should not be afraid of developing a marker/stratification system that is, at this point, seemingly not in easy technological reach. Engineers can and do constantly evolve new technologies – for example the new desktop mass spec machines and miniature sequencers – which put previously infeasible stratifying markers within implementable reach.

The field should also be mindful of the potential for stratification to exacerbate health inequalities. A 2009 opinion paper<sup>76</sup> discusses how while advances in molecular testing and genomic technology offer promise, not all of those who might benefit from such technologies have access to the benefits of these advances. Given the inequities in our health-care system, there is no assurance that expanding research into molecular and genomic testing will benefit everyone equally.

A 2014 paper from a multidisciplinary group of the European Federation of Clinical Chemistry and Laboratory Medicine<sup>77</sup> describes a pathway of development for a laboratory assay measuring a biomarker to become a medically useful test.

### **Population Applicability**

Investigators should be clear regarding in which population the marker/stratum was verified. Is the verification population representative of the population in which you wish to apply the stratification technique? Investigators should carefully consider whether there are factors in the clinical population which were not present in the verification population which could confound use of the stratifier.

For example, in rheumatoid arthritis, cigarette smoking may render a patient less responsive to therapies in general<sup>78,79</sup>. In contrast, moderate alcohol consumption appears to improve prognosis<sup>80,81</sup> and might also impact treatment response. Whilst neither factor will necessarily change the nature of the relationship between stratifier and response, such confounders may impact on the stratifier's utility in a particular patient or population.

Similarly, previous treatments might confound stratification; by interfering with relevant biological pathways treatments themselves may alter or dilute the relationship between theragnostic markers and

future treatment response. Thus a theragnostic marker for a treatment that is second or third line may be evident in treatment-naïve patients at baseline but be less clear in a patient already receiving first-line therapy<sup>6</sup>. Similarly patients receiving background therapies may not show the same biomarker relationships as patients not receiving such therapies. On the other hand it may be that the response to a first-line therapy may 'reveal' or emphasise theragnostic biomarkers for a second therapy. For example, and following on from the previous illustrations, in a complex immune/inflammatory disease, blockade of one pathway may highlight an alternative pathogenic pathway<sup>82,83</sup>. Because demographic factors, can influence prognosis, such as level of education, these may also confound theragnostic markers in distinct populations.

### Supporting Reproducibility

It is crucial that studies are designed **and reported** in such a way that the findings will be reproducible. Part of this is making sure the data are stored, curated and open. A recent publication on The Economics of Reproducibility in Preclinical Research<sup>85</sup> summarises various reasons for lack of reproducibility, and describes how irreproducibility undermines progress towards clinical impact.

With this aim, scientific findings, alongside all relevant data and statistical analysis scripts, should be reported in such a way as to facilitate reproducibility; conceptually the Investigator should put in place the documentation, including details of informatics pipelines, needed by an external investigator wishing to reproduce or refute the main conclusions. The reporting should clearly identify the procedures behind strata discovery including any changes to the original study design that were necessary or opportunistically exploited along the way, for example if the strata being reported are for a primary outcome defined *a priori*, or whether secondary outcomes were developed and reported. It is important for the advancement of scientific knowledge that all findings are reported in the public domain, including negative results or studies that fail to find evidence of stratification. Studies that are well-motivated but ultimately unsuccessful may be highly informative to other Investigators in their own right. This enables the community to maximally benefit from all stratified medicine research.

The reader is referred to helpful guidance developed by the STROBE partnership on Molecular Epidemiology, which discusses reporting standards in this area<sup>86</sup>.

In building stratification models intended to ultimately guide clinical practice, thorough and clear reporting on the means by which the model was constructed is of the utmost importance. Investigators are referred to the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement<sup>87</sup>.

### Unexpected Consequences

When planning for clinical implementation studies, investigators should be sure that they will look for and document any unexpected or adverse effects of implementing the stratification clinically. Investigators need to be mindful of the potential emergence of unexpected/adverse events in those stratified **not** to receive the therapy, if the stratification is theragnostic. It is also possible that an adverse effect that is related to 'on-target' pharmacodynamics becomes more prevalent in a theragnostically stratified population.

### Changing Practice

A significant component in achieving clinical impact is changing healthcare policy and clinical guidelines/practice, which is complex and challenging. When developing a means of stratification intended to affect clinical practice, investigators must be mindful that publishing the results of a study is not the end of the journey. Investigators should consider at an earlier stage whether the design of their study will generate the requisite level of evidence to support such change. Investigators are strongly encouraged to engage with experts in service delivery and change and with implementation scientists (as well as with regulatory and policy agencies where appropriate) to explore these concepts.

One perspective on translation of innovations into practice focuses on the characteristics of the new interventions themselves and the organisational context of change; a classic text here is the book by

Everett Rogers<sup>88</sup>. A review of the field of diffusion of innovations in service organisations is provided by Greenhalgh et al<sup>89</sup>.

Another perspective with a focus on both organisational and clinician perspectives is termed implementation research. An early articulation of this perspective was published in the Lancet in 2003<sup>90</sup>.

There may be a tacit assumption that dissemination of clinical guidelines automatically changes practice; a systematic review of the evidence around that assumption will be of interest<sup>91</sup>, and describes the need for effective strategies to actively facilitate dissemination, uptake and impact from guidelines.

A review of clinical prediction models which seek to inform prediction of disease course/response to therapy in cardiovascular disease articulates the challenges in developing such models and the reasons for their limited diffusion into real world practice. “Barriers including poor statistical performance for new populations; lack of a clear decision to be influenced by the [predictive model] output; poor usability; and the reality of a dynamic set of clinical variables all limit implementation. Model development should begin with a formal understanding of the clinical decision that might be supported by prediction; from this the population and outcome selection should follow.” “New utility-based measures of model performance (such as decision curve analysis) hold promise for focusing more attention on the decisional context in which models will be applied at an earlier stage of model development”.

### **Questions Investigators Should Consider Regarding Progression Towards Clinical Utility:**

- Is the verified association sufficiently strong to warrant investment in the development (and lock down) of an assay with appropriate characteristics for clinical deployment?
- What characteristics will the clinical assay require?
  - Manufacturability
  - Performance - precision/reproducibility, linearity/assay reportable range, etc.
  - Ease of use - number of steps, ease of assay deployment/sample access, requirement(s) for pre-processing in the clinic, amount of diagnostic sample required etc.
  - Conditions of use and storage – range of temperatures for storage/operation, length of shelf life, etc.
  - Speed
  - Pricing/cost of goods
- How might your markers/strata be used to inform the design of a stratified trial of new therapies/diagnostic technologies in your population?
  - What is the required level of evidence of the validity of your markers/strata that would give confidence in their use in a trial?
  - What are the performance characteristics of your markers and associated assay that will be required for use in a trial? Have they proven robust in (e.g. geographically) distinct populations, or a broad cross-section, of the disease under study?
- Have you considered trial designs which may allow assessment of multiple treatments and bio markers simultaneously? What would the requirements be for your marker/strata to inform such a trial?
- Has thought been given to a stratified trial design which could incorporate further discovery studies into the trial which could enrich knowledge around your strata, e.g. to identify new markers of emerging drug resistance, or surrogate markers that are predictive of response to therapy?
- What will the requirements be to demonstrate that stratification by your marker(s) will affect patient outcomes given the wider medical, organisational and health economic context?
  - How can you engage early with the requisite clinical/service delivery/regulatory experts to ensure your study will generate the necessary level of evidence to affect clinical practice?

- What steps have you taken to ensure that the results of your work will be reproducible? Have you reported in sufficient detail to support this?
- Have you collaborated appropriately with experts in service delivery and change, with implementation scientists and with representatives of regulatory bodies in order to give your stratification tool the highest possible chance of entering and impacting on clinical practice?

## 5. Concluding Comments

The ability of the stratified medicine approach to change clinical practice and positively impact on human health depends not only on the methodological rigour which the above Framework seeks to enable, but also on effective engagement and communication with the wider stakeholders involved. Care should be taken by investigators to articulate the rationale for stratification not only in terms of patients benefitting from the most targeted therapy with the best chance of success, but also in sparing patients the wasted time and potential adverse effects of a course of therapy which will not be beneficial. It is crucial that stratification should not be perceived as being motivated by rationing of resource. In clinical practice, it might not be an issue of whether or not a patient gets a therapy at all, but in what order therapies are administered in an individual, when there are multiple therapies available. In such scenarios, stratification will enable prediction of which will not work, will spare patients extended courses of uncontrolled disease/adverse effects due to ineffective therapy, and will allow patients to receive the most appropriate therapy more quickly.

Clinical utility of stratification approaches ultimately and intimately involves patient acceptability and experience. Stratification is about identifying what the special features are of the individual patient that will help clinicians guide their therapy.

# Annex 1. Case Studies

The following case studies provide brief descriptions of scenarios faced by investigators working on complex stratification projects which illustrate the principles outlined in this Framework.

## **Case Study 1: The Biomarcare study – planning for verification alongside discovery**

The [Biomarker for Cardiovascular Risk Assessment in Europe](http://www.biomarcare.eu/)<sup>T</sup> (BiomarCaRE) study is a European collaborative research project with the primary objective to assess the value of established and emerging biomarkers for cardiovascular risk prediction<sup>93</sup>, which exemplifies a number of the principles outlined in this Framework and illustrates an example of a project planning for verification alongside marker/stratum discovery. It comprises 21 well-established prospective European population-based cohort studies (from which most baseline data and outcomes were previously harmonized in the MORGAM Project<sup>U</sup>), four cohorts of diseased subjects (disease cohorts, secondary prevention) and four clinical trials, totalling over 300,000 participants with a follow-up of over three million person years.

The BiomarCaRE project was designed as a multi-modular study including: biomarker selection based on omics discovery studies as well as literature, and assay development (Module 1); data harmonization of large scale studies and biomarker determination, analyses of prognostic models and validation (Module 2); and biomarker assessment in clinical trials and with economic evaluation (Module 3).

In phase 1 of the project, established and novel biomarkers are initially selected according to their association with the risk of CVD in a case cohort design of 20,000 subjects including 4,500 incident cardiovascular events. In a second phase the most promising biomarkers from phase 1 are further determined in 130,000 subjects from European population cohorts. The risk models are developed using Cox regression analysis, taking the case-cohort sampling into account. The modelling steps include initial checking of model assumptions and the stability and parsimony of biomarker selection. As a complementary approach for biomarker selection, a random survival forest machine learning approach is employed to determine important variables for the prediction of individual survival times. Performance of the risk estimation models is being cross-validated using calibration, discrimination indices and graphs and net reclassification improvement. The developed predictive models are compared with established risk score models. However the impact of the biomarkers and the predictive models developed in phase 1 is externally validated using the prospective population-based cohorts and biomarker data from phase 2 with all biomarker determinations being conducted in a single laboratory.

Furthermore, joint modelling of repeated measurements of the biomarkers and survival data is being performed for selected cohorts, while to analyse the effects of various pharmacological interventions on the BiomarCaRE panel of selected biomarkers and clinical outcomes, stepwise statistical strategies are being applied within the clinical trials. In a first step, the predictive value of the BiomarCaRE panel is being assessed in samples from the clinical trials, adjusted for medication and stratified by intervention group. In a second step, the change in biomarker levels from baseline to (on average) 1 year on treatment will be assessed and then the treatment effect will be analysed according to biomarker levels at baseline (high vs. low). Finally a decision-analytic state-transition (Markov) model is being used to perform a cohort simulation and deterministic and probabilistic sensitivity analyses performed to determine the long-term effectiveness of biomarker-guided strategies for primary and secondary prevention expressed in life years and quality-adjusted life years (QALY).

## **Case Study 2: U-BIOPRED – early consideration of intellectual property and partner agreement**

U-BIOPRED (Unbiased BIOmarkers in PREdiction of respiratory disease outcomes) is a research project in respiratory disease markers and stratification involving academia and industry partners. It was funded by the 2008 European Commission/European Federation of Pharmaceutical Industries and Associations Innovative Medicines Initiative (IMI).

<sup>T</sup><http://www.biomarcare.eu/>

<sup>U</sup><https://www.thl.fi/morgam/>

U-BIOPRED used an approach to intellectual property that was consistent with the U-BIOPRED project's ethos of collaboration, and with the principle outlined in Theme 1 above. The main principle was that access to the outcomes of the project should be minimally restricted.

The main principles that were framed in the agreement were:

1. Whoever generates a particular piece of IP is the owner of that IP either solely or jointly in the case of multiple contributing parties.
2. Access to IP for research use is granted on a royalty free basis with no time limit with research use broadly defined as any use other than commercial exploitation.

The legal counsel assisting in the preparation of the project agreement constructed a clause that specified that datasets generated by U-BIOPRED would be made open for public use as soon as possible. However, a requirement was documented for a registration process for those accessing the data, in order to avoid redundancy and to ensure that quality analyses were being performed. Grants and contracts offices of some of the academic partner organisations were concerned that royalty-free access of time for research of the IP could limit their ability to license the IP. This created a significant challenge. All of the principle investigators of the partner organisations agreed with the principle of royalty free access.

As the project agreement was required before funding would be distributed, the resulting impasse threatened the project. The impasse was resolved by agreeing with all the principle investigators that the project agreement with the royalty free access use was the only option and partners who did not agree to those terms would be asked to leave the project. All partner organisations promptly signed the agreement.

### **Case Study 3: The STELAR Consortium - Discovery from Data and Multivariate Stratification**

This Case Study shows how bringing together different methodological perspectives (statistical machine learning and biostatistical modelling) in a disease-focused informatics framework can improve discovery and stratification from multivariate data.

The STELAR (Study Team for Early Life Asthma Research) consortium is a multidisciplinary, multicentre research team, combining methodological expertise in health informatics, biostatistics and computer science (statistical machine learning) with asthma-specific epidemiology and genomics to multiply discovery efforts across its five affiliated birth cohorts.

In order to generate and evolve hypotheses using as much of the relevant data as possible the STELAR team took a graphical modelling approach. A 'graph' was drawn to put existing knowledge about asthma biology into a computable framework where the known dependencies between factors are respected when searching complex data for 'unknown' patterns. To achieve this, the STELAR consortium created a secure web-based research environment, [Asthma e-Lab<sup>W</sup>](#), to support consistent recording, description and sharing of data, computational statistical methods and emerging findings across the cohorts<sup>94</sup>.

An initial study using graphical modelling and machine learning in this way employed data from a birth cohort study of a thousand children in Manchester to address the clinical suspicion that the diagnosis of atopy, or allergic tendency, over-aggregates discrete conditions that might be managed differently. The investigators considered two indirect measures of allergic tendency (skin prick and blood tests) and looked at the data with an approximate model of the probability of gaining or losing sensitisation at ages 1, 3, 5 and 8, asking the 'machine' to find subgroups from the structure in the data. Allergic sensitisation patterns were 'learned' from the data, leading to the creation of a five-class model of latent atopic vulnerability, stratifying (through a complex model) what would have been a single clinical diagnosis of atopy.

The STELAR consortium then replicated the five-class atopy model across an equivalent cohort from the Isle of Wight<sup>95</sup>. The investigators also demonstrated that certain of these strata were associated with functional effects - the multiple early sensitisations group was shown to have a different trajectory of lung development (reflected in specific airways resistance)<sup>96</sup>.

<sup>W</sup><http://www.europeanlung.org/en/projects-and-research/projects/u-biopred/home>



By working in this way the STELAR consortium has been able to tackle some controversial topics quickly, such as the received wisdom of the so called “atopic march”. This is a perceived pattern of allergy progressing over time from skin (dermatitis), to lung (asthma), to the nasal mucosa (rhinitis). It was presumed to be a biologically-driven progression by speculating from population-level results from cross-sectional analyses of disease prevalence at different points in the life course. The STELAR consortium, working in Asthma e-Lab, used data from two birth cohorts and a combination of machine learning and biostatistical methods to examine patterns of allergy over time at the individual level. They found that the “atopic march” pattern of progression within individuals over time was present in less than 5% of children<sup>96</sup>. The atopic march was an ecologic fallacy which was disproved through these data-intensive approaches.

#### **Case Study 4: MIMOmics (Methods for Integrative analysis of Multiple Omics datasets in epidemiological studies) – Team Composition, Multimodal Data Integration and Data Availability**

**MIMOmics**<sup>Y</sup> is a **FP7**<sup>Z</sup>-funded project which aims to investigate and develop methods for integrated analysis of multiple ‘omics datasets. The concept is that datasets are complementary in scale, origin, biological interpretation etc. and therefore integrative use of these datasets should be advantageous over single ‘omic methods. The consortium brought together quantitative methodological experts from biophysics, machine learning, bioinformatics and biostatistics at the outset, to facilitate appropriate study design. To ensure correct analysis of novel ‘omics datasets the consortium includes experts in high throughput methods for glycomics and metabolomics. For providing relevant questions and datasets as well as correct interpretation of the results, biologists and epidemiologists play an important role.

The project focusses on two types of methodological approaches: network methods and prediction models. Network analyses aim to obtain more biological insight in the datasets by identification of clusters of correlated variables. Multiple omics datasets can be considered jointly by introducing a layer for each dataset which are connected via the nodes. With regard to prediction methods one of the research questions is to assess the augmented value of a new ‘omics dataset on top of another dataset. The team also investigate hybrids of network and prediction methods to identify models with good prediction performance and with a biologically sound interpretation.

The consortium has access to datasets from two population cohorts: Dilgom and 10001 Dalmatians. **Dilgom**<sup>AA</sup> (Dietary, Lifestyle and Genetic determinants of Obesity and Metabolic syndrome) comprises members of the population aged 20-75 from the region of Helsinki, Finland. For 518 participants, metabolomics, gene expression and genome wide SNP datasets are available. For a subset of 191 participants exome sequencing datasets is available. The second population cohort is a subset of 1745 individuals of the 10001 Dalmatians cohort/biobank<sup>98</sup>. Glycomics and genome wide SNP datasets can be used for method development.

The datasets were made available via **BC Platforms**<sup>BB</sup> which offers a secure data environment with several methodological tools. The first plan was to use this environment. However it appeared to be too restrictive for method development. It was not sufficiently flexible to upload and test the codes and results from analyses can only be downloaded after permission of the data owners. Therefore we agreed on an alternative procedure, namely partners were allowed to download the data from BC Platforms after permission of the data owners. Before getting access to the data, each partner institute and/or researcher (depending on the data owner) had to sign a data agreement form. Data agreement forms were cohort specific. The most important rules were that datasets are not allowed to leave the institutes and should be stored on secured servers within the institutes and all publications should be approved by the data owners before submission.

To initiate communication across the consortium the investigators formulated a case study with Body Mass Index (BMI) as outcome. BMI was proposed because it is available in most cohorts and it is a strong predictor for survival and metabolic health<sup>99</sup>, which will be the outcomes of interest in the second phase of the project. The research question of the case study was to identify variables from glycomics, metabolomics, gene expression, genome wide association and exome sequencing datasets associated

<sup>W</sup><http://www.asthmaelab.org>

<sup>X</sup><http://www.maas.org.uk/>

with BMI using various methodological techniques for integrated analysis. The investigators showed that adding metabolomics to a model with gene expression improved the prediction accuracy (archived manuscript available<sup>CC</sup>). Adding transcriptomics to a model with metabolomics only slightly improved the performance of the model. Further Glycomics was of added value on top of clinical parameters such as cholesterol and glucose levels. With regard to network methods, there are several novel findings which need to be linked to the results of the prediction models and to be further investigated by replication in other cohorts and by causal inference.

### **Case Study 5: PROMISERA – investigating response to therapy through multimodal data, understanding effect of adherence and benefits of early industrial engagement**

The PROMISERA project, funded by the [Stratified Medicine Scotland Innovation Centre<sup>DD</sup>](#), is a pharmacogenomics project with the objective of understanding the basis of response to methotrexate (MTX) in rheumatoid arthritis (RA). This SMS-IC Exemplar Project represents a significant collaboration between the Universities of Glasgow and Edinburgh, NHS Scotland, the SMEs Sitemic and Pharmatics and industrial partners Thermo Fisher Scientific and Aridhia.

Our ability to predict drug response from clinical data is limited. There is preliminary evidence that measurements of gene expression, metabolic factors and genotype are predictive of MTX response in RA, but no clinically useful prediction tool exists. The study approach adopted develops predictive biomarker profiles of MTX response and toxicity by utilising a systems biology approach that integrates a detailed clinical phenotype with polyomic datasets. From patients enrolled in the existing pan-Scottish Early RA (SERA) inception cohort, the team have collected samples, demographic details and longitudinal information on drug therapy, disease outcome and adverse events.

450 SERA patients, having the inclusion criteria of i) having been recruited and consented to the SERA cohort, ii) having been initially prescribed MTX, iii) having follow up data for at least six months, iv) the absence of another rheumatic disease being diagnosed, and representing the full spectrum of response to MTX were selected from the SERA cohort. Baseline peripheral blood samples, were accessed for DNA and RNA extraction and sequencing.

Whole genome sequence data at 0.5X coverage has been generated from 450 SERA patients. Imputation of the full genome sequence has been undertaken. In parallel RNA sequence data of globin and rRNA-depleted samples has been generated. The sequence datasets are maintained in a project analytic workspace within the SMS-IC research data hub. This collaborative research platform is accessible to all members of the collaboration and operates on a “Safe Haven” basis. Pseudo-anonymised patient phenotypic datasets have been extracted from the SERA CRF database to support an on-going multivariate analysis of the contributory factors to MTX response/non-response including clinical bioassay factors, lifestyle parameters, gene sequence and RNA sequence correlates.

Understanding patient adherence to MTX was considered a key element of this study as potentially this represents a significant confounder of the analysis. Using the Community Health Index (CHI) the SERA database has been linked to Scottish Morbidity Records (SMR) and national prescribing data held by the Information Services Division (ISD) of NHS Scotland. Using data linkage, a measure of each patient's adherence to prescribed MTX will be calculated. DNA and RNA sequences performed using peripheral whole blood, will be integrated with the clinical, demographic and adherence data, to allow us to apply probabilistic machine learning methods to classify RA into subtypes that predict drug response. The development of the use of the ISD prescription encashment datasets represents a novel use of these data in pharmacogenomics studies.

The identification of a predictive RA panel and associated algorithms can be commercialised by SMS-IC industry partners or licensed to third parties. These biomarkers and algorithms will subsequently

<sup>Y</sup>[www.mimomics.eu](http://www.mimomics.eu)

<sup>Z</sup><http://cordis.europa.eu/fp7/>

<sup>AA</sup><http://www.nationalbiobanks.fi/index.php/studies2/7-finrisk>

<sup>BB</sup><http://bcplatforms.com/>

<sup>CC</sup><http://arxiv.org/abs/1601.08197>

<sup>DD</sup>[www.stratmed.co.uk](http://www.stratmed.co.uk)



be used to attract an industry-led adaptive trial, carried out at SMS-IC targeting patients who are unlikely to respond to MTX therapy and are, therefore, candidates for early biologic therapy. The bioinformatics solutions arising will have broad application in the area of chronic disease prediction across disease areas with potential application across industry disease platforms.

### **Case Study 6: Extending the applicability of PARP inhibitor therapy in ovarian cancer – standardised data across sites, and designing a stratification study with future trials in mind**

In ovarian cancer, PARP inhibitors prevent the repair of single strand DNA breaks in tumours with homologous repair (HR) deficiencies, leading to cell death. Identification of HR gene deficiencies can therefore be used as a stratifier; currently this is used clinically for germline BRCA1 and BRCA2 mutations. This [Stratified Medicine Scotland-Innovation Centre<sup>EE</sup>](http://www.stratmed.co.uk) (SMS-IC) Exemplar Project is undertaking sequencing on the tumours of 550 High Grade Serious Ovarian Cancer (HGSOC) patients in order to identify other HR gene defects that will inform future studies of PARP inhibitors in anticipation that this will allow more patients to benefit from the use of these agents.

The Project's objectives are, firstly, to learn more about rarer mutational events in HGSOC and whether any of them are hereditary and, secondly, to improve and standardise clinical data collection across Scotland in order that the investigators can robustly monitor patient outcome, minimising geographical differences and maximising excellence at a national level. The project operates over four Scottish NHS Boards and involves investigators from the Universities of Edinburgh and Glasgow as well as two Industrial Partners Thermo Fisher Scientific and Aridhia.

The Investigators, in conjunction with the SMS-IC Core Laboratory, will perform targeted deep sequencing (at 100X/30X coverage in tumour and germ line samples respectively) in 550 high grade serous ovarian carcinomas (germline and somatic tissue pairs) from across Scotland in order to identify patients who would be suitable for future PARP inhibitor studies. This will require the identification of mutations in over 40 genes.

The patient cohort is initially retrospective, with patients who still have platinum-sensitive ovarian cancer consenting to collection of archival blocks for the first sequencing run. Subsequently this will include prospective patients as they present with newly diagnosed HGSOC. In this way, the Investigators anticipate that it will be possible to derive sequence data to guide entry into future clinical trials within about a year of initiating the project. The bioinformatics analysis will be undertaken by staff at the Edinburgh Cancer Research Centre working through the SMS-IC Core Laboratory data hub in order to optimise the collection of the required clinical data in a robust, secure and ethical fashion.

The intent is to identify a group of HGSOC patients predominantly from the Retrospective Patient Cohort who would be suitable for inclusion in future PARP inhibitor clinical trials.

PARP inhibitors have shown exciting efficacy in BRCA1/2 mutation carriers but the logistical challenges involved with identifying somatic BRCA1/2 mutations or mutations in other genes involved in homologous recombination has meant that pharma are focusing their studies on only germline BRCA1/2 mutation carriers or unselected high grade serous ovarian cancer. This project aims to remove the logistical barrier, allowing trials of these agents to be performed in a much broader patient population.

### **Case Study 7: Stratifying severe asthma into $Th^{High}$ and $Th^{Low}$ to identify a treatable trait linked to IL-5 – considering new paradigms of outcome measurement and mechanistic stratification**

It has been known for some time that interleukin (IL)-5 is a sentinel cytokine involved in the maturation, priming and recruitment of eosinophils in inflammatory diseases dominated by this inflammatory cell, such as asthma, eosinophilic gastroenteritis, oesophagitis and hypereosinophilic syndrome. In a range of animal models of eosinophilic pulmonary inflammation, deletion or blockade of the IL-5 gene or its product was shown to have a profound effect in ablating allergen-induced inflammation and related inflammatory models<sup>100</sup>.

<sup>EE</sup>[www.stratmed.co.uk](http://www.stratmed.co.uk)

However, initial studies of an anti-IL-5 human neutralising monoclonal antibody (mAb), targeting on allergen induced-early and late phase bronchoconstrictor responses in human asthma failed to reveal efficacy<sup>101</sup>. A subsequent Phase II clinical trial in moderate-severe asthma requiring high dose inhaled corticosteroid treatment was then conducted that again failed to reveal efficacy on any of the standard asthma outcome measures<sup>102</sup>. A related pilot study in severe asthma utilizing the anti-IL-5 mAb SCH55700 (later renamed reslizumab) also failed to reveal efficacy apart from a suggestion in reducing exacerbations<sup>103</sup>. In all three studies, mepolizumab and SCH55700 had a profound effect in reducing circulating eosinophils demonstrating systemic bioavailability over a prolonged period following intravenous injection.

At the same time as this negative clinical trial, there was increasing interest in defining a subgroup of moderate-severe asthmatic patients who exhibited gene signatures in their airway epithelium of activation by T helper lymphocyte Type 2 (Th2) cytokines occurring with IL-4/IL-13 specifically. IL-3, -4, -5, -9, -13 and GM-CSF are cytokines encoded in a cluster on human chromosome 5q31 (designated the IL-4 gene cluster) that are co-ordinately regulated<sup>104,105</sup>. A treatable trait linked to Th2 involvement includes sputum (>3%) and blood eosinophilia (>300 cells/ $\mu$ L), forced exhaled Nitric oxide (FeNO) and total serum IgE (>100 ku/L)<sup>106</sup>. Thus, when patients with moderate-severe asthma treated with high dose inhaled corticosteroids and despite this had persistent sputum eosinophilia, then clinical efficacy of mepolizumab was now revealed with a 43% reduction in exacerbations and a 3-fold improvement in asthma quality of life questionnaire (AQLQ) score<sup>107</sup>. However, there were no significant improvements in standard asthma outcomes of lung function, bronchodilator use or airway methacholine hyper-responsiveness and asthma control questionnaire (ACQ). Similar findings were observed in a subsequent large multi-centre study (DREAM)<sup>108</sup>.

As greater understanding of the role of persisting eosinophilia in the presence of high dose corticosteroids strengthened, further trials were conducted on an even smaller subgroup of eosinophilic asthmatic subjects, but this time the entry criteria included even higher doses of inhaled corticosteroids and oral corticosteroids<sup>109,110</sup>. The results were even better, expanding to efficacy not only on exacerbations, but also on a wide range of asthma outcomes including lung function. In 2015, mepolizumab has been licensed in the US for the indication of "add-on maintenance treatment of patients with severe asthma aged 12 years and older, and with an eosinophilic phenotype" and subsequently by the European Commission<sup>99</sup> "as the first approved biologic therapy targeting interleukin-5 (IL-5) treatment for patients with severe refractory eosinophilic asthma in adult patients".

Other anti-IL-5 mAbs are also being developed for this indication including reslizumab<sup>111</sup> and an antibody-dependent cell cytotoxic antibody (benralizumab<sup>112</sup>. Eosinophilia should be regarded as a treatable trait in a wide range of disorders and it is likely that with time other diseases where eosinophilia dominates the cellular response such as rhinosinusitis, eosinophilic gastroenteritis,

Churg Strauss Syndrome etc.<sup>113</sup> will also become amenable to this therapy. Moreover, biologics that attenuate or block the IL-4/-13 pathway are also now in clinical development targeting aspects of the Th2 trait<sup>114</sup>.

### **Case Study 8: Targeted and tailored interventions for smoking cessation**

Interventions designed to change health-related behaviours such as smoking, physical activity and diet may be categorised as generic (the same intervention is used for all members of the target population), targeted (different interventions are used for different subgroups) and tailored (each individual receives an intervention that is matched to his or her unique characteristics)<sup>115</sup>. These fall on a continuum of increasing stratification of the target population into a greater number of numerically smaller and more specific subgroups accompanied by increasing customisation of the interventions. It is assumed that more tailored interventions will be more effective in changing behaviour than less tailored ones, for two main reasons: (1) more tailored interventions will be seen by recipients as more personally relevant, so they will be more likely to attend to, read, understand and act on them; and (2) more tailored interventions are designed to change determinants of the target behaviour that are relevant to particular individuals (or to small subgroups of individuals).

This case study uses smoking cessation as an example of targeted and tailored stratified interventions for behaviour change.

## Targeted interventions

There is substantial evidence that nicotine replacement therapy (NRT) in all its forms (gum, transdermal patch, nasal spray, inhaler and sublingual tablets/lozenges) increases quit rates regardless of setting. It has been hypothesised that highly dependent smokers may benefit more from high-dose NRT whereas low dependence smokers may benefit equally from low- and high-dose NRT. In a rare example in this field of a treatment matching study, Garvey and colleagues<sup>117</sup> tested this hypothesis by classifying smokers as low or high in nicotine dependence using the Heaviness of Smoking Index (HSI)<sup>118</sup>. The HSI asks the number of cigarettes smoked per day and the time from waking to the first cigarette of the day. Participants were 608 smokers who were planning a quit attempt. Within each level of dependence, participants were randomly assigned to placebo, 2-mg, or 4-mg nicotine gum; all three groups also received brief behavioural counselling. At 1 year, quit rates were 11.2, 19.5 and 18.4% for low-dependence smokers receiving placebo, 2-mg gum and 4-mg gum, respectively. The corresponding quit rates for high-dependence smokers were 6.1, 15.7 and 20.7%. Although the pattern of findings was as predicted, the interaction of nicotine gum dose and dependence group was not statistically significant ( $p=0.42$ ).

The authors suggest that some high-dependence smokers receiving high-dose gum may still be under-dosed and others may be overdosed, both of which effects would reduce the benefit of high-dose gum in this subgroup. It is not clear whether the study was powered to detect the predicted interaction effect.

## Tailored interventions

Tailored interventions for smoking cessation are usually explicitly based on theories of the psychological and other determinants of smoking and smoking cessation. They thus provide a good example of hypothesis-driven research and mechanistically-plausible stratification mentioned under Theme 1. Theories are used to define small subgroups (potentially as small as  $N = 1$ ) of the target population and to customise the intervention for each of these subgroups.

For example, the iQuit in Practice intervention consists of a cessation advice report and a three-month programme of text messages. The content of the report and text messages was based on relevant theories of smoking cessation and behaviour change, including social cognitive theory and the

perspectives on change model. Both components of the intervention are tailored using more than 20 items from the online baseline questionnaire including: smoking habits and history; motivation and determination to quit; reasons for quitting; dependence; self-image; pros and cons of quitting; perceived difficult situations; children; living with other smokers; social support; and current health problems. A trial of the short-term effectiveness of the intervention compared with usual care showed promising results at six months<sup>119</sup>.

With complex tailored interventions, treatment matching studies become impracticable. Tailored interventions are therefore usually compared with generic interventions or usual care. Alternative designs could compare interventions that use different degrees of tailoring or tailor on different variables.

## Case Study 9: STRATA – Starting with clinical implementation in mind, multimodal stratification and responsiveness to emerging data.

Schizophrenia: Treatment Resistance and Therapeutic Advances (STRATA) is a UK-based consortium (and one of [MRC's Stratified Medicine consortia](#)) working with industry partners, NGOs, and researchers outside of the UK. Around a third of patients with schizophrenia fail to respond to two or more antipsychotic drugs (all of which work by blocking dopamine receptors) and are termed "treatment non-responsive". Previous work has suggested that such patients have a relatively normal dopamine system, implying that they have a separate, potentially treatable, pathology. Only one drug, clozapine, has an evidence base for this group, but its mechanism of action is unknown and it is associated with

<sup>117</sup><https://www.gsk.com/en-gb/media/press-releases/2015/gsk-s-nucala-mepolizumab-receives-approval-from-us-fda/>

<sup>119</sup><http://lungdiseasenews.com/2015/12/07/european-commission-grants-marketing-authorization-gsks-nucala-mepolizumab-adult-asthma/>

severe adverse effects. The main objectives of STRATA are (1) to develop an imaging and/or genetic based stratification strategy that will help identify patients who will be non-responsive to standard treatments, in order to (a) clarify the pathology underlying treatment non-response, (b) reduce delays to initiation of clozapine and (c) identify patients who could potentially benefit from stratified clinical trials; (2) to develop a multi-centre clinical trials network to deliver standardised stratified clinical trials; and (3) to develop a quantitative understanding of the health economic implications, and the patient and clinician acceptability of adopting a stratified approach in schizophrenia, including quantifying the predictive accuracy beyond which a test will become clinically and economic beneficial.

In the discovery phase, STRATA is establishing procedures for multi-site, standardised stratified patient recruitment and evaluation, multi-modal imaging, genomic discovery, health economic implications, and assessing patient opinion on the acceptability of stratified medicine in the treatment of schizophrenia. Equal numbers of treatment responsive and non-responsive patients are being invited to take part in a hypothesis-driven imaging study which uses PET and MRS to test whether dopamine and glutamate levels in the brain can be used to distinguish responders and non-responders. Aiming to replicate pilot studies in this area, the imaging study will provide insight into the mechanistic plausibility and predictive validity of a dopamine/glutamate stratifier. A parallel genetics study uses pathway analysis to test the dopamine/glutamate hypothesis, while also using hypothesis-neutral approaches (GWAS, polygenic risk score) to discover previously unexpected associations which could potentially elicit new therapeutic targets. The genetics project has built collaborative links with researchers internationally in order to collate existing data from more than ten longitudinal first episode cohorts. This work illustrates the principle that stratification of complex, heterogeneous disease groups is most likely to be possible through multimodal analysis, through linked assay techniques (e.g. MRS/PET) and through integration of geno- and pheno-typic measures (e.g. GWAS and imaging).

The consortium also set out to develop a multi-centre clinical trials network, primarily by recruiting patients into STRATA's imaging study, and in doing so build and maintain the infrastructure and training necessary to deliver standardised stratified clinical trials. These trials are not part of the initial funding package of the consortium – this is therefore a clear illustration of the principle of starting with the end in mind, and considering the path to clinical impact, leveraging the infrastructure put in place to create a framework for future trials to clinically test the stratification paradigms emanating from the discovery work. Furthermore, while STRATA will look for the underlying pathobiological mechanisms which define the treatment-resistant patient group (and so help to identify new treatment targets in the longer term), in the near term the resultant stratifying markers will help to guide early treatment decisions, sparing non-responsive patients lengthy courses of ineffective and costly dopamine-blocking therapies to which they would otherwise be subject. As such, STRATA is pursuing deeper mechanistic understanding while keeping clinical impact centrally in mind.

The consortium involves internationally recognised imaging and genetic centres as well as other relevant expertise, divided into appropriate work streams e.g. patient and public involvement and biostatistics. Other experts from outside the consortium are also invited to offer insight into areas that have become relevant since the initial grant proposal. One example is a face-to-face team meeting day in June 2016 which gathered STRATA principle investigators, researchers, and statisticians with invited experts in genomic methodology, epigenetics, proteomics, and cannabinoids, to discuss our interim findings and scope out possible future directions, taking into account emerging new data. This marked the first phase of an on-going discussion on the most productive future directions of the consortium under different scenarios, depending on the results of the discovery phase, emerging data from outside the consortium and changes in the pharmaceutical landscape. This illustrates the principle of being responsive to emerging data, and not rigidly adhering to an initial hypothesis when new challenges or opportunities arise.

### ***Case Study 10: Deep and Frequent Phenotyping – identification of markers for clinical trials in preclinical Alzheimer's disease***

The challenge of Alzheimer's disease (AD) and other disorders that cause dementia is enormous – these disorders represent some of the largest unmet need in healthcare today. It is therefore of huge concern that clinical trials are failing at such a high rate. Between 2002 and 2012 an astonishing 99.6% of trials failed<sup>120</sup>, almost certainly surpassing any other therapeutic area. Clearly this mismatch between unmet need and enormously high attrition of extremely costly trials is of considerable concern and there is a growing consensus in the field that simply continuing with current approaches in the hope of eventual success is an unsustainable strategy.

Most compounds being developed for disease modification of AD target the formation or clearance of Amyloid and most clinical trials are conducted in people with clinically diagnosed dementia. However, evidence from post mortem and from biomarker studies suggests a prodromal or preclinical period of at least a decade<sup>121,122</sup> a period in which effective treatment would be, in effect, secondary prevention. Indeed one possible reason for trials failure of putative disease modification trials is that by the time clinical disease is apparent, the therapeutic window has past. Recent pre-planned post-hoc analysis of substantial phase III trials of such a putative disease modification agent (Solanezumab) was in line with such reasoning as cognitive benefits were observed only in those with very mild disease, albeit reaching significance only in pooled studies (<http://www.alzforum.org/therapeutics/solanezumab>). However, performing studies in people with very early, even prodromal or preclinical, disease risks inclusion of participants with a range of pathologies or even none, something recently emphasized as the high prevalence of amyloid negative yet symptomatic people becomes apparent<sup>123</sup>. If clinical trials are performed too late to have an effect and if in any case a large proportion of people entering these trials do not even have the pathology the therapy is designed to counter then it is perhaps no surprise that the trials are not succeeding. There is a third reason, beyond stage of disease and heterogeneity of pathology, which is hindering clinical development. Proof of concept and experimental medicine trials are difficult to the point of not being possible when outcomes are clinical and where the disease has a long and variable time-course. This becomes even more critical when the therapeutic window is in the pre-clinical phase where, by definition, clinical outcomes are redundant.

In order to counter these obstacles to progress increasing efforts are being made to identify biomarkers for preclinical detection and hence selection and stratification of participants for trials. One particular type of marker needed to accelerate disease modification therapeutics in AD is a marker of disease progression in the early or preclinical phase of disease. Such a marker would have utility for selection or stratification and is an essential pre-requisite for proof of concept or experimental medicine trials including as an intermediate phenotype in adaptive trials such as the European Prevention of Alzheimer's Disease trial ([www.ep-ad.org](http://www.ep-ad.org)). The Deep and Frequent Phenotyping study seeks to identify such a marker by adopting an agnostic approach to modality; performing a large number of potential biomarkers repeatedly over a one-year period in order to identify a combination set that show change and predict longer term outcomes.

The concept, which grew out of an MRC hosted workshop with Pharma and academic participants, is simple enough but the challenge to conduct such a trial is considerable. Will elderly people agree to have repeated and invasive investigations including lumbar puncture? Are these highly specialized investigations, including MEG for example, practicable in a multi-site study? Can data of enormous complexity be managed and analysed effectively? How can participants in a pre-clinical phase be identified? Tackling some of these issues, we performed a pilot study asking principally whether such an approach would be practicable and acceptable. A consortium including all the NIHR Biomedical Research Centres and Units in the Translational Research Collaboration in Dementia was formed and partnered with the Alzheimer's Society for participant acceptability studies and with Astra Zeneca/MedImmune to establish the study protocol. Some key learnings from the pilot and the plans for the full study are listed below but the bottom line was that the protocol was effectively established in six sites across England and acceptability to participants was enormously high with very few failed assessments. As a consequence of the pilot study, an application to MRC as part of the Dementias Platform UK was successful and a full study will start recruitment in 2017.

- The Alzheimer's Society conducted focus groups, interviews and questionnaire assessments with carers and participants before, during and after the study. As a consequence of these the protocol was adapted including timing and order of procedures. Some adaptations – such as performing EEG as an end of day procedure because of the effect on hair styling – are important to participants but were never likely to be obvious to the study professionals
- Professionals, including site leads and research facility nurses, expressed concern about lumbar puncture but participants found the procedure acceptable and less uncomfortable than anticipated. As a consequence, the Alzheimer's Society is leading on developing an information package including videos to accompany the full study. This package will be of value for professionals in clinical research facilities as much as potential participants.
- Developing and executing very complex procedures in the context of a multi-site study was a challenge but was achieved with a high degree of efficiency and effectiveness by forming modality-groups with clear leadership and governance arrangements



- Recruitment targets were achieved but only with very considerable efforts by site leads; suggesting that a fully powered study would face difficulties. As a consequence, for the full study we will utilize the opportunity presented by the Dementias Platform UK that is bringing multiple existing cohorts to dementia research. We plan to focus on two or three major cohorts, including UK BioBank for example, inviting people to participate in a staged process including use of pre-determined data such as family history and cognition and then performing triage screens including genotyping and PET imaging resulting in a balanced group some with, and some without, preclinical AD.
- Collection, curation and analysis of data in the pilot study alone was demanding and will become even more so in the full study where the quantities of highly disparate datasets will be very considerable. As a consequence we have dedicated significant resource to manage data including intent to obtain full consent for data-sharing within the study group and beyond, and engaging info-tech and a not-for-profit world leader in Open Data projects from planning stage.
- Given the highly focused goal of the study – to identify biomarkers for proof of concept studies in AD – it was obvious that it would be necessary to work in partnership with industry. We have gone beyond this and nested the Deep and Frequent Phenotyping study in the largest proof of concept study, the European Prevention of Alzheimer's Disease. This IMI programme, a public-private partnership funded jointly by the EU and multiple Pharma companies ([www.ep-ad.org](http://www.ep-ad.org)) will include both a longitudinal trials ready cohort and a rolling adaptive proof of concept trial. To ensure compatibility, the selection criteria and baseline assessment measures of EPAD will be used in Deep and Frequent Phenotyping and at the end of the MRC funded component, participants will be invited to then join the EPAD trials ready cohort and potentially therefore have the opportunity for inclusion in the adaptive trial for disease modification agents.

The Deep and Frequent Phenotyping study has the potential to discover markers ranging from PET imaging, through electrophysiology to molecular markers in CSF or blood as well as to act as a platform for the development of innovative markers including, for example ophthalmological measures and connected devices. Although an enormously complex programme, we have a singular objective – markers for use in selection/stratification and for measures of change in preclinical phase – and a study design carefully matched to this objective, we have adapted our protocols and approach based on a successful pilot, we are exploiting opportunities provided by DPUK and EPAD and have built teams of experts including in academia, in industry including Pharma, biotech and information technology and with patient and carer representative organisations. Already data from the pilot are yielding new information of value, with the first papers being written before the full study starts, and the energy and excitement generated within the study and with international supporters and colleagues is tangible. The task ahead is challenging but the importance of the goal of supporting and even enabling clinical trials for dementia prevention, cannot be overstated.

### **Case Study 11: Data Warehouse infrastructure for the MRC-PSORT Stratified Medicine Consortium – open-source shared data storage**

The benefits of a standardized data warehouse infrastructure, both for analysis and integration of data (within and between projects), are extensive. In complex multi-disciplinary research consortia, as exemplified by MRC stratified medicine projects, these benefits also extend to coordinated data sharing. With these objectives in mind, several active projects in the MRC Stratified Medicine initiative have adopted complementary data warehouse infrastructure for multi-omic and clinical data integration and analysis.

Taking the MRC PSORT (Psoriasis Stratification to Optimise Relevant Therapy - <http://www.psорт.org.uk/>) project as an exemplar, two open source data warehouse tools, i2b2 ([www.i2b2.org](http://www.i2b2.org)) and TranSMART (<http://transmartfoundation.org/about-the-pta/>) have been customized and deployed as virtual machines in the MRC-eMedLab cloud ([www.emedlab.org](http://www.emedlab.org)) to create an integrated clinical/ multi-omic data warehouse for the consortium. This suite of tools enables clinicians and researchers to securely access de-identified clinical and omic data generated by the project alongside curated public domain data. The PSORT project has established two complementary data warehouses, the i2b2 based PSORT Cohort Discovery and TranSMART based PSORT Analytics (Figure 1). The infrastructure used in



PSORT has also been directly replicated by reuse of virtual machine configurations in other stratified medicine projects, including the arthritis stratification consortia MRC-MATURA (<http://www.matura.whri.qmul.ac.uk/>) and MRC-RA-Map (<https://research.ncl.ac.uk/mrgnew-castle/translationalprojects/ramap/>), facilitating future data sharing and meta-analysis between these complementary research consortia.

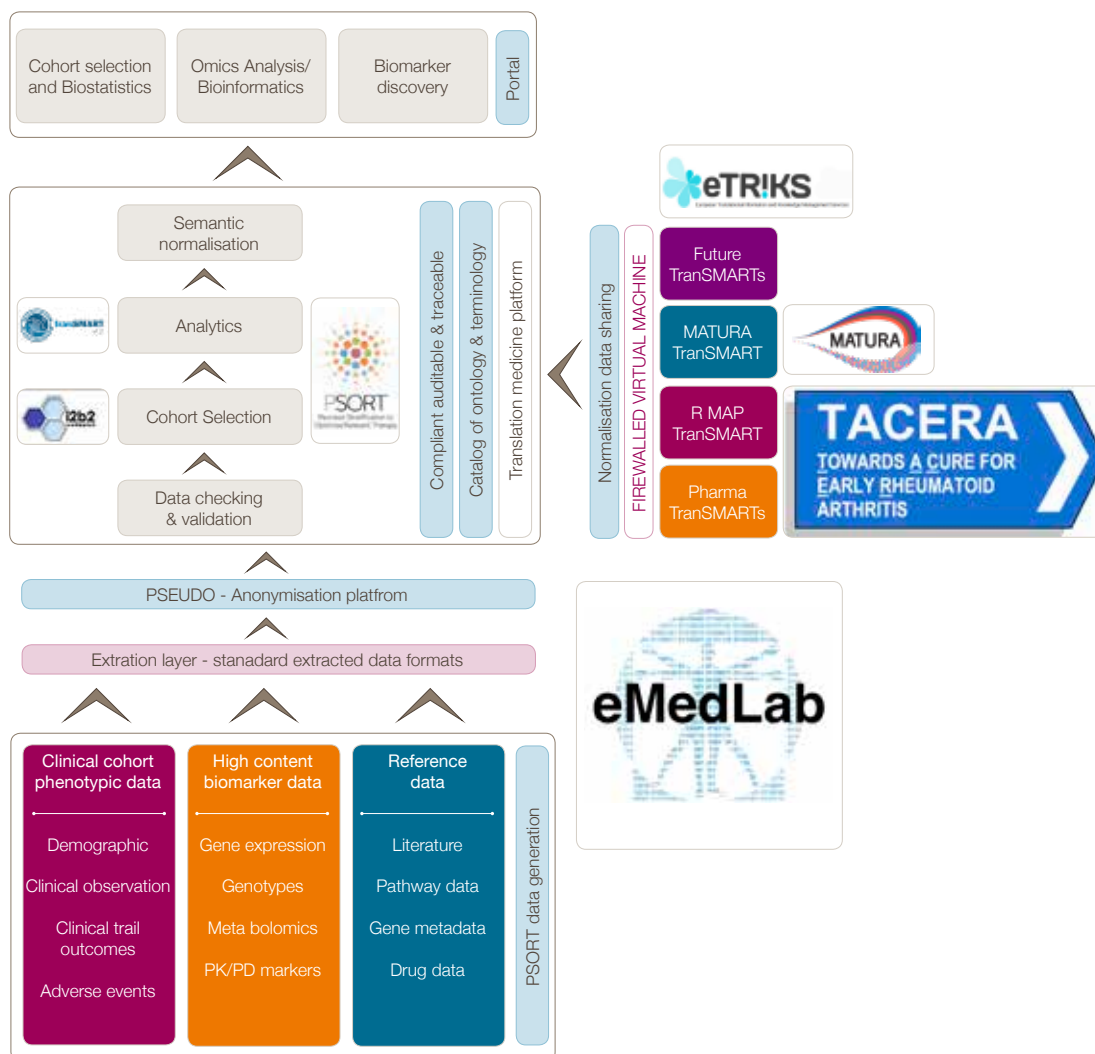


Figure 1. MRC PSORT Data Warehouse Infrastructure

Clinical and Omic data generated by the PSORT project, and also supporting public domain data, requires extensive curation, harmonization and standardization before loading to the i2b2 and TranSMART data warehouses. This undertaking is significant and should not be underestimated, but it also brings considerable intrinsic value by standardizing data into a form that is versioned and analysis ready and suitable for sharing across the consortium and between consortia. Following pre-processing and de-identification as required the data is loaded to the two complementary platforms.

PSORT Cohort Discovery implements the open source i2b2 data warehouse to enable researchers to configure complex longitudinal clinical queries to select patient cohorts matching these criteria. Plugins in the i2b2 environment allow; data review and analysis, e.g. longitudinal data reports; and structured cohort data export to other analysis environments, such as R.

PSORT Analytics implements the open source TranSMART platform enabling PSORT researchers to perform integrated analysis of clinical data with Omic data, using on board analytic plugins such as Smart R. TranSMART also offers an Application Programming Interface (API) to the R package, allowing direct command line analysis of TranSMART data tables.

Collectively i2b2 and TranSMART are hosted as virtual machines in the MRC-eMedLab3 cloud computing environment, allowing flexible resourcing of compute and storage to data warehouse and command line environments to suit different analysis requirements.

## Annex 2: MRC Methodology for Stratified Medicine Workshop Attendees

Dr Michael Barnes, Queen Mary University of London  
Mr Rodrigo Barnes, Aridhia  
Professor Peter Brocklehurst, University College London  
Professor Iain Buchan, University of Manchester  
Professor Chris Chamberlain, UCB  
Professor Ton Coolen, King's College London  
Dr David Crosby, the Medical Research Council  
Professor George Davey-Smith, University of Bristol  
Dr Timothy Davison, Almac  
Mrs Carla Deakin, NICE  
Professor Jon Deeks, University of Birmingham  
Professor Ratko Djukanovic, University of Southampton  
Dr Richard Emsley, University of Manchester  
Professor Jill Francis, City University London  
Professor David Gavaghan, University of Oxford  
Professor Andrew Hattersley, University of Exeter  
Professor Aroon Hingorani, University College London  
Professor Stephen Holgate, University of Southampton  
Professor Peter Holmans, Cardiff University  
Professor Chris Holmes, University of Oxford  
Professor Elaine Holmes, Imperial College London  
Professor Jeanine Houwing-Duistermaat, Leiden University Medical Centre  
Professor John Isaacs, Newcastle University  
Dr Louise Jones, the Medical Research Council  
Dr Andrea Jorgensen, University of Liverpool  
Professor Frank Kee, Queens University Belfast  
Professor Richard Kennedy, Queens University Belfast  
Dr Saskia Litière, European Organisation for Research and Treatment of Cancer  
Professor Fiona Matthews, MRC Biostatistics Unit and University of Newcastle  
Dr Karin Oien, University of Glasgow  
Professor Max Parmar, MRC Clinical Trials Unit, University College London  
Professor Ian Pavord, University of Oxford  
Dr Jonathan Pearce, the Medical Research Council  
Dr Krishna Prasad, MHRA  
Professor David Rand, University of Warwick  
Professor Martin Rossor, University College London  
Mr Antony Sabin, GSK  
Dr Beverley Shields, University of Exeter  
Dr Brian Tom, MRC Biostatistics Unit  
Dr Rhos Walker, the Medical Research Council  
Dr Christine Wells, University of Glasgow  
Professor David Westhead, University of Leeds  
Professor Tony Whetton, University of Manchester  
Professor John Whittaker, GSK and the London School of Hygiene and Tropical Medicine  
Mr Phil Woodward, Pfizer

# Annex 3: Acknowledgements of Contribution

All workshop attendees listed in Annex 2 were consulted in the drafting of this document, and the authors acknowledge and thank them for their insightful contributions to the discussions at the workshop, and to the drafting of this Framework.

Additionally, the authors would also like to acknowledge and thank the following individuals for their helpful comments and contributions during drafting:

Dr Rosemary Barber, University of Sheffield  
Dr Jack Bowden, MRC Biostatistics Unit  
Dr Francesca Buffa, University of Oxford  
Dr Sarah Butcher, Imperial College London  
Professor Peter Diggle, Lancaster University  
Professor Anna Dominiczak, University of Glasgow  
Professor Shitij Kapur, King's College London  
Professor Simon Lovestone, Oxford  
Dr James MacCabe, King's College London  
Professor Sir Munir Pirmohamed, University of Liverpool  
Professor Stephen Sutton, University of Cambridge  
Dr James Wason, MRC Biostatistics Unit

## References

1. Statistical learning and selective inference. Taylor, Jonathan; Tibshirani, Robert J. PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA Volume: 112 Issue: 25 Pages: 7629-7634 Published: JUN 23 2015
2. Asthma phenotyping: a necessity for improved therapeutic precision and new targeted therapies. Chung KF. J INTERN MED. 2016 Feb;279(2):192-204.
3. Priorities and standards in pharmacogenetic research. Need, AC; Motulsky, AG; Goldstein, DB. NATURE GENETICS Volume: 37 Issue: 7 Pages: 671-681 Published: JUL 2005
4. Mepolizumab for severe eosinophilic asthma (DREAM): a multicentre, double-blind, placebo-controlled trial. Pavord ID, Korn S, Howarth P, Bleecker ER, Buhl R, Keene ON, Ortega H, Chanez P. LANCET. 2012 Aug 18;380(9842):651-9
5. Oral glucocorticoid-sparing effect of mepolizumab in eosinophilic asthma. Bel EH, Wenzel SE, Thompson PJ, Prazma CM, Keene ON, Yancey SW, Ortega HG, Pavord ID; SIRIUS Investigators. N ENGL J MED. 2014 Sep 25;371(13):1189-97.
6. Mepolizumab treatment in patients with severe eosinophilic asthma. Ortega HG, Liu MC, Pavord ID, Brusselle GG, FitzGerald JM, Chetta A, Humbert M, Katz LE, Keene ON, Yancey SW, Chanez P; MENSA Investigators. N ENGL J MED. 2014 Sep 25;371(13):1198-207.
7. The phenotype standardization project: improving pharmacogenetic studies of serious adverse drug reactions. Pirmohamed M, Aithal GP, Behr E, Daly A, Roden D. CLIN PHARMACOL THER. 89(6):784-5. 2011 Jun
8. Development of antidrug antibodies against adalimumab and association with disease activity and treatment failure during long-term follow-up. Bartelds GM, Krieckaert CL, Nurmohamed MT, van Schouwenburg PA, Lems WF, Twisk JW, Dijkmans BA, Aarden L, Wolbink GJ. JAMA. Apr 13;305(14):1460-8. 2011
9. Oral anticoagulation: a critique of recent advances and controversies. Pirmohamed M, Kamali F, Daly AK, Wadelius M. Trends Pharmacol Sci. 2015 Mar;36(3):153-63
10. New genetic findings lead the way to a better understanding of fundamental mechanisms of drug hypersensitivity. Pirmohamed M, Ostrov DA, Park BK. J Allergy Clin Immunol. 2015 Aug;136(2):236-44. doi: 10.1016/j.jaci.2015.06.022.
11. Asymmetric rheumatoid vasculitis in a hemiplegic patient. Dolan AL. ANN RHEUM DIS. Jun;54(6):532. 1995
12. Neurovascular mechanisms as a possible cause of remission of rheumatoid arthritis in hemiparetic limbs. Keyszer G, Langer T, Kornhuber M, Taute B, Horneff G. ANN RHEUM DIS. 2004 Oct;63(10):1349-51.
13. The development of rheumatoid arthritis after recombinant hepatitis B vaccination. Pope JE1, Stevens A, Howson W, Bell DA. J RHEUMATOL. Sep;25(9):1687-93. 1998
14. Effect of baseline rheumatoid factor and anticitrullinated peptide antibody serotype on rituximab clinical response: a meta-analysis. Isaacs JD, Cohen SB, Emery P, Tak PP, Wang J, Lei G, Williams S, Lal P, Read SJ. ANN RHEUM DIS;72:329-36. 2013
15. What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. Babyak MA. PSYCHOSOM MED. May-Jun;66(3):411-21. 2004

16. Sample size requirements for training high-dimensional risk predictors. Dobbin, Kevin K.; Song, Xiao  
BIOSTATISTICS Volume: 14 Issue: 4 Pages: 639-652 Published: SEP 2013
17. Survival analysis with high-dimensional covariates. Witten, Daniela M.; Tibshirani, Robert.  
STATISTICAL METHODS IN MEDICAL RESEARCH Volume: 19 Issue: 1 Pages: 29-51 Published: FEB 2010
18. Effect of baseline rheumatoid factor and anticitrullinated peptide antibody serotype on rituximab clinical response: a meta-analysis. Isaacs JD, Cohen SB, Emery P, Tak PP, Wang J, Lei G, Williams S, Lal P, Read SJ.  
ANN RHEUM DIS 2013;72:329-36
19. Credibility of claims of subgroup effects in randomised controlled trials: systematic review. Sun, Xin; Briel, Matthias; Busse, Jason W.; et al.  
BRITISH MEDICAL JOURNAL Volume: 344 Article Number: e1553 Published: MAR 15 2012
20. Adherence to Medication. Osterberg, L and Blaschke, T.  
N ENGL J MED; 353:487-497. 2005
21. Adherence to candesartan and placebo and outcomes in chronic heart failure in the CHARM programme: double-blind, randomized, controlled clinical trial. Granger BB, Swedberg K, Ekman I, et al.  
LANCET;366:2005-2011. 2005
22. A systematic literature review of psychosocial and behavioral factors associated with initial medication adherence: a report of the ISPOR medication adherence & persistence special interest group. Zeber JE, Manias E, Williams AF, et al.  
VALUE HEALTH 2013;16:891-900
23. Adherence and persistence in the use of warfarin after hospital discharge among patients with heart failure and atrial fibrillation. Eapen Z, Mi X, Qualls LG, et al.  
J CARD FAIL 2014;20:23-30.
24. Meta-analysis of trials of interventions to improve medication adherence. Peterson AM, Takiya L, Finley R.  
Am J Health-Syst Pharm 2003;60:657-665.
25. BiomarCaRE: rationale and design of the European BiomarCaRE project including 300,000 participants from 13 European countries. Zeller, Tanja; Hughes, Maria; Tuovinen, Tarja; et al.  
EUROPEAN JOURNAL OF EPIDEMIOLOGY Volume: 29 Issue: 10 Pages: 777-790 Published: OCT 2014
26. Tackling the widespread and critical impact of batch effects in high-throughput data. Leek JT, Scharpf RB, Bravo HC, Simcha D, Langmead B, Johnson WE, et al.  
Nat Rev Genet. 2010;11: 733-739.
27. Noise and bias in microarray analysis of tumor specimens. Febbo PG, Kantoff PW.  
J Clin Oncol. 2006 Aug 10;24(23):3719-21. Epub 2006 Jul 5.
28. A reanalysis of mouse ENCODE comparative gene expression data. Gilad Y, Mizrahi-Man O.  
F1000Res. 2015 May 19;4:121.
29. Some imminent but overlooked preanalytical and analytical challenges currently facing biomarkers and companion diagnostics. Halim AB.  
ANN N Y ACAD SCI. 2015 Jun;1346(1):63-70. doi: 10.1111/nyas.12707. Epub 2015 Mar 10

30. Development and independent validation of a prognostic assay for stage II colon cancer using formalin-fixed paraffin-embedded tissue. Kennedy RD1, Bylesjo M, Kerr P, Davison T, Black JM, Kay EW, Holt RJ, Proutski V, Ahdesmaki M, Farztdinov V, Goffard N, Hey P, McDyer F, Mulligan K, Mussen J, O'Brien E, Oliver G, Walker SM, Mulligan JM, Wilson C, Winter A, O'Donoghue D, Mulcahy H, O'Sullivan J, Sheahan K, Hyland J, Dhir R, Bathe OF, Winqvist O, Manne U, Shanmugam C, Ramaswamy S, Leon EJ, Smith WI Jr, McDermott U, Wilson RH, Longley D, Marshall J, Cummins R, Sargent DJ, Johnston PG, Harkin DP. J CLIN ONCOL. 10;29(35):4620-6. doi: 10.1200/JCO.2011.35.4498. Epub 2011 Nov 7.
31. An optimized whole blood method for flow cytometric measurement of ZAP-70 protein expression in chronic lymphocytic leukemia. Shankey TV, Forman M, Scibelli P, Cobb J, Smith CM, Mills R, Holdaway K, Bernal-Hoyos E, Van Der Heiden M, Popma J, Keeney M. Cytometry B Clin Cytom. 2006 Jul 15;70(4):259-69.
32. Psychometric evaluation and calibration of health-related quality of life item banks - Plans for the patient-reported outcomes measurement information system (PROMIS). Reeve, Bryce B.; Hays, Ron D.; Bjorner, Jakob B.; et al. Group Author(s): PROMIS Cooperative Group  
MEDICAL CARE Volume: 45 Issue: 5 Supplement: 1 Pages: S22-S31 Published: MAY 2007
33. Overview of quantitative measurement methods - Equivalence, invariance, and differential item functioning in health applications. Teresi, Jeanne A.  
MEDICAL CARE Volume: 44 Issue: 11 Supplement: 3 Pages: S39-S49 Published: NOV 2006
34. Developing tailored instruments: item banking and computerized adaptive assessment. Bjorner, Jakob Bue; Chang, Chih-Hung; Thissen, David; et al.  
QUALITY OF LIFE RESEARCH Volume: 16 Supplement: 1 Pages: 95-108 Published: 2007
35. Development and calibration of an item bank for the assessment of activities of daily living in cardiovascular patients using Rasch analysis. Baumeister, Harald; Abberger, Birgit; Haschke, Anne; et al.  
HEALTH AND QUALITY OF LIFE OUTCOMES Volume: 11 Article Number: 133 AUG 2 2013
36. Design options for molecular epidemiology research within cohort studies. Rundle AG, Vineis P, Ahsan H. Cancer Epidemiol Biomarkers Prev. 2005 Aug;14(8):1899-907.
37. STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, Lijmer JG, Moher D, Rennie D, de Vet HC, Kressel HY, Rifai N, Golub RM, Altman DG, Hooft L, Korevaar DA, Cohen JF; STARD Group. Clin Chem. 2015 Dec;61(12):1446-52.
38. Setting analytical performance specifications based on outcome studies - is it possible? Horvath AR, Bossuyt PM, Sandberg S, John AS, Monaghan PJ, Verhagen-Kamerbeek WD, Lennartz L, Cobbaert CM, Ebert C, Lord SJ; Test Evaluation Working Group of the European Federation of Clinical Chemistry and Laboratory Medicine. Clin Chem Lab Med. May;53(6):841-8. doi: 10.1515/cclm-2015-0214. 2015
39. Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. Marco Gerlinger, Stuart Horswell, James Larkin, Andrew J Rowan, Max P Salm, Ignacio Varela, Rosalie Fisher, Nicholas McGranahan, Nicholas Matthews, Claudio R Santos, Pierre Martinez, Benjamin Phillimore, Sharmin Begum, Adam Rabinowitz, Bradley Spencer-Dene, Sakshi Gulati, Paul A Bates, Gordon Stamp, Lisa Pickering, Martin Gore, David L Nicol, Steven Hazell, P Andrew Futreal, Aengus Stewart & Charles Swanton. Nature Genetics 46, 225–233 (2014) doi:10.1038/ng.2891. 2014
40. Immunohistochemical expression of biomarkers: a comparative study between diagnostic bronchial biopsies and surgical specimens of non-small-cell lung cancer. Taillade, L.; Penault-Llorca, F.; Boulet, T.; et al.  
ANNALS OF ONCOLOGY Volume: 18 Issue: 6 Pages: 1043-1050 Published: JUN 2007



41. 2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Aletaha D, Neogi T, Silman AJ et al. *ARTHRITIS RHEUM.* Sep;62(9):2569-81. 2010
42. The Study Team for Early Life Asthma Research (STELAR) consortium 'Asthma e-lab': team science bringing data, methods and investigators together. Adnan Custovic, John Ainsworth, Hasan Arshad, Christopher Bishop, Iain Buchan, Paul Cullinan, Graham Devereux, John Henderson, John Holloway, Graham Roberts, Steve Turner, Ashley Woodcock, Angela Simpson. *THORAX.* 2015 August; 70(8): 799–801
43. The Randomised Assessment of Treatment using Panel Assay of Cardiac Markers (RATPAC) trial: a randomised controlled trial of point-of-care cardiac markers in the emergency department. Goodacre, Steve W.; Bradburn, Mike; Cross, Elizabeth; et al. Group Author(s): RATPAC Res Team  
*HEART* Volume: 97 Issue: 3 Pages: 190-196 Published: FEB 2011
44. Evolution of Translational Omics: Lessons Learned and the Path Forward. Christine M. Micheel, Sharly J. Nass, and Gilbert S. Omenn, Editors; Committee on the Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials; Board on Health Care Services; Board on Health Sciences Policy; Institute of Medicine. ISBN: 978-0-309-22418-5. DOI: 10.17226/13297. 2012.
45. Network-based stratification of tumor mutations. Hofree, Matan; Shen, John P.; Carter, Hannah; et al. *NATURE METHODS* Volume: 10 Issue: 11 Pages: 1108-1115 Published: NOV 2013
46. Integration of Genomic Data Enables Selective Discovery of Breast Cancer Drivers. Sanchez-Garcia, Felix; Villagrasa, Patricia; Matsui, Junji; et al.  
*CELL* Volume: 159 Issue: 6 Pages: 1461-1475 Published: DEC 4 2014
47. An Integrated Approach to Uncover Drivers of Cancer  
Akavia, Uri David; Litvin, Oren; Kim, Jessica; et al.  
*CELL* Volume: 143 Issue: 6 Pages: 1005-1017 Published: DEC 10 2010
48. Patient-specific driver gene prediction and risk assessment through integrated network analysis of cancer omics profiles. Bertrand, Denis; Chng, Kern Rei; Sherbaf, Faranak Ghazi; et al.  
*NUCLEIC ACIDS RESEARCH* Volume: 43 Issue: 7 Article Number: e44 Published: APR 20 2015
49. PET/MRI in pancreatic and periampullary cancer: correlating diffusion-weighted imaging, MR spectroscopy and glucose metabolic activity with clinical stage and prognosis. Chen BB, Tien YW, Chang MC, Cheng MF, Chang YT, Wu CH, Chen XJ, Kuo TC, Yang SH, Shih IL, Lai HS, Shih TT.  
*Eur J Nucl Med Mol Imaging.* 2016 Mar [Epub ahead of print] PubMed PMID: 26993317.
50. Evaluation of Markers and Risk Prediction Models: Overview of Relationships between NRI and Decision-Analytic Measures. Van Calster, Ben; Vickers, Andrew J.; Pencina, Michael J.; et al.  
*MEDICAL DECISION MAKING* Volume: 33 Issue: 4 Pages: 490-501 Published: MAY 2013
51. Why most discovered true associations are inflated. Ioannidis, John P. A.  
*EPIDEMIOLOGY* Volume: 19 Issue: 5 Pages: 640-648 Published: SEP 2008
52. The lasso method for variable selection in the cox model. Tibshirani, R  
*STATISTICS IN MEDICINE* Volume: 16 Issue: 4 Pages: 385-395 Published: FEB 28 1997
53. Eosinophilic phenotypes of airway disease. Pavord, Ian D  
*Annals of the American Thoracic Society* Volume: 10 Suppl Pages: S143-9 Published: 2013-Dec

54. Comparative Performance of Four Methods for High-throughput Glycosylation Analysis of Immunoglobulin G in Genetic and Epidemiological Research. Huffman, Jennifer E.; Pucic-Bakovic, Maja; Klaric, Lucija; et al.  
MOLECULAR & CELLULAR PROTEOMICS Volume: 13 Issue: 6 Pages: 1598-1610 Published: JUN 2014
55. Bring on the biomarkers. Poste, George  
NATURE Volume: 469 Issue: 7329 Pages: 156-157 Published: JAN 13 2011
56. Upward bias in odds ratio estimates from genome-wide association studies. Garner, Chad  
GENETIC EPIDEMIOLOGY Volume: 31 Issue: 4 Pages: 288-295 Published: MAY 2007
57. Unbiased Estimation of Odds Ratios: Combining Genomewide Association Scans with Replication Studies. Bowden, Jack; Dudbridge, Frank  
GENETIC EPIDEMIOLOGY Volume: 33 Issue: 5 Pages: 406-418 Published: JUL 2009
58. Overcoming the winner's curse: estimating penetrance parameters from case-control data. Zöllner, S. and Pritchard, J.K., 2007.  
THE AMERICAN JOURNAL OF HUMAN GENETICS, 80(4), pp.605-615
59. A flexible genome-wide bootstrap method that accounts for ranking and threshold-selection bias in GWAS interpretation and replication study design. Faye, L.L., Sun, L., Dimitromanolakis, A. and Bull, S.B., 2011.  
STATISTICS IN MEDICINE, 30(15), pp.1898-1912
60. Conditional estimation of sensitivity and specificity from a phase 2 biomarker study allowing early termination for futility. Pepe, M.S., Feng, Z., Longton, G. and Koopmeiners, J., 2009.  
STATISTICS IN MEDICINE, 28(5), pp.762-779
61. Correcting for bias in the selection and validation of informative diagnostic tests. D. Robertson, T. Prevost and J. Bowden.  
STATISTICS IN MEDICINE (2015). 34: 1417–1437
62. Pivotal Evaluation of the Accuracy of a Biomarker Used for Classification or Prediction: Standards for Study Design. Pepe, Margaret S.; Feng, Ziding; Janes, Holly; et al.  
JOURNAL OF THE NATIONAL CANCER INSTITUTE Volume: 100 Issue: 20 Pages: 1432-1438  
Published: OCT 15 2008
63. Editorial: Updated Guidance on Human Genome Epidemiology (HuGE) Reviews and Meta-Analyses of Genetic Associations. Gwinn, Marta; Ioannidis, John P.; Little, Julian; et al.  
AMERICAN JOURNAL OF EPIDEMIOLOGY Volume: 180 Issue: 6 Pages: 559-561 Published: SEP 15 2014
64. Effectiveness of quality improvement strategies on the management of diabetes: a systematic review and meta-analysis. Tricco, Andrea C.; Ivers, Noah M.; Grimshaw, Jeremy M.; et al.  
LANCET Volume: 379 Issue: 9833 Pages: 2252-2261 Published: JUN 16 2012
65. Measuring the Performance of Markers for Guiding Treatment Decisions. Janes, Holly; Pepe, Margaret S.; Bossuyt, Patrick M.; et al.  
ANNALS OF INTERNAL MEDICINE Volume: 154 Issue: 4 Pages: 253-W80 Published: FEB 15 2011
66. A statistician's perspective on biomarkers in drug development. Jenkins, Martin; Flynn, Aiden; Smart, Trevor; et al. for the PSI Biomarker Special Interest Group  
PHARMACEUTICAL STATISTICS Volume: 10 Issue: 6 Special Issue: SI Pages: 494-507 Published: NOV-DEC 2011
67. Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research. Ewout W. Steyerberg, Karel G. M. Moons, Danielle A. van der Windt, Jill A. Hayden, Pablo Perel, Sara Schroter, Richard D. Riley, Harry Hemingway, Douglas G. Altman, for the PROGRESS Group..  
PLOS MEDICINE February 2013 | Volume 10 | Issue 2 | e1001381.

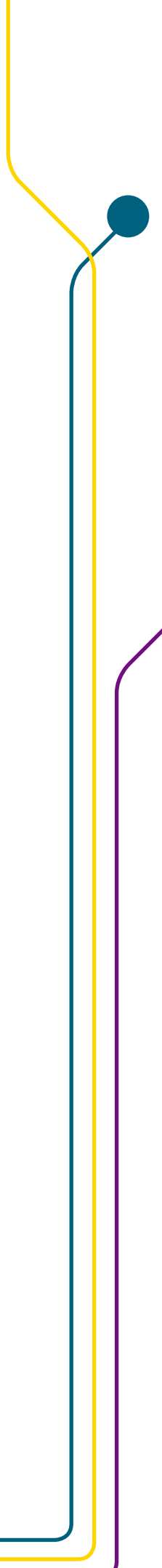
68. Prognosis research strategy (PROGRESS) 4: Stratified medicine research. Aroon D Hingorani, Daniëlle A van der Windt, Richard D Riley, Keith Abrams, Karel G M Moons, Ewout W Steyerberg, Sara Schroter, Willi Sauerbrei, Douglas G Altman, Harry Hemingway for the PROGRESS Group.  
BMJ 2013;346:e5793 doi: 10.1136/bmj.e5793
69. Building the foundation for genomics in precision medicine. Aronson, Samuel J.; Rehm, Heidi L.  
NATURE Volume: 526 Issue: 7573 Pages: 336-342 Published: OCT 15 2015
70. Screening and prostate cancer mortality: Results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of follow-up. Schröder FH, Hugosson J, Roobol MJ, et al.  
Lancet;384:2027-2035 2014
71. Screening for prostate cancer: U.S. Preventive Services Task Force recommendation statement. Moyer VA.  
Ann Intern Med 2012;157:120-134.
72. Radical prostatectomy versus observation for localized prostate cancer. Wilt TJ, Brawer MK, Jones KM, et al  
N Engl J Med;367:203-213. 2012
73. Aligning incentives to fulfil the promise of personalised medicine. Dzau, Victor J.; Ginsburg, Geoffrey S.; Van Nuys, Karen; et al.  
LANCET Volume: 385 Issue: 9982 Pages: 2118-2119 Published: MAY 23 2015
74. The value of positron emission tomography in patients with non-small cell lung cancer.  
Kee, Frank; Erridge, Sara; Bradbury, Ian; et al.  
EUROPEAN JOURNAL OF RADIOLOGY Volume: 73 Issue: 1 Pages: 50-58 Published: JAN 2010
75. A review of quantitative risk-benefit methodologies for assessing drug safety and efficacy-report of the ISPOR risk-benefit management working group. Guo JJ, Pandey S, Doyle J, Bian B, Lis Y, Raisch DW.  
Value Health. 2010 Aug;13(5):657-66. doi: 10.1111/j.1524-4733.2010.00725.x. Epub 2010 Apr 15. Review.
76. Personalized Medicine: A Perk of Privilege? Griggs, J. J.  
CLINICAL PHARMACOLOGY & THERAPEUTICS Volume: 86 Issue: 1 Pages: 21-23 Published: JUL 2009
77. From biomarkers to medical tests: The changing landscape of test evaluation  
Horvath, Andrea R.; Lord, Sarah J.; StJohn, Andrew; et al. Group Author(s): Test Evaluation Working Grp Europe  
CLINICA CHIMICA ACTA Volume: 427 Special Issue: SI Pages: 49-57 Published: JAN 1 2014
78. Patients with early rheumatoid arthritis who smoke are less likely to respond to treatment with methotrexate and tumor necrosis factor inhibitors: observations from the Epidemiological Investigation of Rheumatoid Arthritis and the Swedish Rheumatology Register cohorts. Saevarsdottir S, Wedrén S, Seddighzadeh M, Bengtsson C, Wesley A, Lindblad S, Askling J, Alfredsson L, Klareskog L.  
ARTHRITIS RHEUM.63(1):26-36. doi: 10.1002/art.27758. 2011 Jan
79. Predictors of response to anti-TNF-alpha therapy among patients with rheumatoid arthritis: results from the British Society for Rheumatology Biologics Register. Hyrich KL, Watson KD, Silman AJ, Symmons DP; British Society for Rheumatology Biologics Register.  
RHEUMATOLOGY (Oxford). 2006 Dec;45(12):1558-65. Epub 2006 May 16.
80. The effect of alcohol on radiographic progression in rheumatoid arthritis. Nissen MJ, Gabay C, Scherer A, Finckh A; Swiss Clinical Quality Management Project in Rheumatoid Arthritis.  
ARTHRITIS RHEUM.62(5):1265-72. doi: 10.1002/art.27388. 2010 May

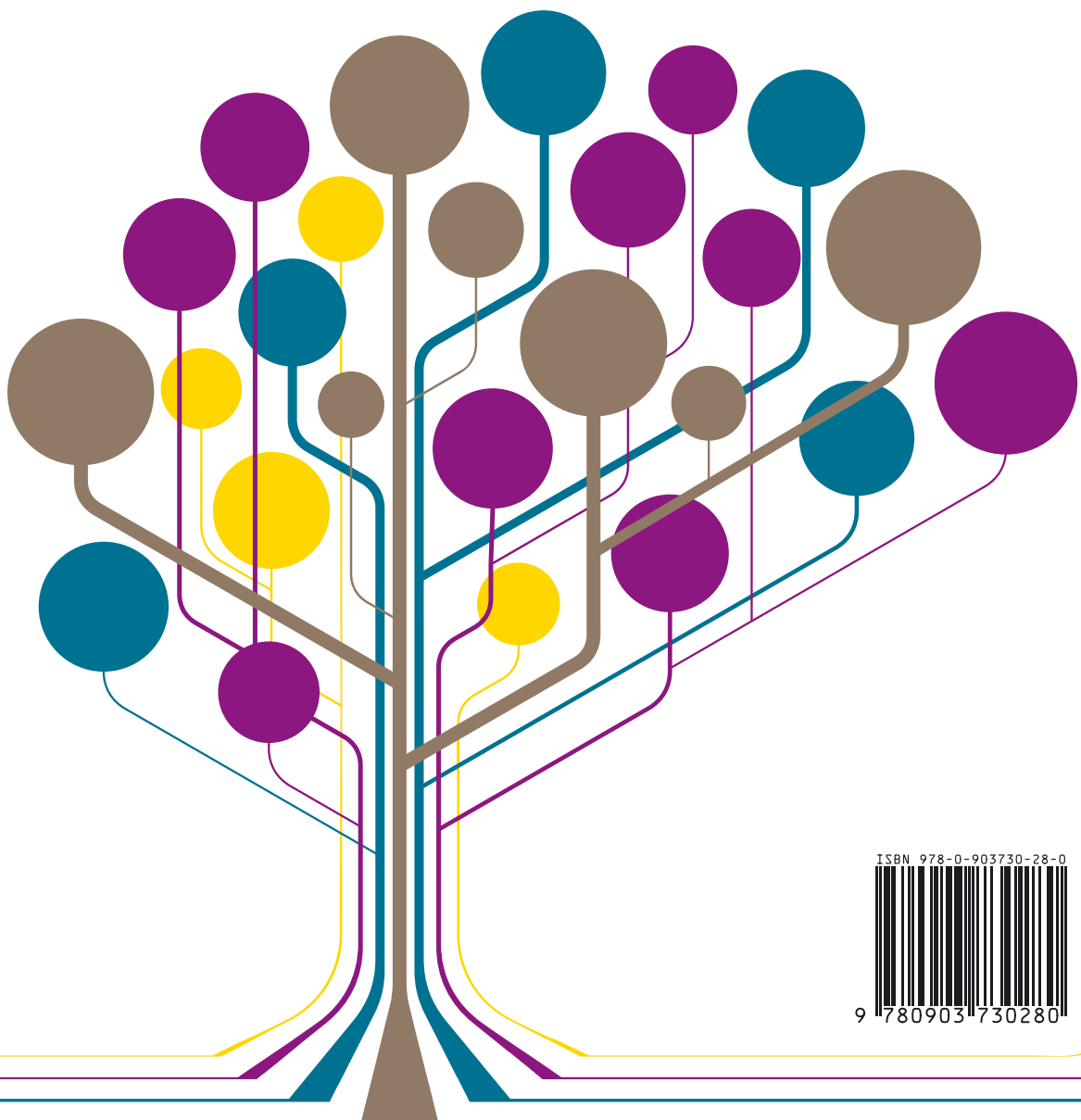
81. Alcohol consumption is associated with lower self-reported disease activity and better health-related quality of life in female rheumatoid arthritis patients in Sweden: data from BARFOT, a multicenter study on early RA. Bergman S, Symeonidou S, Andersson ML, Söderlin MK; BARFOT study group. BMC MUSCULOSKELET DISORD. 24;14:218. doi: 10.1186/1471-2474-14-218. 2013 Jul
82. Increased IL-17 production by peripheral T helper cells after tumour necrosis factor blockade in rheumatoid arthritis is accompanied by inhibition of migration-associated chemokine receptor expression. Aerts NE, de Knop KJ, Leysen J, Ebo DG, Bridts CH, Weyler JJ, et al. RHEUMATOLOGY (Oxford). 49:2264–72. 2010
83. Incomplete response of inflammatory arthritis to TNF blockade is associated with the Th17 pathway. Alzabin S, Abraham SM, Taher TE, Palfreeman A, Hull D, McNamee K, et al. ANN RHEUM DIS; 71:1741–8. 2012
84. Higher education is associated with a better rheumatoid arthritis outcome concerning for pain and function but not disease activity: results from the EIRA cohort and Swedish rheumatology register. Jiang X, Sandberg ME, Saevarsdottir S, Klareskog L, Alfredsson L, Bengtsson C. ARTHRITIS RES THER. 6;17:317. doi: 10.1186/s13075-015-0836-6. 2015 Nov
85. The Economics of Reproducibility in Preclinical Research. Freedman, Leonard P.; Cockburn, Iain M.; Simcoe, Timothy S. PLOS BIOLOGY Volume: 13 Issue: 6 Article Number: e1002165 Published: JUN 2015
86. STrengthening the Reporting of OBservational studies in Epidemiology - Molecular Epidemiology (STROBE-ME): an extension of the STROBE statement. Gallo V1, Egger M, McCormack V, Farmer PB, Ioannidis JP, Kirsch-Volders M, Matullo G, Phillips DH, Schoket B, Stromberg U, Vermeulen R, Wild C, Porta M, Vineis P. EUR J CLIN INVEST. 2012 Apr;42(4):462.
87. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. Gary S. Collins, PhD; Johannes B. Reitsma, MD, PhD; Douglas G. Altman, DSc; and Karel G.M. Moons, PhD. Ann Intern Med. 2015;162:55-63. doi:10.7326/M14-0697
88. Everett M. Rogers. Diffusion of Innovations, Free Press, 4th ed. 1995.
89. Diffusion of innovations in service organizations: systematic review and recommendations. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Milbank Q. 2004;82(4):581-629.
90. From best evidence to best practice: effective implementation of change in patients' care. Grol, R; Grimshaw, J. LANCET Volume: 362 Issue: 9391 Pages: 1225-1230 OCT 11 2003
91. Effectiveness and efficiency of guideline dissemination and implementation strategies. Grimshaw, JM; Thomas, RE; MacLennan, G; et al. HEALTH TECHNOLOGY ASSESSMENT Volume: 8 Issue: 6 Pages: 1-+ Published: FEB 2004
92. Clinical Prediction Models for Cardiovascular Disease Tufts Predictive Analytics and Comparative Effectiveness Clinical Prediction Model Database. Wessler, Benjamin S.; Lai, Lana Y. H.; Kramer, Whitney; et al. CIRCULATION-CARDIOVASCULAR QUALITY AND OUTCOMES Volume: 8 Issue: 4 Pages: 368-375 Published: JUL 2015
93. BiomarCaRE: rationale and design of the European BiomarCaRE project including 300,000 participants from 13 European countries. Zeller, Tanja; Hughes, Maria; Tuovinen, Tarja; et al. EUROPEAN JOURNAL OF EPIDEMIOLOGY Volume: 29 Issue: 10 Pages: 777-790 Published: OCT 2014

94. The Study Team for Early Life Asthma Research (STELAR) consortium 'Asthma e-lab': team science bringing data, methods and investigators together. Custovic, Adnan; Ainsworth, John; Arshad, Hasan; et al. THORAX Volume: 70 Issue: 8 Pages: 799-801 Published: AUG 2015
95. Multiple atopy phenotypes and their associations with asthma: similar findings from two birth cohorts. Lazic, N.; Roberts, G.; Custovic, A.; et al. ALLERGY Volume: 68 Issue: 6 Pages: 764-770 Published: JUN 2013
96. Trajectories of Lung Function during Childhood. Belgrave, Danielle C. M.; Buchan, Iain; Bishop, Christopher; et al. AMERICAN JOURNAL OF RESPIRATORY AND CRITICAL CARE MEDICINE Volume: 189 Issue: 9 Pages: 1101-1109 Published: MAY 1 2014
97. Developmental Profiles of Eczema, Wheeze, and Rhinitis: Two Population-Based Birth Cohort Studies. Belgrave, Danielle C. M.; Granell, Raquel; Simpson, Angela; et al. PLOS MEDICINE Volume: 11 Issue: 10 Article Number: e1001748 Published: OCT 2014
98. "10001 Dalmatians": Croatia launches its national biobank. Rudan, I., Marusic, A., Jankovic, S., Rotim, K., Boban, M., Lauc, G., Grkovic, I., Dogas, Z., Zemunik, T., Vatauvuk, Z., Bencic, G., Rudan, D., Mulic, R., Krzelj, V., Terzic, J., Stojanovic, D., Puntaric, D., Bilic, E., Ropac, D., Vorko-Jovic, A., Znaor, A., Stevanovic, R., Biloglav, Z., and Polasek, O. CROATIAN MEDICAL JOURNAL Volume: 50 Issue: 1 Pages: 4-6 Published: FEB 2009
99. Metabolic mediators of the effects of body-mass index, overweight, and obesity on coronary heart disease and stroke: a pooled analysis of 97 prospective cohorts with 1.8 million participants. Lu, Yuan; Hajifathalian, Kaveh; Ezzati, Majid; et al. LANCET Volume: 383 Issue: 9921 Pages: 970-983 Published: MAR 15 2014
100. Inhibition of pulmonary eosinophilia and hyperreactivity by antibodies to interleukin-5. Egan RW, Athwahl D, Chou CC, Emtage S, Jehn CH, Kung TT, Mauser PJ, Murgolo NJ, Bodmer MW. INT ARCH ALLERGY IMMUNOL. 1995 May-Jun;107(1-3):321-2
101. Effects of an interleukin-5 blocking monoclonal antibody on eosinophils, airway hyper-responsiveness, and the late asthmatic response. Leckie MJ, ten Brinke A, Khan J, Diamant Z, O'Connor BJ, Walls CM, Mathur AK, Cowley HC, Chung KF, Djukanovic R, Hansel TT, Holgate ST, Sterk PJ, Barnes PJ. LANCET. 2000 Dec 23-30;356(9248):2144-8.
102. A study to evaluate safety and efficacy of mepolizumab in patients with moderate persistent asthma. Flood-Page P, Swenson C, Faierman I, Matthews J, Williams M, Brannick L, Robinson D, Wenzel S, Busse W, Hansel TT, Barnes NC; International Mepolizumab Study Group. AM J RESPIR CRIT CARE MED. 2007 Dec 1;176(11):1062-71.
103. Effect of SCH55700, a humanized anti-human interleukin-5 antibody, in severe persistent asthma: a pilot study. Kips JC, O'Connor BJ, Langley SJ, Woodcock A, Kerstjens HA, Postma DS, Danzig M, Cuss F, Pauwels RA. AM J RESPIR CRIT CARE MED. 2003 Jun 15;167(12):1655-9
104. Genome-wide profiling identifies epithelial cell genes associated with asthma and with treatment response to corticosteroids. Woodruff PG, Boushey HA, Dolganov GM, Barker CS, Yang YH, Donnelly S, Ellwanger A, Sidhu SS, Dao-Pick TP, Pantoja C, Erle DJ, Yamamoto KR, Fahy JV. PROC NATL ACAD SCI U S A. 2007 Oct 2;104(40):15858-63
105. T-helper type 2-driven inflammation defines major subphenotypes of asthma. Woodruff PG, Modrek B, Choy DF, Jia G, Abbas AR, Ellwanger A, Koth LL, Arron JR, Fahy JV. AM J RESPIR CRIT CARE MED. 2009 Sep 1;180(5):388-95.

106. Role of T2 inflammation biomarkers in severe asthma. Parulekar AD, Diamant Z, Hanania NA. *CURR OPIN PULM MED*. 2016 Jan;22(1):59-68.
107. Mepolizumab and exacerbations of refractory eosinophilic asthma. Haldar P, Brightling CE, Hargadon B, Gupta S, Monteiro W, Sousa A, Marshall RP, Bradding P, Green RH, Wardlaw AJ, Pavord ID. *N ENGL J MED*. 2009 Mar 5;360(10):973-84.
108. Mepolizumab for severe eosinophilic asthma (DREAM): a multicentre, double-blind, placebo-controlled trial. Pavord ID, Korn S, Howarth P, Bleecker ER, Buhl R, Keene ON, Ortega H, Chanez P. *LANCET*. 2012 Aug 18;380(9842):651-9.
109. Oral glucocorticoid-sparing effect of mepolizumab in eosinophilic asthma. Bel EH, Wenzel SE, Thompson PJ, Prazma CM, Keene ON, Yancey SW, Ortega HG, Pavord ID; SIRIUS Investigators. *N ENGL J MED*. 2014 Sep 25;371(13):1189-97.
110. Mepolizumab treatment in patients with severe eosinophilic asthma. Ortega HG, Liu MC, Pavord ID, Brusselle GG, FitzGerald JM, Chetta A, Humbert M, Katz LE, Keene ON, Yancey SW, Chanez P; MENSA Investigators. *N Engl J Med*. 2014 Sep 25;371(13):1198-207.
111. Reslizumab for inadequately controlled asthma with elevated blood eosinophil counts: results from two multicentre, parallel, double-blind, randomised, placebo-controlled, phase 3 trials. Castro M, Zangrilli J, Wechsler ME, Bateman ED, Brusselle GG, Bardin P, Murphy K, Maspero JF, O'Brien C, Korn S. *LANCET RESPIR MED*. 2015 May;3(5):355-66.
112. Benralizumab, an anti-interleukin 5 receptor  $\alpha$  monoclonal antibody, versus placebo for uncontrolled eosinophilic asthma: a phase 2b randomised dose-ranging study. Castro M, Wenzel SE, Bleecker ER, Pizzichini E, Kuna P, Busse WW, Gossage DL, Ward CK, Wu Y, Wang B, Khatri DB, van der Merwe R, Kolbeck R, Molfino NA, Raible DG. *LANCET RESPIR MED*. 2014 Nov;2(11):879-90.14.
113. Inhibition of interleukin-5 for the treatment of eosinophilic diseases. Corren J. *DISCOV MED*. 2012 Apr;13(71):305-12.
114. Anti-IL-4/-13 based therapy in asthma. Walsh GM. *EXPERT OPIN EMERG DRUGS*. 2015 Sep;20(3):349-52.
115. Understanding tailoring in communicating about health. Hawkins, R.P., Kreuter, M., Resnicow, K., Fishbein, M. and Dijkstra, A. *HEALTH EDUCATION RESEARCH* 2008;23:454-466
116. Nicotine replacement therapy for smoking cessation. Stead, L.F., Perera, R., Bullen, C., Mant, D., Hartmann-Boyce, J., Cahill, K. and Lancaster, T. *COCHRANE DATABASE OF SYSTEMATIC REVIEWS* 2012, Issue 11. Art. No.: CD000146. DOI: 10.1002/14651858.CD000146.pub4
117. Effects of nicotine gum dose by level of nicotine dependence. Garvey, A.J., Kinnunen, T., Nordstrom, B.L., Utman, C.H., Doherty, K., Rosner, B. and Vokonas, P.S. *NICOTINE & TOBACCO RESEARCH* 2000;2:53-63.
118. Measuring the Heaviness of Smoking: using self-reported time to the first cigarette of the day and number of cigarettes smoked per day. Heatherton, T.F., Kozlowski, L.T., Frecker, R.C., Rickert, W. and Robinson, J. *BRITISH JOURNAL OF ADDICTION* 1989;84:791-800
119. Randomized controlled trial to assess the short-term effectiveness of tailored web- and text-based facilitation of smoking cessation in primary care (iQuit in Practice). Naughton, F., Jamison, J., Boase, S., Sloan, M., Gilbert, H., Prevost, A.T., Mason, D., Smith, S., Brimicombe, J., Evans, R. and Sutton, S. *ADDICTION* 2014;109:1184-1193.



- 
120. Alzheimer's disease drug-development pipeline: few candidates, frequent failures. J. L. Cummings, T. Morstorf, K. Zhong.  
ALZHEIMER'S RESEARCH & THERAPY (2014) 6, 37.
121. Proceedings of the Meeting of the International Working, A. D. the American Alzheimer's Association on "The Preclinical State of, July, U. S. A. Washington Dc, Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria. B. Dubois, H. Hampel, H. H. Feldman, P. Scheltens, P. Aisen, S. Andrieu, H. Bakardjian, H. Benali, L. Bertram, K. Blennow, K. Broich, E. Cavedo, S. Crutch, J. F. Dartigues, C. Duyckaerts, S. Epelbaum, G. B. Frisoni, S. Gauthier, R. Genthon, A. A. Gouw, M. O. Habert, D. M. Holtzman, M. Kivipelto, S. Lista, J. L. Molinuevo, S. E. O'Bryant, G. D. Rabinovici, C. Rowe, S. Salloway, L. S. Schneider, R. Sperling, M. Teichmann, M. C. Carrillo, J. Cummings, C. R. Jack, Jr., G.  
ALZHEIMERS DEMENT (2016) 12, 292-323
122. The evolution of preclinical Alzheimer's disease: implications for prevention trials. R. Sperling, E. Mormino, K. Johnson.  
NEURON (2014) 84, 608-622.
123. Suspected non-Alzheimer disease pathophysiology--concept and controversy. C. R. Jack, Jr., D. S. Knopman, G. Chetelat, D. Dickson, A. M. Fagan, G. B. Frisoni, W. Jagust, E. C. Mormino, R. C. Petersen, R. A. Sperling, W. M. van der Flier, V. L. Villemagne, P. J. Visser, S. J. Vos.  
NAT REV NEUROL (2016) 12, 117-124.



ISBN 978-0-903730-28-0



9 780903 730280